In *The Quantum Theory of Fields*, Nobel Laureate Steven Weinberg combines his exceptional physical insight with his gift for clear exposition to provide a self-contained, comprehensive, and up-to-date introduction to quantum field theory. Volume I introduces the foundations of the subject. The development is fresh and logical throughout, with each step carefully motivated by what has gone before, and emphasizing the reasons why such a theory should describe nature. After a brief historical outline, the book begins anew with the principles about which we are most certain, relativity and quantum mechanics, and the properties of particles that follow from these principles. Quantum field theory then emerges from this as a natural consequence. The classic calculations of quantum electrodynamics are presented in a thoroughly modern way, showing the use of path integrals and dimensional regularization. The account of renormalization theory reflects the changes in our view of quantum field theory since the advent of effective field theories. The book's scope extends beyond quantum electrodynamics to elementary particle physics and nuclear physics. It contains much original material, and is peppered with examples and insights drawn from the author's experience as a leader of elementary particle research. Problems are included at the end of each chapter. A second volume (published September, 1996) describes the modern applications of quantum field theory in today's standard model of elementary particles, and in some areas of condensed matter physics.

From reviews of Volume I

'...an impressively lucid and thorough presentation of the subject... Weinberg manages to present difficult topics with richness of meaning and marvellous clarity. Full of valuable insights, his treatise is sure to become a classic, doing for quantum field theory what Dirac's *Quantum Mechanics* did for quantum mechanics. I eagerly await the publication of the second volume.'

S. S. Schweber, *Nature*

'For over twenty years there has been no good modern textbook on the subject. For all that time, Steven Weinberg has been promising to write one. That he has finally done it is cause for celebration among those who try to teach and try to learn the subject. Weinberg's book is for serious students of field theory... it is the first textbook to treat quantum field theory the way it is used by physicists today.'

Howard Georgi, *Science*

'Steven Weinberg, who contributed to the development of quantum chromodynamics and shared the Nobel Prize in Physics for his contributions to the electroweak theory, has written a definitive text on the physical foundations of quantum field theory. His book differs significantly from the long line of previous books on quantum field theory... To summarize, *Foundations* builds the structure of quantum field theory on the sure footing of physical insight. It is beautifully produced and meticulously edited... and it is a real bargain in price. If you want to learn quantum field theory, or have already learned it and want to have a definitive reference at hand, purchase this book.'

O. W. Greenberg, *Physics Today*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

'In addition to a superb treatment of all the conventional topics there are numerous sections covering areas that are not normally emphasized, such as the subject of field redefinitions, higher-rank tensor fields and an unusually clear and thorough treatment of infrared effects... this latest book reinforces his high scholarly standards. It provides a unique exposition that will prove invaluable both to new research students as well as to experienced research workers. Together with Volume II, this will become a classic text on a subject of central importance to a wide area of theoretical physics.'

M. B. Green, *CERN Courier*

'I believe that what readers will find particularly helpful in this volume is the consistency of the whole approach, and the emphasis on quantities and properties that are directly useful to particle physicists. This is particularly true for those who are interested in the more phenomenological aspects. The reader only needs limited background knowledge, and a clear line is followed throughout the book, making it easy to follow. The author presents extremely thorough but elementary discusssions of important physical questions, some of which seem to be an original way of addressing the subject.'

J. Zinn-Justin, *Physics World*

'This is a well-written book by one of the masters of the subject... it is certainly destined to become a standard text book and should find its way to the shelves of every physics library.'

J. Madore, *Classical and Quantum Gravity*

'The book starts out with an excellent historical introduction, not found anywhere else, giving citations to many by now classic papers... a valuable reference work as well as a textbook for graduate students.'

G. Roepstorff, *Zentralblatt für Mathematik*

From reviews of Volume II

'It is a majestic exposition. The two volumes are structured in a logical way. Everything is explained with incisive clarity. Weinberg always goes to the heart of any argument, and includes many things that cannot be found elsewhere in the literature. Often I find myself thinking: "Ah! Now I understand that properly."... I find it hard to imagine a better treatment of quantum field theory than Weinberg's. All serious students and researchers will want to have these volumes on their shelves.'

John C. Taylor, *Nature*

'Weinberg's *Modern Applications* goes to the boundaries of our present understanding of field theory. It is unmatched by any other book on quantum field theory for its depth, generality and definitive character, and it will be an essential reference for serious students and researchers in elementary particle physics.'

O. W. Greenberg, *Physics Today*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

'... Steven Weinberg is one of our most gifted makers of theoretical tools as well as a virtuoso in their use. His new book conveys both the satisfaction of understanding nature and the feel of the atelier, for the "modern applications" of its subtitle include both the derivation of physical consequences and the development of new tools for understanding and applying field theory itself... *Modern Applications* is a splendid book, with abundant useful references to the original literature. It is a very interesting read from cover to cover, for the wholeness Weinberg's personal perspective gives to quantum field theory and particle physics.'

Chris Quigg, *Science*

'Experienced researchers and beginning graduate students will delight in the gems of wisdom to be found in these pages. This book combines exposition of technical detail with physical insight in a unique manner that confirms the promise of Volume I and I have no doubt that these two volumes will rapidly constitute the classic treatment of this important subject.'

M. B. Green, *CERN Courier*

'... a valued reference and a mine of useful information for professional field theorists.'

Tom Kibble, *New Scientist*

'... a clear presentation of the subject, explaining the underlying concepts in much depth and in an accessible style. I expect that these volumes will become the first source we turn to when trying to answer the challenging questions asked by bright postgraduates when they first encounter quantum field theory... I have no doubt that *The Quantum Theory of Fields* will soon be found on the bookshelves of most particle theorists, and that it will be one of the main sources used in the preparation of lectures on the subject for postgraduate students.'

C. T. C. Sachrajda, *The Times Higher Education Supplement*

'... Weinberg has produced a masterpiece that will be a standard reference on the field for a long time to come.'

B. E. Y. Svensson, *Elementa*

From reviews of Volume III

'... has produced a treatise that many of us had long awaited, perhaps without fully realizing it... with the publication of *The Quantum Theory of Fields*, Vol. III, has performed an analogous service for supersymmetry... Although this volume is the third in a trilogy, it is quite different from its two predecessors, and it stands on its own... May a new generation of students imbibe its content and spirit.'

Physics Today

'The third volume of *The Quantum Theory of Fields* is a self-contained introduction to the world of supersymmetry and supergravity. It will be useful both for experienced researchers in the field and for students who want to take the first steps towards learning about supersymmetry. Unlike other books in this field, it covers the wide spectrum of possible applications of supersymmetry in physics.'

Hans Peter Nilles, *Nature*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

'Weinberg is of course one of the creators of modern quantum field theory, as well as of its physical culmination, the standard model of all (nongravitational) interactions. It is... very timely that this latest part of his monograph, devoted to supersymmetry and supergravity, has just appeared. As a text, it has been pretested by Weinberg for a freestanding one-year graduate course; as a clear organizing reference to this extremely vast field, it will help the experts as well... Weinberg's style of presentation is as clear and meticulous as in his previous works.'

Stanley Deser, *Journal of General Relativity and Gravitation*

'Weinberg tries to be as elementary and clear as possible and steers clear of more sophisticated mathematical tools. Together with the previous volumes, this volume will serve as an invaluable reference to researchers and a textbook for graduate students.'

G. Roepstorff, *Zentralblatt MATH*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

The Quantum Theory of Fields

Volume I
Foundations

Steven Weinberg
University of Texas at Austin

CAMBRIDGE
UNIVERSITY PRESS

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo,
Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521550017

© S. Weinberg 1995, 2005

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1995
Eighth printing 2004
Paperback edition first published 2005
7th printing 2012

Printed and bound by MPG Books Group, UK

A catalogue record for this publication is available from the British Library.

ISBN 978-0-521-55001-7 Volume 1 hardback
ISBN 978-0-521-67053-1 Volume 1 paperback
ISBN 978-0-521-55002-4 Volume 2 hardback
ISBN 978-0-521-67054-8 Volume 2 paperback
ISBN 978-0-521-66000-6 Volume 3 hardback
ISBN 978-0-521-67055-5 Volume 3 paperback
ISBN 978-0-521-67056-2 three-volume set

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

To Louise

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Contents

Sections marked with an asterisk are somewhat out of the book's main line of development and may be omitted in a first reading.

PREFACE xx

NOTATION xxv

1 HISTORICAL INTRODUCTION 1

1.1 Relativistic Wave Mechanics 3

De Broglie waves □ Schrödinger–Klein–Gordon wave equation □ Fine structure □ Spin □ Dirac equation □ Negative energies □ Exclusion principle □ Positrons □ Dirac equation reconsidered

1.2 The Birth of Quantum Field Theory 15

Born, Heisenberg, Jordan quantized field □ Spontaneous emission □ Anticommutators □ Heisenberg–Pauli quantum field theory □ Furry–Oppenheimer quantization of Dirac field □ Pauli–Weisskopf quantization of scalar field □ Early calculations in quantum electrodynamics □ Neutrons □ Mesons

1.3 The Problem of Infinities 31

Infinite electron energy shifts □ Vacuum polarization □ Scattering of light by light □ Infrared divergences □ Search for alternatives □ Renormalization □ Shelter Island Conference □ Lamb shift □ Anomalous electron magnetic moment □ Schwinger, Tomonaga, Feynman, Dyson formalisms □ Why not earlier?

Bibliography 39

References 40

2 RELATIVISTIC QUANTUM MECHANICS 49

2.1 Quantum Mechanics 49

Rays □ Scalar products □ Observables □ Probabilities

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

2.2 Symmetries 50

Wigner's theorem □ Antilinear and antiunitary operators □ Observables □ Group structure □ Representations up to a phase □ Superselection rules □ Lie groups □ Structure constants □ Abelian symmetries

2.3 Quantum Lorentz Transformations 55

Lorentz transformations □ Quantum operators □ Inversions

2.4 The Poincaré Algebra 58

$J^{\mu\nu}$ and P^{μ} □ Transformation properties □ Commutation relations □ Conserved and non-conserved generators □ Finite translations and rotations □ Inönü–Wigner contraction □ Galilean algebra

2.5 One-Particle States 62

Transformation rules □ Boosts □ Little groups □ Normalization □ Massive particles □ Massless particles □ Helicity and polarization

2.6 Space Inversion and Time-Reversal 74

Transformation of $J^{\mu\nu}$ and P^{μ} □ P is unitary and T is antiunitary □ Massive particles □ Massless particles □ Kramers degeneracy □ Electric dipole moments

2.7 Projective Representations* 81

Two-cocyles □ Central charges □ Simply connected groups □ No central charges in the Lorentz group □ Double connectivity of the Lorentz group □ Covering groups □ Superselection rules reconsidered

Appendix A The Symmetry Representation Theorem 91

Appendix B Group Operators and Homotopy Classes 96

Appendix C Inversions and Degenerate Multiplets 100

Problems 104

References 105

3 SCATTERING THEORY 107

3.1 'In' and 'Out' States 107

Multi-particle states □ Wave packets □ Asymptotic conditions at early and late times □ Lippmann– Schwinger equations □ Principal value and delta functions

3.2 The *S*-matrix 113

Definition of the S-matrix □ The T-matrix □ Born approximation □ Unitarity of the S-matrix

3.3 Symmetries of the *S*-Matrix 116

Lorentz invariance □ Sufficient conditions □ Internal symmetries □ Electric charge, strangeness, isospin, $SU(3)$ □ Parity conservation □ Intrinsic parities □

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Pion parity □ Parity non-conservation □ Time-reversal invariance □ Watson's theorem □ PT non-conservation □ C, CP, CPT □ Neutral K-mesons □ CP non-conservation

3.4 Rates and Cross-Sections 134

Rates in a box □ Decay rates □ Cross-sections □ Lorentz invariance □ Phase space □ Dalitz plots

3.5 Perturbation Theory 141

Old-fashioned perturbation theory □ Time-dependent perturbation theory □ Time-ordered products □ The Dyson series □ Lorentz-invariant theories □ Distorted wave Born approximation

3.6 Implications of Unitarity 147

Optical theorem □ Diffraction peaks □ CPT relations □ Particle and antiparticle decay rates □ Kinetic theory □ Boltzmann H-theorem

3.7 Partial-Wave Expansions* 151

Discrete basis □ Expansion in spherical harmonics □ Total elastic and inelastic cross-sections □ Phase shifts □ Threshold behavior: exothermic, endothermic, and elastic reactions □ Scattering length □ High-energy elastic and inelastic scattering

3.8 Resonances* 159

Reasons for resonances: weak coupling, barriers, complexity □ Energy-dependence □ Unitarity □ Breit–Wigner formula □ Unresolved resonances □ Phase shifts at resonance □ Ramsauer–Townsend effect

Problems 165

References 166

4 THE CLUSTER DECOMPOSITION PRINCIPLE 169

4.1 Bosons and Fermions 170

Permutation phases □ Bose and Fermi statistics □ Normalization for identical particles

4.2 Creation and Annihilation Operators 173

Creation operators □ Calculating the adjoint □ Derivation of commutation/anticommutation relations □ Representation of general operators □ Free-particle Hamiltonian □ Lorentz transformation of creation and annihilation operators □ C, P, T properties of creation and annihilation operators

4.3 Cluster Decomposition and Connected Amplitudes 177

Decorrelation of distant experiments □ Connected amplitudes □ Counting delta functions

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

4.4 Structure of the Interaction 182

Condition for cluster decomposition □ Graphical analysis □ Two-body scattering implies three-body scattering

Problems 189

References 189

5 QUANTUM FIELDS AND ANTIPARTICLES 191

5.1 Free Fields 191

Creation and annihilation fields □ Lorentz transformation of the coefficient functions □ Construction of the coefficient functions □ Implementing cluster decomposition □ Lorentz invariance requires causality □ Causality requires antiparticles □ Field equations □ Normal ordering

5.2 Causal Scalar Fields 201

Creation and annihilation fields □ Satisfying causality □ Scalar fields describe bosons □ Antiparticles □ P, C, T transformations □ π^0

5.3 Causal Vector Fields 207

Creation and annihilation fields □ Spin zero or spin one □ Vector fields describe bosons □ Polarization vectors □ Satisfying causality □ Antiparticles □ Mass zero limit □ P, C, T transformations

5.4 The Dirac Formalism 213

Clifford representations of the Poincaré algebra □ Transformation of Dirac matrices □ Dimensionality of Dirac matrices □ Explicit matrices □ γ_5 □ Pseudounitarity □ Complex conjugate and transpose

5.5 Causal Dirac Fields 219

Creation and annihilation fields □ Dirac spinors □ Satisfying causality □ Dirac fields describe fermions □ Antiparticles □ Space inversion □ Intrinsic parity of particle–antiparticle pairs □ Charge-conjugation □ Intrinsic C-phase of particle–antiparticle pairs □ Majorana fermions □ Time-reversal □ Bilinear covariants □ Beta decay interactions

5.6 General Irreducible Representations of the Homogeneous Lorentz Group* 229

Isomorphism with $SU(2) \otimes SU(2)$ □ (A, B) representation of familiar fields □ Rarita–Schwinger field □ Space inversion

5.7 General Causal Fields* 233

Constructing the coefficient functions □ Scalar Hamiltonian densities □ Satisfying causality □ Antiparticles □ General spin-statistics connection □ Equivalence of different field types □ Space inversion □ Intrinsic parity of general particle–antiparticle pairs □ Charge-conjugation □ Intrinsic C-phase of antiparticles □

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Self-charge-conjugate particles and reality relations □ Time-reversal □ Problems for higher spin?

5.8 The CPT Theorem 244

CPT transformation of scalar, vector, and Dirac fields □ CPT transformation of scalar interaction density □ CPT transformation of general irreducible fields □ CPT invariance of Hamiltonian

5.9 Massless Particle Fields 246

Constructing the coefficient functions □ No vector fields for helicity ± 1 □ Need for gauge invariance □ Antisymmetric tensor fields for helicity ± 1 □ Sums over helicity □ Constructing causal fields for helicity ± 1 □ Gravitons □ Spin ≥ 3 □ General irreducible massless fields □ Unique helicity for (A, B) fields

Problems 255

References 256

6 THE FEYNMAN RULES 259

6.1 Derivation of the Rules 259

Pairings □ Wick's theorem □ Coordinate space rules □ Combinatoric factors □ Sign factors □ Examples

6.2 Calculation of the Propagator 274

Numerator polynomial □ Feynman propagator for scalar fields □ Dirac fields □ General irreducible fields □ Covariant propagators □ Non-covariant terms in time-ordered products

6.3 Momentum Space Rules 280

Conversion to momentum space □ Feynman rules □ Counting independent momenta □ Examples □ Loop suppression factors

6.4 Off the Mass Shell 286

Currents □ Off-shell amplitudes are exact matrix elements of Heisenberg-picture operators □ Proof of the theorem

Problems 290

References 291

7 THE CANONICAL FORMALISM 292

7.1 Canonical Variables 293

Canonical commutation relations □ Examples: real scalars, complex scalars, vector fields, Dirac fields □ Free-particle Hamiltonians □ Free-field Lagrangian □ Canonical formalism for interacting fields

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

7.2 The Lagrangian Formalism 298

Lagrangian equations of motion □ Action □ Lagrangian density □ Euler–Lagrange equations □ Reality of the action □ From Lagrangians to Hamiltonians □ Scalar fields revisited □ From Heisenberg to interaction picture □ Auxiliary fields □ Integrating by parts in the action

7.3 Global Symmetries 306

Noether's theorem □ Explicit formula for conserved quantities □ Explicit formula for conserved currents □ Quantum symmetry generators □ Energy–momentum tensor □ Momentum □ Internal symmetries □ Current commutation relations

7.4 Lorentz Invariance 314

Currents $\mathcal{M}^{\rho\mu\nu}$ □ Generators $J^{\mu\nu}$ □ Belinfante tensor □ Lorentz invariance of S-matrix

7.5 Transition to Interaction Picture: Examples 318

Scalar field with derivative coupling □ Vector field □ Dirac field

7.6 Constraints and Dirac Brackets 325

Primary and secondary constraints □ Poisson brackets □ First and second class constraints □ Dirac brackets □ Example: real vector field

7.7 Field Redefinitions and Redundant Couplings* 331

Redundant parameters □ Field redefinitions □ Example: real scalar field

Appendix Dirac Brackets from Canonical Commutators 332

Problems 337

References 338

8 ELECTRODYNAMICS 339

8.1 Gauge Invariance 339

Need for coupling to conserved current □ Charge operator □ Local symmetry □ Photon action □ Field equations □ Gauge-invariant derivatives

8.2 Constraints and Gauge Conditions 343

Primary and secondary constraints □ Constraints are first class □ Gauge fixing □ Coulomb gauge □ Solution for A^0

8.3 Quantization in Coulomb Gauge 346

Remaining constraints are second class □ Calculation of Dirac brackets in Coulomb gauge □ Construction of Hamiltonian □ Coulomb interaction

8.4 Electrodynamics in the Interaction Picture 350

Free-field and interaction Hamiltonians □ Interaction picture operators □ Normal mode decomposition

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

8.5 The Photon Propagator 353

Numerator polynomial □ Separation of non-covariant terms □ Cancellation of non-covariant terms

8.6 Feynman Rules for Spinor Electrodynamics 355

Feynman graphs □ Vertices □ External lines □ Internal lines □ Expansion in $\alpha/4\pi$ □ Circular, linear, and elliptic polarization □ Polarization and spin sums

8.7 Compton Scattering 362

S-matrix □ Differential cross-section □ Kinematics □ Spin sums □ Traces □ Klein–Nishina formula □ Polarization by Thomson scattering □ Total cross-section

8.8 Generalization : p-form Gauge Fields* 369

Motivation □ p-forms □ Exterior derivatives □ Closed and exact p-forms □ p-form gauge fields □ Dual fields and currents in D spacetime dimensions □ p-form gauge fields equivalent to $(D-p-2)$-form gauge fields □ Nothing new in four spacetime dimensions

Appendix Traces 372

Problems 374

References 375

9 PATH-INTEGRAL METHODS 376

9.1 The General Path-Integral Formula 378

Transition amplitudes for infinitesimal intervals □ Transition amplitudes for finite intervals □ Interpolating functions □ Matrix elements of time-ordered products □ Equations of motion

9.2 Transition to the S-Matrix 385

Wave function of vacuum □ $i\epsilon$ terms

9.3 Lagrangian Version of the Path-Integral Formula 389

Integrating out the 'momenta' □ Derivatively coupled scalars □ Non-linear sigma model □ Vector field

9.4 Path-Integral Derivation of Feynman Rules 395

Separation of free-field action □ Gaussian integration □ Propagators: scalar fields, vector fields, derivative coupling

9.5 Path Integrals for Fermions 399

Anticommuting c-numbers □ Eigenvectors of canonical operators □ Summing states by Berezin integration □ Changes of variables □ Transition amplitudes for infinitesimal intervals □ Transition amplitudes for finite intervals □ Derivation of Feynman rules □ Fermion propagator □ Vacuum amplitudes as determinants

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

9.6 Path-Integral Formulation of Quantum Electrodynamics 413

Path integral in Coulomb gauge □ Reintroduction of a^0 □ Transition to covariant gauges

9.7 Varieties of Statistics* 418

Preparing 'in' and 'out' states □ Composition rules □ Only bosons and fermions in ≥ 3 dimensions □ Anyons in two dimensions

Appendix Gaussian Multiple Integrals 420

Problems 423

References 423

10 NON-PERTURBATIVE METHODS 425

10.1 Symmetries 425

Translations □ Charge conservation □ Furry's theorem

10.2 Polology 428

Pole formula for general amplitudes □ Derivation of the pole formula □ Pion exchange

10.3 Field and Mass Renormalization 436

LSZ reduction formula □ Renormalized fields □ Propagator poles □ No radiative corrections in external lines □ Counterterms in self-energy parts

10.4 Renormalized Charge and Ward Identities 442

Charge operator □ Electromagnetic field renormalization □ Charge renormalization □ Ward–Takahashi identity □ Ward identity

10.5 Gauge Invariance 448

Transversality of multi-photon amplitudes □ Schwinger terms □ Gauge terms in photon propagator □ Structure of photon propagator □ Zero photon renormalized mass □ Calculation of Z_3 □ Radiative corrections to choice of gauge

10.6 Electromagnetic Form Factors and Magnetic Moment 452

Matrix elements of J^0 □ Form factors of J^μ: spin zero □ Form factors of J^μ: spin $\frac{1}{2}$ □ Magnetic moment of a spin $\frac{1}{2}$ particle □ Measuring the form factors

10.7 The Källen–Lehmann Representation* 457

Spectral functions □ Causality relations □ Spectral representation □ Asymptotic behavior of propagators □ Poles □ Bound on field renormalization constant □ $Z = 0$ for composite particles

10.8 Dispersion Relations* 462

History □ Analytic properties of massless boson forward scattering amplitude

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

□ Subtractions □ Dispersion relation □ Crossing symmetry □ Pomeranchuk's theorem □ Regge asymptotic behavior □ Photon scattering

Problems 469

References 470

11 ONE-LOOP RADIATIVE CORRECTIONS IN QUANTUM ELECTRODYNAMICS 472

11.1 Counterterms 472

Field, charge, and mass renormalization □ Lagrangian counterterms

11.2 Vacuum Polarization 473

One-loop integral for photon self-energy part □ Feynman parameters □ Wick rotation □ Dimensional regularization □ Gauge invariance □ Calculation of Z_3 □ Cancellation of divergences □ Vacuum polarization in charged particle scattering □ Uehling effect □ Muonic atoms

11.3 Anomalous Magnetic Moments and Charge Radii 485

One-loop formula for vertex function □ Calculation of form factors □ Anomalous lepton magnetic moments to order α □ Anomalous muon magnetic moment to order $\alpha^2 \ln(m_\mu/m_e)$ □ Charge radius of leptons

11.4 Electron Self-Energy 493

One-loop formula for electron self-energy part □ Electron mass renormalization □ Cancellation of ultraviolet divergences

Appendix Assorted Integrals 497

Problems 498

References 498

12 GENERAL RENORMALIZATION THEORY 499

12.1 Degrees of Divergence 500

Superficial degree of divergence □ Dimensional analysis □ Renormalizability □ Criterion for actual convergence

12.2 Cancellation of Divergences 505

Subtraction by differentiation □ Renormalization program □ Renormalizable theories □ Example: quantum electrodynamics □ Overlapping divergences □ BPHZ renormalization prescription □ Changing the renormalization point: ϕ^4 theory

12.3 Is Renormalizability Necessary? 516

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Renormalizable interactions cataloged □ No renormalizable theories of gravitation □ Cancellation of divergences in non-renormalizable theories □ Suppression of non-renormalizable interactions □ Limits on new mass scales □ Problems with higher derivatives? □ Detection of non-renormalizable interactions □ Low-energy expansions in non-renormalizable theories □ Example: scalar with only derivative coupling □ Saturation or new physics? □ Effective field theories

12.4 The Floating Cutoff* 525

Wilson's approach □ Renormalization group equation □ Polchinski's theorem □ Attraction to a stable surface □ Floating cutoff vs renormalization

12.5 Accidental Symmetries* 529

General renormalizable theory of charged leptons □ Redefinition of the lepton fields □ Accidental conservation of lepton flavors, P, C, and T

Problems 531

References 532

13 INFRARED EFFECTS 534

13.1 Soft Photon Amplitudes 534

Single photon emission □ Negligible emission from internal lines □ Lorentz invariance implies charge conservation □ Single graviton emission □ Lorentz invariance implies equivalence principle □ Multi-photon emission □ Factorization

13.2 Virtual Soft Photons 539

Effect of soft virtual photons □ Radiative corrections on internal lines

13.3 Real Soft Photons; Cancellation of Divergences 544

Sum over helicities □ Integration over energies □ Sum over photon number □ Cancellation of infrared cutoff factors □ Likewise for gravitation

13.4 General Infrared Divergences 548

Massless charged particles □ Infrared divergences in general □ Jets □ Lee–Nauenberg theorem

13.5 Soft Photon Scattering* 553

Poles in the amplitude □ Conservation relations □ Universality of the low-energy limit

13.6 The External Field Approximation* 556

Sums over photon vertex permutations □ Non-relativistic limit □ Crossed ladder exchange

Problems 562

References 562

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

14 BOUND STATES IN EXTERNAL FIELDS 564

14.1 The Dirac Equation 565

Dirac wave functions as field matrix elements □ Anticommutators and completeness □ Energy eigenstates □ Negative energy wave functions □ Orthonormalization □ 'Large' and 'small' components □ Parity □ Spin- and angle-dependence □ Radial wave equations □ Energies □ Fine structure □ Non-relativistic approximations

14.2 Radiative Corrections in External Fields 572

Electron propagator in an external field □ Inhomogeneous Dirac equation □ Effects of radiative corrections □ Energy shifts

14.3 The Lamb Shift in Light Atoms 578

Separating high and low energies □ High-energy term □ Low-energy term □ Effect of mass renormalization □ Total energy shift □ $\ell = 0$ □ $\ell \neq 0$ Numerical results □ Theory vs experiment for classic Lamb shift □ Theory vs experiment for $1s$ energy shift

Problems 594

References 595

AUTHOR INDEX 597

SUBJECT INDEX 602

OUTLINE OF VOLUME II

15 NON-ABELIAN GAUGE THEORIES

16 EXTERNAL FIELD METHODS

17 RENORMALIZATION OF GAUGE THEORIES

18 RENORMALIZATION GROUP METHODS

19 SPONTANEOUSLY BROKEN GLOBAL SYMMETRIES

20 OPERATOR PRODUCT EXPANSIONS

21 SPONTANEOUSLY BROKEN LOCAL SYMMETRIES

22 ANOMALIES

23 EXTENDED FIELD CONFIGURATIONS

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Preface To Volume I

Why another book on quantum field theory? Today the student of quantum field theory can choose from among a score of excellent books, several of them quite up-to-date. Another book will be worth while only if it offers something new in content or perspective.

As to content, although this book contains a good amount of new material, I suppose the most distinctive thing about it is its generality; I have tried throughout to discuss matters in a context that is as general as possible. This is in part because quantum field theory has found applications far removed from the scene of its old successes, quantum electrodynamics, but even more because I think that this generality will help to keep the important points from being submerged in the technicalities of specific theories. Of course, specific examples are frequently used to illustrate general points, examples that are chosen from contemporary particle physics or nuclear physics as well as from quantum electrodynamics.

It is, however, the perspective of this book, rather than its content, that provided my chief motivation in writing it. I aim to present quantum field theory in a manner that will give the reader the clearest possible idea of *why* this theory takes the form it does, and why in this form it does such a good job of describing the real world.

The traditional approach, since the first papers of Heisenberg and Pauli on general quantum field theory, has been to take the existence of fields for granted, relying for justification on our experience with electromagnetism, and 'quantize' them — that is, apply to various simple field theories the rules of canonical quantization or path integration. Some of this traditional approach will be found here in the historical introduction presented in Chapter 1. This is certainly a way of getting rapidly into the subject, but it seems to me that it leaves the reflective reader with too many unanswered questions. Why should we believe in the rules of canonical quantization or path integration? Why should we adopt the simple field equations and Lagrangians that are found in the literature? For that matter, why have fields at all? It does not seem satisfactory to me to appeal to experience; after all, our purpose in theoretical physics is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.001

not just to describe the world as we find it, but to explain — in terms of a few fundamental principles — why the world is the way it is.

The point of view of this book is that quantum field theory is the way it is because (aside from theories like string theory that have an infinite number of particle types) it is the only way to reconcile the principles of quantum mechanics (including the cluster decomposition property) with those of special relativity. This is a point of view I have held for many years, but it is also one that has become newly appropriate. We have learned in recent years to think of our successful quantum field theories, including quantum electrodynamics, as 'effective field theories,' low-energy approximations to a deeper theory that may not even be a field theory, but something different like a string theory. On this basis, the reason that quantum field theories describe physics at accessible energies is that *any* relativistic quantum theory will look at sufficiently low energy like a quantum field theory. It is therefore important to understand the rationale for quantum field theory in terms of the principles of relativity and quantum mechanics. Also, we think differently now about some of the problems of quantum field theories, such as non-renormalizability and 'triviality,' that used to bother us when we thought of these theories as truly fundamental, and the discussions here will reflect these changes. This is intended to be a book on quantum field theory for the era of effective field theories.

The most immediate and certain consequences of relativity and quantum mechanics are the properties of particle states, so here particles come first — they are introduced in Chapter 2 as ingredients in the representation of the inhomogeneous Lorentz group in the Hilbert space of quantum mechanics. Chapter 3 provides a framework for addressing the fundamental dynamical question: given a state that in the distant past looks like a certain collection of free particles, what will it look like in the future? Knowing the generator of time-translations, the Hamiltonian, we can answer this question through the perturbative expansion for the array of transition amplitudes known as the S-matrix. In Chapter 4 the principle of cluster decomposition is invoked to describe how the generator of time-translations, the Hamiltonian, is to be constructed from creation and annihilation operators. Then in Chapter 5 we return to Lorentz invariance, and show that it requires these creation and annihilation operators to be grouped together in causal quantum fields. As a spin-off, we deduce the CPT theorem and the connection between spin and statistics. The formalism is used in Chapter 6 to derive the Feynman rules for calculating the S-matrix.

It is not until Chapter 7 that we come to Lagrangians and the canonical formalism. The rationale here for introducing them is not that they have proved useful elsewhere in physics (never a very satisfying explanation)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.001

but rather that this formalism makes it easy to choose interaction Hamiltonians for which the S-matrix satisfies various assumed symmetries. In particular, the Lorentz invariance of the Lagrangian density ensures the existence of a set of ten operators that satisfy the algebra of the Poincaré group and, as we show in Chapter 3, this is the key condition that we need to prove the Lorentz invariance of the S-matrix. Quantum electrodynamics finally appears in Chapter 8. Path integration is introduced in Chapter 9, and used to justify some of the hand-waving in Chapter 8 regarding the Feynman rules for quantum electrodynamics. This is a somewhat later introduction of path integrals than is fashionable these days, but it seems to me that although path integration is by far the best way of rapidly deriving Feynman rules from a given Lagrangian, it rather obscures the quantum mechanical reasons underlying these calculations.

Volume I concludes with a series of chapters, 10–14, that provide an introduction to the calculation of radiative corrections, involving loop graphs, in general field theories. Here too the arrangement is a bit unusual; we start with a chapter on non-perturbative methods, in part because the results we obtain help us to understand the necessity for field and mass renormalization, without regard to whether the theory contains infinities or not. Chapter 11 presents the classic one-loop calculations of quantum electrodynamics, both as an opportunity to explain useful calculational techniques (Feynman parameters, Wick rotation, dimensional and Pauli–Villars regularization), and also as a concrete example of renormalization in action. The experience gained in Chapter 11 is extended to all orders and general theories in Chapter 12, which also describes the modern view of non-renormalizability that is appropriate to effective field theories. Chapter 13 is a digression on the special problems raised by massless particles of low energy or parallel momenta. The Dirac equation for an electron in an external electromagnetic field, which historically appeared almost at the very start of relativistic quantum mechanics, is not seen here until Chapter 14, on bound state problems, because this equation should not be viewed (as Dirac did) as a relativistic version of the Schrödinger equation, but rather as an approximation to a true relativistic quantum theory, the quantum field theory of photons and electrons. This chapter ends with a treatment of the Lamb shift, bringing the confrontation of theory and experiment up to date.

The reader may feel that some of the topics treated here, especially in Chapter 3, could more properly have been left to textbooks on nuclear or elementary particle physics. So they might, but in my experience these topics are usually either not covered or covered poorly, using specific dynamical models rather than the general principles of symmetry and quantum mechanics. I have met string theorists who have never heard of the relation between time-reversal invariance and final-state phase shifts,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.001

and nuclear theorists who do not understand why resonances are governed by the Breit–Wigner formula. So in the early chapters I have tried to err on the side of inclusion rather than exclusion.

Volume II will deal with the advances that have revived quantum field theory in recent years: non-Abelian gauge theories, the renormalization group, broken symmetries, anomalies, instantons, and so on.

I have tried to give citations both to the classic papers in the quantum theory of fields and to useful references on topics that are mentioned but not presented in detail in this book. I did not always know who was responsible for material presented here, and the mere absence of a citation should not be taken as a claim that the material presented here is original. But some of it is. I hope that I have improved on the original literature or standard textbook treatments in several places, as for instance in the proof that symmetry operators are either unitary or antiunitary; the discussion of superselection rules; the analysis of particle degeneracy associated with unconventional representations of inversions; the use of the cluster decomposition principle; the derivation of the reduction formula; the derivation of the external field approximation; and even the calculation of the Lamb shift.

I have also supplied problems for each chapter except the first. Some of these problems aim simply at providing exercise in the use of techniques described in the chapter; others are intended to suggest extensions of the results of the chapter to a wider class of theories.

In teaching quantum field theory, I have found that each of the two volumes of this book provides enough material for a one-year course for graduate students. I intended that this book should be accessible to students who are familiar with non-relativistic quantum mechanics and classical electrodynamics. I assume a basic knowledge of complex analysis and matrix algebra, but topics in group theory and topology are explained where they are introduced.

This is not a book for the student who wants immediately to begin calculating Feynman graphs in the standard model of weak, electromagnetic, and strong interactions. Nor is this a book for those who seek a higher level of mathematical rigor. Indeed, there are parts of this book whose lack of rigor will bring tears to the eyes of the mathematically inclined reader. Rather, I hope it will suit the physicists and physics students who want to understand *why* quantum field theory is the way it is, so that they will be ready for whatever new developments in physics may take us beyond our present understandings.

* * *

Much of the material in this book I learned from my interactions over

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.001

the years with numerous other physicists, far too many to name here. But I must acknowledge my special intellectual debt to Sidney Coleman, and to my colleagues at the University of Texas: Arno Bohm, Luis Boya, Phil Candelas, Bryce DeWitt, Cecile DeWitt-Morette, Jacques Distler, Willy Fischler, Josh Feinberg, Joaquim Gomis, Vadim Kaplunovsky, Joe Polchinski, and Paul Shapiro. I owe thanks for help in the preparation of the historical introduction to Gerry Holton, Arthur Miller, and Sam Schweber. Thanks are also due to Alyce Wilson, who prepared the illustrations and typed the LaTeX input files until I learned how to do it, and to Terry Riley for finding countless books and articles. For finding various errors in the first printing of this volume, I am greatly indebted to numerous students and colleagues, especially Stephen Adler, Hideaki Aoyama, Kevin Cahill, Amir Kashani-Poor, Achim Kupferoth, Michio Masujima, Herbert Neuberger, Fabio Siringo, and San Fu Tuan. I am grateful to Maureen Storey and Alison Woollatt of Cambridge University Press for helping to ready this book for publication, and especially to my editor, Rufus Neal, for his friendly good advice.

STEVEN WEINBERG

Austin, Texas
October, 1994

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.001

Notation

Latin indices i, j, k, and so on generally run over the three spatial coordinate labels, usually taken as 1, 2, 3.

Greek indices μ, ν, etc. generally run over the four spacetime coordinate labels 1, 2, 3, 0, with x^0 the time coordinate.

Repeated indices are generally summed, unless otherwise indicated.

The spacetime metric $\eta_{\mu\nu}$ is diagonal, with elements $\eta_{11} = \eta_{22} = \eta_{33} = 1$, $\eta_{00} = -1$.

The d'Alembertian is defined as $\Box \equiv \eta^{\mu\nu}\partial^2/\partial x^\mu \partial x^\nu = \nabla^2 - \partial^2/\partial t^2$, where ∇^2 is the Laplacian $\partial^2/\partial x^i \partial x^i$.

The 'Levi–Civita tensor' $\epsilon^{\mu\nu\rho\sigma}$ is defined as the totally antisymmetric quantity with $\epsilon^{0123} = +1$.

Spatial three-vectors are indicated by letters in boldface.

A hat over any vector indicates the corresponding unit vector: Thus, $\hat{\mathbf{v}} \equiv \mathbf{v}/|\mathbf{v}|$.

A dot over any quantity denotes the time-derivative of that quantity.

Dirac matrices γ_μ are defined so that $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\eta_{\mu\nu}$. Also, $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$, and $\beta = i\gamma^0$.

The step function $\theta(s)$ has the value $+1$ for $s > 0$ and 0 for $s < 0$.

The complex conjugate, transpose, and Hermitian adjoint of a matrix or vector A are denoted A^*, A^{T}, and $A^\dagger = A^{*\mathrm{T}}$, respectively. The Hermitian adjoint of an operator O is denoted $O^\dagger$, except where an asterisk is used to emphasize that a vector or matrix of operators is not transposed. +H.c. or +c.c. at the end of an equation indicates the addition of the Hermitian

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.002

adjoint or complex conjugate of the foregoing terms. A bar on a Dirac spinor u is defined by $\bar{u} = u^\dagger \beta$.

Except in Chapter 1, we use units with $\hbar$ and the speed of light taken to be unity. Throughout $-e$ is the rationalized charge of the electron, so that the fine structure constant is $\alpha = e^2/4\pi \simeq 1/137$.

Numbers in parenthesis at the end of quoted numerical data give the uncertainty in the last digits of the quoted figure. Where not otherwise indicated, experimental data are taken from 'Review of Particle Properties,' *Phys. Rev.* **D50**, 1173 (1994).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.002

1

Historical Introduction

Our immersion in the present state of physics makes it hard for us to understand the difficulties of physicists even a few years ago, or to profit from their experience. At the same time, a knowledge of our history is a mixed blessing — it can stand in the way of the logical reconstruction of physical theory that seems to be continually necessary.

I have tried in this book to present the quantum theory of fields in a logical manner, emphasizing the deductive trail that ascends from the physical principles of special relativity and quantum mechanics. This approach necessarily draws me away from the order in which the subject in fact developed. To take one example, it is historically correct that quantum field theory grew in part out of a study of relativistic wave equations, including the Maxwell, Klein–Gordon, and Dirac equations. For this reason it is natural that courses and treatises on quantum field theory introduce these wave equations early, and give them great weight. Nevertheless, it has long seemed to me that a much better starting point is Wigner's definition of particles as representations of the inhomogeneous Lorentz group, even though this work was not published until 1939 and did not have a great impact for many years after. In this book we start with particles and get to the wave equations later.

This is not to say that particles are necessarily more fundamental than fields. For many years after 1950 it was generally assumed that the laws of nature take the form of a quantum theory of fields. I start with particles in this book, not because they are more fundamental, but because what we know about particles is more *certain*, more directly derivable from the principles of quantum mechanics and relativity. If it turned out that some physical system could not be described by a quantum field theory, it would be a sensation; if it turned out that the system did not obey the rules of quantum mechanics and relativity, it would be a cataclysm.

In fact, lately there has been a reaction against looking at quantum field theory as fundamental. The underlying theory might not be a theory of fields *or* particles, but perhaps of something quite different, like strings.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

From this point of view, quantum electrodynamics and the other quantum field theories of which we are so proud are mere 'effective field theories,' low-energy approximations to a more fundamental theory. The reason that our field theories work so well is not that they are fundamental truths, but that any relativistic quantum theory will look like a field theory when applied to particles at sufficiently low energy. On this basis, if we want to know why quantum field theories are the way they are, we have to start with particles.

But we do not want to pay the price of altogether forgetting our past. This chapter will therefore present the history of quantum field theory from earliest times to 1949, when it finally assumed its modern form. In the remainder of the book I will try to keep history from intruding on physics.

One problem that I found in writing this chapter is that the history of quantum field theory is from the beginning inextricably entangled with the history of quantum mechanics itself. Thus, the reader who is familiar with the history of quantum mechanics may find some material here that he or she already knows, especially in the first section, where I discuss the early attempts to put together quantum mechanics with special relativity. In this case I can only suggest that the reader should skip on to the less familiar parts.

On the other hand, readers who have no prior familiarity with quantum field theory may find parts of this chapter too brief to be altogether clear. I urge such readers not to worry. This chapter is not intended as a self-contained introduction to quantum field theory, and is not needed as a basis for the rest of the book. Some readers may even prefer to start with the next chapter, and come back to the history later. However, for many readers the history of quantum field theory should serve as a good introduction to quantum field theory itself.

I should add that this chapter is not intended as an original work of historical scholarship. I have based it on books and articles by real historians, plus some historical reminiscences and original physics articles that I have read. Most of these are listed in the bibliography given at the end of this chapter, and in the list of references. The reader who wants to go more deeply into historical matters is urged to consult these listed works.

A word about notation. In order to keep some of the flavor of past times, in this chapter I will show explicit factors of $\hbar$ and c (and even h), but in order to facilitate comparison with modern physics literature, I will use the more modern *rationalized* electrostatic units for charge, so that the fine structure constant $\alpha \simeq 1/137$ is $e^2/4\pi\hbar c$. In subsequent chapters I will mostly use the 'natural' system of units, simply setting $\hbar = c = 1$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

1.1 Relativistic Wave Mechanics

Wave mechanics started out as relativistic wave mechanics. Indeed, as we shall see, the founders of wave mechanics, Louis de Broglie and Erwin Schrödinger, took a good deal of their inspiration from special relativity. It was only later that it became generally clear that relativistic wave mechanics, in the sense of a relativistic quantum theory of a fixed number of particles, is an impossibility. Thus, despite its many successes, relativistic wave mechanics was ultimately to give way to quantum field theory. Nevertheless, relativistic wave mechanics survived as an important element in the formal apparatus of quantum field theory, and it posed a challenge to field theory, to reproduce its successes.

The possibility that material particles can like photons be described in terms of waves was first suggested[1] in 1923 by Louis de Broglie. Apart from the analogy with radiation, the chief clue was Lorentz invariance: if particles are described by a wave whose phase at position $\mathbf{x}$ and time t is of the form $2\pi(\boldsymbol{\kappa} \cdot \mathbf{x} - \nu t)$, and if this phase is to be Lorentz invariant, then the vector $\boldsymbol{\kappa}$ and the frequency ν must transform like $\mathbf{x}$ and t, and hence like $\mathbf{p}$ and E. In order for this to be possible $\boldsymbol{\kappa}$ and ν must have the same velocity dependence as $\mathbf{p}$ and E, and therefore must be proportional to them, with the same constant of proportionality. For photons, one had the Einstein relation $E = h\nu$, so it was natural to assume that, for material particles,

$$\boldsymbol{\kappa} = \mathbf{p}/h \ , \qquad \nu = E/h \ , \tag{1.1.1}$$

just as for photons. The group velocity $\partial\nu/\partial\boldsymbol{\kappa}$ of the wave then turns out to equal the particle velocity, so wave packets just keep up with the particle they represent.

By assuming that any closed orbit contains an integral number of particle wavelengths $\lambda = 1/|\boldsymbol{\kappa}|$, de Broglie was able to derive the old quantization conditions of Niels Bohr and Arnold Sommerfeld, which though quite mysterious had worked well in accounting for atomic spectra. Also, both de Broglie and Walter Elsasser[2] suggested that de Broglie's wave theory could be tested by looking for interference effects in the scattering of electrons from crystals; such effects were established a few years later by Clinton Joseph Davisson and Lester H. Germer.[3] However, it was still unclear how the de Broglie relations (1.1.1) should be modified for non-free particles, as for instance for an electron in a general Coulomb field.

Wave mechanics was by-passed in the next step in the history of quantum mechanics, the development of matrix mechanics[4] by Werner Heisenberg, Max Born, Pascual Jordan and Wolfgang Pauli in the years 1925–1926. At least part of the inspiration for matrix mechanics was the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

insistence that the theory should involve only observables, such as the energy levels, or emission and absorption rates. Heisenberg's 1925 paper opens with the manifesto: 'The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities that in principle are observable.' This sort of positivism was to reemerge at various times in the history of quantum field theory, as for instance in the introduction of the S-matrix by John Wheeler and Heisenberg (see Chapter 3) and in the revival of dispersion theory in the 1950s (see Chapter 10), though modern quantum field theory is very far from this ideal. It would take us too far from our subject to describe matrix mechanics in any detail here.

As everyone knows, wave mechanics was revived by Erwin Schrödinger. In his 1926 series of papers,[5] the familiar non-relativistic wave equation is suggested first, and then used to rederive the results of matrix mechanics. Only later, in the sixth section of the fourth paper, is a relativistic wave equation offered. According to Dirac,[6] the history is actually quite different: Schrödinger first derived the relativistic equation, then became discouraged because it gave the wrong fine structure for hydrogen, and then some months later realized that the non-relativistic approximation to his relativistic equation was of value even if the relativistic equation itself was incorrect! By the time that Schrödinger came to publish his relativistic wave equation, it had already been independently rediscovered by Oskar Klein[7] and Walter Gordon,[8] and for this reason it is usually called the 'Klein–Gordon equation.'

Schrödinger's relativistic wave equation was derived by noting first that, for a 'Lorentz electron' of mass m and charge e in an external vector potential $\mathbf{A}$ and Coulomb potential ϕ, the Hamiltonian H and momentum $\mathbf{p}$ are related by*

$$0 = (H + e\phi)^2 - c^2(\mathbf{p} + e\mathbf{A}/c)^2 - m^2c^4 \,. \tag{1.1.2}$$

For a *free* particle described by a plane wave $\exp\left\{2\pi i(\boldsymbol{\kappa}\cdot\mathbf{x} - \nu t)\right\}$, the de Broglie relations (1.1.1) can be obtained by the identifications

$$\mathbf{p} = h\boldsymbol{\kappa} \to -i\hbar\nabla \;, \qquad E = h\nu \to i\hbar\frac{\partial}{\partial t} \;, \tag{1.1.3}$$

where $\hbar$ is the convenient symbol (introduced later by Dirac) for $h/2\pi$. By an admittedly formal analogy, Schrödinger guessed that an electron in the external fields $\mathbf{A}, \phi$ would be described by a wave function $\psi(\mathbf{x}, t)$ satisfying the equation obtained by making the same replacements in

* This is Lorentz invariant, because the quantities $\mathbf{A}$ and ϕ have the same Lorentz transformation property as $c\mathbf{p}$ and E. Schrödinger actually wrote H and $\mathbf{p}$ in terms of partial derivatives of an action function, but this makes no difference to our present discussion.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

(1.1.2):

$$0 = \left[\left(i\hbar\frac{\partial}{\partial t} + e\phi\right)^2 - c^2\left(-i\hbar\nabla + \frac{e\mathbf{A}}{c}\right)^2 - m^2c^4\right]\psi(\mathbf{x},t)\,. \tag{1.1.4}$$

In particular, for the stationary states of hydrogen we have $\mathbf{A} = 0$ and $\phi = e/4\pi r$, and ψ has the time-dependence $\exp(-iEt/\hbar)$, so (1.1.4) becomes

$$0 = \left[\left(E + \frac{e^2}{4\pi r}\right)^2 - c^2\hbar^2\nabla^2 - m^2c^4\right]\psi(\mathbf{x})\,. \tag{1.1.5}$$

Solutions satisfying reasonable boundary conditions can be found for the energy values[9]

$$E = mc^2\left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4}\left(\frac{n}{\ell+\frac{1}{2}} - \frac{3}{4}\right) + \cdots\right], \tag{1.1.6}$$

where $\alpha \equiv e^2/4\pi\hbar c$ is the 'fine structure constant,' roughly $1/137$; n is a positive-definite integer, and ℓ, the orbital angular momentum in units of $\hbar$, is an integer with $0 \le \ell \le n-1$. The α^2 term gave good agreement with the gross features of the hydrogen spectrum (the Lyman, Balmer, etc. series) and, according to Dirac,[6] it was this agreement that led Schrödinger eventually to develop his non-relativistic wave equation. On the other hand, the α^4 term gave a fine structure in disagreement with existing accurate measurements of Friedrich Paschen.[10]

It is instructive here to compare Schrödinger's result with that of Arnold Sommerfeld,[10a] obtained using the rules of the old quantum theory:

$$E = mc^2\left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4}\left(\frac{n}{k} - \frac{3}{4}\right) + \cdots\right]. \tag{1.1.7}$$

where m is the electron mass. Here k is an integer between 1 and n, which in Sommerfeld's theory is given in terms of the orbital angular momentum $\ell\hbar$ as $k = \ell + 1$. This gave a fine structure splitting in agreement with experiment: for instance, for $n = 2$ Eq. (1.1.7) gives two levels ($k = 1$ and $k = 2$), split by the observed amount $\alpha^4mc^2/32$, or 4.53×10^{-5} eV. In contrast, Schrödinger's result (1.1.6) gives an $n = 2$ fine structure splitting $\alpha^4mc^2/12$, considerably larger than observed.

Schrödinger correctly recognized that the source of this discrepancy was his neglect of the spin of the electron. The splitting of atomic energy levels by non-inverse-square electric fields in alkali atoms and by weak external magnetic fields (the so-called anomalous Zeeman effect) had revealed a multiplicity of states larger than could be accounted for by the Bohr–Sommerfeld theory; this led George Uhlenbeck and Samuel Goudsmit[11] in 1925 to suggest that the electron has an intrinsic angular

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

momentum $\hbar/2$. Also, the magnitude of the Zeeman splitting[12] allowed them to estimate further that the electron has a magnetic moment

$$\mu = \frac{e\hbar}{2mc} . \tag{1.1.8}$$

It was clear that the electron's spin would be coupled to its orbital angular momentum, so that Schrödinger's relativistic equation should not be expected to give the correct fine structure splitting.

Indeed, by 1927 several authors[13] had been able to show that the spin–orbit coupling was able to account for the discrepancy between Schrödinger's result (1.1.6) and experiment. There are really two effects here: one is a direct coupling between the magnetic moment (1.1.8) and the magnetic field felt by the electron as it moves through the electrostatic field of the atom; the other is the relativistic 'Thomas precession' caused (even in the absence of a magnetic moment) by the circular motion of the spinning electron.[14] Together, these two effects were found to lift the level with total angular momentum $j = \ell + \frac{1}{2}$ to the energy (1.1.7) given by Sommerfeld for $k = \ell + 1 = j + \frac{1}{2}$, while the level with $j = \ell - \frac{1}{2}$ was lowered to the value given by Sommerfeld for $k = \ell = j + \frac{1}{2}$. Thus the energy was found to depend only on n and j, but not separately on ℓ:

$$E = mc^2 \left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4} \left(\frac{n}{j+\frac{1}{2}} - \frac{3}{4} \right) + \cdots \right] . \tag{1.1.9}$$

By accident Sommerfeld's theory had given the correct magnitude of the splitting in hydrogen ($j + \frac{1}{2}$ like k runs over integer values from 1 to n) though it was wrong as to the assignment of orbital angular momentum values ℓ to these various levels. In addition, the multiplicity of the fine structure levels in hydrogen was now predicted to be 2 for $j = \frac{1}{2}$ and $2(2j + 1)$ for $j > \frac{1}{2}$ (corresponding to ℓ values $j \pm \frac{1}{2}$), in agreement with experiment.

Despite these successes, there still was not a thorough relativistic theory which incorporated the electron's spin from the beginning. Such a theory was discovered in 1928 by Paul Dirac. However, he did not set out simply to make a relativistic theory of the spinning electron; instead, he approached the problem by posing a question that would today seem very strange. At the beginning of his 1928 paper,[15] he asks 'why Nature should have chosen this particular model for the electron, instead of being satisfied with the point charge.' To us today, this question is like asking why bacteria have only one cell: having spin $\hbar/2$ is just one of the properties that define a particle as an electron, rather than one of the many other types of particles with various spins that are known today. However, in 1928 it was possible to believe that all matter consisted of electrons, and perhaps something similar with positive charge in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

atomic nucleus. Thus, in the spirit of the times in which it was asked, Dirac's question can be restated: 'Why do the fundamental constituents of matter have to have spin $\hbar/2$?'

For Dirac, the key to this question was the requirement that probabilities must be positive. It was known[16] that the probability density for the non-relativistic Schrödinger equation is $|\psi|^2$, and that this satisfies a continuity equation of the form

$$\frac{\partial}{\partial t}(|\psi|^2) - \frac{i\hbar}{2m}\nabla \cdot (\psi^*\nabla\psi - \psi\nabla\psi^*) = 0$$

so the space-integral of $|\psi|^2$ is time-independent. On the other hand, the only probability density ρ and current $\mathbf{J}$, which can be formed from solutions of the relativistic Schrödinger equation and which satisfy a conservation law,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0\,, \tag{1.1.10}$$

are of the form

$$\rho = N \operatorname{Im} \psi^* \left(\frac{\partial}{\partial t} - \frac{ie\phi}{\hbar}\right) \psi\,, \tag{1.1.11}$$

$$\mathbf{J} = N c^2 \operatorname{Im} \psi^* \left(\nabla + \frac{ie\mathbf{A}}{\hbar c}\right) \psi\,, \tag{1.1.12}$$

with N an arbitrary constant. It is not possible to identify ρ as the probability density, because (with or without an external potential ϕ) ρ does not have definite sign. To quote Dirac's reminiscences[17] about this problem

> I remember once when I was in Copenhagen, that Bohr asked me what I was working on and I told him I was trying to get a satisfactory relativistic theory of the electron, and Bohr said 'But Klein and Gordon have already done that!' That answer first rather disturbed me. Bohr seemed quite satisfied by Klein's solution, but I was not because of the negative probabilities that it led to. I just kept on with it, worrying about getting a theory which would have only positive probabilities.

According to George Gamow,[18] Dirac found the answer to this problem on an evening in 1928 while staring into a fireplace at St John's College, Cambridge. He realized that the reason that the Klein–Gordon (or relativistic Schrödinger) equation can give negative probabilities is that the ρ in the conservation equation (1.1.10) involves a time-derivative of the wave function. This in turn happens because the wave function satisfies a differential equation of *second* order in the time. The problem therefore

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

was to replace this wave equation with another one of first order in time derivatives, like the non-relativistic Schrödinger equation.

Suppose the electron wave function is a multi-component quantity $\psi_n(x)$, which satisfies a wave equation of the form,

$$i\hbar \frac{\partial \psi}{\partial t} = \mathscr{H} \psi \,, \tag{1.1.13}$$

where $\mathscr{H}$ is some matrix function of space derivatives. In order to have a chance at a Lorentz-invariant theory, we must suppose that because the equation is linear in time-derivatives, it is also linear in space-derivatives, so that $\mathscr{H}$ takes the form:

$$\mathscr{H} = -i\hbar c \boldsymbol{\alpha} \cdot \nabla + \alpha_4 mc^2 \,, \tag{1.1.14}$$

where α_1, α_2, α_3, and α_4 are constant matrices. From (1.1.13) we can derive the second-order equation

$$-\hbar^2 \frac{\partial^2 \psi}{\partial t^2} = \mathscr{H}^2 \psi = -\hbar^2 c^2 \alpha_i \alpha_j \frac{\partial^2 \psi}{\partial x_i \partial x_j} - i\hbar mc^3 (\alpha_i \alpha_4 + \alpha_4 \alpha_i) \frac{\partial \psi}{\partial x_i} + m^2 c^4 \alpha_4^2 \psi \,.$$

(The summation convention is in force here; i and j run over the values 1, 2, 3, or x, y, z.) But this must agree with the free-field form of the relativistic Schrödinger equation (1.1.4), which just expresses the relativistic relation between momentum and energy. Therefore, the matrices $\boldsymbol{\alpha}$ and α_4 must satisfy the relations

$$\alpha_i \alpha_j + \alpha_j \alpha_i = 2\delta_{ij} 1 \,, \tag{1.1.15}$$

$$\alpha_i \alpha_4 + \alpha_4 \alpha_i = 0 \,, \tag{1.1.16}$$

$$\alpha_4^2 = 1 \,, \tag{1.1.17}$$

where δ_{ij} is the Kronecker delta (unity for $i = j$; zero for $i \neq j$) and 1 is the unit matrix. Dirac found a set of 4×4 matrices which satisfy these relations

$$\alpha_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \alpha_2 = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix},$$

$$\alpha_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \qquad \alpha_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{1.1.18}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

To show that this formalism is Lorentz-invariant, Dirac multiplied Eq. (1.1.13) on the left with α_4, so that it could be put in the form

$$\left[\hbar c\gamma^\mu \frac{\partial}{\partial x^\mu} + mc^2\right]\psi = 0\,, \tag{1.1.19}$$

where

$$\boldsymbol{\gamma} \equiv -i\alpha_4\boldsymbol{\alpha}\,, \qquad \gamma^0 \equiv -i\alpha_4\,. \tag{1.1.20}$$

(The Greek indices μ, ν, etc. will now run over the values 1, 2, 3, 0, with $x^0 = ct$. Dirac used $x_4 = ict$, and correspondingly $\gamma_4 = \alpha_4$.) The matrices γ^μ satisfy the anticommutation relations

$$\frac{1}{2}(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu) = \eta^{\mu\nu} \equiv \begin{cases} +1 & \mu = \nu = 1, 2, 3 \\ -1 & \mu = \nu = 0 \\ 0 & \mu \neq \nu \end{cases}\,. \tag{1.1.21}$$

Dirac noted that these anticommutation relations are Lorentz-invariant, in the sense that they are also satisfied by the matrices $\Lambda^\mu{}_\nu\gamma^\nu$, where Λ is any Lorentz transformation. He concluded from this that $\Lambda^\mu{}_\nu\gamma^\nu$ must be related to γ^μ by a similarity transformation:

$$\Lambda^\mu{}_\nu\gamma^\nu = S^{-1}(\Lambda)\gamma^\mu S(\Lambda)\,.$$

It follows that the wave equation is invariant if, under a Lorentz transformation $x^\mu \to \Lambda^\mu{}_\nu x^\nu$, the wave function undergoes the matrix transformation $\psi \to S(\Lambda)\psi$. (These matters are discussed more fully, from a rather different point of view, in Chapter 5.)

To study the behavior of electrons in an arbitrary external electromagnetic field, Dirac followed the 'usual procedure' of making the replacements

$$i\hbar\frac{\partial}{\partial t} \to i\hbar\frac{\partial}{\partial t} + e\phi \qquad -i\hbar\nabla \to -i\hbar\nabla + \frac{e}{c}\mathbf{A} \tag{1.1.22}$$

as in Eq. (1.1.4). The wave equation (1.1.13) then takes the form

$$\left(i\hbar\frac{\partial}{\partial t} + e\phi\right)\psi = (-i\hbar c\nabla + e\mathbf{A})\cdot\boldsymbol{\alpha}\psi + mc^2\alpha_4\psi\,. \tag{1.1.23}$$

Dirac used this equation to show that in a central field, the conservation of angular momentum takes the form

$$[\mathscr{H},\ -i\hbar\mathbf{r}\times\nabla + \hbar\boldsymbol{\sigma}/2] = 0\,, \tag{1.1.24}$$

where $\mathscr{H}$ is the matrix differential operator (1.1.14) and $\boldsymbol{\sigma}$ is the 4×4 version of the spin matrix introduced earlier by Pauli[19]

$$\boldsymbol{\sigma} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}\boldsymbol{\alpha}\,. \tag{1.1.25}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

Since each component of $\boldsymbol{\sigma}$ has eigenvalues ± 1, the presence of the extra term in (1.1.24) shows that the electron has intrinsic angular momentum $\hbar/2$.

Dirac also iterated Eq. (1.1.23), obtaining a second-order equation, which turned out to have just the same form as the Klein–Gordon equation (1.1.4) except for the presence on the right-hand-side of two additional terms

$$[-e\hbar c\boldsymbol{\sigma}\cdot\mathbf{B} - ie\hbar c\boldsymbol{\alpha}\cdot\mathbf{E}]\,\psi\,. \tag{1.1.26}$$

For a slowly moving electron, the first term dominates, and represents a magnetic moment in agreement with the value (1.1.8) found by Goudsmit and Uhlenbeck.[11] As Dirac recognized, this magnetic moment, together with the relativistic nature of the theory, guaranteed that this theory would give a fine structure splitting in agreement (to order $\alpha^4 mc^2$) with that found by Heisenberg, Jordan, and Charles G. Darwin.[13] A little later, an 'exact' formula for the hydrogen energy levels in Dirac's theory was derived by Darwin[20] and Gordon[21]

$$E = mc^2\left(1+\frac{\alpha^2}{\left\{n-j-\frac{1}{2}+\left[\left(j+\frac{1}{2}\right)^2-\alpha^2\right]^{\frac{1}{2}}\right\}^2}\right)^{-1/2}\,. \tag{1.1.27}$$

The first three terms of a power series expansion in α^2 agree with the approximate result (1.1.9).

This theory achieved Dirac's primary aim: a relativistic formalism with positive probabilities. From (1.1.13) we can derive a continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0 \tag{1.1.28}$$

with

$$\rho = |\psi|^2\,, \qquad \mathbf{J} = c\psi^\dagger\boldsymbol{\alpha}\psi\,, \tag{1.1.29}$$

so that the positive quantity $|\psi|^2$ can be interpreted as a probability density, with constant total probability $\int|\psi|^2 d^3x$. However, there was another difficulty which Dirac was not immediately able to resolve.

For a given momentum $\mathbf{p}$, the wave equation (1.1.13) has *four* solutions of the plane wave form

$$\psi \propto \exp\left[\frac{i}{\hbar}(\mathbf{p}\cdot\mathbf{x} - Et)\right]\,. \tag{1.1.30}$$

Two solutions with $E = +\sqrt{\mathbf{p}^2c^2+m^2c^4}$ correspond to the two spin states of an electron with $J_z = \pm\hbar/2$. The other two solutions have $E =$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

$-\sqrt{\mathbf{p}^2c^2+m^2c^4}$, and no obvious physical interpretation. As Dirac pointed out, this problem arises also for the relativistic Schrödinger equation: for each $\mathbf{p}$, there are two solutions of the form (1.1.30), one with positive E and one with negative E.

Of course, even in classical physics, the relativistic relation $E^2 = \mathbf{p}^2c^2 + m^2c^4$ has two solutions, $E = \pm\sqrt{\mathbf{p}^2c^2+m^2c^4}$. However, in classical physics we can simply assume that the only physical particles are those with positive E. Since the positive solutions have $E > mc^2$ and the negative ones have $E < -mc^2$, there is a finite gap between them, and no continuous process can take a particle from positive to negative energy.

The problem of negative energies is much more troublesome in relativistic quantum mechanics. As Dirac pointed out in his 1928 paper,[15] the interaction of electrons with radiation can produce transitions in which a positive-energy electron falls into a negative-energy state, with the energy carried off by two or more photons. Why then is matter stable?

In 1930 Dirac offered a remarkable solution.[22] Dirac's proposal was based on the exclusion principle, so a few words about the history of this principle are in order here.

The periodic table of the elements and the systematics of X-ray spectroscopy had together by 1924 revealed a pattern in the population of atomic energy levels by electrons:[23] The maximum number N_n of electrons in a shell characterized by principal quantum number n is given by twice the number of orbital states with that n

$$N_n = 2\sum_{\ell=0}^{n-1}(2\ell+1) = 2n^2 = 2,\ 8,\ 18,\ldots\ . \tag{1.1.31}$$

Wolfgang Pauli[24] in 1925 suggested that this pattern could be understood if N_n is the total number of possible states in the nth shell, and if in addition there is some mysterious 'exclusion principle' which forbids more than one electron from occupying the same state. He explained the puzzling factor 2 in (1.1.31) as due to a 'peculiar, classically non-describable duplexity' of the electron states, and as we have seen this was understood a little later as due to the spin of the electron.[11] The exclusion principle answered a question that had remained obscure in the old atomic theory of Bohr and Sommerfeld: why do not all the electrons in heavy atoms fall down into the shell of lowest energy? Subsequently Pauli's exclusion principle was formalized by a number of authors[25] as the requirement that the wave function of a multi-electron system is antisymmetric in the coordinates, orbital and spin, of all the electrons. This principle was incorporated into statistical mechanics by Enrico Fermi[26] and Dirac,[27] and for this reason particles obeying the exclusion principle are generally called 'fermions,'

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

just as particles like photons for which the wave function is symmetric and which obey the statistics of Bose and Einstein are called 'bosons.' The exclusion principle has played a fundamental role in the theory of metals, white dwarf and neutron stars, etc., as well as in chemistry and atomic physics, but a discussion of these matters would take us too far afield here.

Dirac's proposal was that the positive energy electrons cannot fall down into negative energy states because 'all the states of negative energy are occupied except perhaps a few of small velocity.' The few vacant states, or 'holes,' in the sea of negative energy electrons behave like particles with opposite quantum numbers: positive energy and positive charge. The only particle with positive charge that was known at that time was the proton, and as Dirac later recalled,[27a] 'the whole climate of opinion at that time was against new particles' so Dirac identified his holes as protons; in fact, the title of his 1930 article[22] was 'A Theory of Electrons and Protons.'

The hole theory faced a number of immediate difficulties. One obvious problem was raised by the infinite charge density of the ubiquitous negative-energy electrons: where is their electric field? Dirac proposed to reinterpret the charge density appearing in Maxwell's equations as 'the departure from the normal state of electrification of the world.' Another problem has to do with the huge dissimilarity between the observed masses and interactions of the electrons and protons. Dirac hoped that Coulomb interactions between electrons would somehow account for these differences but Hermann Weyl[28] showed that the hole theory was in fact entirely symmetric between negative and positive charge. Finally, Dirac[22] predicted the existence of an electron–proton annihilation process in which a positive-energy electron meets a hole in the sea of negative-energy electrons and falls down into the unoccupied level, emitting a pair of gamma ray photons. By itself this would not have created difficulties for the hole theory; it was even hoped by some that this would provide an explanation, then lacking, of the energy source of the stars. However, it was soon pointed out[29] by Julius Robert Oppenheimer and Igor Tamm that electron–proton annihilation in atoms would take place at much too fast a rate to be consistent with the observed stability of ordinary matter. For these reasons, by 1931 Dirac had changed his mind, and decided that the holes would have to appear not as protons but as a new sort of positively charged particle, of the same mass as the electron.[29a]

The second and third of these problems were eliminated by the discovery of the positron by Carl D. Anderson,[30] who apparently did not know of this prediction by Dirac. On August 2, 1932, a peculiar cosmic ray track was observed in a Wilson cloud chamber subjected to a 15 kG magnetic field. The track was observed to curve in a direction that would be expected for a *positively* charged particle, and yet its range was at least

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

ten times greater than the expected range of a proton! Both the range and the specific ionization of the track were consistent with the hypothesis that this was a new particle which differs from the electron only in the sign of its charge, as would be expected for one of Dirac's holes. (This discovery had been made earlier by P.M.S. Blackett, but not immediately published by him. Anderson quotes press reports of evidence for light positive particles in cosmic ray tracks, obtained by Blackett and Giuseppe Occhialini.) Thus it appeared that Dirac was wrong only in his original identification of the hole with the proton.

The discovery of the more-or-less predicted positron, together with the earlier successes of the Dirac equation in accounting for the magnetic moment of the electron and the fine structure of hydrogen, gave Dirac's theory a prestige that it has held for over six decades. However, although there seems little doubt that Dirac's theory will survive in some form in any future physical theory, there are serious reasons for being dissatisfied with its original rationale:

(i) Dirac's analysis of the problem of negative probabilities in Schrödinger's relativistic wave equation would seem to rule out the existence of any particle of zero spin. Yet even in the 1920s particles of zero spin were known — for instance, the hydrogen atom in its ground state, and the helium nucleus. Of course, it could be argued that hydrogen atoms and alpha particles are not elementary, and therefore do not need to be described by a relativistic wave equation, but it was not (and still is not) clear how the idea of elementarity is incorporated in the formalism of relativistic quantum mechanics. Today we know of a large number of spin zero particles — π mesons, K mesons, and so on — that are no less elementary than the proton and neutron. We also know of spin one particles — the $W^{\pm}$ and Z^0 — which seem as elementary as the electron or any other particle. Further, apart from effects of the strong interactions, we would today calculate the fine structure of 'mesonic atoms,' consisting of a spinless negative π or K meson bound to an atomic nucleus, from the stationary solutions of the relativistic Klein–Gordon–Schrödinger equation! Thus, it is difficult to agree that there is anything fundamentally wrong with the relativistic equation for zero spin that *forced* the development of the Dirac equation — the problem simply is that the electron happens to have spin $\hbar/2$, not zero.

(ii) As far as we now know, for *every* kind of particle there is an 'antiparticle' with the same mass and opposite charge. (Some purely neutral particles, such as the photon, are their own antiparticles.) But how can we interpret the antiparticles of charged *bosons*, such as the $\pi^{\pm}$ mesons or $W^{\pm}$ particles, as holes in a sea of negative energy states? For particles quantized according to the rules of Bose–Einstein statistics,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

there is no exclusion principle, and hence nothing to keep positive-energy particles from falling down into the negative-energy states, occupied or not. And if the hole theory does not work for bosonic antiparticles, why should we believe it for fermions? I asked Dirac in 1972 how he then felt about this point; he told me that he did not regard bosons like the pion or $W^{\pm}$ as 'important.' In a lecture[27a] a few years later, Dirac referred to the fact that for bosons 'we no longer have the picture of a vacuum with negative energy states filled up', and remarked that in this case 'the whole theory becomes more complicated.' The next section will show how the development of quantum field theory made the interpretation of antiparticles as holes unnecessary, even though unfortunately it lingers on in many textbooks. To quote Julian Schwinger,[30a] 'The picture of an infinite sea of negative energy electrons is now best regarded as a historical curiosity, and forgotten.'

(iii) One of the great successes of the Dirac theory was its correct prediction of the magnetic moment of the electron. This was particularly striking, as the magnetic moment (1.1.8) is twice as large as would be expected for the orbital motion of a charged point particle with angular momentum $\hbar/2$; this factor of 2 had remained mysterious until Dirac's theory. However, there is really nothing in Dirac's line of argument that leads unequivocally to this particular value for the magnetic moment. At the point where we brought electric and magnetic fields into the wave equation (1.1.23), we could just as well have added a 'Pauli term'[31]

$$\kappa\alpha_4[\gamma^\mu, \gamma^\nu]\psi F_{\mu\nu} \tag{1.1.32}$$

with arbitrary coefficient κ. (Here $F_{\mu\nu}$ is the usual electromagnetic field strength tensor, with $F^{12} = B_3$, $F^{01} = E_1$, etc.) This term could be obtained by first adding a term to the free-field equations proportional to $[\gamma^\mu, \gamma^\nu](\partial^2/\partial x^\mu \partial x^\nu)\psi$, which of course equals zero, and then making the substitutions (1.1.22) as before. A more modern approach would be simply to remark that the term (1.1.32) is consistent with all accepted invariance principles, including Lorentz invariance and gauge invariance, and so there is no reason why such a term should *not* be included in the field equations. (See Section 12.3.) This term would give an additional contribution proportional to κ to the magnetic moment of the electron, so apart from the possible demand for a purely formal simplicity, there was no reason to expect any particular value for the magnetic moment of the electron in Dirac's theory.

As we shall see in this book, these problems were all eventually to be solved (or at least clarified) through the development of quantum field theory.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

1.2 The Birth of Quantum Field Theory

The photon is the only particle that was known as a field before it was detected as a particle. Thus it is natural that the formalism of quantum field theory should have been developed in the first instance in connection with radiation and only later applied to other particles and fields.

In 1926, in one of the central papers on matrix mechanics, Born, Heisenberg, and Jordan[32] applied their new methods to the free radiation field. For simplicity, they ignored the polarization of electromagnetic waves and worked in one space dimension, with coordinate x running from 0 to L; the radiation field $u(x,t)$ if constrained to vanish at these endpoints thus has the same behavior as the displacement of a string with ends fixed at $x = 0$ and $x = L$. By analogy with either the case of a string or the full electromagnetic field, the Hamiltonian was taken to have the form

$$H = \frac{1}{2}\int_0^L \left\{ \left(\frac{\partial u}{\partial t}\right)^2 + c^2 \left(\frac{\partial u}{\partial x}\right)^2 \right\} dx \,. \tag{1.2.1}$$

In order to reduce this expression to a sum of squares, the field u was expressed as a sum of Fourier components with $u = 0$ at both $x = 0$ and $x = L$:

$$u(x,t) = \sum_{k=1}^{\infty} q_k(t) \sin\left(\frac{\omega_k x}{c}\right) , \tag{1.2.2}$$

$$\omega_k \equiv k\pi c/L \,, \tag{1.2.3}$$

so that

$$H = \frac{L}{4}\sum_{k=1}^{\infty} \left\{ \dot{q}_k^2(t) + \omega_k^2 q_k^2(t) \right\} . \tag{1.2.4}$$

Thus the string or field behaves like a sum of independent harmonic oscillators with angular frequencies ω_k, as had been anticipated 20 years earlier by Paul Ehrenfest.[32a]

In particular, the 'momentum' $p_k(t)$ canonically conjugate to $q_k(t)$ is determined, as in particle mechanics, by the condition that if H is expressed as a function of the ps and qs, then

$$\dot{q}_k(t) = \frac{\partial}{\partial p_k(t)} H(p(t), q(t)) \,.$$

This yields a 'momentum'

$$p_k(t) = \frac{L}{2}\dot{q}_k(t) \tag{1.2.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

so the canonical commutation relations may be written

$$\left[\dot{q}_k(t), q_j(t)\right] = \frac{2}{L}\left[p_k(t), q_j(t)\right] = \frac{-2i\hbar}{L}\,\delta_{kj}\,, \tag{1.2.6}$$

$$\left[q_k(t), q_j(t)\right] = 0\,. \tag{1.2.7}$$

Also, the time-dependence of $q_k(t)$ is governed by the Hamiltonian equation of motion

$$\ddot{q}_k(t) = \frac{2}{L}\,\dot{p}_k(t) = -\frac{2}{L}\,\frac{\partial H}{\partial q_k(t)} = -\omega_k^2 q_k(t)\,. \tag{1.2.8}$$

The form of the matrices defined by Eqs. (1.2.6)–(1.2.8) was already known to Born, Heisenberg, and Jordan through previous work on the harmonic oscillator. The q-matrix is given by

$$q_k(t) = \sqrt{\frac{\hbar}{L\omega_k}}\left[a_k \exp(-i\omega_k t) + a_k^\dagger \exp(+i\omega_k t)\right] \tag{1.2.9}$$

with a_k a time-independent matrix and $a_k^\dagger$ its Hermitian adjoint, satisfying the commutation relations

$$\left[\, a_k\ ,\ a_j^\dagger \,\right] = \delta_{kj}\,, \tag{1.2.10}$$

$$\left[\, a_k\ ,\ a_j \,\right] = 0\,. \tag{1.2.11}$$

The rows and columns of these matrices are labelled with a set of positive integers $n_1, n_2, \ldots$, one for each normal mode. The matrix elements are

$$(a_k)_{n_1', n_2', \ldots, n_1, n_2 \ldots} = \sqrt{n_k}\,\delta_{n_k', n_k - 1} \prod_{j \neq k} \delta_{n_j' n_j}\,, \tag{1.2.12}$$

$$(a_k^\dagger)_{n_1', n_2', \ldots, n_1, n_2 \ldots} = \sqrt{n_k + 1}\,\delta_{n_k', n_k + 1} \prod_{j \neq k} \delta_{n_j' n_j}\,. \tag{1.2.13}$$

For a single normal mode, these matrices may be written explicitly as

$$a = \begin{bmatrix} 0 & \sqrt{1} & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & \cdots \\ 0 & 0 & 0 & \sqrt{3} & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}, \quad a^\dagger = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ \sqrt{1} & 0 & 0 & 0 & \cdots \\ 0 & \sqrt{2} & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}.$$

It is straightforward to check that (1.2.12) and (1.2.13) do satisfy the commutation relations (1.2.10) and (1.2.11).

The physical interpretation of a column vector with integer components $n_1, n_2, \ldots$ is that it represents a state with n_k quanta in each normal mode k. The matrix a_k or $a_k^\dagger$ acting on such a column vector will respectively

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

lower or raise n_k by one unit, leaving all n_ℓ with $\ell \neq k$ unchanged; they may therefore be interpreted as operators which annihilate or create one quantum in the kth normal mode. In particular, the vector with all n_k equal to zero represents the vacuum; it is annihilated by any a_k.

This interpretation is further borne out by inspection of the Hamiltonian. Using (1.2.9) and (1.2.10) in (1.2.4) gives

$$H = \sum_k \hbar\omega_k \left(a_k^\dagger a_k + \tfrac{1}{2}\right) . \tag{1.2.14}$$

The Hamiltonian is then diagonal in the n-representation

$$(H)_{n_1',n_2',\ldots,n_1,n_2\ldots} = \sum_k \hbar\omega_k \left(n_k + \tfrac{1}{2}\right) \prod_j \delta_{n_j' n_j} . \tag{1.2.15}$$

We see that the energy of the state is just the sum of energies $\hbar\omega_k$ for each quantum present in the state, plus an infinite zero-point energy $E_0 = \frac{1}{2}\sum_k \hbar\omega_k$. Applied to the radiation field, this formalism justified the Bose method of counting radiation states according to the numbers n_k of quanta in each normal mode.

Born, Heisenberg, and Jordan used this formalism to derive an expression for the r.m.s. energy fluctuations in black-body radiation. (For this purpose they actually only used the commutation relations (1.2.6)–(1.2.7).) However, this approach was soon applied to a more urgent problem, the calculation of the rates for spontaneous emission of radiation.

In order to appreciate the difficulties here, it is necessary to go back in time a bit. In one of the first papers on matrix mechanics, Born and Jordan[33] had assumed in effect that an atom, in dropping from a state β to a lower state α, would emit radiation just like a classical charged oscillator with displacement

$$\mathbf{r}(t) = \mathbf{r}_{\beta\alpha} \exp(-2\pi i\nu t) + \mathbf{r}_{\beta\alpha}{}^* \exp(2\pi i\nu t) , \tag{1.2.16}$$

where

$$h\nu = E_\beta - E_\alpha \tag{1.2.17}$$

and $\mathbf{r}_{\beta\alpha}$ is the β,α element of the matrix associated with the electron position. The energy E of such an oscillator is

$$E = \frac{1}{2}m\left(\dot{\mathbf{r}}^2 + (2\pi\nu)^2\mathbf{r}^2\right) = 8\pi^2 m\nu^2 |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.18}$$

A straightforward classical calculation then gives the radiated power, and dividing by the energy $h\nu$ per photon gives the rate of photon emission

$$A(\beta \to \alpha) = \frac{16\pi^3 e^2 \nu^3}{3hc^3} |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

However, it was not at all clear why the formulas for emission of radiation by a classical dipole should be taken over in this manner in dealing with spontaneous emission.

A little later a more convincing though even less direct derivation was given by Dirac.[34] By considering the behavior of quantized atomic states in an oscillating *classical* electromagnetic field with energy density per frequency interval u at frequency (1.2.17), he was able to derive formulas for the rates $uB(\alpha \to \beta)$ and $uB(\beta \to \alpha)$ for absorption or induced emission:

$$B(\alpha \to \beta) = B(\beta \to \alpha) \simeq \frac{2\pi^2 e^2}{3h^2} |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.20}$$

(Note that the expression on the right is symmetric between states α and β, because $\mathbf{r}_{\alpha\beta}$ is just $\mathbf{r}_{\beta\alpha}{}^*$.) Einstein[34a] had already shown in 1917 that the possibility of thermal equilibrium between atoms and black-body radiation imposes a relation between the rate $A(\beta \to \alpha)$ of spontaneous emission and the rates uB for induced emission or absorption:

$$A(\beta \to \alpha) = \left(\frac{8\pi h\nu^3}{c^3}\right) B(\beta \to \alpha) . \tag{1.2.21}$$

Using (1.2.20) in this relation immediately yields the Born–Jordan result (1.2.19) for the rate of spontaneous emission. Nevertheless, it still seemed unsatisfactory that thermodynamic arguments should be needed to derive formulas for processes involving a single atom.

Finally, in 1927 Dirac[35] was able to give a thoroughly quantum mechanical treatment of spontaneous emission. The vector potential $\mathbf{A}(\mathbf{x}, t)$ was expanded in normal modes, as in Eq. (1.2.2), and the coefficients were shown to satisfy commutation relations like (1.2.6). In consequence, each state of the free radiation field was specified by a set of integers n_k, one for each normal mode, and the matrix elements of the electromagnetic interaction $e\dot{\mathbf{r}} \cdot \mathbf{A}$ took the form of a sum over normal modes, with matrix coefficients proportional to the matrices a_k and $a_k^\dagger$ defined in Eqs. (1.2.10)–(1.2.13). The crucial result here is the factor $\sqrt{n_k + 1}$ in Eq. (1.2.13); the probability for a transition in which the number of photons in a normal mode k rises from n_k to $n_k + 1$ is proportional to the square of this factor, or $n_k + 1$. But in a radiation field with n_k photons in a normal mode k, the energy density u per frequency interval is

$$u(\nu_k) = \left(\frac{8\pi\nu_k^2}{c^3}\right) n_k \times h\nu_k ,$$

so the rate for emission of radiation in normal mode k is proportional to

$$n_k + 1 = \frac{c^3 u(\nu_k)}{8\pi h\nu_k^3} + 1 .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

The first term is interpreted as the contribution of induced emission, and the second term as the contribution of spontaneous emission. Hence, without any appeal to thermodynamics, Dirac could conclude that the ratio of the rates uB for induced emission and A for spontaneous emission is given by the Einstein relation, Eq. (1.2.21). Using his earlier result (1.2.20) for B, Dirac was thus able to rederive the Born–Jordan formula[33] (1.2.19) for spontaneous emission rate A. A little later, similar methods were used by Dirac to give a quantum mechanical treatment of the scattering of radiation and the lifetime of excited atomic states,[36] and by Victor Weisskopf and Eugene Wigner to make a detailed study of spectral line shapes.[36a] Dirac in his work was separating the electromagnetic potential into a radiation field $\mathbf{A}$ and a static Coulomb potential A^0, in a manner which did not preserve the manifest Lorentz and gauge invariance of classical electrodynamics. These matters were put on a firmer foundation a little later by Enrico Fermi.[36b] Many physicists in the 1930s learned their quantum electrodynamics from Fermi's 1932 review.

The use of canonical commutation relations for q and p or a and $a^\dagger$ also raised a question as to the Lorentz invariance of the quantized theory. Jordan and Pauli[37] in 1928 were able to show that the commutators of fields at different spacetime points were in fact Lorentz-invariant. (These commutators are calculated in Chapter 5.) Somewhat later, Bohr and Leon Rosenfeld[38] used a number of ingenious thought experiments to show that these commutation relations express limitations on our ability to measure fields at spacetime points separated by time-like intervals.

It was not long after the successful quantization of the electromagnetic field that these techniques were applied to other fields. At first this was regarded as a 'second quantization'; the fields to be quantized were the wave functions used in one-particle quantum mechanics, such as the Dirac wave function of the electron. The first step in this direction seems to have been taken in 1927 by Jordan.[39] In 1928 an essential element was supplied by Jordan and Wigner.[40] They recognized that the Pauli exclusion principle prevents the occupation number n_k of electrons in any normal mode k (counting spin as well as position variables) from taking any values other than 0 or 1. The electron field therefore cannot be expanded as a superposition of operators satisfying the commutation relations (1.2.10), (1.2.11), because these relations require n_k to take all integer values from 0 to ∞. Instead, they proposed that the electron field should be expanded in a sum of operators a_k, $a_k^\dagger$ satisfying the *anticommutation* relations

$$a_k\, a_j^\dagger + a_j^\dagger a_k = \delta_{jk}\,, \tag{1.2.22}$$

$$a_k\, a_j + a_j a_k = 0\,. \tag{1.2.23}$$

The relations can be satisfied by matrices labelled by a set of integers

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

$n_1, n_2, \cdots$, one for each normal mode, each integer taking just the values zero and one:

$$(a_k)_{n_1',n_2',\ldots,n_1,n_2,\ldots} = \begin{cases} 1 & n_k' = 0,\ n_k = 1,\ n_j' = n_j \text{ for } j \neq k \\ 0 & \text{otherwise}, \end{cases} \tag{1.2.24}$$

$$(a_k^\dagger)_{n_1',n_2',\ldots,n_1,n_2,\ldots} = \begin{cases} 1 & n_k' = 1,\ n_k = 0,\ n_j' = n_j \text{ for } j \neq k \\ 0 & \text{otherwise}. \end{cases} \tag{1.2.25}$$

For instance, for a single normal mode we have just two rows and two columns, corresponding to the values unity and zero of n' and n; the a and $a^\dagger$ matrices take the form

$$a = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad a^\dagger = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The reader may check that (1.2.24) and (1.2.25) do satisfy the anticommutation relations (1.2.22) and (1.2.23).

The interpretation of a column vector characterized by integers $n_1, n_2, \ldots$ is that it represents a state with n_k quanta in each normal mode k, just as for bosons. The difference is, of course, that since each n_k takes only the values 0 and 1, there can be at most one quantum in each normal mode, as required by the Pauli exclusion principle. Again, a_k destroys a quantum in normal mode k if there is one there already, and otherwise gives zero; also, $a_k^\dagger$ creates a quantum in normal mode k *unless* there is one there already, in which case it gives zero. Much later it was shown by Fierz and Pauli[40a] that the choice between commutation and anticommutation relations is dictated solely by the particle's spin: commutators must be used for particles with integer spin like the photon, and anticommutators for particles with half-integer spin like the electron. (This will be shown in a different way in Chapter 5.)

The theory of general quantum fields was first laid out in 1929, in a pair of comprehensive articles by Heisenberg and Pauli.[41] The starting point of their work was the application of the canonical formalism to the fields themselves, rather than to the coefficients of the normal modes appearing in the fields. Heisenberg and Pauli took the Lagrangian L as the space-integral of a local function of fields and spacetime derivatives of fields; the field equations were then determined from the principle that the action $\int L\,dt$ should be stationary when the fields are varied; and the commutation relations were determined from the assumption that the variational derivative of the Lagrangian with respect to any field's time-derivative behaves like a 'momentum' conjugate to that field (except that commutation relations become anticommutation relations for fermion fields). They also went on to apply this general formalism to the electromagnetic and Dirac fields, and explored the various invariance and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

conservation laws, including the conservation of charge, momentum, and energy, and Lorentz and gauge invariance.

The Heisenberg–Pauli formalism is essentially the same as that described in our Chapter 7, and so for the present we can limit ourselves to a single example which will turn out to be useful later in this section. For a free complex scalar field $\phi(x)$ the Lagrangian is taken as

$$L = \int d^3x \left[\dot{\phi}^\dagger \dot{\phi} - c^2 (\nabla\phi)^\dagger \cdot (\nabla\phi) - \left(\frac{mc^2}{\hbar}\right)^2 \phi^\dagger \phi \right] . \tag{1.2.26}$$

If we subject $\phi(x)$ to an infinitesimal variation $\delta\phi(x)$, the Lagrangian is changed by the amount

$$\delta L = \int d^3x \left[\dot{\phi}^\dagger \delta\dot{\phi} + \dot{\phi}\delta\dot{\phi}^\dagger - c^2 \nabla\phi^\dagger \cdot \nabla\delta\phi - c^2 \nabla\phi \cdot \nabla\delta\phi^\dagger \right.$$
$$\left. - \left(\frac{mc^2}{\hbar}\right)^2 \phi^\dagger \delta\phi - \left(\frac{mc^2}{\hbar}\right)^2 \phi\delta\phi^\dagger \right] . \tag{1.2.27}$$

It is assumed in using the principle of stationary action that the variation in the fields should vanish on the boundaries of the spacetime region of integration. Thus, in computing the change in the action $\int L\,dt$, we can immediately integrate by parts, and write

$$\delta \int L\,dt = c^2 \int d^4x \left[\delta\phi^\dagger \left(\Box - \left(\frac{mc}{\hbar}\right)^2 \right) \phi + \delta\phi \left(\Box - \left(\frac{mc}{\hbar}\right)^2 \right) \phi^\dagger \right] .$$

But this must vanish for any $\delta\phi$ and $\delta\phi^\dagger$, so ϕ must satisfy the familiar relativistic wave equation

$$\left[\Box - \left(\frac{mc}{\hbar}\right)^2 \right] \phi = 0 \tag{1.2.28}$$

and its adjoint. The 'momenta' canonically conjugate to the fields ϕ and $\phi^\dagger$ are given by the variational derivatives of L with respect to $\dot{\phi}$ and $\dot{\phi}^\dagger$, which we can read off from (1.2.27) as

$$\pi \equiv \frac{\delta L}{\delta\dot{\phi}} = \dot{\phi}^\dagger , \tag{1.2.29}$$

$$\pi^\dagger \equiv \frac{\delta L}{\delta\dot{\phi}^\dagger} = \dot{\phi} . \tag{1.2.30}$$

These field variables satisfy the usual canonical commutation relations,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

with a delta function in place of a Kronecker delta

$$\left[\pi(\mathbf{x},t),\phi(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = -i\hbar\delta^3(\mathbf{x}-\mathbf{y})\,, \tag{1.2.31}$$

$$\left[\pi(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\phi(\mathbf{y},t)\right] = 0\,, \tag{1.2.32}$$

$$\left[\pi(\mathbf{x},t),\pi(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\pi^\dagger(\mathbf{y},t)\right] = \left[\pi(\mathbf{x},t),\pi^\dagger(\mathbf{y},t)\right] = 0\,, \tag{1.2.33}$$

$$\left[\phi(\mathbf{x},t),\phi(\mathbf{y},t)\right] = \left[\phi^\dagger(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = \left[\phi(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = 0\,. \tag{1.2.34}$$

The Hamiltonian here is given (just as in particle mechanics) by the 'sum' of all canonical momenta times the time-derivatives of the corresponding fields, minus the Lagrangian:

$$H = \int d^3x\left[\pi\dot{\phi} + \pi^\dagger\dot{\phi}^\dagger\right] - L \tag{1.2.35}$$

or, using (1.2.26), (1.2.29), and (1.2.30):

$$H = \int d^3x\left[\pi^\dagger\pi + c^2(\nabla\phi)^\dagger\cdot(\nabla\phi) + \left(\frac{m^2c^4}{\hbar^2}\right)\phi^\dagger\phi\right]\,. \tag{1.2.36}$$

After the papers by Heisenberg and Pauli one element was still missing before quantum field theory could reach its final pre-war form: a solution to the problem of the negative-energy states. We saw in the last section that in 1930, at just about the time of the Heisenberg–Pauli papers, Dirac had proposed that the negative-energy states of the electron were all filled, but with only the holes in the negative-energy sea observable, rather than the negative-energy electrons themselves. After Dirac's idea was seemingly confirmed by the discovery of the positron in 1932, his 'hole theory' was used to calculate a number of processes to the lowest order of perturbation theory, including electron–positron pair production and scattering.

At the same time, a great deal of work was put into the development of a formalism whose Lorentz invariance would be explicit. The most influential effort was the 'many-time' formalism of Dirac, Vladimir Fock, and Boris Podolsky,[42] in which the state vector was represented by a wave function depending on the spacetime and spin coordinates of all electrons, positive-energy and negative-energy. In this formalism, the total number of electrons of either positive or negative energy is conserved; for instance, production of an electron–positron pair is described as the excitation of a negative-energy electron to a positive-energy state, and the annihilation of an electron and positron is described as the corresponding deexcitation. This many-time formalism had the advantage of manifest Lorentz invariance, but it had a number of disadvantages: In particular, there was a profound difference between the treatment of the photon, described in terms of a quantized electromagnetic field, and that of the electron and positron. Not all physicists felt this to be a disadvantage;

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

the electron field unlike the electromagnetic field did not have a classical limit, so there were doubts about its physical significance. Also, Dirac[42a] conceived of fields as the means by which we observe particles, so that he did not expect particles and fields to be described in the same terms. Though I do not know whether it bothered anyone at the time, there was a more practical disadvantage of the many-time formalism: it would have been difficult to use it to describe a process like nuclear beta decay, in which an electron and antineutrino are created without an accompanying positron or neutrino. The successful calculation by Fermi[43] of the electron energy distribution in beta decay deserves to be counted as one of the early triumphs of quantum field theory.

The essential idea that was needed to demonstrate the equivalence of the Dirac hole theory with a quantum field theory of the electron was provided by Fock[43a] and by Wendell Furry and Oppenheimer[44] in 1933–4. To appreciate this idea from a more modern standpoint, suppose we try to construct an electron field in analogy with the electromagnetic field or the Born–Heisenberg–Jordan field (1.2.2). Since electrons carry a charge, we would not like to mix annihilation and creation operators, so we might try to write the field as

$$\psi(x) = \sum_k u_k(\mathbf{x}) e^{-i\omega_k t} a_k \,, \tag{1.2.37}$$

where $u_k(\mathbf{x})e^{-i\omega_k t}$ are a complete set of orthonormal plane-wave solutions of the Dirac equation (1.1.13) (with k now labelling the three-momentum, spin, and sign of the energy):

$$\mathscr{H} u_k = \hbar\omega_k u_k \,, \tag{1.2.38}$$

$$\mathscr{H} \equiv -i\hbar c \boldsymbol{\alpha} \cdot \boldsymbol{\nabla} + \alpha_4 mc^2 \,, \tag{1.2.39}$$

$$\int u_k^\dagger u_\ell \, d^3x = \delta_{k\ell} \,, \tag{1.2.40}$$

and a_k are the corresponding annihilation operators, satisfying the Jordan–Wigner anticommutation relations (1.2.22)–(1.2.23). According to the ideas of 'second quantization' or the canonical quantization procedure of Heisenberg and Pauli,[41] the Hamiltonian is formed by calculating the 'expectation value' of $\mathscr{H}$ with a 'wave function' replaced by the quantized field (1.2.37)

$$H = \int d^3x \, \psi^\dagger \mathscr{H} \psi = \sum_k \hbar\omega_k a_k^\dagger a_k \,. \tag{1.2.41}$$

The trouble is, of course, that this is not a positive operator — half the ω_k are negative while the operators $a_k^\dagger a_k$ take only the positive eigenvalues 1 and 0. (See Eqs. (1.2.24) and (1.2.25).) In order to cure this disease, Furry and Oppenheimer picked up Dirac's idea[42] that the positron is the absence of a negative-energy electron; the anticommutation relations are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

symmetric between creation and annihilation operators, so they defined the positron creation and annihilation operators as the corresponding annihilation and creation operators for negative-energy electrons

$$b_k^\dagger \equiv a_k\,, \qquad b_k \equiv a_k^\dagger \qquad (\text{for } \omega_k < 0) \tag{1.2.42}$$

where the label k on b denotes a positive-energy positron mode with momenta and spin opposite to those of the electron mode k. The Dirac field (1.2.37) may then be written

$$\psi(x) = \sum_k{}^{(+)} a_k u_k(x) + \sum_k{}^{(-)} b_k^\dagger u_k(x)\,, \tag{1.2.43}$$

where $(+)$ and $(-)$ indicate sums over normal modes k with $\omega_k > 0$ and $\omega_k < 0$, respectively, and $u_k(x) \equiv u_k(\mathbf{x})e^{-i\omega_k t}$. Similarly, using the anticommutation relations for the bs, we can rewrite the energy operator (1.2.41) as

$$H = \sum_k{}^{(+)} \hbar\omega_k a_k^\dagger a_k + \sum_k{}^{(-)} \hbar|\omega_k| b_k^\dagger b_k + E_0\,, \tag{1.2.44}$$

where E_0 is the infinite c-number

$$E_0 = -\sum_k{}^{(-)} \hbar|\omega_k|\,. \tag{1.2.45}$$

In order for this redefinition to be more than a mere formality, it is necessary also to specify that the physical vacuum is a state Ψ_0 containing no positive-energy electrons or positrons:

$$a_k\Psi_0 = 0 \qquad (\omega_k > 0)\,, \tag{1.2.46}$$

$$b_k\Psi_0 = 0 \qquad (\omega_k < 0)\,. \tag{1.2.47}$$

Hence (1.2.44) gives the energy of the vacuum as just E_0. If we measure all energies relative to the vacuum energy E_0, then the physical energy operator is $H - E_0$; and Eq. (1.2.44) shows that this is a *positive* operator.

The problem of negative-energy states for a charged spin zero particle was also resolved in 1934, by Pauli and Weisskopf,[45] in a paper written in part to challenge Dirac's picture of filled negative-energy states. Here the creation and annihilation operators satisfy commutation rather than anticommutation relations, so it is not possible to interchange the roles of these operators freely, as was the case for fermions. Instead we must return to the Heisenberg–Pauli canonical formalism[41] to decide which coefficients of the various normal modes are creation or annihilation operators.

Pauli and Weisskopf expanded the free charged scalar field in plane waves in a cube of spatial volume $V \equiv L^3$:

$$\phi(\mathbf{x}, t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} q(\mathbf{k}, t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{1.2.48}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

with the wave numbers restricted by the periodicity condition, that the quantities $k_j L/2\pi$ for $j = 1,2,3$ should be a set of three positive or negative integers. Similarly the canonically conjugate variable (1.2.29) was expanded as

$$\pi(\mathbf{x},t) \equiv \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} p(\mathbf{k},t) e^{-i\mathbf{k}\cdot\mathbf{x}} . \tag{1.2.49}$$

The minus sign is put into the exponent here so that (1.2.29) now becomes:

$$p(\mathbf{k},t) = \dot{q}^\dagger(\mathbf{k},t) . \tag{1.2.50}$$

The Fourier inversion formula gives

$$q(\mathbf{k},t) = \frac{1}{\sqrt{V}} \int d^3x \, \phi(\mathbf{x},t) e^{-i\mathbf{k}\cdot\mathbf{x}} , \tag{1.2.51}$$

$$p(\mathbf{k},t) = \frac{1}{\sqrt{V}} \int d^3x \, \pi(\mathbf{x},t) e^{+i\mathbf{k}\cdot\mathbf{x}} , \tag{1.2.52}$$

and therefore the canonical commutation relations (1.2.31)–(1.2.34) yield for the qs and ps:

$$\left[p(\mathbf{k},t), q(\mathbf{l},t)\right] = \frac{-i\hbar}{V} \int d^3x \, e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{l}\cdot\mathbf{x}} = -i\hbar\delta_{\mathbf{kl}} \tag{1.2.53}$$

$$\begin{aligned} \left[p(\mathbf{k},t), q^\dagger(\mathbf{l},t)\right] &= \left[p(\mathbf{k},t), p(\mathbf{l},t)\right] = \left[p(\mathbf{k},t), p^\dagger(\mathbf{l},t)\right] \\ &= \left[q(\mathbf{k},t), q(\mathbf{l},t)\right] = \left[q(\mathbf{k},t), q^\dagger(\mathbf{l},t)\right] = 0 \end{aligned} \tag{1.2.54}$$

together with other relations that may be derived from these by taking their Hermitian adjoints. By inserting (1.2.48) and (1.2.49) in the formula (1.2.36) for the Hamiltonian, we can also write this operator in terms of ps and qs:

$$H = \sum_{\mathbf{k}} \left[p^\dagger(\mathbf{k},t) p(\mathbf{k},t) + \omega_{\mathbf{k}}^2 q^\dagger(\mathbf{k},t) q(\mathbf{k},t)\right] , \tag{1.2.55}$$

where

$$\omega_{\mathbf{k}}^2 \equiv c^2\mathbf{k}^2 + \left(\frac{mc^2}{\hbar}\right)^2 . \tag{1.2.56}$$

The time-derivatives of the ps are then given by the Hamiltonian equation

$$\dot{p}(\mathbf{k},t) = -\frac{\partial H}{\partial q(\mathbf{k},t)} = -\omega_{\mathbf{k}}^2 q^\dagger(\mathbf{k},t) \tag{1.2.57}$$

(and its adjoint), a result which in the light of Eq. (1.2.50) is just equivalent to the Klein–Gordon–Schrödinger wave equation (1.2.28).

We see that, just as in the case of the 1926 model of Born, Heisenberg, and Jordan,[4] the free field behaves like an infinite number of coupled

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

harmonic oscillators. Pauli and Weisskopf could construct p and q operators which satisfy the commutation relations (1.2.53)–(1.2.54) and the 'equations of motion' (1.2.50) and (1.2.57), by introducing annihilation and creation operators a, b, $a^\dagger$, $b^\dagger$ of *two* different kinds, corresponding to particles and antiparticles:

$$q(\mathbf{k},t) = i\sqrt{\frac{\hbar}{2\omega_\mathbf{k}}}\left[a(\mathbf{k})\exp(-i\omega_\mathbf{k}t) - b^\dagger(\mathbf{k})\exp(i\omega_\mathbf{k}t)\right] \tag{1.2.58}$$

$$p(\mathbf{k},t) = \sqrt{\frac{\hbar\omega_\mathbf{k}}{2}}\left[b(\mathbf{k})\exp(-i\omega_\mathbf{k}t) + a^\dagger(\mathbf{k})\exp(+i\omega_\mathbf{k}t)\right] \tag{1.2.59}$$

where

$$\left[a(\mathbf{k}),a^\dagger(\mathbf{l})\right] = \left[b(\mathbf{k}),b^\dagger(\mathbf{l})\right] = \delta_{\mathbf{kl}}\,, \tag{1.2.60}$$

$$\left[a(\mathbf{k}),a(\mathbf{l})\right] = \left[b(\mathbf{k}),b(\mathbf{l})\right] = 0\,, \tag{1.2.61}$$

$$\left[a(\mathbf{k}),b(\mathbf{l})\right] = \left[a(\mathbf{k}),b^\dagger(\mathbf{l})\right] = \left[a^\dagger(\mathbf{k}),b(\mathbf{l})\right] = \left[a^\dagger(\mathbf{k}),b^\dagger(\mathbf{l})\right] = 0\,. \tag{1.2.62}$$

It is straightforward to check that these operators do satisfy the desired relations (1.2.53), (1.2.54), (1.2.50), and (1.2.57). The field (1.2.48) may be written

$$\phi(\mathbf{x},t) = \frac{i}{\sqrt{V}}\sum_\mathbf{k}\sqrt{\frac{\hbar}{2\omega_\mathbf{k}}}\left[a(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{x} - i\omega_\mathbf{k}t) - b^\dagger(-\mathbf{k})\exp(-i\mathbf{k}\cdot\mathbf{x} + i\omega_\mathbf{k}t)\right] \tag{1.2.63}$$

and the Hamiltonian (1.2.55) takes the form

$$H = \sum_\mathbf{k}\frac{1}{2}\hbar\omega_\mathbf{k}\left[b^\dagger(\mathbf{k})b(\mathbf{k}) + b(\mathbf{k})b^\dagger(\mathbf{k}) + a^\dagger(\mathbf{k})a(\mathbf{k}) + a(\mathbf{k})a^\dagger(\mathbf{k})\right]$$

or, using (1.2.60)–(1.2.62)

$$H = \sum_\mathbf{k}\hbar\omega_\mathbf{k}\left[b^\dagger(\mathbf{k})b(\mathbf{k}) + a^\dagger(\mathbf{k})a(\mathbf{k})\right] + E_0\,, \tag{1.2.64}$$

where E_0 is the infinite c-number

$$E_0 \equiv \sum_\mathbf{k}\hbar\omega_\mathbf{k}\,. \tag{1.2.65}$$

The existence of two different kinds of operators a and b, which appear in precisely the same way in the Hamiltonian, shows that this is a theory with *two* kinds of particles with the same mass. As emphasized by Pauli and Weisskopf, these two varieties can be identified as particles and the corresponding antiparticles, and if charged have opposite charges. Thus,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

as we stressed above, bosons of spin zero as well as fermions of spin 1/2 can have distinct antiparticles, which for bosons cannot be identified as holes in a sea of negative energy particles.

We now can tell whether a and b or $a^\dagger$ and $b^\dagger$ are the annihilation operators by taking the expectation values of commutation relations in the vacuum state Ψ_0. For instance, if $a_{\mathbf{k}}^\dagger$ were an annihilation operator it would give zero when applied to the vacuum state, so the vacuum expectation value of (1.2.60) would give

$$-||a(\mathbf{k})\Psi_0||^2 = \left(\Psi_0, \left[a(\mathbf{k}), a^\dagger(\mathbf{k})\right]\Psi_0\right) = +1 \tag{1.2.66}$$

in conflict with the requirement that the left-hand side must be negative-definite. In this way we can conclude that it is a_k and b_k that are the annihilation operators, and therefore

$$a(\mathbf{k})\Psi_0 = b(\mathbf{k})\Psi_0 = 0\,. \tag{1.2.67}$$

This is consistent with all commutation relations. Thus, the canonical formalism forces the coefficient of the $e^{+i\omega t}$ in the field (1.2.58) to be a creation operator, as it also is in the Furry–Oppenheimer formalism[44] for spin 1/2.

Equations (1.2.64) and (1.2.67) now tell us that E_0 is the energy of the vacuum state. If we measure all energies relative to E_0, then the physical energy operator is $H - E_0$, and (1.2.64) shows that this again is positive.

What about the problem that served Dirac as a starting point, the problem of negative probabilities? As Dirac had recognized, the only probability density ρ, which can be formed from solutions of the Klein–Gordon–Schrödinger free scalar wave equation (1.2.28), and which satisfies a conservation law of the form (1.1.10), must be proportional to the quantity

$$\rho = 2\,\mathrm{Im}\left[\phi^\dagger \frac{\partial\phi}{\partial t}\right] \tag{1.2.68}$$

and therefore is not necessarily a positive quantity. Similarly, in the 'second-quantized' theory, where ϕ is given by Eq. (1.2.63), ρ is not a positive operator. Since $\phi^\dagger(x)$ does not commute with $\dot{\phi}(x)$ here, we can write (1.2.68) in various forms, which differ by infinite c-numbers; it proves convenient to write it as

$$\rho = \frac{i}{\hbar}\left[\frac{\partial\phi}{\partial t}\phi^\dagger - \frac{\partial\phi^\dagger}{\partial t}\phi\right]\,. \tag{1.2.69}$$

The space-integral of this operator is then easily calculated to be

$$N \equiv \int \rho\, d^3x = \sum_{\mathbf{k}} \left(a^\dagger(\mathbf{k})a(\mathbf{k}) - b^\dagger(\mathbf{k})b(\mathbf{k})\right) \tag{1.2.70}$$

and clearly has eigenvalues of either sign.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

However, in a sense this problem appears in quantum field theory for spin 1/2 as well as spin zero. The density operator $\psi^\dagger\psi$ of Dirac is indeed a positive operator, but in order to construct a physical density we ought to subtract the contribution of the filled electron states. In particular, using the plane-wave decomposition (1.2.43), we may write the total number operator as

$$N \equiv \int d^3x\, \psi^\dagger\psi = \sum_{\mathbf{k}} {}^{(+)}a^\dagger(\mathbf{k})a(\mathbf{k}) + \sum_{\mathbf{k}} {}^{(-)}b(\mathbf{k})b^\dagger(\mathbf{k})\,.$$

The anticommutation relations for the bs allow us to rewrite this as

$$N - N_0 = \sum_{\mathbf{k}} {}^{(+)}a^\dagger_{\mathbf{k}}a_{\mathbf{k}} - \sum_{\mathbf{k}} {}^{(-)}b^\dagger_{\mathbf{k}}b_{\mathbf{k}}\,, \tag{1.2.71}$$

where N_0 is the infinite constant

$$N_0 = \sum_{\mathbf{k}} {}^{(-)}1\,. \tag{1.2.72}$$

According to Eqs. (1.2.46) and (1.2.47), N_0 is the number of particles in the vacuum, so Furry and Oppenheimer reasoned that the physical number operator is $N - N_0$, and this now has both negative and positive eigenvalues, just as for spin zero.

The solution to this problem provided by quantum field theory is that neither the ψ of Furry and Oppenheimer nor the ϕ of Pauli and Weisskopf are probability amplitudes, which would have to define conserved positive probability densities. Instead, the physical Hilbert space is spanned by states defined as containing definite numbers of particles and/or antiparticles in each mode. If Φ_n are a complete orthonormal set of such states, then a measurement of particle numbers in an arbitrary state Ψ will yield a probability for finding the system in state Φ_n, given by

$$P_n = |(\Phi_n, \Psi)|^2\,, \tag{1.2.73}$$

where (Φ_n, Ψ) is the usual Hilbert space scalar product. Hence, no question as to the possibility of negative probabilities will arise for any spin. The wave fields ϕ, ψ, etc, are not probability amplitudes at all, but *operators* which create or destroy particles in the various normal modes. It would be a good thing if the misleading expression 'second quantization' were permanently retired.

In particular, the operators N and $N - N_0$ of Eqs. (1.2.70) and (1.2.71) are not to be interpreted as total probabilities, but as number operators: specifically, the number of particles *minus* the number of antiparticles. For charged particles, the conservation of charge forces the charge operators to be proportional to these number operators, so the minus signs in (1.2.70) and (1.2.71) allow us immediately to conclude that particles and antiparticles have opposite charge. In this field-theoretic formalism, interactions

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

contribute terms to the Hamiltonian which are of third, fourth, or higher order in field variables, and the rates of various processes are given by using these interaction operators in a time-dependent perturbation theory. The conceptual framework described in the above brief remarks will serve as the basis for much of the work in this book.

Despite its apparent advantages, quantum field theory did not immediately supplant hole theory; rather, the two points of view coexisted for a while, and various combinations of field-theoretic and hole-theoretic ideas were used in calculations of physical reaction rates. This period saw a number of calculations of cross sections to lowest order in powers of e^2 for various processes, such as $e^- + \gamma \to e^- + \gamma$ in 1929 by Klein and Nishina;[46] $e^+ + e^- \to 2\gamma$ in 1930 by Dirac;[47] $e^- + e^- \to e^- + e^-$ in 1932 by Møller;[48] $e^- + Z \to e^- + \gamma + Z$ and $\gamma + Z \to e^+ + e^- + Z$ (where Z denotes the Coulomb field of a heavy atom) in 1934 by Bethe and Heitler;[49] and $e^+ + e^- \to e^+ + e^-$ in 1936 by Bhabha.[50] (Rules for the calculation of such processes are given in Chapter 8, and worked out in detail there for the case of electron–photon scattering.) These lowest-order calculations gave finite results, in reasonable agreement with the experimental data.

Nevertheless, a general feeling of dissatisfaction with quantum field theory (whether or not in the form of hole theory) persisted throughout the 1930s. One of the reasons for this was the apparent failure of quantum electrodynamics to account for the penetrating power of the charged particles in cosmic ray showers, noted in 1936 by Oppenheimer and Franklin Carlson.[50a] Another cause of dissatisfaction that turned out to be related to the first was the steady discovery of new kinds of particles and interactions. We have already mentioned the electron, photon, positron, neutrino, and, of course, the nucleus of hydrogen, the proton. Throughout the 1920s it was generally believed that heavier nuclei are composed of protons and electrons, but it was hard to see how a light particle like the electron could be confined in the nucleus. Another severe difficulty with this picture was pointed out in 1931 by Ehrenfest and Oppenheimer:[51] the nucleus of ordinary nitrogen, N^{14}, in order to have atomic number 7 and atomic weight 14, would have to be composed of 14 protons and 7 electrons, and would therefore have to be a fermion, in conflict with the result of molecular spectroscopy[52] that N^{14} is a boson. This problem (and others) were solved in 1932 with the discovery of the neutron,[53] and by Heisenberg's subsequent suggestion[54] that nuclei are composed of protons and neutrons, not protons and electrons. It was clear that a strong non-electromagnetic force of short range would have to operate between neutrons and protons to hold nuclei together.

After the success of the Fermi theory of beta decay, several authors[54a] speculated that nuclear forces might be explained in this theory as due to the exchange of electrons and neutrinos. A few years later, in 1935,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

Hideki Yukawa proposed a quite different quantum field theory of the nuclear force.[55] In an essentially classical calculation, he found that the interaction of a scalar field with nucleons (protons or neutrons) would produce a nucleon–nucleon potential, with a dependence on the nucleon separation r given by

$$V(r) \propto \frac{1}{r} \exp(-\lambda r) \tag{1.2.74}$$

instead of the $1/r$ Coulomb potential produced by electric fields. The quantity λ was introduced as a parameter in Yukawa's scalar field equation, and when this equation was quantized, Yukawa found that it described particles of mass $\hbar\lambda/c$. The observed range of the strong interactions within nuclei led Yukawa to estimate that $\hbar\lambda/c$ is of the order of 200 electron masses. In 1937 such 'mesons' were discovered in cloud chamber experiments[56] by Seth Neddermeyer and Anderson and by Jabez Curry Street and Edward Carl Stevenson, and it was generally believed that these were the hypothesized particles of Yukawa.

The discovery of mesons revealed that the charged particles in cosmic ray showers are not all electrons, and thus cleared up the problem with these showers that had bothered Oppenheimer and Carlson. At the same time, however, it created new difficulties. Lothar Nordheim[56a] pointed out in 1939 that the same strong interactions by which the mesons are copiously produced at high altitudes (and which are required in Yukawa's theory) should have led to the mesons' absorption in the atmosphere, a result contradicted by their copious appearance at lower altitudes. In 1947 it was shown in an experiment by Marcello Conversi, Ettore Pancini, and Oreste Piccioni[57] that the mesons which predominate in cosmic rays at low altitude actually interact weakly with nucleons, and therefore could not be identified with Yukawa's particle. This puzzle was cleared up by a theoretical suggestion,[58] and its subsequent experimental confirmation[59] by Cesare Lattes, Occhialini, and Cecil Powell — there are two kinds of mesons with slightly different masses: the heavier (now called the π meson or pion) has strong interactions and plays the role in nuclear force envisaged by Yukawa; the lighter (now called the μ meson, or muon) has only weak and electromagnetic interactions, and predominates in cosmic rays at sea level, being produced by the decay of π mesons. In the same year, 1947, entirely new kinds of particles (now known as K mesons and hyperons) were found in cosmic rays by George Rochester and Clifford Butler.[60] From 1947 until the present particles have continued to be discovered in a bewildering variety, but to pursue this story would take us outside the bounds of our present survey. These discoveries showed clearly that any conceptual framework which was limited to photons, electrons, and positrons would be far too narrow to be taken seriously as

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

a fundamental theory. But an even more important obstacle was presented by a purely theoretical problem — the problem of infinities.

1.3 The Problem of Infinities

Quantum field theory deals with fields $\psi(x)$ that destroy and create particles at a spacetime point x. Earlier experience with classical electron theory provided a warning that a point electron will have infinite electromagnetic self-mass; this mass is $e^2/6\pi ac^2$ for a surface distribution of charge with radius a, and therefore blows up for $a \to 0$. Disappointingly this problem appeared with even greater severity in the early days of quantum field theory, and although greatly ameliorated by subsequent improvements in the theory, it remains with us to the present day.

The problem of infinities in quantum field theory was apparently first noted in the 1929–30 papers of Heisenberg and Pauli.[41] Soon after, the presence of infinities was confirmed in calculations of the electromagnetic self-energy of a bound electron by Oppenheimer,[61] and of a free electron by Ivar Waller.[62] They used ordinary second-order perturbation theory, with an intermediate state consisting of an electron and a photon: for instance, the shift of the energy E_n of an electron in the nth energy level of hydrogen is given by

$$\Delta E_n = \sum_{m,\lambda} \int d^3k \, \frac{|<m;\mathbf{k},\lambda|H'|n>|^2}{E_n - E_m - |\mathbf{k}|c} , \tag{1.3.1}$$

where the sums and integral are over all intermediate electron states m, photon helicities λ, and photon momenta $\mathbf{k}$, and H' is the term in the Hamiltonian representing the interaction of radiation and electrons. This calculation gave a self-energy that is formally infinite; further; if this infinity is removed by discarding all intermediate states with photon wave numbers greater than $1/a$, then the self-energy behaves like $1/a^2$ as $a \to 0$. Infinities of this sort are often called ultraviolet divergences, because they arise from intermediate states containing particles of very short wavelength.

These calculations treated the electron according to the rules of the original Dirac theory, without filled negative-electron states. A few years later Weisskopf repeated the calculation of the electron self-mass in the new hole theory, with all negative-energy states full. In this case another term appears in second-order perturbation theory, which in a non-hole-theory language can be described as arising from processes in which the electron in its final state first appears out of the vacuum together with a photon and a positron which then annihilate along with the initial

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

electron. Initially Weisskopf found a $1/a^2$ dependence on the photon wave-number cutoff $1/a$. The same calculation was being carried out (at the suggestion of Bohr) at that time by Carlson and Furry. After seeing Weisskopf's results, Furry realized that while Weisskopf had included an electrostatic term that he and Carlson had neglected, Weisskopf had made a new mistake in the calculation of the magnetic self-energy. After hearing from Furry and correcting his own error, Weisskopf found that the $1/a^2$ terms in the total mass shift cancelled! However, despite this cancellation, an infinity remained: with a wave-number cutoff $1/a$, the self-mass was found to be[63]

$$m_{em} = \frac{3\alpha}{2\pi} m \ln\left(\frac{\hbar}{mca}\right), \tag{1.3.2}$$

The weakening of the cut-off dependence, to $\ln a$ as compared with the classical $1/a$ or the early quantum $1/a^2$, was mildly encouraging at the time and turned out to be of great importance later, in the development of renormalization theory.

An infinity of quite a different kind was encountered in 1933, apparently first by Dirac.[64] He considered the effect of an external static nearly uniform charge density $\varepsilon(\mathbf{x})$ on the vacuum, i.e., on the negative-energy electrons in the filled energy levels of hole theory. The Coulomb interaction between $\varepsilon(\mathbf{x})$ and the charge density of the negative-energy electrons produces a 'vacuum polarization,' with induced charge density

$$\delta\varepsilon = A\varepsilon + B\left(\frac{\hbar}{mc}\right)^2 \nabla^2\varepsilon + \cdots . \tag{1.3.3}$$

The constant B is finite, and of order α. On the other hand, A is logarithmically divergent, of order $\alpha \ln a$, where $1/a$ is the wave-number cutoff.

Infinities also seemed to occur in a related problem, the scattering of light by light. Hans Euler, Bernard Kockel, and Heisenberg[65] showed in 1935–6 that these infinities could be eliminated by using a more-or-less arbitrary prescription suggested earlier by Dirac[66] and Heisenberg[67]. They calculated an effective Lagrangian density for the non-linear electrodynamic effects produced by virtual electron–positron pairs:

$$\mathscr{L} = \frac{1}{2}\left(\mathbf{E}^2 - \mathbf{B}^2\right) + \frac{e^4\hbar}{360\pi^2 m_e^4 c^7}\left[\left(\mathbf{E}^2 - \mathbf{B}^2\right)^2 + 7\left(\mathbf{E}\cdot\mathbf{B}\right)^2\right] + \cdots , \tag{1.3.4}$$

valid for frequencies $\nu \ll m_e c^2/h$. Soon after, Nicholas Kemmer and Weisskopf[68] presented arguments that in this case the infinities are spurious, and that Eq. (1.3.4) can be derived without any subtraction prescription.

One bright spot in the struggle with infinities was the successful treat-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

ment of *infrared* divergences, those that arise from the low-energy rather than the high-energy part of the range of integration. In 1937 it was shown by Felix Bloch and Arne Nordsieck[68a] that these infinities cancel provided one includes processes in which arbitrary numbers of low-energy photons are produced. This will be discussed in modern terms in Chapter 13.

Yet another infinity turned up in a calculation by Sidney Michael Dancoff[69] in 1939 of the radiative corrections to the scattering of electrons by the static Coulomb field of an atom. The calculation contained a mistake (one of the terms was omitted), but this was not realized until later.[69a]

Throughout the 1930s, these various infinities were seen not merely as failures of specific calculations. Rather, they seemed to indicate a gap in the understanding of relativistic quantum field theory on the most fundamental level, an opinion reinforced by the problems with cosmic rays mentioned in the previous section.

One of the symptoms of this uneasy pessimism was the continued exploration throughout the 1930s and 1940s of alternative formalisms. As Julian Schwinger[69b] later recalled, 'The preoccupation of the majority of involved physicists was not with analyzing and carefully applying the known relativistic theory of coupled electron and electromagnetic fields but with changing it.' Thus in 1938 Heisenberg[70] proposed the existence of a fundamental length L, analogous to the fundamental action h and fundamental velocity c. Field theory was supposed to work only for distances larger than L, so that all divergent integrals would effectively be cut off at distances L, or momenta h/L. Several specific proposals[70a] were made for giving field theory a non-local structure. Some theorists began to suspect that the formalism of state-vectors and quantum fields should be replaced by one based solely on observable quantities, such as the S-matrix introduced by John Archibald Wheeler[71] in 1937 and Heisenberg[72] in 1943, whose elements are the amplitudes for various scattering processes. As we shall see, the concept of the S-matrix has now become a vital part of modern quantum field theory, and for some theorists a pure S-matrix theory became an ideal, especially as a possible solution to the problems of the strong interactions.[73] In yet another direction, Wheeler and Richard Feynman[74] in 1945 attempted to eliminate the electromagnetic field, deriving electromagnetic interactions in terms of an interaction at a distance. They were able to show that a pure retarded (or pure advanced) potential could be obtained by taking into account the interaction not only between source and test charges, but also between these charges and all the other charges in the universe. Perhaps the most radical modification of quantum mechanics suggested during this period was the introduction by Dirac[75] of states of negative probability,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

as a means of cancelling infinities in sums over states. This idea, of an 'indefinite metric' in Hilbert space, has also flourished in quantum field theory, though not in the form originally suggested.

A more conservative idea for dealing with the infinities was also in the air during the 1930s. Perhaps these infinities could all be absorbed into a redefinition, a 'renormalization' of the parameters of the theory. For instance, it was already known that in any Lorentz-invariant classical theory the electromagnetic self-energy and self-momentum of an electron *must* take the form of corrections to the mass of the electron; hence the infinities in these quantities can be cancelled by a negative infinity in the 'bare' non-electromagnetic mass of the electron, leaving a finite measurable 'renormalized' mass. Also, Eq. (1.3.3) shows that the vacuum polarization changes the charge of the electron, from $e \equiv \int d^3x\, \varepsilon$, to

$$e_{\text{TOTAL}} = \int d^3x(\varepsilon + \delta\varepsilon) = (1 + A)e\,. \tag{1.3.5}$$

Vacuum polarization gives finite results in lowest order if observables like scattering cross-sections are expressed in terms of e_{TOTAL} rather than e. The question was, whether all infinities in quantum field theory could be dealt with in this way. In 1936 Weisskopf[76] suggested that this is the case, and verified that known infinities could be eliminated by renormalization of physical parameters in a variety of sample calculations. However, it was impossible with the calculational techniques then available to show that infinities could always be eliminated in this way, and Dancoff's calculation[69] seemed to show that they could not.

Another effect of the appearance of infinities was a tendency to believe that any effect which turned out to be infinite in quantum field theory was actually not there at all. In particular, the 1928 Dirac theory had predicted complete degeneracy of the $2s_{1/2}$–$2p_{1/2}$ levels of hydrogen to all orders in α; any attempt at a quantum electromagnetic calculation of the splitting of these two levels ran into the problem of the infinite self-energy of a bound electron; therefore the existence of such a splitting was generally not taken seriously. Later Bethe[80] recalled that 'This shift comes out infinite in all existing theories, and has therefore always been ignored.' This attitude persisted even in the late 1930s, when spectroscopic experiments[77] began to indicate the presence of a $2s_{1/2}$–$2p_{1/2}$ splitting of order 1000 MHz. One notable exception was Edwin Albrecht Uehling,[78] who realized that the vacuum polarization effect mentioned earlier would produce a $2s_{1/2}$–$2p_{1/2}$ splitting; unfortunately, as we shall see in Chapter 14, this contribution to the splitting is much smaller than 1000 MHz, and of the wrong sign.

The gloom surrounding quantum field theory began to lift soon after World War II. On June 1–4, 1947, the Conference on the Foundations of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

Quantum Mechanics at Shelter Island, NY brought theoretical physicists who had been working on the problems of quantum field theory through the 1930s together with a younger generation of theorists who had started scientific work during the war, and — of crucial importance — a few experimental physicists. The discussion leaders were Hans Kramers, Oppenheimer, and Weisskopf. One of the experimentalists (or rather theorist turned experimentalist), Willis Lamb, described a decisive measurement[79] of the $2s_{1/2}$–$2p_{1/2}$ shift in hydrogen. A beam of hydrogen atoms from an oven, many in $2s$ and $2p$ states, was aimed at a detector sensitive only to atoms in excited states. The atoms in $2p$ states can decay very rapidly to the $1s$ ground state by one-photon (Lyman α) emission, while the $2s$ states decay only very slowly by two-photon emission, so in effect the detector was measuring the number of atoms in the metastable $2s$ state. The beam was passed through a magnetic field, which added a known Zeeman splitting to any $2s_{1/2}$–$2p_{1/2}$ splitting naturally present. The beam was also exposed to a microwave-frequency electromagnetic field, with a fixed frequency $\nu \sim 10$ GHz. At a certain magnetic field strength the detector signal was observed to be quenched, indicating that the microwave field was producing resonant transitions from the metastable $2s$ state to the $2p$ state and thence by a rapid Lyman α emission to the ground state. The total (Zeeman plus intrinsic) $2s$–$2p$ splitting at this value of the magnetic field strength would have to be just $h\nu$, from which the intrinsic splitting could be inferred. A preliminary value of 1000 MHz was announced, in agreement with the earlier spectroscopic measurements.[77] The impact of this discovery can be summarized in a saying that was current in Copenhagen when I was a graduate student there in 1954: 'Just because something is infinite does not mean it is zero!'

The discovery of the Lamb shift aroused intense interest among the theorists at Shelter Island, many of whom had already been working on improved formalisms for calculation in quantum electrodynamics. Kramers described his work on mass renormalization in the classical electrodynamics of an extended electron,[79a] which showed that the difficulties associated with the divergence of the self-energy in the limit of zero radius do not appear explicitly if the theory is reexpressed so that the mass parameter in the formalism is identified with the experimental electron mass. Schwinger and Weisskopf (who had already heard rumors of Lamb's result, and discussed the matter on the trip to Shelter Island) suggested that since the inclusion of intermediate states involving positrons was known to reduce the divergence in energy level shifts from $1/a^2$ to $\ln a$, perhaps the *differences* of the shifts in atomic energy levels might turn out to be finite when these intermediate states were taken into account. (In fact, in 1946, before he learned of Lamb's experiment, Weisskopf had already assigned this problem to a graduate student, Bruce French.) Almost im-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

mediately after the conference, during a train ride to Schenectady, Hans Bethe[80] carried out a non-relativistic calculation, still without including the effects of intermediate states containing positrons, but using a simple cutoff at virtual photon momenta of order $m_e c^2$ to eliminate infinities. He obtained the encouraging approximate value of 1040 MHz. Fully relativistic calculations using the renormalization idea to eliminate infinities were soon thereafter carried out by a number of other authors,[81] with excellent agreement with experiment.

Another exciting experimental result was reported at Shelter Island by Isidor I. Rabi. Measurements in his laboratory of the hyperfine structure of hydrogen and deuterium had suggested[82] that the magnetic moment of the electron is larger than the Dirac value $e\hbar/2mc$ by a factor of about 1.0013, and subsequent measurements of the gyromagnetic ratios in sodium and gallium had given a precise value[83]

$$\mu = \frac{e\hbar}{2mc}\,[1.00118 \pm 0.00003]\,.$$

Learning of these results, Gregory Breit suggested[83a] that they arose from an order α radiative correction to the electron magnetic moment. At Shelter Island, both Breit and Schwinger described their efforts to calculate this correction. Shortly after the conference Schwinger completed a successful calculation of the anomalous magnetic moment of the electron[84]

$$\mu = \frac{e\hbar}{2mc}\left[1 + \frac{\alpha}{2\pi}\right] = \frac{e\hbar}{2mc}\,[1.001162]$$

in excellent agreement with observation. This, together with Bethe's calculation of the Lamb shift, at last convinced physicists of the reality of radiative corrections.

The mathematical methods used in this period presented a bewildering variety of concepts and formalisms. One approach developed by Schwinger[85] was based on operator methods and the action principle, and was presented by him at a conference at Pocono Manor in 1948, the successor to the Shelter Island Conference. Another Lorentz-invariant operator formalism had been developed earlier by Sin-Itiro Tomonaga[86] and his co-workers in Japan, but their work was not at first known in the West. Tomonaga had grappled with infinities in Yukawa's meson theory in the 1930s. In 1947 he and his group were still out of the loop of scientific communication; they learned about Lamb's experiment from an article in *Newsweek*.

An apparently quite different approach was invented by Feynman,[87] and described briefly by him at the Pocono Conference. Instead of introducing quantum field operators, Feynman represented the S-matrix as a functional integral of $\exp\left(iW\right)$, where W is the action integral for a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

set of Dirac particles interacting with a *classical* electromagnetic field, integrated over all Dirac particle trajectories satisfying certain initial and final conditions for $t \to \pm\infty$. One result of great practical importance that came out of Feynman's work was a set of graphical rules for calculating S-matrix elements to any desired order of perturbation theory. Unlike the old perturbation theory of the 1920s and 1930s, these Feynman rules automatically lumped together particle creation and antiparticle annihilation processes, and thereby gave results that were Lorentz-invariant at every stage. We have already seen in Weisskopf's early calculation[63] of the electron self-energy, that it is only in such calculations, including particles and antiparticles on the same footing, that the nature of the infinities becomes transparent.

Finally, in a pair of papers in 1949, Freeman Dyson[88] showed that the operator formalisms of Schwinger and Tomonaga would yield the same graphical rules that had been found by Feynman. Dyson also carried out an analysis of the infinities in general Feynman diagrams, and outlined a proof that these are always precisely the sort which could be removed by renormalization. One of the most striking results that could be inferred from Dyson's analysis was a criterion for deciding which quantum field theories are 'renormalizable', in the sense that all infinities can be absorbed into a redefinition of a *finite* number of coupling constants and masses. In particular, an interaction like the Pauli term (1.1.32), which would have changed the predicted magnetic moment of the electron, would spoil the renormalizability of quantum electrodynamics. With the publication of Dyson's papers, there was at last a general and systematic formalism that physicists could easily learn to use, and that would provide a common language for the subsequent applications of quantum field theory to the problems of physics.

I cannot leave the infinities without taking up a puzzling aspect of this story. Oppenheimer[61] in 1930 had already noticed that most of the ultraviolet divergence in the self-energy of a bound electron cancels when one takes the difference between the shifts of two atomic energy levels, and Weisskopf[63] in 1934 had found that most of the divergence in the self-energy of a free electron cancels when one includes intermediate states containing positrons. It would have been natural even in 1934 to guess that including positron intermediate states *and* subtracting the energy shifts of pairs of atomic states would eliminate the ultraviolet divergence in their relative energy shift.* There was even experimental evidence[77] for

* In fact, this guess would have been wrong. As discussed in Section 14.3, radiative corrections to the electron mass affect atomic energy levels not only through a shift in the electron rest energy, which is the same in all atomic energy levels, but also through a change in the electron kinetic energy, that varies from one level to another.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

a $2s_{1/2}$–$2p_{1/2}$ energy difference of order 1000 MHz. So why did no one before 1947 attempt an *numerical* estimate of this energy difference?

Strictly speaking, there was one such attempt[88a] in 1939, but it focused on the wrong part of the problem, the charge radius of the *proton*, which has only a tiny effect on hydrogen energy levels. The calculation gave a result in rough agreement with the early experiments.[77] This was a mistake, as shown in 1939 by Lamb.[88b]

A fully relativistic calculation of the Lamb shift including positrons in intermediate states could have been attempted during the 1930s, using the old non-relativistic perturbation theory. As long as one keeps all terms up to a given order, old-fashioned non-relativistic perturbation theory gives the same results as the manifestly relativistic formalisms of Feynman, Schwinger, and Tomonaga. In fact, after Bethe's work, the first precise calculations[81] of the Lamb shift in the USA by French and Weisskopf and Norman Kroll and Lamb were done in just this way, though Tomonoga's group [81] in Japan was already using covariant methods to solve this and other problems.

The one missing element was confidence in renormalization as a means of dealing with infinities. As we have seen, renormalization was widely discussed in the late 1930s. But it had become accepted wisdom in the 1930s, and a point of view especially urged by Oppenheimer,[89] that quantum electrodynamics could not be taken seriously at energies of more than about 100 MeV, and that the solution to its problems could be found only in really adventurous new ideas.

Several things happened at Shelter Island to change this expectation. One was news that the problems concerning cosmic rays discussed in the previous section were beginning to be resolved; Robert Marshak presented the hypothesis[58] that there were two types of 'meson' with similar masses; the muons that had actually been observed, and the pions responsible for nuclear forces. More important was the fact that now there were reliable experimental values for the Lamb shift and the anomalous magnetic moment that forced physicists to think carefully about radiative corrections. Probably equally important was the fact that the conference brought together theorists who had in their own individual ways been thinking about renormalization as a solution to the problem of infinities. When the revolution came in the late 1940s, it was made by physicists who though mostly young were playing a conservative role, turning away from the search by their predecessors for a radical solution.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

Bibliography

- S. Aramaki, 'Development of the Renormalization Theory in Quantum Electrodynamics,' *Historia Scientiarum* **36**, 97 (1989); *ibid.* **37**, 91 (1989). [Section 1.3.]
- R. T. Beyer, ed., *Foundations of Nuclear Physics* (Dover Publications, Inc., New York, 1949). [Section 1.2.]
- L. Brown, 'Yukawa's Prediction of the Meson,' *Centauros* **25**, 71 (1981). [Section 1.2.]
- L. M. Brown and L. Hoddeson, eds., *The Birth of Particle Physics* (Cambridge University Press, Cambridge, 1983). [Sections 1.1, 1.2, 1.3.]
- T. Y. Cao and S. S. Schweber, 'The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory,' *Synthèse* **97**, 33 (1993). [Section 1.3.]
- P. A. M. Dirac, *The Development of Quantum Theory* (Gordon and Breach Science Publishers, New York, 1971). [Section 1.1.]
- E. Fermi, 'Quantum Theory of Radiation,' *Rev. Mod. Phys.* **4**, 87 (1932). [Sections 1.2 and 1.3.]
- G. Gamow, *Thirty Years that Shook Physics* (Doubleday and Co., Garden City, New York, 1966). [Section 1.1.]
- M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill Book Co., New York, 1966). [Section 1.1.]
- J. Mehra, 'The Golden Age of Theoretical Physics: P.A.M. Dirac's Scientific Work from 1924 to 1933,' in *Aspects of Quantum Theory*, ed. by A. Salam and E. P. Wigner, (Cambridge University Press, Cambridge, 1972). [Section 1.1.]
- A. I. Miller, *Early Quantum Electrodynamics – A Source Book* (Cambridge University Press, Cambridge, UK, 1994). [Sections 1.1, 1.2, 1.3.]
- A. Pais, *Inward Bound* (Clarendon Press, Oxford, 1986). [Sections 1.1, 1.2, 1.3.]
- S. S. Schweber, 'Feynman and the Visualization of Space-Time Processes,' *Rev. Mod. Phys.* **58**, 449 (1986). [Section 1.3.]
- S. S. Schweber, 'Some Chapters for a History of Quantum Field Theory: 1938–1952,' in *Relativity, Groups, and Topology II*, ed. by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984). [Sections 1.1, 1.2, 1.3.]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

- S. S. Schweber, 'A Short History of Shelter Island I,' in *Shelter Island II*, ed. by R. Jackiw, S. Weinberg, and E. Witten (MIT Press, Cambridge, MA, 1985). [Section 1.3.]
- S. S. Schweber, *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton University Press, Princeton, 1994). [Section 1.1, 1.2, 1.3.]
- J. Schwinger, ed., *Selected Papers in Quantum Electrodynamics* (Dover Publications Inc., New York, 1958). [Sections 1.2 and 1.3.]
- S.-I. Tomonaga, in *The Physicist's Conception of Nature* (Reidel, Dordrecht, 1973). [Sections 1.2 and 1.3.]
- S. Weinberg, 'The Search for Unity: Notes for a History of Quantum Field Theory,' *Daedalus*, Fall 1977. [Sections 1.1, 1.2, 1.3.]
- V. F. Weisskopf, 'Growing Up with Field Theory: The Development of Quantum Electrodynamics in Half a Century,' 1979 Bernard Gregory Lecture at CERN, published in L. Brown and L. Hoddeson, *op. cit.*. [Sections 1.1, 1.2, 1.3.]
- G. Wentzel, 'Quantum Theory of Fields (Until 1947)' in *Theoretical Physics in the Twentieth Century*, ed. by M. Fierz and V. F. Weisskopf (Interscience Publishers Inc., New York, 1960). [Sections 1.2 and 1.3.]
- E. Whittaker, *A History of the Theories of Aether and Electricity* (Humanities Press, New York, 1973). [Section 1.1.]

References

1. L. de Broglie, *Comptes Rendus* **177**, 507, 548, 630 (1923); *Nature* **112**, 540 (1923); Thèse de doctorat (Masson et Cie, Paris, 1924); *Annales de Physique* **3**, 22 (1925) [reprinted in English in *Wave Mechanics*, ed. by G. Ludwig, (Pergamon Press, New York, 1968)]; *Phil. Mag.* **47**, 446 (1924).
2. W. Elsasser, *Naturwiss.* **13**, 711 (1925).
3. C. J. Davisson and L. H. Germer, *Phys. Rev.* **30**, 705 (1927).
4. W. Heisenberg, *A. Phys.* **33**, 879 (1925); M. Born and P. Jordan, *Z. f. Phys.* **34**, 858 (1925); P. A. M. Dirac, *Proc. Roy. Soc.* **A109**, 642 (1925); M. Born, W. Heisenberg, and P. Jordan, *Z. f. Phys.* **35**, 557 (1926); W. Pauli, *Z. f. Phys.* **36**, 336 (1926). These papers are reprinted in *Sources of Quantum Mechanics*, ed. by B. L. van der Waerden (Dover Publications, Inc., New York, 1968).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

5. E. Schrödinger, *Ann. Phys.* **79**, 361, 489; **80**, 437; **81**, 109 (1926). These papers are reprinted in English, unfortunately in a somewhat abridged form, in *Wave Mechanics*, Ref. 1. Also see *Collected Papers on Wave Mechanics*, trans. by J. F. Schearer and W. M. Deans (Blackie and Son, London, 1928).

6. See, e.g., P. A. M. Dirac, *The Development of Quantum Theory* (Gordon and Breach, New York, 1971). Also see Dirac's obituary of Schrödinger, *Nature* **189**, 355 (1961), and his article in *Scientific American* **208**, 45 (1963).

7. O. Klein, *Z. f. Phys.* **37**, 895 (1926). Also see V. Fock, *Z. f. Phys.* **38**, 242 (1926); *ibid*, **39**, 226 (1926).

8. W. Gordon, *Z. f. Phys.* **40**, 117 (1926).

9. For the details of the calculation, see, e.g., L. I. Schiff, *Quantum Mechanics*, 3rd edn, (McGraw-Hill, Inc. New York, 1968): Section 51.

10. F. Paschen, *Ann. Phys.* **50**, 901 (1916). These experiments were actually carried out using He^+ because its fine structure splitting is 16 times larger than for hydrogen. The fine structure of spectral lines was first discovered interferometrically by A. A. Michelson, *Phil. Mag.* **31**, 338 (1891); *ibid.*, **34**, 280 (1892).

10*a*. A. Sommerfeld, *Münchner Berichte* 1915, pp. 425, 429; *Ann. Phys.* **51**, 1, 125 (1916). Also see W. Wilson, *Phil. Mag.* **29**, 795 (1915).

11. G. E. Uhlenbeck and S. Goudsmit, *Naturwiss.* **13**, 953 (1925); *Nature* **117**, 264 (1926). The electron spin had been earlier suggested for other reasons by A. H. Compton, *J. Frank. Inst.* **192**, 145 (1921).

12. The general formula for Zeeman splitting in one-electron atoms had been discovered empirically by A. Landé, *Z. f. Phys.* **5**, 231 (1921); *ibid.*, **7**, 398 (1921); *ibid.*, **15**, 189 (1923); *ibid.*, **19**, 112 (1923). At the time, the extra non-orbital angular momentum appearing in this formula was thought to be the angular momentum of the atomic 'core;' A. Sommerfeld, *Ann. Phys.* **63**, 221 (1920); *ibid.*, **70**, 32 (1923). It was only later that the extra angular momentum was recognized, as in Ref. 11, to be due to the spin of the electron.

13. W. Heisenberg and P. Jordan, *Z. f. Phys.* **37**, 263 (1926); C. G. Darwin, *Proc. Roy. Soc.* **A116**, 227 (1927). Darwin says that several authors did this work at about the same time, while Dirac quotes only Darwin.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

14. L. H. Thomas, *Nature* **117**, 514 (1926). Also see S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972): Section 5.1.

15. P. A. M. Dirac, *Proc. Roy. Soc.* **A117**, 610 (1928). Also see Dirac, *ibid.*, **A118**, 351 (1928), for the application of this theory to the calculation of the Zeeman and Paschen–Back effects and the relative strengths of lines within fine-structure multiplets.

16. For the probabilistic interpretation of non-relativistic quantum mechanics, see M. Born *Z. f. Phys.* **37**, 863 (1926); *ibid*, **38**, 803 (1926) (reprinted in an abridged English version in *Wave Mechanics*, Ref. 1); G. Wentzel, *Z. f. Phys.* **40**, 590 (1926); W. Heisenberg, *Z. f. Phys.* **43**, 172 (1927). N. Bohr. *Nature* **121**, 580 (1928); Naturwissenchaften **17**, 483 (1929); *Electrons et Photons – Rapports et Discussions du Ve Conseil de Physique Solvay* (Gauthier-Villars, Paris, 1928).

17. Conversation between Dirac and J. Mehra, March 28, 1969, quoted by Mehra in *Aspects of Quantum Theory*, ed. by A. Salam and E. P. Wigner (Cambridge University Press, Cambridge, 1972).

18. G. Gamow, *Thirty Years that Shook Physics*, (Doubleday and Co., Garden City, NY, 1966): p. 125.

19. W. Pauli, *Z. f. Phys.* **37**, 263 (1926); **43**, 601 (1927).

20. C. G. Darwin, *Proc. Roy. Soc.* **A118**, 654 (1928); *ibid.*, **A120**, 621 (1928).

21. W. Gordon, *Z. f. Phys.* **48**, 11 (1928).

22. P. A. M. Dirac, *Proc. Roy. Soc.* **A126**, 360 (1930); also see Ref. 47.

23. E. C. Stoner, *Phil. Mag.* **48**, 719 (1924).

24. W. Pauli, *Z. f. Phys.* **31**, 765 (1925).

25. W. Heisenberg, *Z. f. Phys.* **38**, 411 (1926); *ibid.*, **39**, 499 (1926); P. A. M. Dirac, *Proc. Roy. Soc.* **A112**, 661 (1926); W. Pauli, *Z. f. Phys.* **41**, 81 (1927); J. C. Slater, *Phys. Rev.* **34**, 1293 (1929).

26. E. Fermi, *Z. f. Phys.* **36**, 902 (1926); *Rend. Accad. Lincei* **3**, 145 (1926).

27. P. A. M. Dirac, Ref. 25.

27*a*. P. A. M. Dirac, First W. R. Crane Lecture at the University of Michigan, April 17, 1978, unpublished.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

28. H. Weyl, *The Theory of Groups and Quantum Mechanics*, translated from the second (1931) German edition by H. P. Robertson (Dover Publications, Inc., New York): Chapter IV, Section 12. Also see P. A. M. Dirac, *Proc. Roy. Soc.* **A133**, 61 (1931).

29. J. R. Oppenheimer, *Phys. Rev.* **35**, 562 (1930); I. Tamm, *Z. f. Phys.* **62**, 545 (1930).

29*a*. P. A. M. Dirac, *Proc. Roy. Soc.* **133**, 60 (1931).

30. C. D. Anderson, *Science* **76**, 238 (1932); *Phys. Rev.* **43**, 491 (1933). The latter paper is reprinted in *Foundations of Nuclear Physics*, ed. by R. T. Beyer (Dover Publications, Inc., New York, 1949).

30*a*. J. Schwinger, 'A Report on Quantum Electrodynamics,' in *The Physicist's Conception of Nature* (Reidel, Dordrecht, 1973): p.415.

31. W. Pauli, *Handbuch der Physik* (Julius Springer, Berlin, 1932–1933); *Rev. Mod. Phys.* **13**, 203 (1941).

32. Born, Heisenberg, and Jordan, Ref. 4, Section 3.

32*a*. P. Ehrenfest, *Phys. Z.* **7**, 528 (1906).

33. Born and Jordan, Ref. 4. Unfortunately the relevant parts of this paper are not included in the reprint collection *Sources of Quantum Mechanics*, cited in Ref. 4.

34. P. A. M. Dirac, *Proc. Roy. Soc.* **A112**, 661 (1926): Section 5. For a more accessible derivation, see L. I. Schiff, *Quantum Mechanics*, 3rd edn. (McGraw-Hill Book Company, New York, 1968): Section 44.

34*a*. A. Einstein, *Phys. Z.* **18**, 121 (1917); reprinted in English in van der Waerden, Ref. 4.

35. P. A. M. Dirac, *Proc. Roy. Soc.* **A114**, 243 (1927); reprinted in *Quantum Electrodynamics*, ed. by J. Schwinger (Dover Publications, Inc., New York, 1958).

36. P. A. M. Dirac, *Proc. Roy. Soc.* **A114**, 710 (1927).

36*a*. V. F. Weisskopf and E. Wigner, *Z. f. Phys.* **63**, 54 (1930)

36*b*. E. Fermi, *Lincei Rend.* **9**, 881 (1929); **12**, 431 (1930); *Rev. Mod. Phys.* **4**, 87 (1932).

37. P. Jordan and W. Pauli, *Z. f. Phys.* **47**, 151 (1928).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

38. N. Bohr and L. Rosenfeld, *Kon. dansk. vid. Selsk., Mat.-Fys. Medd.* **XII**, No. 8 (1933) (translation in *Selected Papers of Leon Rosenfeld*, ed. by R. S. Cohen and J. Stachel (Reidel, Dordrecht, 1979)); *Phys. Rev.* **78**, 794 (1950).

39. P. Jordan, *Z. f. Phys.* **44**, 473 (1927). Also see P. Jordan and O. Klein, *Z. f. Phys.* **45**, 751 (1929); P. Jordan, *Phys. Zeit.* **30**, 700 (1929).

40. P. Jordan and E. Wigner, *Z. f. Phys.* **47**, 631 (1928). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

40*a*. M. Fierz, *Helv. Phys. Acta* **12**, (1939); W. Pauli, *Phys. Rev.* **58**, 716 (1940); W. Pauli and F. J. Belinfante, *Physica* **7**, 177 (1940).

41. W. Heisenberg and W. Pauli, *Z. f. Phys.* **56**, 1 (1929); *ibid.*, **59**, 168 (1930).

42. P. A. M. Dirac, *Proc. Roy. Soc.* **A136**, 453 (1932); P. A. M. Dirac, V. A. Fock, and B. Podolsky, *Phys. Zeit. der Sowjetunion* **2**, 468 (1932); P. A. M. Dirac, Phys. *Zeit. der Sowjetunion* **3**, 64 (1933). The latter two articles are reprinted in *Quantum Electrodynamics*, Ref. 35, pp. 29 and 312. Also see L. Rosenfeld, *Z. f. Phys.* **76**, 729 (1932).

42*a*. P. A. M. Dirac, *Proc. Roy. Soc.* London **A136**, 453 (1932).

43. E. Fermi, *Z. f. Phys.* **88**, 161 (1934). Fermi quotes unpublished work of Pauli for the proposition that an unobserved neutral particle is emitted along with the electron in beta decay. This particle was called the neutrino to distinguish it from the recently discovered neutron.

43*a*. V. Fock, *C. R. Leningrad* 1933, p. 267.

44. W. H. Furry and J. R. Oppenheimer, *Phys. Rev.* **45**, 245 (1934). This paper uses a density matrix formalism developed by P. A. M. Dirac, Proc. Camb. *Phil. Soc.* **30**, 150 (1934). Also see R. E. Peierls, *Proc. Roy. Soc.* **146**, 420 (1934); W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934); L. Rosenfeld, *Z. f. Phys.* **76**, 729 (1932).

45. W. Pauli and V. Weisskopf, *Helv. Phys. Acta* **7**, 709 (1934), reprinted in English translation in A. I. Miller, *Early Quantum Electrodynamics* (Cambridge University Press, Cambridge, 1994). Also see W. Pauli, *Ann. Inst. Henri Poincaré* **6**, 137 (1936).

46. O. Klein and Y. Nishina, *Z. f. Phys.* **52**, 853 (1929); Y. Nishina, *ibid.*, 869 (1929); also see I. Tamm, *Z. f. Phys.* **62**, 545 (1930).

47. P. A. M. Dirac, *Proc. Camb. Phil. Soc.* **26**, 361 (1930).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

48. C. Møller, *Ann. d. Phys.* **14**, 531, 568 (1932).

49. H. Bethe and W. Heitler, *Proc. Roy. Soc.* **A146**, 83 (1934); also see G. Racah, *Nuovo Cimento* **11**, No. 7 (1934); *ibid.*, **13**, 69 (1936).

50. H. J. Bhabha, *Proc. Roy. Soc.* **A154**, 195 (1936).

50*a*. J. F. Carlson and J. R. Oppenheimer, *Phys. Rev.* **51**, 220 (1937).

51. P. Ehrenfest and J. R. Oppenheimer, *Phys. Rev.* **37**, 333 (1931).

52. W. Heitler and G. Herzberg, *Naturwiss.* **17**, 673 (1929); F. Rasetti, *Z. f. Phys.* **61**, 598 (1930).

53. J. Chadwick, *Proc. Roy. Soc.* **A136**, 692 (1932). This article is reprinted in *The Foundations of Nuclear Physics*, Ref. 30.

54. W. Heisenberg, *Z. f. Phys.* **77**, 1 (1932); also see I. Curie-Joliot and F. Joliot, *Compt. Rend.* **194**, 273 (1932).

54*a*. For references, see L. M. Brown and H. Rechenberg, *Hist. Stud. in Phys. and Bio. Science*, **25**, 1 (1994).

55. H. Yukawa, *Proc. Phys.-Math. Soc. (Japan)* (3) **17**, 48 (1935). This article is reprinted in *The Foundations of Nuclear Physics*, Ref. 30.

56. S. H. Neddermeyer and C. D. Anderson, *Phys. Rev.* **51**, 884 (1937); J. C. Street and E. C. Stevenson, *Phys. Rev.* **52**, 1003 (1937).

56*a*. L. Nordheim and N. Webb, *Phys. Rev.* **56**, 494 (1939).

57. M. Conversi, E. Pancini, and O. Piccioni, *Phys. Rev.* **71**, 209L (1947).

58. S. Sakata and T. Inoue, *Prog. Theor. Phys.* **1**, 143 (1946); R. E. Marshak and H. A. Bethe, *Phys. Rev.* **77**, 506 (1947).

59. C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature* **160**, 453, 486 (1947).

60. G. D. Rochester and C. C. Butler, *Nature* **160**, 855 (1947).

61. J. R. Oppenheimer, *Phys. Rev.* **35**, 461 (1930).

62. I. Waller, *Z. f. Phys.* **59**, 168 (1930); *ibid.*, **61**, 721, 837 (1930); *ibid.*, **62**, 673 (1930).

63. V. F. Weisskopf, *Z. f. Phys.* **89**, 27 (1934), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45; *ibid.*, **90**, 817 (1934). The electromagnetic self energy is calculated only to lowest order in α in these references; the proof that the divergence

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

is only logarithmic in all orders of perturbation theory was given by Weisskopf; *Phys. Rev.* **56**, 72 (1939). (This last article is reprinted in *Quantum Electrodynamics*, Ref. 35).

64. P. A. M. Dirac, XVII Conseil Solvay de Physique, p. 203 (1933), reprinted in *Early Quantum Electrodynamics*, Ref. 45. For subsequent calculations based on less restrictive assumptions, see W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934); *Sachs. Akad. Wiss.* **86**, 317 (1934); R. Serber, *Phys. Rev.* **43**, 49 (1935); E. A. Uehling. *Phys. Rev.* **48**, 55 (1935); W. Pauli and M. Rose, *Phys. Rev.* **49**, 462 (1936). Also see Furry and Oppenheimer, Ref. 44; Peierls. Ref. 44; Weisskopf, Ref. 63.

65. H. Euler and B. Kockel, *Naturwiss.* **23**, 246 (1935); W. Heisenberg and H. Euler, *Z. f. Phys.* **98**, 714 (1936).

66. P. A. M. Dirac, *Proc. Camb. Phil. Soc.* **30**, 150 (1934).

67. W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934).

68. N. Kemmer and V. F. Weisskopf. *Nature* **137**, 659 (1936).

68*a*. F. Bloch and A. Nordsieck, *Phys. Rev.* **52**, 54 (1937). Also see W. Pauli and M. Fierz, *Nuovo Cimento* **15**, 167 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45.

69. S. M. Dancoff, *Phys. Rev.* **55**, 959 (1939).

69*a*. H. W. Lewis, *Phys. Rev.* **73**, 173 (1948); S. Epstein, *Phys. Rev.* **73**, 177 (1948). Also see J. Schwinger, Ref. 84; Z. Koba and S. Tomonaga, *Prog. Theor. Phys.* **3**/3, 290 (1948).

69*b*. J. Schwinger, in *The Birth of Particle Physics*, ed. by L. Brown and L. Hoddeson (Cambridge University Press, Cambridge, 1983): p. 336.

70. W. Heisenberg, *Ann. d. Phys.* **32**, 20 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45.

70*a*. G. Wentzel, *Z. f. Phys.* **86**, 479, 635 (1933); *Z. f. Phys.* **87**, 726 (1034); M. Born and L. Infeld, *Proc. Roy. Soc.* **A150**. 141 (1935); W. Pauli, *Ann. Inst. Henri Poincaré* **6**, 137 (1936).

71. J. A. Wheeler, *Phys. Rev.* **52**, 1107 (1937).

72. W. Heisenberg, *Z. f. Phys.* **120**, 513, 673 (1943); *Z. Naturforsch.* **1**, 608 (1946). Also see C. Møller, *Kon. Dansk. Vid. Sels. Mat.-Fys. Medd.* **23**, No. 1 (1945); *ibid.* **23**, No. 19, (1946).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

73. See, e.g., G. Chew, *The S-Matrix Theory of Strong Interactions* (W. A. Benjamin, Inc. New York, 1961).

74. J. A. Wheeler and R. P. Feynman, *Rev. Mod. Phys.* **17**, 157 (1945), *ibid.*, **21**, 425 (1949). For further references and a discussion of the application of action-at-a-distance theories in cosmology, see S. Weinberg *Gravitation and Cosmology*, (Wiley, 1972): Section 16.3.

75. P. A. M. Dirac, *Proc. Roy. Soc.* **A180**, 1 (1942). For a criticism, see W. Pauli, *Rev. Mod. Phys.* **15**, 175 (1943). For a review of classical theories of this type, and of yet other attempts to solve the problem of infinities, see R. E. Peierls in *Rapports du 8^{m}e Conseil de Physique Solvay 1948* (R. Stoops, Brussels, 1950): p. 241.

76. V. F. Weisskopf, *Kon. Dan. Vid. Sel., Mat.-fys. Medd.* **XIV**, No. 6 (1936), especially p. 34 and pp. 5–6. This article is reprinted in *Quantum Electrodynamics*, Ref. 35, and in English translation in *Early Quantum Electrodynamics*, Ref. 45. Also see W. Pauli and M. Fierz, Ref. 68*a*; H. A. Kramers, Ref. 79*a*.

77. S. Pasternack, *Phys. Rev.* **54**, 1113 (1938). This suggestion was based on experiments of W.V. Houston, *Phys. Rev.* **51**, 446 (1937); R. C. Williams, *Phys. Rev.* **54**, 558 (1938). For a report of contrary data, see J. W. Drinkwater, O. Richardson, and W. E. Williams, *Proc. Roy. Soc.* **174**, 164 (1940).

78. E. A. Uehling, Ref. 64.

79. W. E. Lamb, Jr and R. C. Retherford, *Phys. Rev.* **72**, 241 (1947). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

79*a*. H. A. Kramers, *Nuovo Cimento* **15**, 108 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45; *Ned. T. Natwink.* **11**, 134 (1944); *Rapports du 8^{m}e Conseil de Physique Solvay 1948* (R. Stoops, Brussels, 1950).

80. H. A. Bethe, *Phys. Rev.* **72**, 339 (1947). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

81. J. B. French and V. F. Weisskopf; *Phys. Rev.* **75**, 1240 (1949); N. M. Kroll and W. E. Lamb, *ibid.*, **75**, 388 (1949); J. Schwinger, *Phys. Rev.* **75**, 898 (1949); R. P. Feynman, *Rev. Mod. Phys.* **20**, 367 (1948); *Phys. Rev.*, **74**, 939, 1430 (1948); **76**, 749, 769 (1949); **80**, 440 (1950); H. Fukuda, Y. Miyamoto, and S. Tomonaga, *Prog. Theor. Phys. Rev. Mod. Phys.* **4**, 47, 121 (1948). The article by Kroll and Lamb is reprinted in *Quantum Electrodynamics*, Ref. 35.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

82. J. E. Nafe, E. B. Nelson, and I. I. Rabi, *Phys. Rev.* **71**, 914 (1947); D. E. Nagel, R. S. Julian, and J. R. Zacharias, *Phys. Rev.* **72**, 973 (1947).

83. P. Kusch and H. M. Foley, *Phys. Rev.* **72**, 1256 (1947).

83*a*. G. Breit, *Phys. Rev.* **71**, 984 (1947). Schwinger in Ref. 84 includes a corrected version of Breit's results.

84. J. Schwinger, *Phys. Rev.* **73**, 416 (1948). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

85. J. Schwinger, *Phys. Rev.* **74**, 1439 (1948); *ibid.*, **75**, 651 (1949); *ibid.*, **76**, 790 (1949); *ibid.*, **82**, 664, 914 (1951); *ibid.*, **91**, 713 (1953); *Proc. Nat. Acad. Sci.* **37**, 452 (1951). All but the first two of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

86. S. Tomonaga, *Prog. Theor. Phys. Rev. Mod. Phys.* **1**, 27 (1946); Z. Koba, T. Tati, and S. Tomonaga, *ibid.*, **2**, 101 (1947); S. Kanesawa and S. Tomonaga, *ibid.*, **3**, 1, 101 (1948); S. Tomonaga, *Phys. Rev.* **74**, 224 (1948); D. Ito, Z. Koba, and S. Tomonaga, *Prog. Theor. Phys.* **3**, 276 (1948); Z. Koba and S. Tomonaga, *ibid.*, **3**, 290 (1948). The first and fourth of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

87. R. P. Feynman, *Rev. Mod. Phys.* **20**, 367 (1948); *Phys. Rev.* **74**, 939, 1430 (1948); *ibid.*, **76**, 749, 769 (1949); *ibid* **80**, 440 (1950). All but the second and third of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

88. F. J. Dyson, *Phys. Rev.* **75**, 486, 1736 (1949). These articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

88*a*. H. Fröhlich, W. Heitler, and B. Kahn, *Proc. Roy. Soc.* **A171**, 269 (1939); *Phys. Rev.* **56**, 961 (1939).

88*b*. W. E. Lamb, Jr, *Phys. Rev.* **56**, 384 (1939); *Phys. Rev.* **57**, 458 (1940).

89. Quoted by R. Serber, in *The Birth of Particle Physics*, Ref. 69*b*, p. 270.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.003

2

Relativistic Quantum Mechanics

The point of view of this book is that quantum field theory is the way it is because (with certain qualifications) this is the only way to reconcile quantum mechanics with special relativity. Therefore our first task is to study how symmetries like Lorentz invariance appear in a quantum setting.

2.1 Quantum Mechanics

First, some good news: quantum field theory is based on the same quantum mechanics that was invented by Schrödinger, Heisenberg, Pauli, Born, and others in 1925–26, and has been used ever since in atomic, molecular, nuclear, and condensed matter physics. The reader is assumed to be already familiar with quantum mechanics; this section provides only the briefest of summaries of quantum mechanics, in the generalized version of Dirac.[1]

(i) Physical states are represented by rays in Hilbert space. A Hilbert space is a kind of complex vector space; that is, if Φ and Ψ are vectors in the space (often called 'state-vectors') then so is $\xi\Phi + \eta\Psi$, for arbitrary complex numbers ξ, η. It has a norm*: for any pair of vectors there is a complex number (Φ, Ψ), such that

$$(\Phi, \Psi) = (\Psi, \Phi)^* , \tag{2.1.1}$$

$$(\Phi, \xi_1\Psi_1 + \xi_2\Psi_2) = \xi_1(\Phi, \Psi_1) + \xi_2(\Phi, \Psi_2) , \tag{2.1.2}$$

$$(\eta_1\Phi_1 + \eta_2\Phi_2, \Psi) = \eta_1^*(\Phi_1, \Psi) + \eta_2^*(\Phi_2, \Psi) . \tag{2.1.3}$$

The norm (Ψ, Ψ) also satisfies a positivity condition: $(\Psi, \Psi) \geq 0$, and vanishes if and only if $\Psi = 0$. (There are also certain technical assumptions that allow us to take limits of vectors within Hilbert space.) A *ray* is a

* We shall often use the Dirac bra-ket notation: instead of (Ψ_1, Ψ_2), we may write $\langle 1|2\rangle$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

set of normalized vectors (i.e., $(\Psi, \Psi) = 1$) with Ψ and Ψ' belonging to the same ray if $\Psi' = \xi\Psi$, where ξ is an arbitrary complex number with $|\xi| = 1$.

(ii) Observables are represented by Hermitian operators. These are mappings $\Psi \to A\Psi$ of Hilbert space into itself, linear in the sense that

$$A(\xi\Psi + \eta\Phi) = \xi A\Psi + \eta A\Phi \tag{2.1.4}$$

and satisfying the reality condition $A^\dagger = A$, where for any linear operator A the adjoint $A^\dagger$ is defined by

$$(\Phi, A^\dagger\Psi) \equiv (A\Phi, \Psi) = (\Psi, A\Phi)^* . \tag{2.1.5}$$

(There are also technical assumptions about the continuity of $A\Psi$ as a function of Ψ.) A state represented by a ray $\mathscr{R}$ has a definite value α for the observable represented by an operator A if vectors Ψ belonging to this ray are eigenvectors of A with eigenvalue α:

$$A\Psi = \alpha\Psi \quad \text{for} \quad \Psi \text{ in } \mathscr{R} . \tag{2.1.6}$$

An elementary theorem tells us that for A Hermitian, α is real, and eigenvectors with different αs are orthogonal.

(iii) If a system is in a state represented by a ray $\mathscr{R}$, and an experiment is done to test whether it is in any one of the different states represented by mutually orthogonal rays $\mathscr{R}_1, \mathscr{R}_2, \ldots$ (for instance, by measuring one or more observables) then the probability of finding it in the state represented by $\mathscr{R}_n$ is

$$P(\mathscr{R} \to \mathscr{R}_n) = |(\Psi, \Psi_n)|^2 , \tag{2.1.7}$$

where Ψ and Ψ_n are any vectors belonging to rays $\mathscr{R}$ and $\mathscr{R}_n$, respectively. (A pair of rays is said to be orthogonal if the state-vectors from the two rays have vanishing scalar products.) Another elementary theorem gives a total probability unity:

$$\sum_n P(\mathscr{R} \to \mathscr{R}_n) = 1 \tag{2.1.8}$$

if the state-vectors Ψ_n form a complete set.

2.2 Symmetries

A symmetry transformation is a change in our point of view that does not change the results of possible experiments. If an observer O sees a system in a state represented by a ray $\mathscr{R}$ or $\mathscr{R}_1$ or $\mathscr{R}_2 \ldots$, then an equivalent observer O' who looks at the *same* system will observe it in a different

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

state, represented by a ray $\mathscr{R}'$ or $\mathscr{R}'_1$ or $\mathscr{R}'_2 \ldots$, respectively, but the two observers must find the same probabilities

$$P(\mathscr{R} \to \mathscr{R}_n) = P(\mathscr{R}' \to \mathscr{R}'_n) \,. \tag{2.2.1}$$

(This is only a necessary condition for a ray transformation to be a symmetry; further conditions are discussed in the next chapter.) An important theorem proved by Wigner[2] in the early 1930s tells us that for any such transformation $\mathscr{R} \to \mathscr{R}'$ of rays we may define an operator U on Hilbert space, such that if Ψ is in ray $\mathscr{R}$ then $U\Psi$ is in the ray $\mathscr{R}'$, with U either *unitary* and *linear*

$$(U\Phi, U\Psi) = (\Phi, \Psi) \,, \tag{2.2.2}$$

$$U(\xi\Phi + \eta\Psi) = \xi U\Phi + \eta U\Psi \tag{2.2.3}$$

or else *antiunitary* and *antilinear*

$$(U\Phi, U\Psi) = (\Phi, \Psi)^* \,, \tag{2.2.4}$$

$$U(\xi\Phi + \eta\Psi) = \xi^* U\Phi + \eta^* U\Psi \,. \tag{2.2.5}$$

Wigner's proof omits some steps. A more complete proof is given at the end of this chapter in Appendix A.

As already mentioned, the adjoint of a linear operator L is defined by

$$(\Phi, L^\dagger\Psi) \equiv (L\Phi, \Psi) \,. \tag{2.2.6}$$

This condition cannot be satisfied for an antilinear operator, because in this case the right-hand side of Eq. (2.2.6) would be linear in Φ, while the left-hand side is antilinear in Φ. Instead, the adjoint of an antilinear operator A is defined by

$$(\Phi, A^\dagger\Psi) \equiv (A\Phi, \Psi)^* = (\Psi, A\Phi) \,. \tag{2.2.7}$$

With this definition, the conditions for unitarity or antiunitarity both take the form

$$U^\dagger = U^{-1} \,. \tag{2.2.8}$$

There is always a trivial symmetry transformation $\mathscr{R} \to \mathscr{R}$, represented by the identity operator $U = 1$. This operator is, of course, unitary and linear. Continuity then demands that any symmetry (like a rotation or translation or Lorentz transformation) that can be made trivial by a continuous change of some parameters (like angles or distances or velocities) must be represented by a linear unitary operator U rather than one that is antilinear and antiunitary. (Symmetries represented by antiunitary antilinear operators are less prominent in physics; they all involve a reversal in the direction of time's flow. See Section 2.6.)

In particular, a symmetry transformation that is infinitesimally close to being trivial can be represented by a linear unitary operator that is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

infinitesimally close to the identity:

$$U = 1 + i\epsilon t \tag{2.2.9}$$

with ϵ a real infinitesimal. For this to be unitary and linear, t must be Hermitian and linear, so it is a candidate for an observable. Indeed, most (and perhaps all) of the observables of physics, such as angular momentum or momentum, arise in this way from symmetry transformations.

The set of symmetry transformations has certain properties that define it as a *group*. If T_1 is a transformation that takes rays $\mathscr{R}_n$ into $\mathscr{R}'_n$, and T_2 is another transformation that takes $\mathscr{R}'_n$ into $\mathscr{R}''_n$, then the result of performing both transformations is another symmetry transformation, which we write T_2T_1, that takes $\mathscr{R}_n$ into $\mathscr{R}''_n$. Also, a symmetry transformation T which takes rays $\mathscr{R}_n$ into $\mathscr{R}'_n$ has an inverse, written T^{-1}, which takes $\mathscr{R}'_n$ into $\mathscr{R}_n$, and there is an identity transformation, $T = 1$, which leaves rays unchanged.

The unitary or antiunitary operators $U(T)$ corresponding to these symmetry transformations have properties that mirror this group structure, but with a complication due to the fact that, unlike the symmetry transformations themselves, the operators $U(T)$ act on vectors in the Hilbert space, rather than on rays. If T_1 takes $\mathscr{R}_n$ into $\mathscr{R}'_n$, then acting on a vector Ψ_n in the ray $\mathscr{R}_n$, $U(T_1)$ must yield a vector $U(T_1)\Psi_n$ in the ray $\mathscr{R}'_n$, and if T_2 takes this ray into $\mathscr{R}''_n$, then acting on $U(T_1)\Psi_n$ it must yield a vector $U(T_2)U(T_1)\Psi_n$ in the ray $\mathscr{R}''_n$. But $U(T_2T_1)\Psi_n$ is also in this ray, so these vectors can differ only by a phase $\phi_n(T_2, T_1)$

$$U(T_2)U(T_1)\Psi_n = e^{i\phi_n(T_2,T_1)}U(T_2T_1)\Psi_n\,. \tag{2.2.10}$$

Furthermore, with one significant exception, the linearity (or antilinearity) of $U(T)$ tells us that these phases are independent of the state Ψ_n. Here is the proof. Consider any two different vectors Ψ_A, Ψ_B, which are not proportional to each other. Then, applying Eq. (2.2.10) to the state $\Psi_{AB} \equiv \Psi_A + \Psi_B$, we have

$$\begin{aligned} e^{i\phi_{AB}}\, U(T_2T_1)(\Psi_A + \Psi_B) &= U(T_2)U(T_1)(\Psi_A + \Psi_B) \\ &= U(T_2)U(T_1)\Psi_A + U(T_2)U(T_1)\Psi_B \\ &= e^{i\phi_A}U(T_2T_1)\Psi_A + e^{i\phi_B}U(T_2T_1)\Psi_B. \end{aligned} \tag{2.2.11}$$

Any unitary or antiunitary operator has an inverse (its adjoint) which is also unitary or antiunitary. Multiplying (2.2.11) on the left with $U^{-1}(T_2T_1)$, we have then

$$e^{\pm i\phi_{AB}}(\Psi_A + \Psi_B) = e^{\pm i\phi_A}\Psi_A + e^{\pm i\phi_B}\Psi_B\,, \tag{2.2.12}$$

the upper and lower signs referring to $U(T_2T_1)$ unitary or antiunitary, respectively. Since Ψ_A and Ψ_B are linearly independent, this is only

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

possible if

$$e^{i\phi_{AB}} = e^{i\phi_A} = e^{i\phi_B} . \tag{2.2.13}$$

So as promised, the phase in Eq. (2.2.10) is independent of the state-vector Ψ_n, and therefore this can be written as an operator relation

$$U(T_2)U(T_1) = e^{i\phi(T_2,T_1)}U(T_2T_1) . \tag{2.2.14}$$

For $\phi = 0$, this would say that the $U(T)$ furnish a representation of the group of symmetry transformations. For general phases $\phi(T_2, T_1)$, we have what is called a projective representation, or a representation 'up to a phase'. The structure of the Lie group cannot by itself tell us whether physical state-vectors furnish an ordinary or a projective representation, but as we shall see, it can tell us whether the group has any intrinsically projective representations at all.

The exception to the argument that led to Eq. (2.2.14) is that it may not be possible to prepare the system in a state represented by $\Psi_A + \Psi_B$. For instance, it is widely believed to be impossible to prepare a system in a superposition of two states whose total angular momenta are integers and half-integers, respectively. In such cases, we say that there is a 'superselection rule' between different classes of states,[3] and the phases $\phi(T_2, T_1)$ *may* depend on which of these classes of states the operators $U(T_2)U(T_1)$ and $U(T_2, T_1)$ act upon. We will have more to say about these phases and projective representations in Section 2.7. As we shall see there, any symmetry group with projective representations can always be enlarged (without otherwise changing its physical implications) in such a way that its representations can all be defined as non-projective, with $\phi = 0$. Until Section 2.7, we will just assume that this has been done, and take $\phi = 0$ in Eq. (2.2.14).

There is a kind of group, known as a *connected Lie group*, of special importance in physics. These are groups of transformations $T(\theta)$ that are described by a finite set of real continuous parameters, say θ^a, with each element of the group connected to the identity by a path within the group. The group multiplication law then takes the form

$$T(\bar{\theta})T(\theta) = T\Big(f(\bar{\theta},\theta)\Big) \tag{2.2.15}$$

with $f^a(\bar{\theta},\theta)$ a function of the $\bar{\theta}$s and θs. Taking $\theta^a = 0$ as the coordinates of the identity, we must have

$$f^a(\theta,0) = f^a(0,\theta) = \theta^a . \tag{2.2.16}$$

As already mentioned, the transformations of such continuous groups must be represented on the physical Hilbert space by unitary (rather than antiunitary) operators $U(T(\theta))$. For a Lie group, these operators can be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

represented in at least a finite neighborhood of the identity by a power series

$$U\Big(T(\theta)\Big) = 1 + i\theta^a t_a + \tfrac{1}{2}\,\theta^b\theta^c t_{bc} + \cdots , \tag{2.2.17}$$

where t_a, $t_{bc} = t_{cb}$, etc. are operators independent of the θs, with t_a Hermitian. Suppose that the $U(T(\theta))$ form an ordinary (i.e., not projective) representation of this group of transformations, i.e.,

$$U\Big(T(\bar\theta)\Big)\,U\Big(T(\theta)\Big) = U\Big(T(f(\bar\theta,\theta))\Big) . \tag{2.2.18}$$

Let us see what this condition looks like when expanded in powers of θ^a and $\bar\theta^a$. According to Eq. (2.2.16), the expansion of $f^a(\bar\theta,\theta)$ to second order must take the form

$$f^a(\bar\theta,\theta) = \theta^a + \bar\theta^a + f^a{}_{bc}\bar\theta^b\theta^c + \cdots \tag{2.2.19}$$

with real coefficients $f^a{}_{bc}$. (The presence of any terms of order θ^2 or $\bar\theta^2$ would violate Eq. (2.2.16).) Then Eq. (2.2.18) reads

$$\begin{aligned}\Big[1 + i\bar\theta^a t_a + \tfrac{1}{2}\,\bar\theta^b\bar\theta^c t_{bc} + \cdots\Big] \times \Big[1 + i\theta^a t_a + \tfrac{1}{2}\theta^b\theta^c t_{bc} + \cdots\Big] \\ = 1 + i\Big(\theta^a + \bar\theta^a + f^a{}_{bc}\bar\theta^b\theta^c + \cdots\Big)\,t_a \\ + \tfrac{1}{2}(\theta^b + \bar\theta^b + \cdots)\,(\theta^c + \bar\theta^c + \cdots)t_{bc} + \cdots\end{aligned} \tag{2.2.20}$$

The terms of order $1, \theta, \bar\theta, \theta^2$, and $\bar\theta^2$ automatically match on both sides of Eq. (2.2.20), but from the $\bar\theta\theta$ terms we obtain a non-trivial condition

$$t_{bc} = -t_b t_c - i f^a{}_{bc}\, t_a . \tag{2.2.21}$$

This shows that if we are given the structure of the group, i.e., the function $f(\theta,\bar\theta)$, and hence its quadratic coefficient $f^a{}_{bc}$, we can calculate the second-order terms in $U(T(\theta))$ from the generators t_a appearing in the first-order terms. However, there is a consistency condition: the operator t_{bc} must be *symmetric* in b and c (because it is the second derivative of $U(T(\theta))$ with respect to θ^b and θ^c) so Eq. (2.2.21) requires that

$$[t_b,\ t_c] = i\,C^a{}_{bc}\,t_a , \tag{2.2.22}$$

where $C^a{}_{bc}$ are a set of real constants known as *structure constants*

$$C^a{}_{bc} \equiv -f^a{}_{bc} + f^a{}_{cb} . \tag{2.2.23}$$

Such a set of commutation relations is known as a *Lie algebra*. In Section 2.7 we will prove in effect that the commutation relation (2.2.22) is the single condition needed to ensure that this process can be continued: the complete power series for $U(T(\theta))$ may be calculated from an infinite sequence of relations like Eq. (2.2.21), provided we know the first-order terms, the generators t_a. This does not necessarily mean that the operators $U(T(\theta))$ are uniquely determined for all θ^a if we know the t_a, but it

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

does mean that the $U(T(\theta))$ are uniquely determined in at least a finite neighborhood of the coordinates $\theta^a = 0$ of the identity, in such a way that Eq. (2.2.15) is satisfied if $\theta, \bar{\theta}$, and $f(\theta, \bar{\theta})$ are in this neighborhood. The extension to all θ^a is discussed in Section 2.7.

There is a special case of some importance, that we will encounter again and again. Suppose that the function $f(\theta, \bar{\theta})$ (perhaps just for some subset of the coordinates θ^a) is simply additive

$$f^a(\theta, \bar{\theta}) = \theta^a + \bar{\theta}^a \,. \tag{2.2.24}$$

This is the case for instance for translations in spacetime, or for rotations about any one fixed axis (though *not* for both together). Then the coefficients $f^a{}_{bc}$ in Eq. (2.2.19) vanish, and so do the structure constants (2.2.23). The generators then all commute

$$[t_b, t_c] = 0 \,. \tag{2.2.25}$$

Such a group is called *Abelian*. In this case, it is easy to calculate $U(T(\theta))$ for all θ^a. From Eqs. (2.2.18) and (2.2.24), we have for any integer N

$$U\Big(T(\theta)\Big) = \left[U\left(T\left(\frac{\theta}{N}\right)\right)\right]^N \,.$$

Letting $N \to \infty$, and keeping only the first-order term in $U(T(\theta/N))$, we have then

$$U\Big(T(\theta)\Big) = \lim_{N\to\infty}\left[1 + \frac{i}{N}\theta^a t_a\right]^N$$

and hence

$$U\Big(T(\theta)\Big) = \exp(it_a\theta^a) \,. \tag{2.2.26}$$

2.3 Quantum Lorentz Transformations

Einstein's principle of relativity states the equivalence of certain 'inertial' frames of reference. It is distinguished from the Galilean principle of relativity, obeyed by Newtonian mechanics, by the transformation connecting coordinate systems in different inertial frames. If x^μ are the coordinates in one inertial frame (with x^1, x^2, x^3 Cartesian space coordinates, and $x^0 = t$ a time coordinate, the speed of light being set equal to unity) then in any other inertial frame, the coordinates x'^μ must satisfy

$$\eta_{\mu\nu} dx'^\mu dx'^\nu = \eta_{\mu\nu} dx^\mu dx^\nu \tag{2.3.1}$$

or equivalently

$$\eta_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x'^\nu}{\partial x^\sigma} = \eta_{\rho\sigma} \,. \tag{2.3.2}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

Here $\eta_{\mu\nu}$ is the diagonal matrix, with elements

$$\eta_{11} = \eta_{22} = \eta_{33} = +1, \quad \eta_{00} = -1 \tag{2.3.3}$$

and the summation convention is in force: we sum over any index like μ and ν in Eq. (2.3.2), which appears twice in the same term, once upstairs and once downstairs. These transformations have the special property that the speed of light is the same (in our units, equal to unity) in all inertial frames;* a light wave travelling at unit speed satisfies $|d\mathbf{x}/dt| = 1$, or in other words $\eta_{\mu\nu}dx^\mu dx^\nu = d\mathbf{x}^2 - dt^2 = 0$, from which it follows that also $\eta_{\mu\nu}dx'^\mu dx'^\nu = 0$, and hence $|d\mathbf{x}'/dt'| = 1$.

Any coordinate transformation $x^\mu \to x'^\mu$ that satisfies Eq. (2.3.2) is *linear*[3a]

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu \tag{2.3.4}$$

with a^μ arbitrary constants, and $\Lambda^\mu{}_\nu$ a constant matrix satisfying the conditions

$$\eta_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma = \eta_{\rho\sigma} \,. \tag{2.3.5}$$

For some purposes, it is useful to write the Lorentz transformation condition in a different way. The matrix $\eta_{\mu\nu}$ has an inverse, written $\eta^{\mu\nu}$, which happens to have the same components: it is diagonal, with $\eta^{00} = -1$, $\eta^{11} = \eta^{22} = \eta^{33} = +1$. Multiplying Eq. (2.3.5) with $\eta^{\sigma\tau}\Lambda^\kappa{}_\tau$ and inserting parentheses judiciously, we have

$$\eta_{\mu\nu}\Lambda^\mu{}_\rho(\Lambda^\nu{}_\sigma\Lambda^\kappa{}_\tau\eta^{\sigma\tau}) = \Lambda^\kappa{}_\rho = \eta_{\mu\nu}\eta^{\nu\kappa}\Lambda^\mu{}_\rho \,.$$

Multiplying with the inverse of the matrix $\eta_{\mu\nu}\Lambda^\mu{}_\rho$ then gives

$$\Lambda^\nu{}_\sigma\Lambda^\kappa{}_\tau\eta^{\sigma\tau} = \eta^{\nu\kappa} \,. \tag{2.3.6}$$

These transformations form a group. If we first perform a Lorentz transformation (2.3.4), and then a second Lorentz transformation $x'^\mu \to x''^\mu$, with

$$x''^\mu = \bar\Lambda^\mu{}_\rho x'^\rho + \bar a^\mu = \bar\Lambda^\mu{}_\rho(\Lambda^\rho{}_\nu x^\nu + a^\rho) + \bar a^\mu$$

then the effect is the same as the Lorentz transformation $x^\mu \to x''^\mu$, with

$$x''^\mu = (\bar\Lambda^\mu{}_\rho\Lambda^\rho{}_\nu)x^\nu + (\bar\Lambda^\mu{}_\rho a^\rho + \bar a^\mu) \,. \tag{2.3.7}$$

(Note that if $\Lambda^\mu{}_\nu$ and $\bar\Lambda^\mu{}_\nu$ both satisfy Eq. (2.3.5), then so does $\bar\Lambda^\mu{}_\rho\Lambda^\rho{}_\nu$, so this *is* a Lorentz transformation. The bar is used here just to distinguish

* There is a larger class of coordinate transformations, known as *conformal transformations*, for which $\eta_{\mu\nu}dx'^\mu dx'^\nu$ is proportional though generally not equal to $\eta_{\mu\nu}dx^\mu dx^\nu$, and which therefore also leave the speed of light invariant. Conformal invariance in two dimensions has proved enormously important in string theory and statistical mechanics, but the physical relevance of these conformal transformations in four spacetime dimensions is not yet clear.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

one Lorentz transformation from the other.) The transformations $T(\Lambda, a)$ induced on physical states therefore satisfy the composition rule

$$T(\bar{\Lambda}, \bar{a})T(\Lambda, a) = T(\bar{\Lambda}\Lambda, \bar{\Lambda}a + \bar{a}) . \tag{2.3.8}$$

Taking the determinant of Eq. (2.3.5) gives

$$(\mathrm{Det}\Lambda)^2 = 1 \tag{2.3.9}$$

so $\Lambda^{\mu}{}_{\nu}$ has an inverse, $(\Lambda^{-1})^{\nu}{}_{\rho}$ which we see from Eq. (2.3.5) takes the form

$$(\Lambda^{-1})^{\rho}{}_{\nu} = \Lambda_{\nu}{}^{\rho} \equiv \eta_{\nu\mu}\eta^{\rho\sigma}\Lambda^{\mu}{}_{\sigma} . \tag{2.3.10}$$

The inverse of the transformation $T(\Lambda, a)$ is seen from Eq. (2.3.8) to be $T(\Lambda^{-1}, -\Lambda^{-1}a)$, and, of course, the identity transformation is $T(1, 0)$.

In accordance with the discussion in the previous section, the transformations $T(\Lambda, a)$ induce a unitary linear transformation on vectors in the physical Hilbert space

$$\Psi \to U(\Lambda, a)\Psi .$$

The operators U satisfy a composition rule

$$U(\bar{\Lambda}, \bar{a})U(\Lambda, a) = U(\bar{\Lambda}\Lambda, \bar{\Lambda}a + \bar{a}) . \tag{2.3.11}$$

(As already mentioned, to avoid the appearance of a phase factor on the right-hand side of Eq. (2.3.11), it is, in general, necessary to enlarge the Lorentz group. The appropriate enlargement is described in Section 2.7.)

The whole group of transformations $T(\Lambda, a)$ is properly known as the *inhomogeneous Lorentz group*, or *Poincaré group*. It has a number of important subgroups. First, those transformations with $a^{\mu} = 0$ obviously form a subgroup, with

$$T(\bar{\Lambda}, 0)\ T(\Lambda, 0) = T(\bar{\Lambda}\Lambda, 0), \tag{2.3.12}$$

known as the *homogeneous Lorentz group*. Also, we note from Eq. (2.3.9) that either $\mathrm{Det}\Lambda = +1$ or $\mathrm{Det}\Lambda = -1$; those transformations with $\mathrm{Det}\Lambda = +1$ obviously form a subgroup of either the homogeneous or the inhomogeneous Lorentz groups. Further, from the 00-components of Eqs. (2.3.5) and (2.3.6), we have

$$(\Lambda^0{}_0)^2 = 1 + \Lambda^i{}_0\Lambda^i{}_0 = 1 + \Lambda^0{}_i\Lambda^0{}_i . \tag{2.3.13}$$

with i summed over the values 1, 2, and 3. We see that either $\Lambda^0{}_0 \geq +1$ or $\Lambda^0{}_0 \leq -1$. Those transformations with $\Lambda^0{}_0 \geq +1$ form a subgroup. Note that if $\Lambda^{\mu}{}_{\nu}$ and $\bar{\Lambda}^{\mu}{}_{\nu}$ are two such Λs, then

$$(\bar{\Lambda}\Lambda)^0{}_0 = \bar{\Lambda}^0{}_0\Lambda^0{}_0 + \bar{\Lambda}^0{}_1\Lambda^1{}_0 + \bar{\Lambda}^0{}_2\Lambda^2{}_0 + \bar{\Lambda}^0{}_3\Lambda^3{}_0\ ;.$$

But Eq. (2.3.13) shows that the three-vector $(\Lambda^1{}_0, \Lambda^2{}_0, \Lambda^3{}_0)$ has length $\sqrt{(\Lambda^0{}_0)^2 - 1}$, and similarly the three-vector $(\bar{\Lambda}^0{}_1, \bar{\Lambda}^0{}_2, \bar{\Lambda}^0{}_3)$ has length

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

$\sqrt{(\bar{\Lambda}^0{}_0)^2 - 1}$, so the scalar product of these two three-vectors is bounded by

$$|\bar{\Lambda}^0{}_1\Lambda^1{}_0 + \bar{\Lambda}^0{}_2\Lambda^2{}_0 + \bar{\Lambda}^0{}_3\Lambda^3{}_0| \leq \sqrt{(\Lambda^0{}_0)^2 - 1}\sqrt{(\bar{\Lambda}^0{}_0)^2 - 1}\,, \qquad (2.3.14)$$

and so

$$(\bar{\Lambda}\Lambda)^0{}_0 \geq \bar{\Lambda}^0{}_0\Lambda^0{}_0 - \sqrt{(\Lambda^0{}_0)^2 - 1}\sqrt{(\bar{\Lambda}^0{}_0)^2 - 1} \quad \geq 1\,.$$

The subgroup of Lorentz transformations with $\text{Det}\,\Lambda = +1$ and $\Lambda^0{}_0 \geq +1$ is known as the *proper orthochronous Lorentz group.* Since it is not possible by a continuous change of parameters to jump from $\text{Det}\,\Lambda = +1$ to $\text{Det}\,\Lambda = -1$, or from $\Lambda^0{}_0 \geq +1$ to $\Lambda^0{}_0 \leq -1$, any Lorentz transformation that can be obtained from the identity by a continuous change of parameters must have $\text{Det}\,\Lambda$ and $\Lambda^0{}_0$ of the same sign as for the identity, and hence must belong to the proper orthochronous Lorentz group.

Any Lorentz transformation is either proper and orthochronous, or may be written as the product of an element of the proper orthochronous Lorentz group with one of the discrete transformations $\mathscr{P}$ or $\mathscr{T}$ or $\mathscr{P}\mathscr{T}$, where $\mathscr{P}$ is the space inversion, whose non-zero elements are

$$\mathscr{P}^0{}_0 = 1,\ \mathscr{P}^1{}_1 = \mathscr{P}^2{}_2 = \mathscr{P}^3{}_3 = -1\,, \qquad (2.3.15)$$

and $\mathscr{T}$ is the time-reversal matrix, whose non-zero elements are

$$\mathscr{T}^0{}_0 = -1,\ \mathscr{T}^1{}_1 = \mathscr{T}^2{}_2 = \mathscr{T}^3{}_3 = 1\,. \qquad (2.3.16)$$

Thus the study of the whole Lorentz group reduces to the study of its proper orthochronous subgroup, plus space inversion and time-reversal. We will consider space inversion and time-reversal separately in Section 2.6. Until then, we will deal only with the homogeneous or inhomogeneous proper orthochronous Lorentz group.

2.4 The Poincaré Algebra

As we saw in Section 2.2, much of the information about any Lie symmetry group is contained in properties of the group elements near the identity. For the inhomogeneous Lorentz group, the identity is the transformation $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu$, $a^\mu = 0$, so we want to study those transformations with

$$\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \omega^\mu{}_\nu\,, \qquad a^\mu = \epsilon^\mu\,, \qquad (2.4.1)$$

both $\omega^\mu{}_\nu$ and ϵ^μ being taken infinitesimal. The Lorentz condition (2.3.5)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

reads here

$$\begin{aligned}\eta_{\rho\sigma} &= \eta_{\mu\nu}(\delta^{\mu}{}_{\rho} + \omega^{\mu}{}_{\rho})\,(\delta^{\nu}{}_{\sigma} + \omega^{\nu}{}_{\sigma}) \\ &= \eta_{\sigma\rho} + \omega_{\sigma\rho} + \omega_{\rho\sigma} + O(\omega^2)\,.\end{aligned}$$

We are here using the convention, to be used throughout this book, that indices may be lowered or raised by contraction with $\eta_{\mu\nu}$ or $\eta^{\mu\nu}$

$$\begin{aligned}\omega_{\sigma\rho} &\equiv \eta_{\mu\sigma}\omega^{\mu}{}_{\rho} \\ \omega^{\mu}{}_{\rho} &\equiv \eta^{\mu\sigma}\omega_{\sigma\rho}\,.\end{aligned}$$

Keeping only the terms of first order in ω in the Lorentz condition (2.3.5), we see that this condition now reduces to the antisymmetry of $\omega_{\mu\nu}$

$$\omega_{\mu\nu} = -\omega_{\nu\mu}\,. \tag{2.4.2}$$

An antisymmetric second-rank tensor in four dimensions has $(4 \times 3)/2 = 6$ independent components, so including the four components of ϵ^{μ}, an inhomogeneous Lorentz transformation is described by $6 + 4 = 10$ parameters.

Since $U(1,0)$ carries any ray into itself, it must be proportional to the unit operator,* and by a choice of phase may be made equal to it. For an infinitesimal Lorentz transformation (2.4.1), $U(1+\omega,\epsilon)$ must then equal **1** plus terms linear in $\omega_{\rho\sigma}$ and ϵ_{ρ}. We write this as

$$U(1+\omega,\epsilon) = 1 + \tfrac{1}{2} i\,\omega_{\rho\sigma}J^{\rho\sigma} - i\epsilon_{\rho}P^{\rho} + \cdots\,. \tag{2.4.3}$$

Here $J^{\rho\sigma}$ and P^{ρ} are ω- and ϵ-independent operators, and the dots denote terms of higher order in ω and/or ϵ. In order for $U(1+\omega,\epsilon)$ to be unitary, the operators $J^{\rho\sigma}$ and P^{ρ} must be Hermitian

$$J^{\rho\sigma\dagger} = J^{\rho\sigma}\quad,\quad P^{\rho\dagger} = P^{\rho}\,. \tag{2.4.4}$$

Since $\omega_{\rho\sigma}$ is antisymmetric, we can take its coefficient $J^{\rho\sigma}$ to be antisymmetric also

$$J^{\rho\sigma} = -J^{\sigma\rho}\,. \tag{2.4.5}$$

As we shall see, P^1, P^2, and P^3 are the components of the momentum operators, J^{23}, J^{31}, and J^{12} are the components of the angular momentum vector, and P^0 is the energy operator, or *Hamiltonian.***

---

* In the absence of superselection rules, the possibility that the proportionality factor may depend on the state on which $U(1,0)$ acts can be ruled out by the same reasoning that we used in Section 2.2 to rule out the possibility that the phases in projective representations of symmetry groups may depend on the states on which the symmetries act. Where superselection rules apply, it may be necessary to redefine $U(1,0)$ by phase factors that depend on the sector on which it acts.

** We will see that this identification of the angular-momentum generators is forced on us by the commutation relations of the $J^{\mu\nu}$. On the other hand, the commutation relations do not allow us to distinguish between P^{μ} and $-P^{\mu}$, so the sign for the $\epsilon_{\rho}P^{\rho}$ term in (2.4.3) is a matter of convention. The consistency of the choice in (2.4.3) with the usual definition of the Hamiltonian P^0 is shown in Section 3.1.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

We now examine the Lorentz transformation properties of $J^{\rho\sigma}$ and P^{ρ}. We consider the product

$$U(\Lambda,a)\,U(1+\omega,\epsilon)U^{-1}(\Lambda,a)\,,$$

where $\Lambda^{\mu}{}_{\nu}$ and a^{μ} are here the parameters of a new transformation, unrelated to ω and ϵ. According to Eq. (2.3.11), the product $U(\Lambda^{-1},-\Lambda^{-1}a)\,U(\Lambda,a)$ equals $U(1,0)$, so $U(\Lambda^{-1},-\Lambda^{-1}a)$ is the inverse of $U(\Lambda,a)$. It follows then from (2.3.11) that

$$U(\Lambda,a)U(1+\omega,\epsilon)U^{-1}(\Lambda,a)=U\left(\Lambda(1+\omega)\Lambda^{-1},\Lambda\epsilon-\Lambda\omega\Lambda^{-1}a\right)\,. \quad (2.4.6)$$

To first order in ω and ϵ, we have then

$$U(\Lambda,a)\left[\tfrac{1}{2}\omega_{\rho\sigma}J^{\rho\sigma}-\epsilon_{\rho}P^{\rho}\right]U^{-1}(\Lambda,a)=\tfrac{1}{2}(\Lambda\omega\Lambda^{-1})_{\mu\nu}J^{\mu\nu} -(\Lambda\epsilon-\Lambda\omega\Lambda^{-1}a)_{\mu}P^{\mu}\,. \quad (2.4.7)$$

Equating coefficients of $\omega_{\rho\sigma}$ and ϵ_{ρ} on both sides of this equation (and using (2.3.10)), we find

$$U(\Lambda,a)J^{\rho\sigma}U^{-1}(\Lambda,a)=\Lambda_{\mu}{}^{\rho}\Lambda_{\nu}{}^{\sigma}(J^{\mu\nu}-a^{\mu}P^{\nu}+a^{\nu}P^{\mu})\,, \quad (2.4.8)$$

$$U(\Lambda,a)P^{\rho}\,U^{-1}(\Lambda,a)=\Lambda_{\mu}{}^{\rho}P^{\mu}\,. \quad (2.4.9)$$

For homogeneous Lorentz transformations (with $a^{\mu}=0$), these transformation rules simply say that $J^{\mu\nu}$ is a tensor and P^{μ} is a vector. For pure translations (with $\Lambda^{\mu}{}_{\nu}=\delta^{\mu}{}_{\nu}$), they tell us that P^{ρ} is translation-invariant, but $J^{\rho\sigma}$ is not. In particular, the change of the space–space components of $J^{\rho\sigma}$ under a spatial translation is just the usual change of the angular momentum under a change of the origin relative to which the angular momentum is calculated.

Next, let's apply rules (2.4.8), (2.4.9) to a transformation that is itself infinitesimal, i.e., $\Lambda^{\mu}{}_{\nu}=\delta^{\mu}{}_{\nu}+\omega^{\mu}{}_{\nu}$ and $a^{\mu}=\epsilon^{\mu}$, with infinitesimals $\omega^{\mu}{}_{\nu}$ and ϵ^{μ} unrelated to the previous ω and ϵ. Using Eq. (2.4.3), and keeping only terms of first order in $\omega^{\mu}{}_{\nu}$ and ϵ^{μ}, Eqs. (2.4.8) and (2.4.9) now become

$$i\left[\tfrac{1}{2}\omega_{\mu\nu}J^{\mu\nu}-\epsilon_{\mu}P^{\mu}\,,\;J^{\rho\sigma}\right]=\omega_{\mu}{}^{\rho}J^{\mu\sigma}+\omega_{\nu}{}^{\sigma}J^{\rho\nu}-\epsilon^{\rho}P^{\sigma}+\epsilon^{\sigma}P^{\rho}\,, \quad (2.4.10)$$

$$i\left[\tfrac{1}{2}\omega_{\mu\nu}J^{\mu\nu}-\epsilon_{\mu}P^{\mu}\,,\;P^{\rho}\right]=\omega_{\mu}{}^{\rho}P^{\mu}\,. \quad (2.4.11)$$

Equating coefficients of $\omega_{\mu\nu}$ and ϵ_{μ} on both sides of these equations, we find the commutation rules

$$i\,[J^{\mu\nu},J^{\rho\sigma}]=\eta^{\nu\rho}J^{\mu\sigma}-\eta^{\mu\rho}J^{\nu\sigma}-\eta^{\sigma\mu}J^{\rho\nu}+\eta^{\sigma\nu}J^{\rho\mu}\,, \quad (2.4.12)$$

$$i\,[P^{\mu}\,,\;J^{\rho\sigma}]=\eta^{\mu\rho}P^{\sigma}-\eta^{\mu\sigma}P^{\rho}\,, \quad (2.4.13)$$

$$[P^{\mu}\,,\;P^{\rho}]=0\,. \quad (2.4.14)$$

This is the Lie algebra of the Poincaré group.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

In quantum mechanics a special role is played by those operators that are *conserved*, i.e., that commute with the energy operator $H = P^0$. Inspection of Eqs. (2.4.13) and (2.4.14) shows that these are the momentum three-vector

$$\mathbf{P} = \left\{P^1, P^2, P^3\right\} \tag{2.4.15}$$

and the angular-momentum three-vector

$$\mathbf{J} = \left\{J^{23}, J^{31}, J^{12}\right\} \tag{2.4.16}$$

and, of course, the energy P^0 itself. The remaining generators form what is called the 'boost' three-vector

$$\mathbf{K} = \left\{J^{01}, J^{02}, J^{03}\right\} . \tag{2.4.17}$$

These are *not* conserved, which is why we do not use the eigenvalues of $\mathbf{K}$ to label physical states. In a three-dimensional notation, the commutation relations (2.4.12), (2.4.13), (2.4.14) may be written

$$[J_i, J_j] = i\epsilon_{ijk}J_k , \tag{2.4.18}$$
$$[J_i, K_j] = i\epsilon_{ijk}K_k , \tag{2.4.19}$$
$$[K_i, K_j] = -i\epsilon_{ijk}J_k , \tag{2.4.20}$$
$$[J_i, P_j] = i\epsilon_{ijk}P_k , \tag{2.4.21}$$
$$[K_i, P_j] = -iH\delta_{ij} , \tag{2.4.22}$$
$$[J_i, H] = [P_i, H] = [H, H] = 0 , \tag{2.4.23}$$
$$[K_i, H] = -iP_i , \tag{2.4.24}$$

where i, j, k, etc. run over the values 1, 2, and 3, and ϵ_{ijk} is the totally antisymmetric quantity with $\epsilon_{123} = +1$. The commutation relation (2.4.18) will be recognized as that of the angular-momentum operator.

The pure translations $T(1, a)$ form a subgroup of the inhomogeneous Lorentz group with a group multiplication rule given by (2.3.8) as

$$T(1, \bar{a})T(1, a) = T(1, \bar{a} + a). \tag{2.4.25}$$

This is additive in the same sense as (2.2.24), so by using (2.4.3) and repeating the same arguments that led to (2.2.26), we find that finite translations are represented on the physical Hilbert space by

$$U(1, a) = \exp(-iP^\mu a_\mu). \tag{2.4.26}$$

In exactly the same way, we can show that a rotation $R_{\boldsymbol{\theta}}$ by an angle $|\boldsymbol{\theta}|$ around the direction of $\boldsymbol{\theta}$ is represented on the physical Hilbert space by

$$U(R_{\boldsymbol{\theta}}, 0) = \exp(i\mathbf{J} \cdot \boldsymbol{\theta}). \tag{2.4.27}$$

It is interesting to compare the Poincaré algebra with the Lie algebra of the symmetry group of Newtonian mechanics, the Galilean group. We

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

could derive this algebra by starting with the transformation rules of the Galilean group and then following the same procedure that was used here to derive the Poincaré algebra. However, since we already have Eqs. (2.4.18)–(2.4.24), it is easier to obtain the Galilean algebra as a low-velocity limit of the Poincaré algebra, by what is known as an *Inönü–Wigner contraction*[4,5]. For a system of particles of typical mass m and typical velocity v, the momentum and the angular-momentum operators are expected to be of order $\mathbf{J} \sim 1$, $\mathbf{P} \sim mv$. On the other hand, the energy operator is $H = M + W$ with a total mass M and non-mass energy W (kinetic plus potential) of order $M \sim m$, $W \sim mv^2$. Inspection of Eqs. (2.4.18)–(2.4.24) shows that these commutation relations have a limit for $v << 1$ of the form

$$
\begin{aligned}
&[J_i, J_j] = i\epsilon_{ijk} J_k\,, \qquad [J_i, K_j] = i\epsilon_{ijk} K_k\,, \qquad [K_i, K_j] = 0\,,\\
&[J_i, P_j] = i\epsilon_{ijk} P_k\,, \qquad [K_i, P_j] = -i M\delta_{ij}\,,\\
&[J_i, W] = [P_i, W] = 0\,, \qquad [K_i, W] = -i P_i\,,\\
&[J_i, M] = [P_i, M] = [K_i, M] = [W, M] = 0\,,
\end{aligned}
$$

with $\mathbf{K}$ of order $1/v$. Note that the product of a translation $\mathbf{x} \to \mathbf{x} + \mathbf{a}$ and a 'boost' $\mathbf{x} \to \mathbf{x} + \mathbf{v}t$ should be the transformation $\mathbf{x} \to \mathbf{x} + \mathbf{v}t + \mathbf{a}$, but this is not true for the action of these operators on Hilbert space:

$$\exp(-i\mathbf{K}\cdot\mathbf{v})\exp(-i\mathbf{P}\cdot\mathbf{a}) = \exp(iM\mathbf{a}\cdot\mathbf{v}/2)\exp\left(-i(\mathbf{K}\cdot\mathbf{v} + \mathbf{P}\cdot\mathbf{a})\right)\,.$$

The appearance of the phase factor $\exp(iM\mathbf{a}\cdot\mathbf{v}/2)$ shows that this is a projective representation, with a superselection rule forbidding the superposition of states of different mass. In this respect, the mathematics of the Poincaré group is simpler than that of the Galilean group. However, there is nothing to prevent us from formally enlarging the Galilean group, by adding one more generator to its Lie algebra, which commutes with all the other generators, and whose eigenvalues are the masses of the various states. In this case physical states provide an ordinary rather than a projective representation of the expanded symmetry group. The difference appears to be a mere matter of notation, except that with this reinterpretation of the Galilean group there is no need for a mass superselection rule.

2.5 One-Particle States

We now consider the classification of one-particle states according to their transformation under the inhomogeneous Lorentz group.

The components of the energy-momentum four-vector all commute with each other, so it is natural to express physical state-vectors in terms of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

eigenvectors of the four-momentum. Introducing a label σ to denote all other degrees of freedom, we thus consider state-vectors $\Psi_{p,\sigma}$ with

$$P^\mu \Psi_{p,\sigma} = p^\mu \Psi_{p,\sigma} \ . \tag{2.5.1}$$

For general states, consisting for instance of several unbound particles, the label σ would have to be allowed to include continuous as well as discrete labels. We take as part of the definition of a *one*-particle state, that the label σ is purely discrete, and will limit ourselves here to that case. (However, a specific bound state of two or more particles, such as the lowest state of the hydrogen atom, is to be considered as a one-particle state. It is not an *elementary* particle, but the distinction between composite and elementary particles is of no relevance here.)

Eqs. (2.5.1) and (2.4.26) tell us how the states $\Psi_{p,\sigma}$ transform under translations:

$$U(1,a)\Psi_{p,\sigma} = e^{-ip\cdot a}\Psi_{p,\sigma} \ .$$

We must now consider how these states transform under homogeneous Lorentz transformations.

Using (2.4.9), we see that the effect of operating on $\Psi_{p,\sigma}$ with a quantum homogeneous Lorentz transformation $U(\Lambda,0) \equiv U(\Lambda)$ is to produce an eigenvector of the four-momentum with eigenvalue Λp

$$\begin{aligned} P^\mu U(\Lambda)\Psi_{p,\sigma} &= U(\Lambda)\left[U^{-1}(\Lambda)P^\mu U(\Lambda)\right]\Psi_{p,\sigma} = U(\Lambda)(\Lambda^{-1\mu}{}_\rho P^\rho)\Psi_{p,\sigma} \\ &= \Lambda^\mu{}_\rho p^\rho U(\Lambda)\Psi_{p,\sigma} \ . \end{aligned} \tag{2.5.2}$$

Hence $U(\Lambda)\Psi_{p,\sigma}$ must be a linear combination of the state-vectors $\Psi_{\Lambda p,\sigma'}$:

$$U(\Lambda)\Psi_{p,\sigma} = \sum_{\sigma'} C_{\sigma'\sigma}(\Lambda,p)\Psi_{\Lambda p,\sigma'} \ . \tag{2.5.3}$$

In general, it may be possible by using suitable linear combinations of the $\Psi_{p,\sigma}$ to choose the σ labels in such a way that the matrix $C_{\sigma'\sigma}(\Lambda,p)$ is block-diagonal; in other words, so that the $\Psi_{p,\sigma}$ with σ within any one block by themselves furnish a representation of the inhomogeneous Lorentz group. It is natural to identify the states of a specific particle type with the components of a representation of the inhomogeneous Lorentz group which is irreducible, in the sense that it cannot be further decomposed in this way.* Our task now is to work out the structure of the

* Of course, different particle species may correspond to representations that are isomorphic, i.e., that have matrices $C_{\sigma'\sigma}(\Lambda,p)$ that are either identical, or identical up to a similarity transformation. In some cases it may be convenient to define particle types as irreducible representations of larger groups that contain the inhomogeneous proper orthochronous Lorentz group as a subgroup; for instance, as we shall see, for massless particles whose interactions respect the symmetry of space inversion it is customary to treat all the components of an irreducible representation of the inhomogeneous Lorentz group including space inversion as a single particle type.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

coefficients $C_{\sigma'\sigma}(\Lambda, p)$ in irreducible representations of the inhomogeneous Lorentz group.

For this purpose, note that the only functions of p^μ that are left invariant by all proper orthochronous Lorentz transformations $\Lambda^\mu{}_\nu$ are the invariant square $p^2 \equiv \eta_{\mu\nu}p^\mu p^\nu$, and for $p^2 \leq 0$, also the sign of p^0. Hence, for each value of p^2, and (for $p^2 \leq 0$) each sign of p^0, we can choose a 'standard' four-momentum, say k^μ, and express any p^μ of this class as

$$p^\mu = L^\mu{}_\nu(p)k^\nu \,, \tag{2.5.4}$$

where $L^\mu{}_\nu$ is some standard Lorentz transformation that depends on p^μ, and also implicitly on our choice of the standard k^μ. We can then *define* the states $\Psi_{p,\sigma}$ of momentum p by

$$\Psi_{p,\sigma} \equiv N(p)\, U(L(p))\Psi_{k,\sigma} \,, \tag{2.5.5}$$

where $N(p)$ is a numerical normalization factor, to be chosen later. Up to this point, we have said nothing about how the σ labels are related for different momenta; Eq. (2.5.5) now fills that gap.

Operating on (2.5.5) with an arbitrary homogeneous Lorentz transformation $U(\Lambda)$, we now find

$$\begin{aligned} U(\Lambda)\Psi_{p,\sigma} &= N(p)\, U(\Lambda L(p))\Psi_{k,\sigma} \\ &= N(p)\, U(L(\Lambda p))\, U(L^{-1}(\Lambda p)\Lambda L(p))\, \Psi_{k,\sigma} \,. \end{aligned} \tag{2.5.6}$$

The point of this last step is that the Lorentz transformation $L^{-1}(\Lambda p)\Lambda L(p)$ takes k to $L(p)k = p$, and then to Λp, and then back to k, so it belongs to the subgroup of the homogeneous Lorentz group consisting of Lorentz transformations $W^\mu{}_\nu$ that leave k^μ invariant:

$$W^\mu{}_\nu k^\nu = k^\mu. \tag{2.5.7}$$

This subgroup is called the *little group*.[5] For any W satisfying Eq. (2.5.7), we have

$$U(W)\Psi_{k,\sigma} = \sum_{\sigma'} D_{\sigma'\sigma}(W)\Psi_{k,\sigma'} \,. \tag{2.5.8}$$

The coefficients $D(W)$ furnish a representation of the little group; that is, for any elements $\bar{W}, W$ we have

$$\begin{aligned} \sum_{\sigma'} D_{\sigma'\sigma}(\bar{W}W)\Psi_{k,\sigma'} &= U(\bar{W}W)\Psi_{k,\sigma} = U(\bar{W})U(W)\Psi_{k,\sigma} \\ &= U(\bar{W})\sum_{\sigma''} D_{\sigma''\sigma}(W)\Psi_{k,\sigma''} = \sum_{\sigma'\sigma''} D_{\sigma''\sigma}(W)D_{\sigma'\sigma''}(\bar{W})\Psi_{k,\sigma'} \end{aligned}$$

and so

$$D_{\sigma'\sigma}(\bar{W}W) = \sum_{\sigma''} D_{\sigma'\sigma''}(\bar{W})D_{\sigma''\sigma}(W) \,. \tag{2.5.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

In particular, we may apply Eq. (2.5.8) to the little-group transformation

$$W(\Lambda, p) \equiv L^{-1}(\Lambda p)\Lambda L(p) \tag{2.5.10}$$

and then Eq. (2.5.6) takes the form

$$U(\Lambda)\Psi_{p,\sigma} = N(p)\sum_{\sigma'} D_{\sigma'\sigma}\left(W(\Lambda, p)\right) U\left(L(\Lambda p)\right)\Psi_{k,\sigma'} ,$$

or, recalling the definition (2.5.5):

$$U(\Lambda)\Psi_{p,\sigma} = \left(\frac{N(p)}{N(\Lambda p)}\right)\sum_{\sigma'} D_{\sigma'\sigma}\left(W(\Lambda, p)\right)\Psi_{\Lambda p,\sigma'} . \tag{2.5.11}$$

Apart from the question of normalization, the problem of determining the coefficients $C_{\sigma'\sigma}$ in the transformation rule (2.5.3) has been reduced to the problem of finding the representations of the little group. This approach, of deriving representations of a group like the inhomogeneous Lorentz group from the representations of a little group, is called the method of induced representations.[6]

Table 2.1 gives a convenient choice of the standard momentum k^μ and the corresponding little group for the various classes of four-momenta.

Of these six classes of four-momenta, only (a), (c), and (f) have any known interpretations in terms of physical states. Not much needs to be said here about case (f) — $p^\mu = 0$; it describes the vacuum, which is simply left invariant by $U(\Lambda)$. In what follows we will consider only cases (a) and (c), which cover particles of mass $M > 0$ and mass zero, respectively.

This is a good place to pause, and say something about the normalization of these states. By the usual orthonormalization procedure of quantum mechanics, we may choose the states with standard momentum k^μ to be orthonormal, in the sense that

$$(\Psi_{k',\sigma'}, \Psi_{k,\sigma}) = \delta^3(\mathbf{k}' - \mathbf{k})\delta_{\sigma'\sigma} . \tag{2.5.12}$$

(The delta function appears here because $\Psi_{k,\sigma}$ and $\Psi_{k',\sigma'}$ are eigenstates of a Hermitian operator with eigenvalues $\mathbf{k}$ and $\mathbf{k}'$, respectively.) This has the immediate consequence that the representation of the little group in Eqs. (2.5.8) and (2.5.11) must be unitary**

$$D^\dagger(W) = D^{-1}(W) . \tag{2.5.13}$$

Now, what about scalar products for arbitrary momenta? Using the unitarity of the operator $U(\Lambda)$ in Eqs. (2.5.5) and (2.5.11), we find for the

** The little groups $SO(2,1)$ and $SO(3,1)$ for $p^2 > 0$ and $p^\mu = 0$ have no non-trivial finite-dimensional unitary representations, so if there were any states with a given momentum p^μ with $p^2 > 0$ or $p^\mu = 0$ that transform non-trivially under the little group, there would have to be an infinite number of them.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

Table 2.1. Standard momenta and the corresponding little group for various classes of four-momenta. Here κ is an arbitrary positive energy, say 1 eV. The little groups are mostly pretty obvious: $SO(3)$ is the ordinary rotation group in three dimensions (excluding space inversions), because rotations are the only proper orthochronous Lorentz transformations that leave at rest a particle with zero momentum, while $SO(2,1)$ and $SO(3,1)$ are the Lorentz groups in $(2+1)$- and $(3+1)$-dimensions, respectively. The group $ISO(2)$ is the group of Euclidean geometry, consisting of rotations and translations in two dimensions. Its appearance as the little group for $p^2 = 0$ is explained below.

	Standard k^μ	Little Group
(a) $p^2 = -M^2 < 0, p^0 > 0$	$(0,0,0,M)$	$SO(3)$
(b) $p^2 = -M^2 < 0, p^0 < 0$	$(0,0,0,-M)$	$SO(3)$
(c) $p^2 = 0, p^0 > 0$	$(0, 0, \kappa, \kappa)$	$ISO(2)$
(d) $p^2 = 0, p^0 < 0$	$(0, 0, \kappa, -\kappa)$	$ISO(2)$
(e) $p^2 = N^2 > 0$	$(0,0,N,0)$	$SO(2,1)$
(f) $p^\mu = 0$	$(0, 0, 0, 0)$	$SO(3,1)$

scalar product:

$$(\Psi_{p',\sigma'}, \Psi_{p,\sigma}) = N(p)\left(U^{-1}(L(p))\Psi_{p',\sigma'},\ \Psi_{k,\sigma}\right)$$
$$= N(p)N^*(p')D\left(W(L^{-1}(p),p')\right)^*_{\sigma\sigma'}\ \delta^3(\mathbf{k}'-\mathbf{k})$$

where $k' \equiv L^{-1}(p)p'$. Since also $k = L^{-1}(p)p$, the delta function $\delta^3(\mathbf{k}-\mathbf{k}')$ is proportional to $\delta^3(\mathbf{p}-\mathbf{p}')$. For $p' = p$, the little-group transformation here is trivial, $W(L^{-1}(p),p) = 1$, and so the scalar product is

$$(\Psi_{p'\sigma'}, \Psi_{p,\sigma}) = |N(p)|^2\delta_{\sigma'\sigma}\delta^3(\mathbf{k}'-\mathbf{k})\,. \tag{2.5.14}$$

It remains to work out the proportionality factor relating $\delta^3(\mathbf{k}-\mathbf{k}')$ and $\delta^3(\mathbf{p}-\mathbf{p}')$. Note that the Lorentz-invariant integral of an arbitrary scalar function $f(p)$ over four-momenta with $-p^2 = M^2 \geq 0$ and $p^0 > 0$ (i.e.,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

cases (a) or (c)) may be written

$$\begin{aligned}\int d^4p\,\delta(p^2+M^2)\theta(p^0)f(p)&\\ =\int d^3\mathbf{p}\,dp^0\;\delta((p^0)^2-\mathbf{p}^2-M^2)\theta(p^0)f(\mathbf{p},p^0)&\\ =\int d^3\mathbf{p}\frac{f(\mathbf{p},\sqrt{\mathbf{p}^2+M^2})}{2\sqrt{\mathbf{p}^2+M^2}}&\end{aligned}$$

($\theta(p^0)$ is the step function: $\theta(x)=1$ for $x\geq 0$, $\theta(x)=0$ for $x<0$.) We see that when integrating on the 'mass shell' $p^2+M^2=0$, the invariant volume element is

$$d^3\mathbf{p}/\sqrt{\mathbf{p}^2+M^2}\,. \tag{2.5.15}$$

The delta function is defined by

$$\begin{aligned}F(\mathbf{p})&=\int F(\mathbf{p}')\delta^3(\mathbf{p}-\mathbf{p}')d^3\mathbf{p}'\\ &=\int F(\mathbf{p}')\left[\sqrt{\mathbf{p}'^2+M^2}\delta^3(\mathbf{p}'-\mathbf{p})\right]\frac{d^3\mathbf{p}'}{\sqrt{\mathbf{p}'^2+M^2}}\end{aligned}$$

so we see that the invariant delta function is

$$\sqrt{\mathbf{p}'^2+M^2}\delta^3(\mathbf{p}'-\mathbf{p})=p^0\delta^3(\mathbf{p}'-\mathbf{p})\,. \tag{2.5.16}$$

Since p' and p are related to k' and k respectively by a Lorentz transformation, $L(p)$, we have then

$$p^0\delta^3(\mathbf{p}'-\mathbf{p})=k^0\delta^3(\mathbf{k}'-\mathbf{k})$$

and therefore

$$(\Psi_{p',\sigma'},\Psi_{p,\sigma})=|N(p)|^2\delta_{\sigma'\sigma}\left(\frac{p^0}{k^0}\right)\delta^3(\mathbf{p}'-\mathbf{p})\,. \tag{2.5.17}$$

The normalization factor $N(p)$ is sometimes chosen to be just $N(p)=1$, but then we would need to keep track of the p^0/k^0 factor in scalar products. Instead, I will here adopt the more usual convention that

$$N(p)=\sqrt{k^0/p^0} \tag{2.5.18}$$

for which

$$(\Psi_{p',\sigma'},\Psi_{p,\sigma})=\delta_{\sigma'\sigma}\delta^3(\mathbf{p}'-\mathbf{p})\,. \tag{2.5.19}$$

We now consider the two cases of physical interest: particles of mass $M>0$, and particles of zero mass.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

Mass Positive-Definite

The little group here is the three-dimensional rotation group. Its unitary representations can be broken up into a direct sum of irreducible unitary representations[7] $D^{(j)}_{\sigma'\sigma}(R)$ of dimensionality $2j+1$, with $j = 0, \frac{1}{2}, 1, \cdots$. These can be built up from the standard matrices for infinitesimal rotations $R_{ik} = \delta_{ik} + \Theta_{ik}$, with $\Theta_{ik} = -\Theta_{ki}$ infinitesimal:

$$D^{(j)}_{\sigma'\sigma}(1+\Theta) = \delta_{\sigma'\sigma} + \frac{i}{2}\Theta_{ik}(J^{(j)}_{ik})_{\sigma'\sigma}\,, \tag{2.5.20}$$

$$\begin{aligned}(J^{(j)}_{23} \pm iJ^{(j)}_{31})_{\sigma'\sigma} &= (J^{(j)}_{1} \pm iJ^{(j)}_{2})_{\sigma'\sigma} \\ &= \delta_{\sigma',\sigma\pm1}\sqrt{(j\mp\sigma)(j\pm\sigma+1)}\,,\end{aligned} \tag{2.5.21}$$

$$(J^{(j)}_{12})_{\sigma'\sigma} = (J^{(j)}_{3})_{\sigma'\sigma} = \sigma\delta_{\sigma'\sigma}\,, \tag{2.5.22}$$

with σ running over the values $j, j-1, \cdots, -j$. For a particle of mass $M > 0$ and spin j, Eq. (2.5.11) now becomes

$$U(\Lambda)\Psi_{p,\sigma} = \sqrt{\frac{(\Lambda p)^0}{p^0}}\sum_{\sigma'} D^{(j)}_{\sigma'\sigma}(W(\Lambda,p))\,\Psi_{\Lambda p,\sigma'}\,, \tag{2.5.23}$$

with the little-group element $W(\Lambda, p)$ (the Wigner rotation[5]) given by Eq. (2.5.10):

$$W(\Lambda,p) = L^{-1}(\Lambda p)\Lambda L(p).$$

To calculate this rotation, we need to choose a 'standard boost' $L(p)$ which carries the four-momentum from $k^\mu = (0,0,0,M)$ to p^μ. This is conveniently chosen as

$$\begin{aligned}L^i{}_k(p) &= \delta_{ik} + (\gamma-1)\hat{p}_i\hat{p}_k \\ L^i{}_0(p) &= L^0{}_i(p) = \hat{p}_i\sqrt{\gamma^2-1}\,, \\ L^0{}_0(p) &= \gamma\,,\end{aligned} \tag{2.5.24}$$

where

$$\hat{p}_i \equiv p_i/|\mathbf{p}|, \qquad \gamma \equiv \sqrt{\mathbf{p}^2+M^2}/M\,.$$

It is very important that when $\Lambda^\mu{}_\nu$ is an arbitrary three-dimensional rotation $\mathscr{R}$, the Wigner rotation $W(\Lambda, p)$ is the same as $\mathscr{R}$ for all p. To see this, note that the boost (2.5.24) may be expressed as

$$L(p) = R(\hat{\mathbf{p}})B(|\mathbf{p}|)R^{-1}(\hat{\mathbf{p}})\,,$$

where $R(\hat{\mathbf{p}})$ is a rotation (to be defined in a standard way below, in Eq.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

(2.5.47)) that takes the three-axis into the direction of $\mathbf{p}$, and

$$B(|\mathbf{p}|) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma & \sqrt{\gamma^2-1} \\ 0 & 0 & \sqrt{\gamma^2-1} & \gamma \end{bmatrix} .$$

Then for an arbitrary rotation $\mathscr{R}$

$$W(\mathscr{R},p) = R(\mathscr{R}\hat{\mathbf{p}})B^{-1}(|\mathbf{p}|)R^{-1}(\mathscr{R}\hat{\mathbf{p}})\mathscr{R}\,R(\hat{\mathbf{p}})B(|\mathbf{p}|)R^{-1}(\hat{\mathbf{p}}) .$$

But the rotation $R^{-1}(\mathscr{R}\hat{\mathbf{p}})\mathscr{R}\,R(\hat{\mathbf{p}})$ takes the three-axis into the direction $\hat{\mathbf{p}}$, and then into the direction $\mathscr{R}\hat{\mathbf{p}}$, and then back to the three-axis, so it must be just a rotation by some angle θ around the three-axis

$$R^{-1}(\mathscr{R}\hat{\mathbf{p}})\mathscr{R}\,R(\hat{\mathbf{p}}) = R(\theta) \equiv \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Since $R(\theta)$ commutes with $B(|\mathbf{p}|)$, this now gives

$$W(\mathscr{R},p) = R(\mathscr{R}\hat{\mathbf{p}})B^{-1}(|\mathbf{p}|)R(\theta)B(|\mathbf{p}|)R^{-1}(\hat{\mathbf{p}}) = R(\mathscr{R}\hat{\mathbf{p}})R(\theta)R^{-1}(\hat{\mathbf{p}})$$

and hence

$$W(\mathscr{R},p) = \mathscr{R}$$

as was to be shown. Thus states of a moving massive particle (and, by extension, multi-particle states) have the same transformation under rotations as in non-relativistic quantum mechanics. This is another piece of good news — the whole apparatus of spherical harmonics, Clebsch–Gordan coefficients, etc. can be carried over wholesale from non-relativistic to relativistic quantum mechanics.

Mass Zero

First, we have to work out the structure of the little group. Consider an arbitrary little-group element $W^\mu{}_\nu$, with $W^\mu{}_\nu k^\nu = k^\mu$, where k^μ is the standard four-momentum for this case, $k^\mu = (0,0,1,1)$. Acting on a time-like four-vector $t^\mu = (0,0,0,1)$, such a Lorentz transformation must yield a four-vector Wt whose length and scalar product with $Wk = k$ are the same as those of t:

$$(Wt)^\mu(Wt)_\mu = t^\mu t_\mu = -1 ,$$
$$(Wt)^\mu k_\mu = t^\mu k_\mu = -1 .$$

Any four-vector that satisfies the second condition may be written

$$(Wt)^\mu = (\alpha,\beta,\zeta,1+\zeta)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

and the first condition then yields the relation

$$\zeta = (\alpha^2 + \beta^2)/2 \,. \tag{2.5.25}$$

It follows that the effect of $W^\mu{}_\nu$ on t^ν is the same as that of the Lorentz transformation

$$S^\mu{}_\nu(\alpha,\beta) = \begin{bmatrix} 1 & 0 & -\alpha & \alpha \\ 0 & 1 & -\beta & \beta \\ \alpha & \beta & 1-\zeta & \zeta \\ \alpha & \beta & -\zeta & 1+\zeta \end{bmatrix} . \tag{2.5.26}$$

This does not mean that W equals $S(\alpha,\beta)$, but it does mean that $S^{-1}(\alpha,\beta)W$ is a Lorentz transformation that leaves the time-like four-vector (0,0,0,1) invariant, and is therefore a pure rotation. Also, $S^\mu{}_\nu$ like $W^\mu{}_\nu$ leaves the light-like four-vector (0,0,1,1) invariant, so $S^{-1}(\alpha,\beta)W$ must be a rotation by some angle θ around the three-axis

$$S^{-1}(\alpha,\beta)W = R(\theta) \,, \tag{2.5.27}$$

where

$$R^\mu{}_\nu(\theta) \equiv \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

The most general element of the little group is therefore of the form

$$W(\theta,\alpha,\beta) = S(\alpha,\beta)R(\theta) \,. \tag{2.5.28}$$

What group is this? We note that the transformations with $\theta = 0$ or with $\alpha = \beta = 0$ form subgroups:

$$S(\bar\alpha,\bar\beta)S(\alpha,\beta) = S(\bar\alpha+\alpha,\bar\beta+\beta) \tag{2.5.29}$$

$$R(\bar\theta)R(\theta) = R(\bar\theta+\theta) \,. \tag{2.5.30}$$

These subgroups are *Abelian* — that is, their elements all commute with each other. Furthermore, the subgroup with $\theta = 0$ is *invariant*, in the sense that its elements are transformed into other elements of the same subgroup by any member of the group

$$R(\theta)S(\alpha,\beta)R^{-1}(\theta) = S(\alpha\cos\theta+\beta\sin\theta, -\alpha\sin\theta+\beta\cos\theta) \,. \tag{2.5.31}$$

From Eqs. (2.5.29)–(2.5.31) we can work out the product of any group elements. The reader will recognize these multiplication rules as those of the group $ISO(2)$, consisting of translations (by a vector (α,β)) and rotations (by an angle θ) in two dimensions.

Groups that do *not* have invariant Abelian subgroups have certain simple properties, and for this reason are called *semi-simple*. As we have

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

seen, the little group $ISO(2)$ like the inhomogeneous Lorentz group is *not* semi-simple, and this leads to interesting complications. First, let's take a look at the Lie algebra of $ISO(2)$. For θ, α, β infinitesimal, the general group element is

$$W(\theta,\alpha,\beta)^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu} ,$$
$$\omega_{\mu\nu} = \begin{bmatrix} 0 & \theta & -\alpha & \alpha \\ -\theta & 0 & -\beta & \beta \\ \alpha & \beta & 0 & 0 \\ -\alpha & -\beta & 0 & 0 \end{bmatrix} .$$

From (2.4.3), we see then that the corresponding Hilbert space operator is

$$U(W(\theta,\alpha,\beta)) = 1 + i\alpha A + i\beta B + i\theta J_3 , \tag{2.5.32}$$

where A and B are the Hermitian operators

$$A = -J^{13} + J^{10} = J_2 + K_1 , \tag{2.5.33}$$
$$B = -J^{23} + J^{20} = -J_1 + K_2 , \tag{2.5.34}$$

and, as before, $J_3 = J_{12}$. Either from (2.4.18)–(2.4.20), or directly from Eqs. (2.5.29)–(2.5.31), we see that these generators have the commutators

$$[J_3, A] = +iB , \tag{2.5.35}$$
$$[J_3, B] = -iA , \tag{2.5.36}$$
$$[A, B] = 0 . \tag{2.5.37}$$

Since A and B are commuting Hermitian operators they (like the momentum generators of the inhomogeneous Lorentz group) can be simultaneously diagonalized by states $\Psi_{k,a,b}$

$$A\Psi_{k,a,b} = a\Psi_{k,a,b} ,$$
$$B\Psi_{k,a,b} = b\Psi_{k,a,b} .$$

The problem is that if we find one such set of non-zero eigenvalues of A, B, then we find a whole continuum. From Eq. (2.5.31), we have

$$U[R(\theta)]A\,U^{-1}[R(\theta)] = A\cos\theta - B\sin\theta ,$$
$$U[R(\theta)]B\,U^{-1}[R(\theta)] = A\sin\theta + B\cos\theta ,$$

and so, for arbitrary θ,

$$A\Psi^{\theta}_{k,a,b} = (a\cos\theta - b\sin\theta)\Psi^{\theta}_{k,a,b} ,$$
$$B\Psi^{\theta}_{k,a,b} = (a\sin\theta + b\cos\theta)\Psi^{\theta}_{k,a,b} ,$$

where

$$\Psi^{\theta}_{k,a,b} \equiv U^{-1}(R(\theta))\,\Psi_{k,a,b} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

Massless particles are not observed to have any continuous degree of freedom like θ; to avoid such a continuum of states, we must require that physical states (now called $\Psi_{k,\sigma}$) are eigenvectors of A and B with $a = b = 0$:

$$A\Psi_{k,\sigma} = B\Psi_{k,\sigma} = 0\,. \tag{2.5.38}$$

These states are then distinguished by the eigenvalue of the remaining generator

$$J_3\Psi_{k,\sigma} = \sigma\Psi_{k,\sigma}\,. \tag{2.5.39}$$

Since the momentum $\mathbf{k}$ is in the three-direction, σ gives the component of angular momentum in the direction of motion, or *helicity*.

We are now in a position to calculate the Lorentz transformation properties of general massless particle states. First note that by use of the general arguments of Section 2.2, Eq. (2.5.32) generalizes for finite α and β to

$$U(S(\alpha,\beta)) = \exp(i\alpha A + i\beta B) \tag{2.5.40}$$

and for finite θ to

$$U(R(\theta)) = \exp(iJ_3\theta)\,. \tag{2.5.41}$$

An arbitrary element W of the little group can be put in the form (2.5.28), so that

$$U(W)\Psi_{k,\sigma} = \exp(i\alpha A + i\beta B)\exp(i\theta J_3)\Psi_{k,\sigma} = \exp(i\theta\sigma)\Psi_{k,\sigma}$$

and therefore Eq. (2.5.8) gives

$$D_{\sigma'\sigma}(W) = \exp(i\theta\sigma)\delta_{\sigma'\sigma}\,,$$

where θ is the angle defined by expressing W as in Eq. (2.5.28). The Lorentz transformation rule for a massless particle of arbitrary helicity is now given by Eqs. (2.5.11) and (2.5.18) as

$$U(\Lambda)\Psi_{p,\sigma} = \sqrt{\frac{(\Lambda p)^0}{p^0}}\exp\left(i\sigma\theta(\Lambda,p)\right)\Psi_{\Lambda p,\sigma} \tag{2.5.42}$$

with $\theta(\Lambda,p)$ defined by

$$W(\Lambda,p) \equiv L^{-1}(\Lambda p)\Lambda L(p) \equiv S\left(\alpha(\Lambda,p),\beta(\Lambda,p)\right)R\left(\theta(\Lambda,p)\right)\,. \tag{2.5.43}$$

We shall see in Section 5.9 that electromagnetic gauge invariance arises from the part of the little group parameterized by α and β.

At this point we have not yet encountered any reason that would forbid the helicity σ of a massless particle from being an arbitrary real number. As we shall see in Section 2.7, there are topological considerations that restrict the allowed values of σ to integers and half-integers, just as for massive particles.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

To calculate the little-group element (2.5.43) for a given Λ and p, (and also to enable us to calculate the effect of space or time inversion on these states in the next section) we need to fix a convention for the standard Lorentz transformation that takes us from $k^\mu = (0, 0, \kappa, \kappa)$ to p^μ. This may conveniently be chosen to have the form

$$L(p) = R(\hat{\mathbf{p}})B(|\mathbf{p}|/\kappa) \tag{2.5.44}$$

where $B(u)$ is a pure boost along the three-direction:

$$B(u) \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (u^2+1)/2u & (u^2-1)/2u \\ 0 & 0 & (u^2-1)/2u & (u^2+1)/2u \end{bmatrix} \tag{2.5.45}$$

and $R(\hat{\mathbf{p}})$ is a pure rotation that carries the three-axis into the direction of the unit vector $\hat{\mathbf{p}}$. For instance, suppose we take $\hat{\mathbf{p}}$ to have polar and azimuthal angles θ and ϕ:

$$\hat{\mathbf{p}} = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)\ . \tag{2.5.46}$$

Then we can take $R(\hat{\mathbf{p}})$ as a rotation by angle θ around the two-axis, which takes $(0, 0, 1)$ into $(\sin\theta,\ 0, \cos\theta)$, followed by a rotation by angle ϕ around the three-axis:

$$U(R(\hat{\mathbf{p}})) = \exp(-i\phi J_3)\exp(-i\theta J_2)\ , \tag{2.5.47}$$

where $0 \leq \theta \leq \pi$, $\quad 0 \leq \phi < 2\pi$. (We give $U(R(\hat{\mathbf{p}}))$ rather than $R(\hat{\mathbf{p}})$, together with a specification of the range of ϕ and θ, because shifting θ or ϕ by 2π would give the same rotation $R(\hat{\mathbf{p}})$, but a different sign for $U(R(\hat{\mathbf{p}}))$ when acting on half-integer spin states.) Since (2.5.47) is a rotation, and does take the three-axis into the direction (2.5.46), any other choice of such an $R(\hat{\mathbf{p}})$ would differ from this one by at most an initial rotation around the three-axis, corresponding to a mere redefinition of the phase of the one-particle states.

Note that the helicity is Lorentz-invariant; a massless particle of a given helicity σ looks the same (aside from its momentum) in all inertial frames. Indeed, we would be justified in thinking of massless particles of each different helicity as different species of particles. However, as we shall see in the next section, particles of opposite helicity are related by the symmetry of space inversion. Thus, because electromagnetic and gravitational forces obey space inversion symmetry, the massless particles of helicity ± 1 associated with electromagnetic phenomena are both called *photons*, and the massless particles of helicity ± 2 that are believed to be associated with gravitation are both called *gravitons*. On the other hand, the supposedly massless particles of helicity $\pm 1/2$ that are emitted in nuclear beta decay have no interactions (apart from gravitation) that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

respect the symmetry of space inversion, so these particles are given different names: *neutrinos* for helicity $-1/2$, and *antineutrinos* for helicity $+1/2$.

Even though the helicity of a massless particle is Lorentz-invariant, the state itself is not. In particular, because of the helicity-dependent phase factor $\exp(i\sigma\theta)$ in Eq. (2.5.42), a state formed as a linear superposition of one-particle states with opposite helicities will be changed by a Lorentz transformation into a different superposition. For instance, a general one-photon state of four-momentum p may be written

$$\Psi_{p;\alpha} = \alpha_+ \Psi_{p,+1} + \alpha_- \Psi_{p,-1} ,$$

where

$$|\alpha_+|^2 + |\alpha_-|^2 = 1 .$$

The generic case is one of *elliptic polarization*, with $|\alpha_\pm|$ both non-zero and unequal. *Circular polarization* is the limiting case where either α_+ or α_- vanishes, and *linear polarization* is the opposite extreme, with $|\alpha_+| = |\alpha_-|$. The overall phase of α_+ and α_- has no physical significance, and for linear polarization may be adjusted so that $\alpha_- = \alpha_+^*$, but the relative phase is still important. Indeed, for linear polarizations with $\alpha_- = \alpha_+^*$, the phase of α_+ may be identified as the angle between the plane of polarization and some fixed reference direction perpendicular to $\mathbf{p}$. Eq. (2.5.42) shows that under a Lorentz transformation $\Lambda^\mu{}_\nu$, this angle rotates by an amount $\theta(\Lambda, p)$. Plane polarized gravitons can be defined in a similar way, and here Eq. (2.5.42) has the consequence that a Lorentz transformation Λ rotates the plane of polarization by an angle $2\theta(\Lambda, p)$.

2.6 Space Inversion and Time-Reversal

We saw in Section 2.3 that any homogeneous Lorentz transformation is either proper and orthochronous (i.e., $\mathrm{Det}\Lambda = +1$ and $\Lambda^0{}_0 \geq +1$) or else equal to a proper orthochronous transformation times either $\mathscr{P}$ or $\mathscr{T}$ or $\mathscr{P}\mathscr{T}$, where $\mathscr{P}$ and $\mathscr{T}$ are the space inversion and time-reversal transformations

$$\mathscr{P}^\mu{}_\nu = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} , \quad \mathscr{T}^\mu{}_\nu = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} .$$

It used to be thought self-evident that the fundamental multiplication rule of the Poincaré group

$$U(\bar{\Lambda}, \bar{a})\, U(\Lambda, a) = U(\bar{\Lambda}\Lambda, \bar{\Lambda}a + \bar{a})$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

would be valid even if Λ and/or $\bar{\Lambda}$ involved factors of $\mathscr{P}$ or $\mathscr{T}$ or $\mathscr{P}\mathscr{T}$. In particular, it was believed that there are operators corresponding to $\mathscr{P}$ and $\mathscr{T}$ themselves:

$$\mathsf{P} \equiv U(\mathscr{P}, 0) \qquad \mathsf{T} \equiv U(\mathscr{T}, 0)$$

such that

$$\mathsf{P}U(\Lambda, a)\mathsf{P}^{-1} = U(\mathscr{P}\Lambda\mathscr{P}^{-1}, \mathscr{P}a)\,, \tag{2.6.1}$$

$$\mathsf{T}\,U(\Lambda, a)\,\mathsf{T}^{-1} = U(\mathscr{T}\Lambda\mathscr{T}^{-1}, \mathscr{T}a) \tag{2.6.2}$$

for any proper orthochronous Lorentz transformation $\Lambda^{\mu}{}_{\nu}$ and translation a^{μ}. These transformation rules incorporate most of what is meant when we say that P or T are 'conserved'.

In 1956–57 it became understood[8] that this is true for P only in the approximation in which one ignores the effects of weak interactions, such as those that produce nuclear beta decay. Time-reversal survived for a while, but in 1964 there appeared indirect evidence[9] that these properties of T are also only approximately satisfied. (See Section 3.3.) In what follows, we will make believe that operators P and T satisfying Eqs. (2.6.1) and (2.6.2) actually exist, but it should be kept in mind that this is only an approximation.

Let us apply Eqs. (2.6.1) and (2.6.2) in the case of an infinitesimal transformation, i.e.,

$$\Lambda^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu} \qquad a^{\mu} = \epsilon^{\mu}$$

with $\omega_{\mu\nu} = -\omega_{\nu\mu}$ and ϵ_{μ} both infinitesimal. Using (2.4.3), and equating coefficients of $\omega_{\rho\sigma}$ and ϵ_{ρ} in Eqs. (2.6.1) and (2.6.2), we obtain the P and T transformation properties of the Poincaré generators

$$\mathsf{P}\,iJ^{\rho\sigma}\mathsf{P}^{-1} = i\mathscr{P}_{\mu}{}^{\rho}\mathscr{P}_{\nu}{}^{\sigma}J^{\mu\nu}\,, \tag{2.6.3}$$

$$\mathsf{P}\,iP^{\rho}\mathsf{P}^{-1} = i\mathscr{P}_{\mu}{}^{\rho}P^{\mu}\,, \tag{2.6.4}$$

$$\mathsf{T}\,iJ^{\rho\sigma}\mathsf{T}^{-1} = i\mathscr{T}_{\mu}{}^{\rho}\mathscr{T}_{\nu}{}^{\sigma}J^{\mu\nu}\,, \tag{2.6.5}$$

$$\mathsf{T}\,iP^{\rho}\mathsf{T}^{-1} = i\mathscr{T}_{\mu}{}^{\rho}P^{\mu}\,. \tag{2.6.6}$$

This is much like Eqs. (2.4.8) and (2.4.9), except that we have not cancelled factors of i on both sides of these equations, because at this point we have not yet decided whether P and T are linear and unitary or antilinear and antiunitary.

The decision is an easy one. Setting $\rho = 0$ in Eq. (2.6.4) gives

$$\mathsf{P}\,iH\mathsf{P}^{-1} = iH\,,$$

where $H \equiv P^0$ is the energy operator. If P were antiunitary and antilinear then it would anticommute with i, so $\mathsf{P}H\mathsf{P}^{-1} = -H$. But then for any state Ψ of energy $E > 0$, there would have to be another state $\mathsf{P}^{-1}\Psi$ of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

energy $-E < 0$. There are no states of negative energy (energy less than that of the vacuum), so we are forced to choose the other alternative: P *is linear and unitary, and commutes rather than anticommutes with* H.

On the other hand, setting $\rho = 0$ in Eq. (2.6.6) yields

$$\mathsf{T}\, i H \mathsf{T}^{-1} = -i H \,.$$

If we supposed that T is linear and unitary then we could simply cancel the is, and find $\mathsf{T}H\mathsf{T}^{-1} = -H$, with the again disastrous conclusion that for any state Ψ of energy E there is another state $\mathsf{T}^{-1}\Psi$ of energy $-E$. To avoid this, we are forced here to conclude that T *is antilinear and antiunitary.*

Now that we have decided that P is linear and T is antilinear, we can conveniently rewrite Eqs. (2.6.3)–(2.6.6) in terms of the generators (2.4.15)–(2.4.17) in a three-dimensional notation

$$\mathsf{P}\mathbf{J}\mathsf{P}^{-1} = +\mathbf{J} \,, \tag{2.6.7}$$

$$\mathsf{P}\mathbf{K}\mathsf{P}^{-1} = -\mathbf{K} \,, \tag{2.6.8}$$

$$\mathsf{P}\mathbf{P}\mathsf{P}^{-1} = -\mathbf{P} \,, \tag{2.6.9}$$

$$\mathsf{T}\mathbf{J}\mathsf{T}^{-1} = -\mathbf{J} \,, \tag{2.6.10}$$

$$\mathsf{T}\mathbf{K}\mathsf{T}^{-1} = +\mathbf{K} \,, \tag{2.6.11}$$

$$\mathsf{T}\mathbf{P}\mathsf{T}^{-1} = -\mathbf{P} \,, \tag{2.6.12}$$

and, as shown before,

$$\mathsf{P}H\mathsf{P}^{-1} = \mathsf{T}H\mathsf{T}^{-1} = H \,. \tag{2.6.13}$$

It is physically sensible that P should preserve the sign of $\mathbf{J}$, because at least the orbital part is a vector product $\mathbf{r} \times \mathbf{p}$ of two vectors, both of which change sign under an inversion of the spatial coordinate system. On the other hand, T reverses $\mathbf{J}$, because after time-reversal an observer will see all bodies spinning in the opposite direction. Note by the way that Eq. (2.6.10) is consistent with the angular-momentum commutation relations $\mathbf{J} \times \mathbf{J} = i\mathbf{J}$, because T reverses not only $\mathbf{J}$, but also i. The reader can easily check that Eqs. (2.6.7)–(2.6.13) are consistent with all the commutation relations (2.4.18)–(2.4.24).

Let us now consider what P and T do to one-particle states:

P: $M > 0$

The one-particle states $\Psi_{k,\sigma}$ are defined as eigenvectors of $\mathbf{P}, H$, and J_3 with eigenvalues $0, M$, and σ, respectively. From Eqs. (2.6.7), (2.6.9), and (2.6.13), we see that the same must be true of the state $\mathsf{P}\Psi_{k,\sigma}$, and therefore (barring degeneracies) these states can only differ by a phase

$$\mathsf{P}\Psi_{k,\sigma} = \eta_\sigma \Psi_{k,\sigma}$$

with a phase factor ($|\eta| = 1$) that may or may not depend on the spin σ.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

To see that η_σ is σ-independent, we note from (2.5.8), (2.5.20), and (2.5.21) that

$$(J_1 \pm iJ_2)\Psi_{k,\sigma} = \sqrt{(j \mp \sigma)(j \pm \sigma + 1)}\, \Psi_{k,\sigma\pm 1}\,, \tag{2.6.14}$$

where j is the particle's spin. Operating on both sides with P, we find

$$\eta_\sigma = \eta_{\sigma\pm 1}$$

and so η_σ is actually independent of σ. We therefore write

$$\mathsf{P}\Psi_{k,\sigma} = \eta\Psi_{k,\sigma} \tag{2.6.15}$$

with η a phase, known as the *intrinsic parity*, that depends only on the species of particle on which P acts.

To get to finite momentum states, we must apply the unitary operator $U(L(p))$ corresponding to the 'boost' (2.5.24):

$$\Psi_{p,\sigma} = \sqrt{M/p^0}\, U(L(p))\, \Psi_{k,\sigma}\,.$$

We note that

$$\mathscr{P}L(p)\mathscr{P}^{-1} = L(\mathscr{P}p)$$
$$\mathscr{P}p = \left(-\mathbf{p}, \sqrt{\mathbf{p}^2 + M^2}\right)$$

so using Eqs. (2.6.1) and (2.6.15), we have

$$\mathsf{P}\Psi_{p,\sigma} = \sqrt{M/p^0}\, U(L(\mathscr{P}p))\eta\Psi_{k,\sigma}$$

or in other words

$$\mathsf{P}\Psi_{p,\sigma} = \eta\Psi_{\mathscr{P}p,\sigma}\,. \tag{2.6.16}$$

T : $M > 0$

From Eqs. (2.6.10), (2.6.12), and (2.6.13), we see that the effect of T on the zero-momentum one-particle state $\Psi_{k,\sigma}$ is to yield a state with

$$\mathbf{P}(\mathsf{T}\Psi_{k,\sigma}) = 0\,,$$
$$H(\mathsf{T}\Psi_{k,\sigma}) = M(\mathsf{T}\Psi_{k,\sigma})\,,$$
$$J_3(\mathsf{T}\Psi_{k,\sigma}) = -\sigma(\mathsf{T}\Psi_{k,\sigma})\,,$$

and so

$$\mathsf{T}\Psi_{k,\sigma} = \zeta_\sigma\Psi_{k,-\sigma}\,,$$

where ζ_σ is a phase factor. Applying the operator T to (2.6.14), and recalling that T anticommutes with $\mathbf{J}$ *and* i, we find

$$(-J_1 \pm iJ_2)\zeta_\sigma\Psi_{k,-\sigma} = \sqrt{(j \mp \sigma)(j \pm \sigma + 1)}\zeta_{\sigma\pm 1}\, \Psi_{k,-\sigma\mp 1}\,.$$

Using Eq. (2.6.14) again on the left, we see that the square-root factors cancel, and so

$$-\zeta_\sigma = \zeta_{\sigma\pm 1}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

We write the solution as $\zeta_\sigma = \zeta(-)^{j-\sigma}$, with ζ another phase that depends only on the species of particle:

$$\mathsf{T}\Psi_{k,\sigma} = \zeta(-)^{j-\sigma}\Psi_{k,-\sigma} \,. \tag{2.6.17}$$

However, unlike the 'intrinsic parity' η, the time-reversal phase ζ has no physical significance. This is because we can redefine the one-particle states by a change of phase

$$\Psi_{k,\sigma} \to \Psi'_{k,\sigma} = \zeta^{1/2}\Psi_{k,\sigma} \,.$$

in such a way that the phase ζ is eliminated from the transformation rule

$$\mathsf{T}\Psi'_{k,\sigma} = \zeta^{*1/2}\mathsf{T}\Psi_{k,\sigma} = \zeta^{*1/2}\zeta(-)^{j-\sigma}\Psi_{k,-\sigma} = (-)^{j-\sigma}\Psi'_{k,-\sigma} \,.$$

In what follows we will keep the arbitrary phase ζ in Eq. (2.6.17), just to keep open our options in choosing the phase of the one-particle states, but it should be kept in mind that this phase is of no real importance.

To deal with states of finite momentum, we again apply the 'boost' (2.5.24). Note that

$$\mathscr{T}L(p)\mathscr{T}^{-1} = L(\mathscr{P}p) \,,$$

$$\mathscr{P}p = \left(-\mathbf{p}, \sqrt{\mathbf{p}^2 + M^2}\right) \,.$$

(That is, changing the sign of each element of $L^\mu{}_\nu$ with an odd number of *time*-indices is the same as changing the signs of elements with an odd number of *space*-indices.) Using Eqs. (2.6.2) and (2.5.5), we have then

$$\mathsf{T}\Psi_{p,\sigma} = \zeta(-)^{j-\sigma}\Psi_{\mathscr{P}p,-\sigma} \,. \tag{2.6.18}$$

P : $M = 0$

Acting on a state $\Psi_{k,\sigma}$, that is defined as an eigenvector of P^μ with eigenvalue $k^\mu = (0, 0, \kappa, \kappa)$ and an eigenvector of J_3 with eigenvalue σ, the parity operator P yields a state with four-momentum $(\mathscr{P}k)^\mu = (0, 0, -\kappa, \kappa)$ and J_3 equal to σ. Thus it takes a state of helicity (the component of spin along the direction of motion) σ into one of helicity $-\sigma$. As mentioned earlier, this shows that the existence of a space-inversion symmetry requires that any species of massless particle with non-zero helicity must be accompanied with another of opposite helicity. Because P does not leave the standard momentum invariant, it is convenient to consider instead the operator $U(R_2^{-1})\mathsf{P}$, where R_2 is a rotation that also takes k to $\mathscr{P}k$, conveniently chosen as a rotation by 180° around the two-axis

$$U(R_2) = \exp(i\pi J_2) \,. \tag{2.6.19}$$

Since $U(R_2^{-1})$ reverses the sign of J_3, we have

$$U(R_2^{-1})\mathsf{P}\Psi_{k,\sigma} = \eta_\sigma\Psi_{k,-\sigma} \tag{2.6.20}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

with η_σ a phase factor. Now, $R_2^{-1}\mathscr{P}$ commutes with the Lorentz 'boost' (2.5.45), and $\mathscr{P}$ commutes with the rotation $R(\hat{p})$ which takes the three-direction into the direction of $\mathbf{p}$, so by operating on (2.5.5) with P, we find for a general four-momentum p^μ

$$\mathsf{P}\Psi_{p,\sigma} = \sqrt{\frac{\kappa}{p^0}}\, U\left(R(\hat{p})R_2 B\left(\frac{|\mathbf{p}|}{\kappa}\right)\right) U(R_2^{-1})\mathsf{P}\Psi_{k,\sigma}$$

$$= \sqrt{\frac{\kappa}{p^0}}\,\eta_\sigma\, U\left(R(\hat{p})R_2\, B\left(\frac{|\mathbf{p}|}{\kappa}\right)\right)\Psi_{k,-\sigma}\,.$$

Note that $R(\hat{\mathbf{p}})R_2$ is a rotation that takes the three-axis into the direction of $-\hat{\mathbf{p}}$, but $U(R(\hat{\mathbf{p}})R_2)$ is not quite equal to $U(R(-\hat{\mathbf{p}}))$. According to (2.5.47),

$$U(R(-\hat{\mathbf{p}})) = \exp\Big(-i(\phi \pm \pi)J_3\Big)\,\exp\Big(-i(\pi-\theta)J_2\Big)$$

with azimuthal angle chosen as $\phi+\pi$ or $\phi-\pi$ according to whether $0 \le \phi < \pi$ or $\pi \le \phi < 2\pi$, so that it remains in the range of 0 to 2π. Then

$$\begin{aligned} U^{-1}\Big(R(-\hat{\mathbf{p}})\Big)U\Big(R(\hat{\mathbf{p}})R_2\Big) &= \exp\Big(i(\pi-\theta)J_2\Big) \\ &\quad\times \exp\Big(i(\phi\pm\pi)J_3\Big)\exp(-i\phi J_3)\exp(-i\theta J_2)\exp(i\pi J_2) \\ &= \exp\Big(i(\pi-\theta)J_2\Big)\exp(\pm i\pi J_3)\exp\Big(i(\pi-\theta)J_2\Big)\,. \end{aligned}$$

But a rotation of $\pm 180°$ around the three-axis reverses the sign of J_2, so

$$U\Big(R(\hat{\mathbf{p}})R_2\Big) = U\Big(R(-\hat{\mathbf{p}})\Big)\exp(\pm i\pi J_3)\,. \tag{2.6.21}$$

Also, $R(-\hat{\mathbf{p}})B(|\mathbf{p}|/k)$ is just the standard boost $L(\mathscr{P}p)$ in the direction $\mathscr{P}p = (-\mathbf{p}, p^0)$. We have then finally

$$\mathsf{P}\Psi_{p,\sigma} = \eta_\sigma \exp(\mp i\pi\sigma)\Psi_{\mathscr{P}p,-\sigma} \tag{2.6.22}$$

with the phase $-\pi\sigma$ or $+\pi\sigma$ according to whether the two-component of $\mathbf{p}$ is positive or negative, respectively. This peculiar change of sign in the operation of parity for massless particles of half-integer spin is due to the convention adopted in Eq. (2.5.47) for the rotation used to define massless particle states of arbitrary momentum. Because the rotation group is not simply connected, some discontinuity of this sort is unavoidable.

T : $M = 0$

Acting on the state $\Psi_{k,\sigma}$, which has values $k^\mu = (0,0,\kappa,\kappa)$ and σ for P^μ and J_3, the time-reversal operator T yields a state which has values $(\mathscr{P}k)^\mu = (0,0,-\kappa,\kappa)$ and $-\sigma$ for P^μ and J_3. Thus T does not change the helicity $\mathbf{J}\cdot\hat{k}$, and by itself has nothing to say about whether massless particles of one helicity σ are accompanied with others of helicity $-\sigma$. Because T like P does not leave the standard four-momentum k invariant,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

it is convenient to consider the generator $U(R_2^{-1})\mathsf{T}$, where R_2 is the rotation (2.6.19), which also takes k into $\mathscr{P}k$. This commutes with J_3, so

$$U(R_2^{-1})\,\mathsf{T}\,\Psi_{k,\sigma} = \zeta_\sigma \Psi_{k,\sigma} \tag{2.6.23}$$

with ζ_σ another phase. Since $R_2^{-1}\mathscr{T}$ commutes with the boost (2.5.45), and $\mathscr{T}$ commutes with the rotation $R(\hat{p})$, operating with T on the state (2.5.5) gives

$$\mathsf{T}\Psi_{p,\sigma} = \sqrt{\frac{\kappa}{p_0}}\, U\left[R(\hat{p})R_2\, B\left(\frac{|\mathbf{p}|}{\kappa}\right)\right]\zeta_\sigma \Psi_{k,\sigma}\,. \tag{2.6.24}$$

Using Eq. (2.6.21), this yields finally

$$\mathsf{T}\Psi_{p,\sigma} = \zeta_\sigma\, \exp(\pm i\pi\sigma)\Psi_{\mathscr{P}p,\sigma}\,. \tag{2.6.25}$$

Again, the top or bottom sign applies according to whether the two-component of $\mathbf{p}$ is positive or negative, respectively.

* * *

It is interesting that the square T^2 of the time-reversal operator has a very simple action on both massive and massless one-particle states. Using Eq. (2.6.18), and recalling that T is antiunitary, we see that for massive one-particle states:

$$\mathsf{T}^2\Psi_{p,\sigma} = \mathsf{T}\zeta(-)^{j-\sigma}\Psi_{\mathscr{P}p,-\sigma} = \zeta^*(-)^{j-\sigma}\zeta(-)^{j+\sigma}\Psi_{p,\sigma}$$

or in other words

$$\mathsf{T}^2\Psi_{p,\sigma} = (-)^{2j}\Psi_{p,\sigma}\,. \tag{2.6.26}$$

We get the same result for massless particles. If the two-component of $\mathbf{p}$ is positive then the two-component of $\mathscr{P}\mathbf{p}$ is negative, and vice-versa, so Eq. (2.6.25) gives

$$\begin{aligned}\mathsf{T}^2\Psi_{p,\sigma} &= \mathsf{T}\zeta_\sigma \exp(\pm i\pi\sigma)\Psi_{\mathscr{P}p,\sigma} = \zeta_\sigma^* \exp(\mp i\pi\sigma)\zeta_\sigma \exp(\mp i\pi\sigma)\Psi_{p,\sigma}\\ &= \exp(\mp 2i\pi\sigma)\Psi_{p,\sigma}\,.\end{aligned}$$

As long as σ is an integer or half-integer, this can be written

$$\mathsf{T}^2\Psi_{p,\sigma} = (-)^{2|\sigma|}\Psi_{p,\sigma}\,. \tag{2.6.27}$$

By the 'spin' of a massless particle, we usually mean the absolute value of the helicity, so Eq. (2.6.27) is the same as Eq. (2.6.26).

This result has an interesting consequence. When T^2 acts on any state Ψ of a system of non-interacting particles, either massive or massless, it yields a factor $(-)^{2j}$ or $(-)^{2|\sigma|}$ for each particle. Hence if the state contains an odd number of particles of half-integer spin or helicity (plus any number of particles of integer spin or helicity), we get an overall change of sign

$$\mathsf{T}^2\Psi = -\Psi\,. \tag{2.6.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

If we now 'turn on' various interactions, this result will be preserved, provided these interactions respect invariance under time-reversal, even if they do not respect rotational invariance. (For instance, these arguments will apply even if our system is subjected to arbitrary static gravitational and electric fields.) Now, suppose that Ψ is an eigenstate of the Hamiltonian. Since T commutes with the Hamiltonian, $\mathsf{T}\Psi$ will also be an eigenstate of the Hamiltonian. Is it the same state? If so, then $\mathsf{T}\Psi$ can differ from Ψ only by a phase

$$\mathsf{T}\Psi = \zeta\Psi ,$$

but then

$$\mathsf{T}^2\Psi = \mathsf{T}(\zeta\Psi) = \zeta^*\mathsf{T}\Psi = |\zeta|^2\Psi = \Psi ,$$

in contradiction with Eq. (2.6.28). We see that any energy eigenstate Ψ satisfying Eq. (2.6.28) must be degenerate with another eigenstate of the same energy. This is known as a 'Kramers degeneracy.'[10] Of course, this conclusion is trivial if the system is in a rotationally invariant environment, because the total angular-momentum j of any state of this system would have to be a half-integer, and there would therefore be $2j + 1 = 2, 4, \cdots$ degenerate states. The surprising result is that at least a two-fold degeneracy persists even if rotational invariance is perturbed by external fields, such as electrostatic fields, as long as these fields are invariant under T. In particular, if any particle had an electric or gravitational dipole moment then the degeneracy among its $2j + 1$ spin states would be entirely removed in a static electric or gravitational field, so such dipole moments are forbidden by time-reversal invariance.

For the sake of completeness, it should be mentioned that P and T can have more complicated effects on multiplets of particles with the same mass. This possibility will be considered in Appendix C of this chapter. No physically relevant examples are known.

2.7 Projective Representations*

We now return to the possibility mentioned in Section 2.2, that a group of symmetries may be represented projectively on physical states; that is, the elements T, $\bar{T}$, etc. of the symmetry group may be represented on the physical Hilbert space by unitary operators $U(T), U(\bar{T})$, etc., which

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

satisfy the composition rule

$$U(T)U(\bar{T}) = \exp\Big(i\phi(T,\bar{T})\Big)U(T\bar{T}) \tag{2.7.1}$$

with ϕ a real phase. (A bar is used here just to distinguish one symmetry operator from another.) The basic requirement that any phase ϕ in Eq. (2.7.1) would have to satisfy is the associativity condition

$$U(T_3)(U(T_2)U(T_1)) = (U(T_3)U(T_2))U(T_1)\,,$$

which imposes on ϕ the corresponding condition

$$\phi(T_2,T_1) + \phi(T_3,T_2T_1) = \phi(T_3,T_2) + \phi(T_3T_2,T_1)\,. \tag{2.7.2}$$

Of course, any phase of the form

$$\phi(T,\bar{T}) = \alpha(T\bar{T}) - \alpha(T) - \alpha(\bar{T}) \tag{2.7.3}$$

will automatically satisfy Eq. (2.7.2), but a projective representation with such a phase can be replaced with an ordinary representation by replacing $U(T)$ with

$$\tilde{U}(T) \equiv U(T)\exp\Big(i\alpha(T)\Big)$$

for which

$$\tilde{U}(T)\tilde{U}(\bar{T}) = \tilde{U}(T\bar{T})\ .$$

Any set of functions $\phi(T,\bar{T})$ that satisfy Eq. (2.7.2), and that differ only by functions $\Delta\phi(T,\bar{T})$ of the form (2.7.3), is called a 'two-cocycle'. A trivial cocycle is one that contains the function $\phi = 0$, and hence consists of functions of the form (2.7.3), which can be eliminated by a redefinition of $U(T)$. We are interested here in whether a symmetry group allows any non-trivial two-cocyles; that is, whether it may have a representation on the physical Hilbert space that is *intrinsically* projective, in the sense that the phase $\phi(T,\bar{T})$ *cannot* be eliminated in this way.

In order to answer this question, it is useful first to consider the effect of a phase ϕ in Eq. (2.7.1) on the commutation relations of the generators of infinitesimal transformations. When either $\bar{T}$ or T is the identity, the phase ϕ must clearly vanish

$$\phi(T,1) = \phi(1,\bar{T}) = 0\,. \tag{2.7.4}$$

When both T and $\bar{T}$ are *near* the identity the phase must be small. Using coordinates θ^a to parameterize group elements (as in Section 2.2), with $T(0) \equiv 1$, Eq. (2.7.4) tells us that the expansion of $\phi(T(\theta),T(\bar{\theta}))$ around $\theta = \bar{\theta} = 0$ must start with terms of order $\theta\bar{\theta}$:

$$\phi\Big(T(\theta),T(\bar{\theta})\Big) = f_{ab}\theta^a\bar{\theta}^b + \cdots\,, \tag{2.7.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

where f_{ab} are real numerical constants. Inserting this expansion in the power series expansion of Eq. (2.7.1), and repeating the steps that led to (2.2.22), we now have

$$[t_b, t_c] = iC^a{}_{bc}t_a + iC_{bc}1 \ , \tag{2.7.6}$$

where C_{bc} is the antisymmetric coefficient

$$C_{bc} = -f_{bc} + f_{cb} \ . \tag{2.7.7}$$

The appearance of terms on the right-hand side of the commutation relation proportional to the unit element (so-called *central charges*) is the counterpart for the Lie algebra of the presence of phases in a projective representation of a group.

The constants C_{bc} as well as $C^a{}_{bc}$ are subject to an important constraint, which follows from the Jacobi identity. Taking the commutator of (2.7.6) with t_d, and adding the same expressions with b, c, d replaced with c, d, b and d, b, c, the sum of the three double-commutators on the left-hand side vanishes identically, and so

$$C^a{}_{bc}C^e{}_{ad} + C^a{}_{cd}C^e{}_{ab} + C^a{}_{db}C^e{}_{ac} = 0 \tag{2.7.8}$$

and also

$$C^a{}_{bc}C_{ad} + C^a{}_{cd}C_{ab} + C^a{}_{db}C_{ac} = 0 \ . \tag{2.7.9}$$

Eq. (2.7.9) always has one obvious class of non-zero solutions for C_{ab}:

$$C_{ab} = C^e{}_{ab}\phi_e \ , \tag{2.7.10}$$

where ϕ_e is an arbitrary set of real constants. For these solutions, we can eliminate the central charges from Eq. (2.7.6) by a redefinition of the generators

$$t_a \to \tilde{t}_a \equiv t_a + \phi_a \ . \tag{2.7.11}$$

The new generators then satisfy the commutation relations without central charges

$$[\tilde{t}_b, \tilde{t}_c] = iC^a{}_{bc}\tilde{t}_a \ . \tag{2.7.12}$$

A given Lie algebra may or may not allow solutions of Eq. (2.7.9) other than Eq. (2.7.10).

We can now state the key theorem that governs the occurrence of intrinsically projective representations. The phase of any representation $U(T)$ of a given group can be chosen so that $\phi = 0$ in Eq. (2.7.1), if two conditions are met:

(a) The generators of the group in this representation can be redefined (as in Eq. (2.7.11)), so as to eliminate all central charges from the Lie algebra.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

(b) The group is simply connected, i.e., any two group elements may be connected by a path lying within the group, and any two such paths may be continuously transformed into one another. (An equivalent statement is that any loop that starts and ends at the same group element may be shrunk continuously to a point.)

This theorem is proved in Appendix B of this chapter, which also offers comments about the case of groups that are not simply connected. It shows that there are just two (not exclusive) ways that intrinsically projective representations may arise: either algebraically, because the group is represented projectively even near the identity, or topologically, because the group is not simply connected, and hence a path from 1 to T and then from T to $\bar{T}$ may not be continuously deformable into some other path from 1 to $T\bar{T}$. In the latter case, the phase ϕ in Eq. (2.7.1) depends on the particular choice of standard paths, leading from the origin to the various group elements, that are used to define the corresponding U-operators.

Let's now consider each of these possibilities in turn for the special case of the inhomogeneous Lorentz group.

(A) Algebra

With central charges, the commutation relations of the generators of the inhomogeneous Lorentz group would read

$$i[J^{\mu\nu}, J^{\rho\sigma}] = \eta^{\nu\rho} J^{\mu\sigma} - \eta^{\mu\rho} J^{\nu\sigma} - \eta^{\sigma\mu} J^{\rho\nu} + \eta^{\sigma\nu} J^{\rho\mu} + C^{\rho\sigma,\mu\nu} , \tag{2.7.13}$$

$$i[P^{\mu}, J^{\rho\sigma}] = \eta^{\mu\rho} P^{\sigma} - \eta^{\mu\sigma} P^{\rho} + C^{\rho\sigma,\mu} , \tag{2.7.14}$$

$$i[J^{\mu\nu}, P^{\rho}] = \eta^{\nu\rho} P^{\mu} - \eta^{\mu\rho} P^{\nu} + C^{\rho,\mu\nu} , \tag{2.7.15}$$

$$i[P^{\mu}, P^{\rho}] = C^{\rho,\mu} \tag{2.7.16}$$

in place of Eqs. (2.4.12)–(2.4.14). We see that the Cs also satisfy the antisymmetry conditions

$$C^{\rho\sigma,\mu\nu} = -C^{\mu\nu,\rho\sigma} , \tag{2.7.17}$$

$$C^{\rho\sigma,\mu} = -C^{\mu,\rho\sigma} , \tag{2.7.18}$$

$$C^{\rho,\mu} = -C^{\mu,\rho} . \tag{2.7.19}$$

We will now show that all these constants have additional algebraic properties that allow them to be eliminated by shifting the definitions of $J^{\mu\nu}$ and P^{μ} by constant terms. (This corresponds to redefining the phase of the operators $U(\Lambda, a)$.) To derive these properties, we apply the Jacobi

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

identities

$$\left[J^{\mu\nu},[P^{\rho},P^{\sigma}]\right]+\left[P^{\sigma},[J^{\mu\nu},P^{\rho}]\right]+\left[P^{\rho},[P^{\sigma},J^{\mu\nu}]\right]=0, \tag{2.7.20}$$

$$\left[J^{\lambda\eta},[J^{\mu\nu},P^{\rho}]\right]+\left[P^{\rho},[J^{\lambda\eta},J^{\mu\nu}]\right]+\left[J^{\mu\nu},[P^{\rho},J^{\lambda\eta}]\right]=0, \tag{2.7.21}$$

$$\left[J^{\lambda\eta},[J^{\mu\nu},J^{\rho\sigma}]\right]+\left[J^{\rho\sigma},[J^{\lambda\eta},J^{\mu\nu}]\right]+\left[J^{\mu\nu},[J^{\rho\sigma},J^{\lambda\eta}]\right]=0\,. \tag{2.7.22}$$

(The Jacobi identity involving three Ps is automatically satisfied, and hence yields no further information.) Using Eqs. (2.7.13)–(2.7.16) in Eqs. (2.7.20)–(2.7.22), we obtain algebraic conditions on the Cs

$$0=\eta^{\nu\rho}C^{\mu,\sigma}-\eta^{\mu\rho}C^{\nu,\sigma}-\eta^{\nu\sigma}C^{\mu,\rho}+\eta^{\mu\sigma}C^{\nu,\rho}\,, \tag{2.7.23}$$

$$\begin{aligned}0=&\eta^{\nu\rho}C^{\mu,\lambda\eta}-\eta^{\mu\rho}C^{\nu,\lambda\eta}-\eta^{\mu\eta}C^{\rho,\lambda\nu}+\eta^{\lambda\mu}C^{\rho,\eta\nu}\\&+\eta^{\lambda\nu}C^{\rho,\mu\eta}-\eta^{\eta\nu}C^{\rho,\mu\lambda}+\eta^{\rho\lambda}C^{\eta,\mu\nu}-\eta^{\rho\eta}C^{\lambda,\mu\nu}\,,\end{aligned} \tag{2.7.24}$$

$$\begin{aligned}0=&\eta^{\nu\rho}C^{\mu\sigma,\lambda\eta}-\eta^{\mu\rho}C^{\nu\sigma,\lambda\eta}-\eta^{\sigma\mu}C^{\rho\nu,\lambda\eta}+\eta^{\sigma\nu}C^{\rho\mu,\lambda\eta}\\&+\eta^{\eta\mu}C^{\lambda\nu,\rho\sigma}-\eta^{\lambda\mu}C^{\eta\nu,\rho\sigma}-\eta^{\nu\lambda}C^{\mu\eta,\rho\sigma}+\eta^{\nu\eta}C^{\mu\lambda,\rho\sigma}\\&+\eta^{\sigma\lambda}C^{\rho\eta,\mu\nu}-\eta^{\rho\lambda}C^{\sigma\eta,\mu\nu}-\eta^{\eta\rho}C^{\lambda\sigma,\mu\nu}+\eta^{\eta\sigma}C^{\lambda\rho,\mu\nu}\,.\end{aligned} \tag{2.7.25}$$

Contracting Eq. (2.7.23) with $\eta_{\nu\rho}$ gives

$$C^{\mu,\sigma}=0\,. \tag{2.7.26}$$

On the other hand, the constants $C^{\mu,\lambda\eta}$ and $C^{\rho\sigma,\mu\nu}$ are not necessarily zero, but their algebraic structure is simple enough so that they can be eliminated by shifting the definitions of P^{μ} and $J^{\mu\nu}$, respectively. Contracting Eq. (2.7.24) with $\eta_{\nu\rho}$ gives

$$C^{\mu,\lambda\eta}=\eta^{\mu\eta}C^{\lambda}-\eta^{\mu\lambda}C^{\eta}\,, \tag{2.7.27}$$

$$C^{\lambda}\equiv\frac{1}{3}\,\eta_{\rho\nu}C^{\rho,\lambda\nu}\,. \tag{2.7.28}$$

Also, contracting (2.7.25) with $\eta_{\nu\rho}$ gives

$$C^{\mu\sigma,\lambda\eta}=\eta^{\eta\mu}C^{\lambda\sigma}-\eta^{\lambda\mu}C^{\eta\sigma}+\eta^{\sigma\lambda}C^{\eta\mu}-\eta^{\eta\sigma}C^{\lambda\mu}\,, \tag{2.7.29}$$

$$C^{\lambda\sigma}\equiv\frac{1}{2}\,\eta_{\nu\rho}C^{\lambda\nu,\sigma\rho}\,. \tag{2.7.30}$$

(These expressions automatically satisfy Eqs. (2.7.24) and (2.7.25), so there is no further information to be gained from the Jacobi identities.) We now see that if the Cs are not zero, they can be eliminated by defining new generators

$$\tilde{P}^{\mu}\equiv P^{\mu}+C^{\mu}\,, \tag{2.7.31}$$

$$\tilde{J}^{\mu\sigma}\equiv J^{\mu\sigma}+C^{\mu\sigma}\,, \tag{2.7.32}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

and the commutation relations are then what they would be for an ordinary representation

$$i[\tilde{J}^{\mu\nu}, \tilde{J}^{\rho\sigma}] = \eta^{\nu\rho}\tilde{J}^{\mu\sigma} - \eta^{\mu\rho}\tilde{J}^{\nu\sigma} - \eta^{\sigma\mu}\tilde{J}^{\rho\nu} + \eta^{\sigma\nu}\tilde{J}^{\rho\mu}\,, \tag{2.7.33}$$

$$i[\tilde{J}^{\mu\nu}, \tilde{P}^{\rho}] = \eta^{\nu\rho}\tilde{P}^{\mu} - \eta^{\mu\rho}\tilde{P}^{\nu}\,, \tag{2.7.34}$$

$$i[\tilde{P}^{\mu}, \tilde{P}^{\rho}] = 0\,. \tag{2.7.35}$$

The commutation relations will always be taken in the form Eqs. (2.7.33)–(2.7.35), but with tildas dropped.

Incidentally, the fact that there is no central charge in the algebra of the $J^{\mu\nu}$ could have been immediately inferred from the fact that this algebra is of the type known as 'semi-simple'. (Semi-simple Lie algebras are those that have no 'invariant Abelian' subalgebra, consisting of generators that commute with each other and whose commutators with any other generators also belong to the subalgebra.) There is a general theorem[11] that any central charges in semi-simple Lie algebras may always be removed by a redefinition of generators, as in Eq. (2.7.32). On the other hand, the full Poincaré algebra spanned by $J^{\mu\nu}$ and P^{μ} is not semi-simple (the P^{μ} form an invariant Abelian subalgebra), and we needed a special argument to show that its central charges can also be eliminated in this way. Indeed, the non-semi-simple Galilean algebra discussed in Section 2.4 does allow a central charge, the mass M.

We see that the inhomogeneous Lorentz group satisfies the first of the two conditions needed to rule out intrinsically projective representations. How about the second?

(B) Topology

To explore the topology of the inhomogeneous Lorentz group, it is very convenient to represent homogeneous Lorentz transformations by 2×2 complex matrices. Any real four-vector V^{μ} can be used to construct an Hermitian 2×2 matrix

$$v \equiv V^{\mu}\sigma_{\mu} = \begin{pmatrix} V^0 + V^3 & V^1 - iV^2 \\ V^1 + iV^2 & V^0 - V^3 \end{pmatrix}\,, \tag{2.7.36}$$

where σ_{μ} are the usual Pauli matrices with $\sigma_0 \equiv 1$. Conversely any 2×2 Hermitian matrix can be put in this form, and therefore defines a real four-vector V^{μ}.

The property of Hermiticity will be preserved under the transformation

$$v \to \lambda v \lambda^{\dagger} \tag{2.7.37}$$

with λ an arbitrary complex 2×2 matrix. Furthermore, the covariant

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

square of the four-vector is

$$V_\mu V^\mu = (V^1)^2 + (V^2)^2 + (V^3)^2 - (V^0)^2 = -\text{Det } v \tag{2.7.38}$$

and this determinant is preserved by the transformation (2.7.37) provided that

$$|\text{Det } \lambda| = 1 \,. \tag{2.7.39}$$

Each complex 2×2 matrix λ satisfying Eq. (2.7.39) thus defines a real linear transformation of V^μ that leaves Eq. (2.7.38) invariant, i.e., a homogeneous Lorentz transformation $\Lambda(\lambda)$:

$$\lambda V^\mu \sigma_\mu \lambda^\dagger = (\Lambda^\mu{}_\nu(\lambda) V^\nu)\sigma_\mu \,. \tag{2.7.40}$$

Furthermore, for two such matrices λ and $\bar{\lambda}$, we have

$$\begin{aligned}(\lambda\bar{\lambda})V^\mu\sigma_\mu(\lambda\bar{\lambda})^\dagger &= \lambda(\bar{\lambda}V^\mu\sigma_\mu\bar{\lambda}^\dagger)\lambda^\dagger \\ &= \lambda\Lambda^\mu{}_\nu(\bar{\lambda})V^\nu\sigma_\mu\lambda^\dagger = \Lambda^\mu{}_\rho(\lambda)\Lambda^\rho{}_\nu(\bar{\lambda})V^\nu\sigma_\mu\end{aligned}$$

and so

$$\Lambda(\lambda\bar{\lambda}) = \Lambda(\lambda)\Lambda(\bar{\lambda}) \,. \tag{2.7.41}$$

However, two λs that differ only by an overall phase have the same effect on v in Eq. (2.7.37), and so correspond to the same Lorentz transformation. It is therefore convenient to adjust the phase of the λs so that

$$\text{Det } \lambda = 1, \tag{2.7.42}$$

which is consistent with Eq. (2.7.41). The 2×2 complex matrices with unit determinant form a group, known as $SL(2, C)$. (The SL stands for 'special linear', with 'special' denoting a unit determinant, while C stands for 'complex'.) The group elements depend on $4 - 1 = 3$ complex parameters, or 6 real parameters, the same number as the Lorentz group. However, $SL(2, C)$ is not the same as the Lorentz group; if λ is a matrix in $SL(2, C)$, then so is $-\lambda$, and both λ and $-\lambda$ produce the same Lorentz transformation in Eq. (2.7.37). Indeed, it is easy to see that the matrix

$$\lambda(\theta) = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}$$

produces a Lorentz transformation $\Lambda(\lambda(\theta))$ which is just a rotation by an angle θ around the three-axis, and hence $\lambda = -1$ produces a rotation by an angle 2π. The Lorentz group is not the same as $SL(2, C)$, but

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

rather** $SL(2,C)/Z_2$, which is the group of complex 2×2 matrices with unit determinant, and with λ identified with $-\lambda$.

Now, what is the topology of the Lorentz group? By the polar decomposition theorem[12], any complex non-singular matrix λ may be written in the form

$$\lambda = u\, e^h ,$$

where u is unitary and h is Hermitian

$$u^\dagger u = 1 , \qquad h^\dagger = h .$$

Since Det u is a phase factor, and Det $\exp h = \exp \operatorname{Tr} h$ is real and positive, the condition (2.7.42) requires both

$$\operatorname{Det} u = 1 ,$$
$$\operatorname{Tr} h = 0.$$

(The factor u simply provides the rotation subgroup of the Lorentz group; if u is unitary then $\operatorname{Tr}(uvu^\dagger) = \operatorname{Tr} v$, so $V^0 = \frac{1}{2}\operatorname{Tr} v$ is left invariant by $\Lambda(u)$.) Furthermore, this decomposition is unique, so $SL(2,C)$ is topologically just the direct product (i.e., the set of pairs of points) of the space of all us and the space of all hs. Any Hermitian traceless 2×2 matrix h can be expressed as

$$h = \begin{pmatrix} c & a - ib \\ a + ib & -c \end{pmatrix}$$

with a, b, c real but otherwise unconstrained, so the space of all hs is topologically the same as ordinary three-dimensional flat space, R_3. On the other hand, any unitary 2×2 matrix with unit determinant can be expressed as

$$u = \begin{pmatrix} d + ie & f + ig \\ -f + ig & d - ie \end{pmatrix}$$

with d, e, f, g subject to the single non-linear constraint

$$d^2 + e^2 + f^2 + g^2 = 1 ,$$

so the space $SU(2)$ of all us is topologically the same as S_3, the three-dimensional surface of a spherical ball in flat four-dimensional space. Thus $SL(2,C)$ is topologically the same as the direct product $R_3 \times S_3$. This is simply connected: any curve connecting two points of R_3 or S_3 can be deformed into any other, and the same is true of the direct product. (All

** The group Z_2 consists of just elements $+1$ and -1. In general, when we write G/H, with H an invariant subgroup of G, we mean the group G with elements g and gh identified if $g \in G$ and $h \in H$. The subgroup Z_2 is trivially invariant because its elements commute with all elements of $SL(2,C)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

spheres S_n except the circle S_1 are simply connected.) However, we are interested in $SL(2,C)/Z_2$, not $SL(2,C)$. Identifying λ with $-\lambda$ is the same as identifying the unitary factors u and $-u$ (because e^h is always positive), so the Lorentz group has the topology of $R_3 \times S_3/Z_2$, where S_3/Z_2 is the three-dimensional spherical surface with opposite points of the sphere identified. This is *not* simply connected; for instance, a path on S_3 from u to u' cannot be continuously deformed into a path on S_3 from u to $-u'$, even though these two paths link the same points of S_3/Z_2. In fact, S_3/Z_2 is *doubly* connected; the paths between any two points fall into two classes, depending on whether or not they involve an inversion $u \to -u$, and any path in one class can be deformed into another path of that class. An equivalent statement is that a double loop, that goes twice over the same path from any element back to itself, may be continuously contracted to a point. (As discussed in Appendix B, this is summarized mathematically in the statement that the fundamental group, or first homotopy group, of S_3/Z_2 is Z_2.) Similarly, the inhomogeneous Lorentz group has the same topology as $R_4 \times R_3 \times S_3/Z_2$, and is therefore also doubly connected.

Because the Lorentz group (homogeneous or inhomogeneous) is not simply connected, it does have intrinsically projective representations. However, because the double loop that goes twice from 1 to Λ to $\Lambda\bar{\Lambda}$ and then back to 1 *can* be contracted to a point, we must have

$$\left[U(\Lambda)U(\bar{\Lambda})U^{-1}(\Lambda\bar{\Lambda})\right]^2 = 1$$

and hence the phase $e^{i\phi(\Lambda,\bar{\Lambda})}$ is just a sign

$$U(\Lambda)U(\bar{\Lambda}) = \pm U(\Lambda\bar{\Lambda})\,. \tag{2.7.43}$$

Likewise, for the inhomogeneous Lorentz group

$$U(\Lambda,a)U(\bar{\Lambda},\bar{a}) = \pm U(\Lambda\bar{\Lambda},\Lambda\bar{a}+a)\,. \tag{2.7.44}$$

These 'representations up to a sign' are familiar; they are just the states of integer spin, for which the signs in Eqs. (2.7.43) and (2.7.44) are always $+1$, and the states of half-integer spin, for which these signs are $+1$ or -1 according to whether the path from 1 to Λ to $\Lambda\bar{\Lambda}$ and then back to 1 is or is not contractible to a point. This difference arises because a rotation of 2π around the three-axis acting on a state with angular-momentum three-component σ produces a phase $e^{2i\pi\sigma}$, and thus has no effect on a state of integer spin and produces a sign change when acting on a state of half-integer spin. (These two cases correspond to the two irreducible representations of the first homotopy group, Z_2.) Thus Eq. (2.7.43) or Eq. (2.7.44) imposes a superselection rule: we must not mix states of integer and half-integer spin.

For finite mass, the limitation to integer or half-integer spin was pre-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

viously derived by purely algebraic means from the well-known representations of the generators of the little group, which here are just the angular-momentum matrices $\vec{J}^{(j)}$ with j integer or half-integer. On the other hand, for zero mass the action of the little group on physical one-particle states is just a rotation around the momentum, and here there is no *algebraic* reason for a limitation to integer or half-integer helicity. There is, however, a *topological* reason: a rotation by an angle 4π around the momentum can be continuously deformed into no rotation at all, so the factor $\exp(4\pi i\sigma)$ must be unity, and hence σ must be an integer or half-integer.

Instead of working with projective representations and imposing a superselection rule, we can just as well expand the Lorentz group, taking it as $SL(2,C)$ itself, instead of $SL(2,C)/Z_2$ as before. Ordinary rotation invariance forbids transitions between states of integer and half-integer total spin, so the only difference is that now the group is simply-connected, and it therefore has only ordinary representations, not projective representations, so that we cannot infer a superselection rule. This does not mean that we actually can prepare physical systems in linear combinations of states of integer and half-integer spin, but only that the observed Lorentz invariance of nature cannot be used to show that such superpositions are impossible.

Similar remarks apply to any symmetry group. If its Lie algebra involves central charges, then we can always expand the algebra to include generators that commute with anything, and whose eigenvalues are the central charges, just as we did when we added a mass operator to the Lie algebra of the Galilean group at the end of Section 2.4. The expanded Lie algebra is then, of course, free of central charges, so the part of the group near the identity has only ordinary representations, and does not require any superselection rule. Likewise, even though a Lie group G may not be simply connected, it can always be expressed as C/H, where C is a simply connected group known as the 'universal covering group' of G, and H is an invariant subgroup[†] of C. In general, we may just as well take the symmetry group as C instead of G, because there is no difference in their consequences, except that G implies a superselection rule, while C does not. In short, the issue of superselection rules is a bit of a red herring; *it may or it may not be possible to prepare physical systems in arbitrary superpositions of states, but one cannot settle the question by*

[†] The first homotopy group of C/H is H. We have seen that the covering group of the homogeneous Lorentz group is $SL(2,C)$, and the covering group of the three-dimensional rotation group is $SU(2)$. This connection with SL and SU groups is special to the case of three, four, or six dimensions; for general dimensions d the covering group of $SO(d)$ is given a special name, '*Spin*(d)'.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

reference to symmetry principles, because whatever one thinks the symmetry group of nature may be, there is always another group whose consequences are identical except for the absence of superselection rules.

Appendix A The Symmetry Representation Theorem

This appendix presents the proof of the fundamental theorem of Wigner[2] that any symmetry transformation can be represented on the Hilbert space of physical states by an operator that is either linear and unitary or antilinear and antiunitary. For our present purposes, the property of symmetry transformations on which we chiefly rely is that they are ray transformations T that preserve transition probabilities, in the sense that if Ψ_1 and Ψ_2 are state-vectors belonging to rays $\mathscr{R}_1$ and $\mathscr{R}_2$ then any state-vectors Ψ_1' and Ψ_2' belonging to the transformed rays $T\mathscr{R}_1$ and $T\mathscr{R}_2$ satisfy

$$|(\Psi_1', \Psi_2')|^2 = |(\Psi_1, \Psi_2)|^2. \tag{2.A.1}$$

We also require that a symmetry transformation should have an inverse that preserves transition probabilities in the same sense.

To start, consider some complete orthonormal set of state-vectors Ψ_k belonging to rays $\mathscr{R}_k$, with

$$(\Psi_k, \Psi_l) = \delta_{kl}, \tag{2.A.2}$$

and let Ψ_k' be some arbitrary choice of state-vectors belonging to the transformed rays $T\mathscr{R}_k$. From Eq. (2.A.1), we have

$$|(\Psi_k', \Psi_l')|^2 = |(\Psi_k, \Psi_l)|^2 = \delta_{kl}.$$

But (Ψ_k', Ψ_k') is automatically real and positive, so this requires that it should have the value unity, and therefore

$$(\Psi_k', \Psi_l') = \delta_{kl}. \tag{2.A.3}$$

It is easy to see that these transformed states Ψ_k' also form a complete set, for if there were any non-zero state-vector Ψ' that was orthogonal to all of the Ψ_k', then the inverse transform of the ray to which Ψ' belongs would consist of non-zero state-vectors Ψ'' for which, for all k:

$$|(\Psi_k, \Psi'')|^2 = |(\Psi_k', \Psi')|^2 = 0,$$

which is impossible since the Ψ_k were assumed to form a complete set.

We must now establish a phase convention for the states Ψ_k'. For this purpose, we single out one of the Ψ_k, say Ψ_1, and consider the state-vectors

$$\Upsilon_k \equiv \frac{1}{\sqrt{2}}[\Psi_1 + \Psi_k] \tag{2.A.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

belonging to some ray $\mathscr{S}_k$, with $k \neq 1$. Any state-vector Υ'_k belonging to the transformed ray $T\mathscr{S}_k$ may be expanded in the state-vectors Ψ'_l,

$$\Upsilon'_k = \sum_l c_{kl}\Psi'_l \,.$$

From Eq. (2.A.1) we have

$$|c_{kk}| = |c_{k1}| = \frac{1}{\sqrt{2}}$$

and for $l \neq k$ and $l \neq 1$:

$$c_{kl} = 0.$$

For any given k, by an appropriate choice of phase of the two state-vectors Υ'_k and Ψ'_k we can clearly adjust the phases of the two non-zero coefficients c_{kk} and c_{k1} so that both coefficients are just $1/\sqrt{2}$. From now on, the state-vectors Υ'_k and Ψ'_k chosen in this way will be denoted $U\Upsilon_k$ and $U\Psi_k$. As we have seen,

$$U\frac{1}{\sqrt{2}}[\Psi_k + \Psi_1] = U\Upsilon_k = \frac{1}{\sqrt{2}}[U\Psi_k + U\Psi_1] \,. \tag{2.A.5}$$

However, it still remains to define $U\Psi$ for general state-vectors Ψ.

Now consider an arbitrary state-vector Ψ belonging to an arbitrary ray $\mathscr{R}$, and expand it in the Ψ_k:

$$\Psi = \sum_k C_k\Psi_k \,. \tag{2.A.6}$$

Any state Ψ' that belongs to the transformed ray $T\mathscr{R}$ may similarly be expanded in the complete orthonormal set $U\Psi_k$:

$$\Psi' = \sum_k C'_k U\Psi_k \,. \tag{2.A.7}$$

The equality of $|(\Psi_k, \Psi)|^2$ and $|(U\Psi_k, \Psi')|^2$ tells us that for all k (including $k = 1$):

$$|C_k|^2 = |C'_k|^2, \tag{2.A.8}$$

while the equality of $|(\Upsilon_k, \Psi)|^2$ and $|(U\Upsilon_k, \Psi')|^2$ tells us that for all $k \neq 1$:

$$|C_k + C_1|^2 = |C'_k + C'_1|^2. \tag{2.A.9}$$

The ratio of Eqs. (2.A.9) and (2.A.8) yields the formula

$$\mathrm{Re}\,(C_k/C_1) = \mathrm{Re}\,(C'_k/C'_1) \tag{2.A.10}$$

which with Eq. (2.A.8) also requires

$$\mathrm{Im}\,(C_k/C_1) = \pm\mathrm{Im}\,(C'_k/C'_1) \tag{2.A.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

and therefore either

$$C_k/C_1 = C'_k/C'_1, \tag{2.A.12}$$

or else

$$C_k/C_1 = (C'_k/C'_1)^*. \tag{2.A.13}$$

Furthermore, we can show that the same choice must be made for each k. (This step in the proof was omitted by Wigner.) To see this, suppose that for some k, we have $C_k/C_1 = C'_k/C'_1$, while for some $l \neq k$, we have instead $C_l/C_1 = (C'_l/C'_1)^*$. Suppose also that both ratios are complex, so that these are really different cases. (This incidentally requires that $k \neq 1$ and $l \neq 1$, as well as $k \neq l$.) We will show that this is impossible.

Define a state-vector $\Phi \equiv \frac{1}{\sqrt{3}}[\Psi_1 + \Psi_k + \Psi_l]$. Since all the ratios of the coefficients in this state-vector are real, we must get the same ratios in any state-vector Φ' belonging to the transformed ray:

$$\Phi' = \frac{\alpha}{\sqrt{3}}[U\Psi_1 + U\Psi_k + U\Psi_l]\,,$$

where α is a phase factor with $|\alpha| = 1$. But then the equality of the transition probabilities $|(\Phi, \Psi)|$ and $|(\Phi', \Psi')|$ requires that

$$\left|1 + \frac{C'_k}{C'_1} + \frac{C'_l}{C'_1}\right|^2 = \left|1 + \frac{C_k}{C_1} + \frac{C_l}{C_1}\right|^2$$

and hence

$$\left|1 + \frac{C_k}{C_1} + \frac{C^*_l}{C^*_1}\right|^2 = \left|1 + \frac{C_k}{C_1} + \frac{C_l}{C_1}\right|^2.$$

This is only possible if

$$\mathrm{Re}\left(\frac{C_k}{C_1}\frac{C^*_l}{C^*_1}\right) = \mathrm{Re}\left(\frac{C_k}{C_1}\frac{C_l}{C_1}\right)$$

or, in other words, if

$$\mathrm{Im}\left(\frac{C_k}{C_1}\right)\mathrm{Im}\left(\frac{C_l}{C_1}\right) = 0\,.$$

Hence either C_k/C_1 or C_l/C_1 must be real for any pair k, l, in contradiction with our assumptions. We see then that for a given symmetry transformation T applied to a given state-vector $\sum_k C_k\Psi_k$, we must have either Eq. (2.A.12) for all k, or else Eq. (2.A.13) for all k.

Wigner ruled out the second possibility, Eq. (2.A.13), because as he showed any symmetry transformation for which this possibility is realized would have to involve a reversal in the time coordinate, and in the proof he presented he was considering only symmetries like rotations that do

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

not affect the direction of time. Here we are treating symmetries involving time-reversal on the same basis as all other symmetries, so we will have to consider that, for each symmetry T and state-vector $\sum_k C_k \Psi_k$, either Eq. (2.A.12) or Eq. (2.A.13) may apply. Depending on which of these alternatives is realized, we will now define $U\Psi$ to be the particular one of the state-vectors Ψ' belonging to the ray $T\mathscr{R}$ with phase chosen so that either $C_1 = C_1'$ or $C_1 = C_1'^*$, respectively. Then either

$$U(\sum_k C_k \Psi_k) = \sum_k C_k U \Psi_k \tag{2.A.14}$$

or else

$$U(\sum_k C_k \Psi_k) = \sum_k C_k^* U \Psi_k \,. \tag{2.A.15}$$

It remains to be proved that for a given symmetry transformation, we must make the same choice between Eqs. (2.A.14) and (2.A.15) for arbitrary values of the coefficients C_k. Suppose that Eq. (2.A.14) applies for a state-vector $\sum_k A_k \Psi_k$ while Eq. (2.A.15) applies for a state-vector $\sum_k B_k \Psi_k$. Then the invariance of transition probabilities requires that

$$|\sum_k B_k^* A_k|^2 = |\sum_k B_k A_k|^2$$

or equivalently

$$\sum_{kl} \mathrm{Im}\,(A_k^* A_l)\,\mathrm{Im}\,(B_k^* B_l) = 0 \,. \tag{2.A.16}$$

We cannot rule out the possibility that Eq. (2.A.16) may be satisfied for a pair of state-vectors $\sum_k A_k \Psi_k$ and $\sum_k B_k \Psi_k$ belonging to different rays. However, for any pair of such state-vectors, with neither A_k nor B_k all of the same phase (so that Eqs. (2.A.14) and (2.A.15) are not the same), we can always find a third state-vector $\sum_k C_k \Psi_k$ for which*

$$\sum_{kl} \mathrm{Im}\,(C_k^* C_l)\,\mathrm{Im}\,(A_k^* A_l) \neq 0 \tag{2.A.17}$$

and also

$$\sum_{kl} \mathrm{Im}\,(C_k^* C_l)\,\mathrm{Im}\,(B_k^* B_l) \neq 0 \,. \tag{2.A.18}$$

* If for some pair k,l both $A_k^* A_l$ and $B_k^* B_l$ are complex, then choose all Cs to vanish except for C_k and C_l, and choose these two coefficients to have different phases. If $A_k^* A_l$ is complex but $B_k^* B_l$ is real for some pair k,l, then there must be some other pair m,n (possibly with either m or n but not both equal to k or l) for which $B_m^* B_n$ is complex. If also $A_m^* A_n$ is complex, then choose all Cs to vanish except for C_m and C_n, and choose these two coefficients to have different phase. If $A_m^* A_n$ is real, then choose all Cs to vanish except for C_k, C_l, C_m, and C_n, and choose these four coefficients all to have different phases. The case where $B_k^* B_l$ is complex but $A_k^* A_l$ is real is handled in just the same way.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

As we have seen, it follows from Eq. (2.A.17) that the same choice between Eqs. (2.A.14) and (2.A.15) must be made for $\sum_k A_k\Psi_k$ and $\sum_k C_k\Psi_k$, and it follows from Eq. (2.A.18) that the same choice between Eqs. (2.A.14) and (2.A.15) must be made for $\sum_k B_k\Psi_k$ and $\sum_k C_k\Psi_k$, so the same choice between Eqs. (2.A.14) and (2.A.15) must also be made for the two state-vectors $\sum_k A_k\Psi_k$ and $\sum_k B_k\Psi_k$ with which we started. We have thus shown that for a given symmetry transformation T either all state-vectors satisfy Eq. (2.A.14) or else they all satisfy Eq. (2.A.15).

It is now easy to show that as we have defined it, the quantum mechanical operator U is either linear and unitary or antilinear and antiunitary. First, suppose that Eq. (2.A.14) is satisfied for all state-vectors $\sum_k C_k\Psi_k$. Any two state-vectors Ψ and Φ may be expanded as

$$\Psi = \sum_k A_k\Psi_k\ , \quad \Phi = \sum_k B_k\Psi_k$$

and so, using Eq. (2.A.14),

$$\begin{aligned} U(\alpha\Psi + \beta\Phi) &= U\sum_k(\alpha A_k + \beta B_k)\Psi_k = \sum_k(\alpha A_k + \beta B_k)U\Psi_k \\ &= \alpha\sum_k A_k U\Psi_k + \beta\sum_k B_k U\Psi_k\ . \end{aligned}$$

Using Eq. (2.A.14) again, this gives

$$U(\alpha\Psi + \beta\Phi) = \alpha U\Psi + \beta U\Phi, \tag{2.A.19}$$

so U is *linear*. Also, using Eqs. (2.A.2) and (2.A.3), the scalar product of the transformed states is

$$(U\Psi, U\Phi) = \sum_{kl} A_k^* B_l (U\Psi_k, U\Psi_l) = \sum_k A_k^* B_k,$$

and hence

$$(U\Psi, U\Phi) = (\Psi, \Phi), \tag{2.A.20}$$

so U is *unitary*.

The case of a symmetry that satisfies Eq. (2.A.15) for all state-vectors may be dealt with in much the same way. The reader can probably supply the arguments without help, but since antilinear operators may be unfamiliar, we shall give the details here anyway. Suppose that Eq. (2.A.15) is satisfied for all state-vectors $\sum_k C_k\Psi_k$. Any two state-vectors Ψ and Φ may be expanded as before, and so:

$$\begin{aligned} U(\alpha\Psi + \beta\Phi) &= U\sum_k(\alpha A_k + \beta B_k)\Psi_k \\ &= \sum_k(\alpha^* A_k^* + \beta^* B_k^*)U\Psi_k = \alpha^*\sum_k A_k^* U\Psi_k + \beta^*\sum_k B_k^* U\Psi_k. \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

Using Eq. (2.A.15) again, this gives

$$U(\alpha\Psi + \beta\Phi) = \alpha^* U\Psi + \beta^* U\Phi, \tag{2.A.21}$$

so U is *antilinear*. Also, using Eqs. (2.A.2) and (2.A.3), the scalar product of the transformed states is

$$(U\Psi, U\Phi) = \sum_{kl} A_k B_l^* (U\Psi_k, U\Psi_l) = \sum_k A_k B_k^*$$

and hence

$$(U\Psi, U\Phi) = (\Psi, \Phi)^*, \tag{2.A.22}$$

so U is *antiunitary*.

Appendix B Group Operators and Homotopy Classes

In this appendix we shall prove the theorem stated in Section 2.7, that the phases of the operators $U(T)$ for finite symmetry transformations T may be chosen so that these operators form a representation of the symmetry group, rather than a projective representation, provided (a) the generators of the group can be defined so that there are no central charges in the Lie algebra, and (b) the group is simply connected. We shall also comment on the projective representations encountered for groups that are not simply connected, and their relation to the homotopy classes of the group.

To prove this theorem, let us recall the method by which we construct the operators corresponding to symmetry transformations. As described in Section 2.2, we introduce a set of real variables θ^a to parameterize these transformations, in such a way that the transformations satisfy the composition rule (2.2.15):

$$T(\bar{\theta})T(\theta) = T\Big(f(\bar{\theta}, \theta)\Big).$$

We want to construct operators $U(T(\theta)) \equiv U[\theta]$ that satisfy the corresponding condition*

$$U[\bar{\theta}]U[\theta] = U\Big[f(\bar{\theta}, \theta)\Big]. \tag{2.B.1}$$

To do this, we lay down arbitrary 'standard' paths $\Theta_\theta^a(s)$ in group parameter space, running from the origin to each point θ, with $\Theta_\theta^a(0) = 0$ and $\Theta_\theta^a(1) = \theta^a$, and define $U_\theta(s)$ along each such path by the differential

* Square brackets are used here to distinguish U operators constructed as functions of the group parameters from those expressed as functions of the group transformations themselves.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

equation

$$\frac{d}{ds}U_\theta(s) = it_a U_\theta(s) h^a_b(\Theta_\theta(s))\frac{d\Theta^b_\theta(s)}{ds} \tag{2.B.2}$$

with the initial condition

$$U_\theta(0) = 1, \tag{2.B.3}$$

where

$$[h^{-1}]^a_b(\theta) \equiv \left[\frac{\partial f^a(\bar\theta,\theta)}{\partial\bar\theta^b}\right]_{\bar\theta=0}. \tag{2.B.4}$$

We are eventually going to identify the operators $U[\theta]$ with $U_\theta(1)$, but first we must establish some of the properties of $U_\theta(s)$.

In order to check the composition rule, consider two points θ_1 and θ_2, and define a path $\mathscr{P}$ that runs from 0 to θ_1 and thence to $f(\theta_2,\theta_1)$:

$$\Theta^a_{\mathscr{P}}(s) \equiv \begin{cases} \Theta^a_{\theta_1}(2s) & 0 \le s \le \frac{1}{2}, \\ f^a(\Theta_{\theta_2}(2s-1),\theta_1) & \frac{1}{2} \le s \le 1. \end{cases} \tag{2.B.5}$$

At the end of the first segment, we are at $U_{\mathscr{P}}(\frac{1}{2}) = U_{\theta_1}(1)$. To evaluate $U_{\mathscr{P}}(s)$ along the second segment, we need the derivative of $f^a(\Theta_{\theta_2}(2s-1),\theta_1)$. For this purpose, we use the fundamental associativity condition:

$$f^a(f(\theta_3,\theta_2),\theta_1) = f^a(\theta_3, f(\theta_2,\theta_1)). \tag{2.B.6}$$

Matching the coefficients of $\theta^c{}_3$ in the limit $\theta_3 \to 0$ yields the result:

$$\frac{\partial f^a(\theta_2,\theta_1)}{\partial\theta^b_2} h^c{}_a(f(\theta_2,\theta_1)) = h^c{}_b(\theta_2). \tag{2.B.7}$$

Along the second segment the differential equation (2.B.2) for $U_{\mathscr{P}}(s)$ is thus the same as the differential equation for $U_{\theta_2}(2s-1)$. They satisfy different initial conditions, but $U_{\mathscr{P}}(s)U^{-1}_{\theta_1}(1)$ also satisfies the same differential equation as $U_{\theta_2}(2s-1)$, and in addition the same initial condition: at $s = \frac{1}{2}$, both are unity. We therefore conclude that for $\frac{1}{2} \le s \le 1$,

$$U_{\mathscr{P}}(s)\, U^{-1}_{\theta_1}(s) = U_{\theta_2}(2s-1)$$

and in particular

$$U_{\mathscr{P}}(1) = U_{\theta_2}(1)\, U_{\theta_1}(1). \tag{2.B.8}$$

However, this does *not* say that $U_\theta(1)$ satisfies the desired composition rule (2.B.1), because although the path $\Theta_{\mathscr{P}}(s)$ runs from $\theta^a = 0$ to $\theta^a = f^a(\theta_2,\theta_1)$, in general it will not be the same as whatever 'standard' path $\Theta_{f(\theta_2,\theta_1)}$ we have chosen to run directly from $\theta^a = 0$ to $\theta^a = f^a(\theta_2,\theta_1)$. We need to show that $U_\theta(1)$ is independent of the path from 0 to θ in order to be able to identify $U[\theta]$ as $U_\theta(1)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

For this purpose, consider the variation δU of $U_\theta(s)$ produced by a variation $\delta\Theta(s)$ in the path from 0 to θ. Taking the variation of Eq. (2.B.2) gives the differential equation

$$\frac{d}{ds}\delta U = it_a \delta U\, h^a_b(\Theta)\frac{d\Theta^b}{ds} + it_a U h^a_{b,c}(\Theta)\delta\Theta^c \frac{d\Theta^b}{ds} + it_a U h^a_b(\Theta)\frac{d\delta\Theta^b}{ds}$$

where $h^a_{b,c} \equiv \partial h^a_b/\partial\Theta^c$. Using the Lie commutation relations (2.2.22) (without central charges) and rearranging a bit, this gives

$$\frac{d}{ds}\left(U^{-1}\delta U\right) = \frac{d}{ds}\left(i\,U^{-1}t_a U h^a_b \delta\Theta^b\right) + i\,U^{-1}t_a U\delta\Theta^b \frac{d\Theta^c}{ds}\left(h^a_{c,b} - h^a_{b,c} + C^a{}_{ed}h^e_b h^d_c\right) . \quad (2.B.9)$$

However, by taking the limit $\theta_3, \theta_2 \to 0$ in the associativity condition (2.B.6), we find for all θ:

$$h(\theta)^a_{b,c} = -f^a{}_{de}h(\theta)^d_b h(\theta)^e_c\,, \quad (2.B.10)$$

where $f^a{}_{de}$ is the coefficient defined by (2.2.19). Antisymmetrizing in b and c shows that the last term in Eq. (2.B.9) vanishes

$$h^a_{c,b} - h^a_{b,c} + C^a{}_{ed}h^e_b h^d_c = 0\,. \quad (2.B.11)$$

Eq. (2.B.9) thus tells us that the quantity

$$U^{-1}\delta U - iU^{-1}t_a U h^a_b \delta\Theta^b$$

is constant along the path $\theta(s)$. It follows that $U_\theta(1)$ is stationary under any infinitesimal variation of the path that leaves the endpoints $\Theta(0) = 0$ and $\Theta(1) = \theta$ (and $U_\theta(0) = 1$) fixed. But assumption (b) tells us that any path from $\Theta(0) = 0$ to $\Theta(1) = \theta$ can be continuously deformed into any other, so we may now regard $U_\theta(1)$ as a path-independent function of θ alone:

$$U_\theta(1) \equiv U[\theta]. \quad (2.B.12)$$

In particular, since the path $\mathscr{P}$ leads from 0 to $f(\theta_2, \theta_1)$, we have

$$U_{\mathscr{P}}(1) = U[f(\theta_2, \theta_1)] \quad (2.B.13)$$

so that Eq. (2.B.8) shows that $U[\theta]$ satisfies the group multiplication law (2.B.1), as was to be proved.

Now that we have constructed a non-projective representation $U[\theta]$, it remains to prove that any projective representation $\tilde{U}[\theta]$ of the same group with the same representation generators t_a can only differ from $U[\theta]$ by a phase:

$$\tilde{U}[\theta] = e^{i\alpha(\theta)}U[\theta]$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

so that the phase ϕ in the multiplication law for $\tilde{U}[\theta]$:

$$\tilde{U}[\theta']\tilde{U}[\theta] = e^{i\phi(\theta',\theta)}\tilde{U}[f(\theta',\theta)]$$

can be removed by a simple change of phase of $\tilde{U}[\theta]$. To see this, consider the operator

$$U[\theta]^{-1}U[\theta']^{-1}\tilde{U}[\theta']\tilde{U}[\theta] = U[f(\theta',\theta)]^{-1}\tilde{U}[f(\theta',\theta)]e^{i\phi(\theta',\theta)}.$$

Because $U[\theta]$ and $\tilde{U}[\theta]$ have the same generators, the derivative of the left-hand side with respect to θ'^a vanishes at $\theta' = 0$, and so

$$0 = \frac{\partial}{\partial\theta^b}\left\{U[\theta]^{-1}\tilde{U}[\theta]\right\} + i\phi_b(\theta)U[\theta]^{-1}\tilde{U}[\theta],$$

where

$$\phi_b(\theta) \equiv h^a_b(\theta)\left[\frac{\partial}{\partial\theta'^b}\phi(\theta',\theta)\right]_{\theta'=0}.$$

Differentiating this result with respect to θ^c and antisymmetrizing in b and c gives immediately

$$0 = \frac{\partial\phi_b(\theta)}{\partial\theta^c} - \frac{\partial\phi_c(\theta)}{\partial\theta^b}.$$

A familiar theorem[13] tells us that in a simply connected space, this requires that ϕ_b is just a gradient of some function β:

$$\phi_b(\theta) = \frac{\partial\beta(\theta)}{\partial\theta^b}.$$

Thus the quantity $U[\theta]^{-1}\tilde{U}[\theta]e^{i\beta(\theta)}$ is actually constant in θ. Setting it equal to its value at $\theta = 0$, we see that $\tilde{U}$ is just proportional to U:

$$\tilde{U}[\theta] = U[\theta]\exp(-i\beta(\theta) + i\beta(0))$$

as claimed above.

* * *

The above analysis provides some information about the nature of the phase factors that can appear in the group multiplication law when the Lie algebra is free of central charges but the group is not simply connected. Suppose that the path $\mathscr{P}$ from zero to θ to $f(\bar{\theta},\theta)$ cannot be deformed into the standard path we have chosen to go from zero to $f(\bar{\theta},\theta)$, or in other words, that the loop from zero to θ to $f(\bar{\theta},\theta)$ and then back to zero is not continuously deformable to a point. Then $U^{-1}(f(\theta_2,\theta_1))U(\theta_2)U(\theta_1)$ can be a phase factor $\exp(i\phi(\theta_2,\theta_1)) \neq 1$, but ϕ will be the same for all other loops into which this can be continuously deformed. The set consisting of all loops that start and end at the origin and that can be continuously deformed into a given loop is known as the *homotopy class*[14] of that loop; we have thus seen that $\phi(\theta_2,\theta_1)$ depends only on the homotopy class of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

the loop from zero to θ to $f(\bar{\theta},\theta)$ and then back to zero. The set of homotopy classes forms a group; the 'product' of the homotopy class for loops $\mathscr{L}_1$ and $\mathscr{L}_2$ is the homotopy class of the loop formed by going around $\mathscr{L}_1$ and then $\mathscr{L}_2$; the 'inverse' of the homotopy class of the loop $\mathscr{L}$ is the homotopy class of the loop obtained by going around $\mathscr{L}$ in the opposite direction; and the 'identity' is the homotopy class of loops that can be deformed into a point at the origin. This group is known as the *first homotopy group* or *fundamental group* of the space in question. It is easy to see that the phase factors form a representation of this group: if going around loop $\mathscr{L}$ gives a phase factor $e^{i\phi}$, and going around loop $\bar{\mathscr{L}}$ gives a phase factor $e^{i\bar{\phi}}$, then going around both loops gives a phase factor $e^{i\phi}e^{i\bar{\phi}}$. Hence we can catalog all the possible types of projective representations of a given group $\mathscr{G}$ (with no central charges) if we know the one-dimensional representations of the first homotopy group of the parameter space of $\mathscr{G}$. Homotopy groups will be discussed in greater detail in Volume II.

Appendix C Inversions and Degenerate Multiplets

It is usually assumed that the inversions T and P take one-particle states into other one-particle states of the same species, perhaps with phase factors that depend on the particle species. In Section 2.6 we noted in passing that inversions might act in a more complicated way than this on degenerate multiplets of one-particle states, a possibility that seems to have been first suggested by Wigner[15] in 1964. This appendix will explore generalized versions of the inversion operators, in which finite matrices appear in place of the inversion phases, but without making some of Wigner's limiting assumptions.

Let us start with time-reversal. Wigner limited the possible action of the inversion operators by assuming that their squares are proportional to the unit operator. Because T is antiunitary, it is easy to see that the corresponding proportionality factor for T^2 can only be ± 1, perhaps with different signs for subspaces separated by superselection rules. When the sign for T^2 on the space of states with even or odd values of $2j$ is opposite to the sign $(-1)^{2j}$ found in Section 2.6, the physical states involved must furnish representations of the operator T that are more complicated than that assumed so far. But if we are willing to admit this possibility, there does not seem to be any good reason to impose Wigner's condition that T^2 is proportional to unity. It is not convincing to appeal to the structure of the extended Poincaré group; the only useful definition of any of the inversion operators is one that makes the operator exactly or

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

approximately conserved, and this may not be the definition that makes T^2 proportional to the unit operator.

To explore more general possibilities for time-reversal, let us assume that on a massive one-particle state it has the action

$$\mathsf{T}\Psi_{\mathbf{p},\sigma,n} = (-1)^{j-\sigma} \sum_m \mathscr{T}_{mn} \Psi_{-\mathbf{p},-\sigma,m} , \tag{2.C.1}$$

where $\mathbf{p}$, j, and σ are the particle's momentum, spin, and spin z-component, and n, m are indices labelling members of a degenerate multiplet of particle species. (The appearance of the factor $(-1)^{j-\sigma}$ and the reversal of $\mathbf{p}$ and σ are deduced in the same way as in Section 2.6.) The matrix $\mathscr{T}_{mn}$ is unknown, except that because T is antiunitary, $\mathscr{T}$ must be unitary.

Now let us see how we can simplify this transformation by an appropriate choice of basis for the one-particle states. Defining new states by the unitary transformation $\Psi'_{\mathbf{p},\sigma,n} = \sum_m \mathscr{U}_{mn} \Psi_{\mathbf{p},\sigma,m}$, we find the same transformation (2.C.1), with the matrix $\mathscr{T}_{mn}$ changed to

$$\mathscr{T}' = \mathscr{U}^{-1} \mathscr{T} \mathscr{U}^* . \tag{2.C.2}$$

We cannot in general make $\mathscr{T}'$ diagonal by such a choice of basis of the one-particle states, as we could if T were unitary. But we can instead make it block-diagonal, with the blocks either 1×1 phases, or 2×2 matrices of the form

$$\begin{pmatrix} 0 & e^{i\phi/2} \\ e^{-i\phi/2} & 0 \end{pmatrix} , \tag{2.C.3}$$

where the ϕ are various real phases.

(Here is the proof. First, note that Eq. (2.C.2) gives

$$\mathscr{T}' \mathscr{T}'^* = \mathscr{U}^{-1} \mathscr{T} \mathscr{T}^* \mathscr{U} .$$

This is a unitary transformation, so it can be chosen to diagonalize the unitary matrix $\mathscr{T}\mathscr{T}^*$. Assuming this to have been done, and dropping primes, we have

$$\mathscr{T} = D \mathscr{T}^T , \tag{2.C.4}$$

where D is a unitary diagonal matrix, say with phases $e^{i\phi_n}$ along the main diagonal. One immediate consequence is that the diagonal component $\mathscr{T}_{nn}$ vanishes unless $e^{i\phi_n} = 1$. Furthermore, if $e^{i\phi_n} = 1$ but $e^{i\phi_m} \neq 1$, then Eq. (2.C.4) tells us that $\mathscr{T}_{nm} = \mathscr{T}_{mn} = 0$. By listing first all rows and columns for which $e^{i\phi_n} = 1$, the matrix $\mathscr{T}$ is put in the form

$$\mathscr{T} = \begin{pmatrix} \mathscr{A} & 0 \\ 0 & \mathscr{B} \end{pmatrix} , \tag{2.C.5}$$

where $\mathscr{A}$ is symmetric as well as unitary, and the diagonal elements of $\mathscr{B}$ all vanish. Because $\mathscr{A}$ is symmetric, it can be expressed as the exponential

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

of a symmetric anti-Hermitian matrix, so it can be diagonalized by a transformation (2.C.2) acting only on $\mathscr{A}$, with the corresponding submatrix of $\mathscr{U}$ real and hence orthogonal. It is therefore only necessary to consider the submatrix $\mathscr{B}$ that connects the rows and columns for which $e^{i\phi_n} \neq 0$. For $n \neq m$, Eq. (2.C.4) gives $\mathscr{T}_{nm} = e^{i\phi_n}\mathscr{T}_{mn}$ and $\mathscr{T}_{mn} = e^{i\phi_m}\mathscr{T}_{nm}$, so $\mathscr{T}_{nm} = e^{i\phi_n}e^{i\phi_m}\mathscr{T}_{nm}$ and also $\mathscr{T}_{mn} = e^{i\phi_n}e^{i\phi_m}\mathscr{T}_{mn}$. Hence $\mathscr{T}_{nm} = \mathscr{T}_{mn} = 0$ unless $e^{i\phi_n}e^{i\phi_m} = 1$. If we list first all rows and columns of $\mathscr{B}$ with a given phase $e^{i\phi_1} \neq 1$, and then all rows and columns with the opposite phase, and then all rows and columns with some other phase $e^{i\phi_2} \neq 1$ not equal to $e^{\pm i\phi_1}$, and then all rows and columns with opposite phase, and so on, the matrix $\mathscr{B}$ becomes of block diagonal form

$$\mathscr{B} = \begin{pmatrix} \mathscr{B}_1 & 0 & \cdots \\ 0 & \mathscr{B}_2 & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}, \tag{2.C.6}$$

where

$$\mathscr{B}_i = \begin{pmatrix} 0 & e^{i\phi_i/2}\mathscr{C}_i \\ e^{-i\phi_i/2}\mathscr{C}_i^T & 0 \end{pmatrix}. \tag{2.C.7}$$

Furthermore, the unitarity of $\mathscr{T}$ and hence of $\mathscr{B}$ requires that $\mathscr{C}_i\mathscr{C}_i^\dagger = \mathscr{C}_i^\dagger\mathscr{C}_i = 1$, and hence $\mathscr{C}_i$ is square and unitary. By applying a transformation (2.C.2) with $\mathscr{U}$ block-diagonal in the same sense as $\mathscr{T}$, and with the matrix in the ith block of form

$$\begin{pmatrix} V_i & 0 \\ 0 & W_i \end{pmatrix}$$

with V_i and W_i unitary, the submatrices $\mathscr{C}_i$ are subjected to the transformations $\mathscr{C}_i \to V_i^{-1}\mathscr{C}_i W_i^*$, so we can clearly choose this transformation to make $\mathscr{C}_i = 1$. This establishes a correspondence between pairs of individual rows and columns within each block with phases $e^{i\phi_i}$ and $e^{-i\phi_i}$. To put the matrix $\mathscr{B}$ into block-diagonal form with 2×2 blocks of form (2.C.3), it is now only necessary to rearrange the rows and columns so that within the ith block we list rows and columns with phase $e^{i\phi_i}$ alternating with the corresponding rows and columns with phase $e^{-i\phi_i}$.)

It is important to note that where $e^{i\phi} \neq 1$, it is not possible to choose states to diagonalize the time-reversal transformation. If we have a pair of states $\Psi_{\mathbf{p},\sigma,\pm}$ on which T acts with a matrix (2.C.3), then

$$\mathsf{T}\Psi_{\mathbf{p},\sigma,\pm} = e^{\pm i\phi/2}(-1)^{j-\sigma}\Psi_{-\mathbf{p},-\sigma,\mp}. \tag{2.C.8}$$

Then on an arbitrary linear combination of these states, time-reversal gives

$$\mathsf{T}(c_+\Psi_{\mathbf{p},\sigma,+} + c_-\Psi_{\mathbf{p},\sigma,-}) = (-)^{j-\sigma}(e^{i\phi/2}c_+^*\Psi_{-\mathbf{p},-\sigma,-} + e^{-i\phi/2}c_-^*\Psi_{-\mathbf{p},-\sigma,+}).$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

For $c_+\Psi_{\mathbf{p},\sigma,+} + c_-\Psi_{\mathbf{p},\sigma,-}$ to be transformed under T by a phase λ, it is necessary that

$$e^{i\phi/2}c_+^* = \lambda c_- \qquad e^{-i\phi/2}c_-^* = \lambda c_+ .$$

But combining these equations gives $e^{\pm i\phi/2}c_\pm^* = |\lambda|^2 c_\pm^* e^{\mp i\phi/2}$, which is impossible unless either $c_+ = c_- = 0$ or $e^{i\phi}$ is unity. Thus for $e^{i\phi} \neq 1$, time-reversal invariance imposes a two-fold degeneracy on these states, beyond that associated with their spin.

Of course, if there is an additional 'internal' symmetry operator S which subjects these states to the transformation

$$\mathsf{S}\Psi_{\mathbf{p},\sigma,\pm} = e^{\pm i\phi/2}\Psi_{-\mathbf{p},\sigma,\mp} ,$$

then we can redefine the time-reversal operator as $\mathsf{T}' \equiv \mathsf{S}^{-1}\mathsf{T}$, and this operator would not mix the states $\Psi_{\mathbf{p},\sigma,\pm}$ with one another. It is only in the case where no such internal symmetry exists that we can attribute the doubling of particle states to time-reversal itself.

Let's come back now to the question of the square of T. Repeating the transformation (2.C.8) gives

$$\mathsf{T}^2\Psi_{\mathbf{p},\sigma,\pm} = (-1)^{2j}e^{\mp i\phi}\Psi_{\mathbf{p},\sigma,\pm} . \tag{2.C.9}$$

If we were to assume with Wigner that T^2 is proportional to the unit operator, then we would have to have $e^{i\phi} = e^{-i\phi}$, and since the phase is then real it would have to be $+1$ or -1. The choice $e^{i\phi} = -1$ would still require a two-fold degeneracy of one-particle states beyond that associated with their spin, and under Wigner's assumptions *all* particles would show this doubling. But there is no reason not to take a general phase ϕ in Eq. (2.C.8), one that may vanish for some particles and not for others. Thus the fact that observed particles do not show the extra two-fold degeneracy does not rule out the possibility that others might.

We may also consider the possibility of more complicated representations of the parity operator P, with

$$\mathsf{P}\Psi_{\mathbf{p},\sigma,n} = \sum_m \mathscr{P}_{nm}\Psi_{-\mathbf{p},\sigma,m} \tag{2.C.10}$$

with a unitary but otherwise unconstrained matrix $\mathscr{P}$. Unlike the case of time-reversal, here we may always diagonalize this matrix by a choice of basis for the states. But this choice of basis may not be the one in which time-reversal acts simply, so, in principle, P and T together can impose additional degeneracies that would not be required by P or T alone.

As discussed in Chapter 5, any quantum field theory is expected to respect a symmetry known as CPT, which acts on one-particle states as

$$\mathsf{CPT}\Psi_{\mathbf{p},\sigma,n} = (-1)^{j-\sigma}\Psi_{\mathbf{p},-\sigma,n^c} , \tag{2.C.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

where n^c denotes the antiparticle (or 'charge-conjugate') of particle n. No phases or matrices are allowed in this transformation (though of course we could always introduce such phases or matrices by combining CPT with good internal symmetries.) It follows that

$$(\mathsf{CPT})^2\Psi_{\mathbf{p},\sigma,n} = (-1)^{2j}\Psi_{\mathbf{p},-\sigma,n} \,, \tag{2.C.12}$$

so the possibility suggested by Wigner of a sign $-(-1)^{2j}$ in the action of $(\mathsf{CPT})^2$ does not arise in quantum field theory.

To the extent that T is a good symmetry of some class of phenomena, so is the inversion $\mathsf{CP} \equiv (\mathsf{CPT})\mathsf{T}^{-1}$. For the states that transform under T in the conventional way

$$\mathsf{T}\Psi_{\mathbf{p},\sigma,n} \propto \Psi_{-\mathbf{p},-\sigma,n} \,, \tag{2.C.13}$$

the CP operator also acts conventionally

$$\mathsf{CP}\Psi_{\mathbf{p},\sigma,n} \propto \Psi_{-\mathbf{p},\sigma,n^c} \,. \tag{2.C.14}$$

The operator $\mathsf{C} \equiv \mathsf{CPP}^{-1}$ then just interchanges particles and antiparticles

$$\mathsf{C}\Psi_{\mathbf{p},\sigma,n} \propto \Psi_{\mathbf{p},\sigma,n^c} \,. \tag{2.C.15}$$

On the other hand, where T has the unconventional representation (2.C.8), Eq. (2.C.11) gives

$$\mathsf{CP}\Psi_{\mathbf{p},\sigma,\pm} = e^{\mp i\phi/2}\Psi_{-\mathbf{p},\sigma,\mp^c} \tag{2.C.16}$$

In particular, it is possible that the degeneracy indicated by the label $\pm$ may be the same as the particle–antiparticle degeneracy, so that the antiparticle (as defined by CPT) of the state $\Psi_\pm$ is $\Psi_\mp$. In this case, CP would have the unconventional property of *not* interchanging particles and antiparticles. As far as these particles are concerned, CP and T would be what are usually called P and CT. But this is not merely a matter of definition; on other particles CP and T would still have their usual effect.

No examples are known of particles that furnish unconventional representations of inversions, so these possibilities will not be pursued further here. From now on, the inversions will be assumed to have the conventional action assumed in Section 2.6.

Problems

1. Suppose that observer $\mathcal{O}$ sees a W-boson (spin one and mass $m \neq 0$) with momentum $\mathbf{p}$ in the y-direction and spin z-component σ. A second observer $\mathcal{O}'$ moves relative to the first with velocity $\mathbf{v}$ in the z-direction. How does $\mathcal{O}'$ describe the W state?

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

2. Suppose that observer $\mathcal{O}$ sees a photon with momentum $\mathbf{p}$ in the y-direction and polarization vector in the z-direction. A second observer $\mathcal{O}'$ moves relative to the first with velocity $\mathbf{v}$ in the z-direction. How does $\mathcal{O}'$ describe the same photon?

3. Derive the commutation relations for the generators of the Galilean group directly from the group multiplication law (without using our results for the Lorentz group). Include the most general set of central charges that cannot be eliminated by redefinition of the group generators.

4. Show that the operators $P_\mu P^\mu$ and $W_\mu W^\mu$ commute with all Lorentz transformation operators $U(\Lambda, a)$, where $W_\mu \equiv \epsilon_{\mu\nu\rho\lambda} J^{\nu\rho} P^\lambda$.

5. Consider physics in two space and one time dimensions, assuming invariance under a 'Lorentz' group $SO(2,1)$. How would you describe the spin states of a single *massive* particle? How do they behave under Lorentz transformations? What about the inversions P and T?

6. As in Problem 5, consider physics in two space and one time dimensions, assuming invariance under a 'Lorentz' group $O(2,1)$. How would you describe the spin states of a single *massless* particle? How do they behave under Lorentz transformations? What about the inversions P and T?

References

1. P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th edn (Oxford University Press, Oxford, 1958).

2. E. P. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Braunschweig, 1931): pp. 251–3 (English translation, Academic Press, Inc, New York, 1959). For massless particles, see also E. P. Wigner, in *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963): p. 64.

3. G. C. Wick, A. S. Wightman, and E. P. Wigner, *Phys. Rev.* **88**, 101 (1952).

3*a*. See, e.g, S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Section 2.1.

4. E. Inönü and E. P. Wigner, *Nuovo Cimento* **IX**, 705 (1952).

5. E. P. Wigner, *Ann. Math.* **40**, 149 (1939).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

6. G. W. Mackey, *Ann. Math.* **55**, 101 (1952); **58**, 193 (1953); *Acta. Math.* **99**, 265 (1958); *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

7. See, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957): Chapter 4; M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957): Chapter IV; L. D. Landau and E. M. Lifshitz, *Quantum Mechanics – Non Relativistic Theory*, 3rd edn. (Pergamon Press, Oxford, 1977): Section 58; Wu-Ki Tung, *Group Theory in Physics* (World Scientific, Singapore, 1985): Sections 7.3 and 8.1.

8. T. D. Lee and C. N. Yang, *Phys. Rev.* **104**, 254 (1956); C. S. Wu *et al.*, *Phys. Rev.* **105**, 1413 (1957); R. Garwin, L. Lederman, and M. Weinrich, *Phys. Rev.* **105**, 1415 (1957); J. I. Friedman and V. L. Telegdi, *Phys. Rev.* **105**, 1681 (1957).

9. J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Phys. Rev. Letters* **13**, 138 (1964).

10. H. A. Kramers, *Proc. Acad. Sci. Amsterdam* **33**, 959 (1930); also see F. J. Dyson, *J. Math. Phys.* **3**, 140 (1962).

11. V. Bargmann, *Ann. Math.* **59**, 1 (1954): Theorem 7.1.

12. See, e.g., H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices* (Dover Publications, New York, 1961): p. 194.

13. See e. g. H. Flanders, *Differential Forms* (Academic Press, New York, 1963): Section 3.6.

14. For an introduction to homotopy classes and groups, see, e.g., J. G. Hocking and G. S. Young, *Topology* (Addison-Wesley, Reading, MA, 1961): Chapter 4; C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic Press, London, 1983): Chapters 3 and 5.

15. E. P. Wigner, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, ed. by F. Gürsey (Gordon and Breach, New York, 1964): p. 37.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.004

3

Scattering Theory

The general principles of relativistic quantum mechanics described in the previous chapter have so far been applied here only to states of a single stable particle. Such one-particle states by themselves are not very exciting — it is only when two or more particles interact with each other that anything interesting can happen. But experiments do not generally follow the detailed course of events in particle interactions. Rather, the paradigmatic experiment (at least in nuclear or elementary particle physics) is one in which several particles approach each other from a macroscopically large distance, and interact in a microscopically small region, after which the products of the interaction travel out again to a macroscopically large distance. The physical states before and after the collision consist of particles that are so far apart that they are effectively non-interacting, so they can be described as direct products of the one-particle states discussed in the previous chapter. In such an experiment, all that is measured is the probability distribution, or 'cross-sections', for transitions between the initial and final states of distant and effectively non-interacting particles. This chapter will outline the formalism[1] used for calculating these probabilities and cross-sections.

3.1 'In' and 'Out' States

A state consisting of several non-interacting particles may be regarded as one that transforms under the inhomogeneous Lorentz group as a direct product of one-particle states. To label the one-particle states we use their four-momenta p^μ, spin z-component (or, for massless particles, helicity) σ, and, since we now may be dealing with more than one species of particle, an additional discrete label n for the particle type, which includes a specification of its mass, spin, charge, etc. The general transformation rule is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

$$U(\Lambda,a)\Psi_{p_1,\sigma_1,n_1;p_2,\sigma_2,n_2;\cdots} = \exp\Big(-ia_\mu((\Lambda p_1)^\mu + (\Lambda p_2)^\mu + \cdots)\Big)$$
$$\times \sqrt{\frac{(\Lambda p_1)^0(\Lambda p_2)^0\cdots}{p_1^0 p_2^0\cdots}} \sum_{\sigma_1'\sigma_2'\cdots} D^{(j_1)}_{\sigma_1'\sigma_1}\Big(W(\Lambda,p_1)\Big) D^{(j_2)}_{\sigma_2'\sigma_2}\Big(W(\Lambda,p_2)\Big)\cdots$$
$$\times \Psi_{\Lambda p_1,\sigma_1',n_1;\Lambda p_2,\sigma_2',n_2;\cdots} \tag{3.1.1}$$

where $W(\Lambda,p)$ is the Wigner rotation (2.5.10), and $D^{(j)}_{\sigma'\sigma}(W)$ are the conventional $(2j+1)$-dimensional unitary matrices representing the three-dimensional rotation group. (This is for massive particles; for any massless particle, the matrix $D^{(j)}_{\sigma'\sigma}(W(\Lambda,p))$ is replaced with $\delta_{\sigma'\sigma}\exp(i\sigma\theta(\Lambda,p))$, where θ is the angle defined by Eq. (2.5.43).) The states are normalized as in Eq. (2.5.19)

$$\Big(\Psi_{p_1',\sigma_1',n_1';p_2',\sigma_2',n_2';\cdots}\ ,\ \Psi_{p_1,\sigma_1,n_1;p_2,\sigma_2,n_2;\cdots}\Big)$$
$$= \delta^3(\mathbf{p}_1' - \mathbf{p}_1)\delta_{\sigma_1'\sigma_1}\delta_{n_1'n_1}\delta^3(\mathbf{p}_2' - \mathbf{p}_2)\delta_{\sigma_2'\sigma_2}\delta_{n_2'n_2}\cdots$$
$$\pm \text{ permutations} \tag{3.1.2}$$

with the term '$\pm$ permutations' included to take account of the possibility that it is some permutation of the particle types $n_1', n_2', \cdots$ that are of the same species as the particle types $n_1, n_2, \cdots$. (As discussed more fully in Chapter 4, its sign is -1 if this permutation includes an odd permutation of half-integer spin particles, and otherwise $+1$. This will not be important in the work of the present chapter.)

We often use an abbreviated notation, letting one Greek letter, say α, stand for the whole collection $p_1,\sigma_1,n_1;p_2,\sigma_2,n_2;\cdots$. In this notation, Eq. (3.1.2) is written simply

$$(\Psi_{\alpha'},\Psi_\alpha) = \delta(\alpha'-\alpha) \tag{3.1.3}$$

with $\delta(\alpha'-\alpha)$ standing for the sum of products of delta functions and Kronecker deltas appearing on the right-hand side of Eq. (3.1.2). Also, in summing over states, we write

$$\int d\alpha\cdots \equiv \sum_{n_1\sigma_1 n_2\sigma_2\cdots} \int d^3p_1\, d^3p_2\cdots, \tag{3.1.4}$$

it being understood that in such sums and integrals, we include only configurations that do not differ merely by the exchange of identical particles. In particular, the completeness relation for states normalized as in Eq. (3.1.3) reads

$$\Psi = \int d\alpha\ \Psi_\alpha(\Psi_\alpha,\Psi)\,. \tag{3.1.5}$$

The transformation rule (3.1.1) is only possible for particles that for one reason or another are not interacting. Setting $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu$ and $a^\mu =$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

$(0,0,0,\tau)$, for which $U(\Lambda,a) = \exp(iH\tau)$, Eq. (3.1.1) requires among other things that Ψ_α be an energy eigenstate

$$H\Psi_\alpha = E_\alpha \Psi_\alpha \tag{3.1.6}$$

with an energy equal to the sum of the one-particle energies

$$E_\alpha = p_1^0 + p_2^0 + \cdots \tag{3.1.7}$$

and with no interaction terms, terms that would involve more than one particle at a time.

On the other hand, the transformation rule (3.1.1) does apply in scattering processes at times $t \to \pm\infty$. As explained at the beginning of this chapter, in the typical scattering experiment we start with particles at time $t \to -\infty$ so far apart that they are not yet interacting, and end with particles at $t \to +\infty$ so far apart that they have ceased interacting. We therefore have not one but two sets of states that transform as in Eq. (3.1.1): *the 'in' and 'out' states** $\Psi_\alpha{}^+$ *and* $\Psi_\alpha{}^-$ *will be found to contain the particles described by the label* α *if observations are made at* $t \to -\infty$ *or* $t \to +\infty$, *respectively.*

Note how this definition is framed. To maintain manifest Lorentz invariance, in the formalism we are using here, state-vectors do not change with time — a state-vector Ψ describes the whole spacetime history of a system of particles. (This is known as the *Heisenberg picture*, in distinction with the Schrödinger picture, where the operators are constant and the states change with time.) Thus we do *not* say that $\Psi_\alpha{}^\pm$ are the limits at $t \to \mp\infty$ of a time-dependent state-vector $\Psi(t)$.

However, implicit in the definition of the states is a choice of the inertial frame from which the observer views the system; different observers see *equivalent* state-vectors, but not the *same* state-vector. In particular, suppose that a standard observer $\mathcal{O}$ sets his or her clock so that $t = 0$ is at some time during the collision process, while some other observer $\mathcal{O}'$ at rest with respect to the first uses a clock set so that $t' = 0$ is at a time $t = \tau$; that is, the two observers' time coordinates are related by $t' = t - \tau$. Then if $\mathcal{O}$ sees the system to be in a state Ψ, $\mathcal{O}'$ will see the system in a state $U(1,-\tau)\Psi = \exp(-iH\tau)\Psi$. Thus the appearance of the state long before or long after the collision (in whatever basis is used by $\mathcal{O}$) is found by applying a time-translation operator $\exp(-iH\tau)$ with $\tau \to -\infty$ or $\tau \to +\infty$, respectively. Of course, if the state is really an energy eigenstate, then it cannot be localized in time — the operator $\exp(-iH\tau)$ yields an inconsequential phase factor $\exp(-iE_\alpha\tau)$. Therefore, we must consider wave-packets, superpositions $\int d\alpha\, g(\alpha)\Psi_\alpha$ of states, with an amplitude $g(\alpha)$

* The labels '+' and '−' for 'in' and 'out' states may seem backward, but they seem to have become traditional. They arise from the signs in Eq. (3.1.16).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

that is non-zero and smoothly varying over some finite range $\triangle E$ of energies. The 'in' and 'out' states are defined so that the superposition

$$\exp(-iH\tau)\int d\alpha\, g(\alpha)\Psi_\alpha{}^{\pm} = \int d\alpha\, e^{-iE_\alpha\tau}g(\alpha)\Psi_\alpha{}^{\pm}$$

has the appearance of a corresponding superposition of free-particle states for $\tau \ll -1/\triangle E$ or $\tau \gg +1/\triangle E$, respectively.

To make this concrete, suppose we can divide the time-translation generator H into two terms, a free-particle Hamiltonian H_0 and an interaction V,

$$H = H_0 + V \tag{3.1.8}$$

in such a way that H_0 has eigenstates Φ_α that have the same appearance as the eigenstates Ψ_α^+ and Ψ_α^- of the complete Hamiltonian

$$H_0\Phi_\alpha = E_\alpha\Phi_\alpha\,, \tag{3.1.9}$$

$$(\Phi_{\alpha'}, \Phi_\alpha) = \delta(\alpha' - \alpha)\,. \tag{3.1.10}$$

Note that H_0 is assumed here to have the same spectrum as the full Hamiltonian H. This requires that the masses appearing in H_0 be the physical masses that are actually measured, which are not necessarily the same as the 'bare' mass terms appearing in H; the difference if there is any must be included in the interaction V, not H_0. Also, any relevant bound states in the spectrum of H should be introduced into H_0 as if they were elementary particles.**

The 'in' and 'out' states can now be defined as eigenstates of H, not H_0,

$$H\Psi_\alpha{}^{\pm} = E_\alpha\Psi_\alpha{}^{\pm} \tag{3.1.11}$$

which satisfy the condition

$$\int d\alpha\, e^{-iE_\alpha\tau}g(\alpha)\Psi_\alpha{}^{\pm} \to \int d\alpha\, e^{-iE_\alpha\tau}g(\alpha)\Phi_\alpha \tag{3.1.12}$$

for $\tau \to -\infty$ or $\tau \to +\infty$, respectively.

Eq. (3.1.12) can be rewritten as the requirement that:

$$\exp(-iH\tau)\int d\alpha\, g(\alpha)\Psi_\alpha{}^{\pm} \to \exp(-iH_0\tau)\int d\alpha\, g(\alpha)\Phi_\alpha$$

for $\tau \to -\infty$ or $\tau \to +\infty$, respectively. This is sometimes rewritten as a formula for the 'in' and 'out' states:

$$\Psi_\alpha{}^{\pm} = \Omega(\mp\infty)\Phi_\alpha\,, \tag{3.1.13}$$

** Alternatively, in non-relativistic problems we can include the binding potential in H_0. In the application of this method to 'rearrangement collisions,' where some bound states appear in the initial state but not the final state, or vice-versa, one must use a different split of H into H_0 and V in the initial and final states.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

where

$$\Omega(\tau) \equiv \exp(+iH\tau)\exp(-iH_0\tau)\,. \tag{3.1.14}$$

However, it should be kept in mind that $\Omega(\mp\infty)$ in Eq. (3.1.13) gives meaningful results only when acting on a smooth superposition of energy eigenstates.

One immediate consequence of the definition (3.1.12) is that the 'in' and 'out' states are normalized just like the free-particle states. To see this, note that since the left-hand side of Eq. (3.1.12) is obtained by letting the unitary operator $\exp(-iH\tau)$ act on a time-independent state, its norm is independent of time, and therefore equals the norm of its limit for $\tau \to \infty$, i.e., the norm of the right-hand side of Eq. (3.1.12):

$$\int d\alpha\, d\beta\ \exp(-i(E_\alpha - E_\beta)\tau)g(\alpha)g^*(\beta)(\Psi_\beta{}^\pm, \Psi_\alpha{}^\pm)$$
$$= \int d\alpha\, d\beta\ \exp(-i(E_\alpha - E_\beta)\tau)g(\alpha)g^*(\beta)(\Phi_\beta, \Phi_\alpha)\,.$$

Since this is supposed to be true for all smooth functions $g(\alpha)$, the scalar products must be equal

$$(\Psi_\beta{}^\pm, \Psi_\alpha{}^\pm) = (\Phi_\beta, \Phi_\alpha) = \delta(\beta - \alpha)\,. \tag{3.1.15}$$

It is useful for some purposes to have an explicit though formal solution of the energy eigenvalue equation (3.1.11) satisfying the conditions (3.1.12). For this purpose, write Eq. (3.1.11) as

$$(E_\alpha - H_0)\Psi_\alpha{}^\pm = V\Psi_\alpha{}^\pm\,.$$

The operator $E_\alpha - H_0$ is not invertible; it annihilates not only the free-particle state Φ_α, but also the continuum of other free-particle states Φ_β of the same energy. Since the 'in' and 'out' states become just Φ_α for $V \to 0$, we tentatively write the formal solutions as Φ_α plus a term proportional to V:

$$\Psi_\alpha{}^\pm = \Phi_\alpha + (E_\alpha - H_0 \pm i\epsilon)^{-1}V\Psi_\alpha{}^\pm\,, \tag{3.1.16}$$

or, expanding in a complete set of free-particle states,

$$\Psi_\alpha{}^\pm = \Phi_\alpha + \int d\beta\ \frac{T_{\beta\alpha}{}^\pm \Phi_\beta}{E_\alpha - E_\beta \pm i\epsilon}\,, \tag{3.1.17}$$

$$T_{\beta\alpha}{}^\pm \equiv (\Phi_\beta, V\Psi_\alpha{}^\pm)\,, \tag{3.1.18}$$

with ϵ a positive infinitesimal quantity, inserted to give meaning to the reciprocal of $E_\alpha - H_0$. These are known as the *Lippmann-Schwinger equations.*[1a] We shall use Eq. (3.1.17) at the end of the next section to give a slightly less unrigorous proof of the orthonormality of the 'in' and 'out' states.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

It remains to be shown that Eq. (3.1.17), with a $+i\epsilon$ or $-i\epsilon$ in the denominator, satisfies the condition (3.1.12) for an 'in' or an 'out' state, respectively. For this purpose, consider the superpositions

$$\Psi_g{}^{\pm}(t) \equiv \int d\alpha\, e^{-iE_\alpha t} g(\alpha) \Psi_\alpha{}^{\pm}\,, \tag{3.1.19}$$

$$\Phi_g(t) \equiv \int d\alpha\, e^{-iE_\alpha t} g(\alpha) \Phi_\alpha\,. \tag{3.1.20}$$

We want to show that $\Psi_g{}^{+}(t)$ and $\Psi_g{}^{-}(t)$ approach $\Phi_g(t)$ for $t \to -\infty$ and $t \to +\infty$, respectively. Using Eq. (3.1.17) in Eq. (3.1.19) gives

$$\Psi_g{}^{\pm}(t) = \Phi_g(t) + \int d\alpha \int d\beta\, \frac{e^{-iE_\alpha t} g(\alpha) T_{\beta\alpha}{}^{\pm} \Phi_\beta}{E_\alpha - E_\beta \pm i\epsilon}\,. \tag{3.1.21}$$

Let us recklessly interchange the order of integration, and consider first the integrals

$$\mathcal{I}_\beta{}^{\pm} \equiv \int d\alpha\, \frac{e^{-iE_\alpha t} g(\alpha) T_{\beta\alpha}{}^{\pm}}{E_\alpha - E_\beta \pm i\epsilon}\,.$$

For $t \to -\infty$, we can close the contour of integration for the energy variable E_α in the upper half-plane with a large semi-circle, with the contribution from this semi-circle killed by the factor $\exp(-iE_\alpha t)$, which is exponentially small for $t \to -\infty$ and $\operatorname{Im} E_\alpha > 0$. The integral is then given by a sum over the singularities of the integral in the upper half-plane. The functions $g(\alpha)$ and $T_{\beta\alpha}{}^{\pm}$ may, in general, be expected to have some singularities at values of E_α with finite positive imaginary parts, but just as for the large semi-circle, their contribution is exponentially damped for $t \to -\infty$. (Specifically, $-t$ must be much greater than both the time-uncertainty in the wave-packet $g(\alpha)$ and the duration of the collision, which respectively govern the location of the singularities of $g(\alpha)$ and $T_{\beta\alpha}{}^{\pm}$ in the complex E_α plane.) This leaves the singularity in $(E_\alpha - E_\beta \pm i\epsilon)^{-1}$, which is in the upper half-plane for $\mathcal{I}_\beta{}^{-}$ but not $\mathcal{I}_\beta{}^{+}$. We conclude then that $\mathcal{I}_\beta{}^{+}$ vanishes for $t \to -\infty$. In the same way, for $t \to +\infty$ we must close the contour of integration in the lower half-plane, and so $\mathcal{I}_\beta{}^{-}$ vanishes in this limit. We conclude that $\Psi_g{}^{\pm}(t)$ approaches $\Phi_g(t)$ for $t \to \mp\infty$, in agreement with the defining condition (3.1.12).

* * *

For future use, we note a convenient representation of the factor $(E_\alpha - E_\beta \pm i\epsilon)^{-1}$ in Eq. (3.1.17). In general, we can write

$$(E \pm i\epsilon)^{-1} = \frac{\mathcal{P}_\epsilon}{E} \mp i\pi\delta_\epsilon(E)\,, \tag{3.1.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

where

$$\frac{\mathscr{P}_\epsilon}{E} \equiv \frac{E}{E^2+\epsilon^2}\,, \tag{3.1.23}$$

$$\delta_\epsilon(E) \equiv \frac{\epsilon}{\pi(E^2+\epsilon^2)}\,. \tag{3.1.24}$$

The function (3.1.23) is just $1/E$ for $|E| \gg \epsilon$, and vanishes for $E \to 0$, so for $\epsilon \to 0$ it behaves just like the 'principal value function' $\mathscr{P}/E$, which allows us to give meaning to integrals of $1/E$ times any smooth function of E, by excluding an infinitesimal interval around $E = 0$. The function (3.1.24) is of order ϵ for $|E| \gg \epsilon$, and gives unity when integrated over all E, so in the limit $\epsilon \to 0$ it behaves just like the familiar delta function $\delta(E)$. With this understanding, we can drop the label ϵ in Eq. (3.1.22), and write simply

$$(E \pm i\epsilon)^{-1} = \frac{\mathscr{P}}{E} \mp i\pi\delta(E)\,. \tag{3.1.25}$$

3.2 The S-matrix

An experimentalist generally prepares a state to have a definite particle content at $t \to -\infty$, and then measures what this state looks like at $t \to +\infty$. If the state is prepared to have a particle content α for $t \to -\infty$, then it is the 'in' state ${\Psi_\alpha}^+$, and if it is found to have the particle content β at $t \to +\infty$, then it is the 'out' state ${\Psi_\beta}^-$. The probability amplitude for the transition $\alpha \to \beta$ is thus the scalar product

$$S_{\beta\alpha} = ({\Psi_\beta}^-, {\Psi_\alpha}^+)\,. \tag{3.2.1}$$

This array of complex amplitudes is known as the *S-matrix*.[2] If there were no interactions then 'in' and 'out' states would be the same, and then $S_{\beta\alpha}$ would be just $\delta(\alpha-\beta)$. The rate for a reaction $\alpha \to \beta$ is thus proportional to $|S_{\beta\alpha} - \delta(\alpha-\beta)|^2$. We shall see in detail in Section 3.4 what $S_{\beta\alpha}$ has to do with measured rates and cross-sections.

Perhaps it should be stressed that 'in' and 'out' states do not inhabit two different Hilbert spaces. They differ only in how they are labelled: by their appearance either at $t \to -\infty$ or $t \to +\infty$. Any 'in' state can be expanded as a sum of 'out' states, with expansion coefficients given by the S-matrix (3.2.1).

Since $S_{\beta\alpha}$ is the matrix connecting two complete sets of orthonormal states, it must be unitary. To see this in greater detail, apply the completeness relation (3.1.5) to the 'out' states, and write

$$\int d\beta\; S^*_{\beta\gamma} S_{\beta\alpha} = \int d\beta\; ({\Psi_\gamma}^+, {\Psi_\beta}^-)({\Psi_\beta}^-, {\Psi_\alpha}^+) = ({\Psi_\gamma}^+, {\Psi_\alpha}^+)\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Using (3.1.15), this gives

$$\int d\beta \; S^*_{\beta\gamma} S_{\beta\alpha} = \delta(\gamma - \alpha) \tag{3.2.2}$$

or in brief, $S^\dagger S = 1$. In the same way, completeness for the 'in' states gives*

$$\int d\beta \; S_{\gamma\beta} S_{\alpha\beta}{}^* = \delta(\gamma - \alpha) \tag{3.2.3}$$

or, in other words, $SS^\dagger = 1$.

It is often convenient instead of dealing with the S-matrix to work with an operator S, defined to have matrix elements between *free-particle* states equal to the corresponding elements of the S-matrix:

$$(\Phi_\beta, S\Phi_\alpha) \equiv S_{\beta\alpha} \, . \tag{3.2.4}$$

The explicit though highly formal expression (3.1.13) for the 'in' and 'out' states yields a formula for the S-operator:

$$S = \Omega(\infty)^\dagger \Omega(-\infty) = U(+\infty, -\infty) \, , \tag{3.2.5}$$

where

$$U(\tau, \tau_0) \equiv \Omega(\tau)^\dagger \Omega(\tau_0) = \exp(iH_0\tau) \exp(-iH(\tau - \tau_0)) \exp(-iH_0\tau_0) \, . \tag{3.2.6}$$

This will be used in the next section to examine the Lorentz invariance of the S-matrix, and in Sec. 3.5 to derive a formula for the S-matrix in time-dependent perturbation theory.

The methods of the previous section can be used to derive a useful alternative formula for the S-matrix. Let's return to Eq. (3.1.21) for the 'in' state Ψ^+, but this time take $t \to +\infty$. We must now close the contour of integration for E_α in the *lower* half-E_α-plane, and although as before the singularities in ${T_{\beta\alpha}}^+$ and $g(\alpha)$ make no contribution for $t \to +\infty$, we now do pick up a contribution from the singular factor $(E_\alpha - E_\beta + i\epsilon)^{-1}$. The contour runs from $E_\alpha = -\infty$ to $E_\alpha = +\infty$, and then back to $E_\alpha = -\infty$ on a large semi-circle in the lower half-plane, so it circles the singularity in a *clockwise* direction. By the method of residues, this contribution to the integral over E_α is given by the value of the integrand at $E_\alpha = E_\beta - i\epsilon$, times a factor $-2i\pi$. That is, in the limit $\epsilon \to 0+$, for $t \to +\infty$ the integral over α in (3.1.21) has the asymptotic behavior

$${\mathscr{I}_\beta}^+ \to -2i\pi e^{-iE_\beta t} \int d\alpha \; \delta(E_\alpha - E_\beta) g(\alpha) {T_{\beta\alpha}}^+$$

* An alternative proof is given at the end of this section. Note that for infinite 'matrices,' the unitarity conditions $S^\dagger S = 1$ and $SS^\dagger = 1$ are not equivalent.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

and hence, for $t \to +\infty$,

$$\Psi_g{}^+(t) \to \int d\beta\, e^{-iE_\beta t}\Phi_\beta\left[g(\beta) - 2i\pi \int d\alpha\, \delta(E_\alpha - E_\beta)g(\alpha)T_{\beta\alpha}{}^+\right].$$

But expanding (3.1.19) for $\Psi_g{}^+$ in a complete set of 'out' states gives

$$\Psi_g{}^+(t) = \int d\alpha\, e^{-iE_\alpha t}g(\alpha) \int d\beta\, \Psi_\beta{}^-\, S_{\beta\alpha}\,.$$

Since $S_{\beta\alpha}$ contains a factor $\delta(E_\beta - E_\alpha)$, this may be rewritten

$$\Psi_g{}^+(t) = \int d\beta\, \Psi_\beta{}^-\, e^{-iE_\beta t} \int d\alpha\, g(\alpha)S_{\beta\alpha}$$

and, using the defining property (3.1.12) for 'out' states, this has the asymptotic behavior for $t \to +\infty$

$$\Psi_g{}^+(t) \to \int d\beta\, \Phi_\beta\, e^{-iE_\beta t} \int d\alpha\, g(\alpha)S_{\beta\alpha}\,.$$

Comparing this with our previous result, we find

$$\int d\alpha\, g(\alpha)S_{\beta\alpha} = g(\beta) - 2i\pi \int d\alpha\, \delta(E_\alpha - E_\beta)g(\alpha)T_{\beta\alpha}{}^+$$

or in other words

$$S_{\beta\alpha} = \delta(\beta - \alpha) - 2i\pi\delta(E_\alpha - E_\beta)T_{\beta\alpha}{}^+\,. \tag{3.2.7}$$

This suggests a simple approximation for the S-matrix: for a weak interaction V, we can neglect the difference between 'in' and free-particle states in (3.1.18), in which case Eq. (3.2.7) gives

$$S_{\beta\alpha} \simeq \delta(\beta - \alpha) - 2i\pi\delta(E_\alpha - E_\beta)(\Phi_\beta, V\Phi_\alpha)\,. \tag{3.2.8}$$

This is known as the *Born approximation*.[3] Higher-order terms are discussed in Section 3.5.

* * *

We can use the Lippmann–Schwinger equations (3.1.16) for the 'in' and 'out' states to give a proof[4] of the orthonormality of these states and the unitarity of the S-matrix, as well as Eq. (3.2.7), without having to deal with limits at $t \to \mp\infty$. First, by using (3.1.16) on either the left- or right-hand side of the matrix element $(\Psi_\beta^\pm, V\Psi_\alpha^\pm)$ and equating the results, we find that

$$\begin{aligned}(\Psi_\beta{}^\pm, V\Phi_\alpha) + (\Psi_\beta{}^\pm, V(E_\alpha - H_0 \pm i\epsilon)^{-1}V\Psi_\alpha{}^\pm) \\ = (\Phi_\beta, V\Psi_\alpha{}^\pm) + (\Psi_\beta{}^\pm, V(E_\beta - H_0 \mp i\epsilon)^{-1}V\Psi_\alpha{}^\pm).\end{aligned}$$

Summing over a complete set Φ_γ of intermediate states, this gives the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

equation:

$$T_{\alpha\beta}{}^{\pm\star} - T_{\beta\alpha}{}^{\pm} = -\int d\gamma\ T_{\gamma\beta}{}^{\pm\star} T_{\gamma\alpha}{}^{\pm} \times \left([E_\alpha - E_\gamma \pm i\epsilon]^{-1} - [E_\beta - E_\gamma \mp i\epsilon]^{-1} \right) . \quad (3.2.9)$$

To prove the orthonormality of the 'in' and 'out' states, divide Eq. (3.2.9) by $E_\alpha - E_\beta \pm 2i\epsilon$. This gives

$$\left(\frac{T_{\alpha\beta}{}^{\pm}}{E_\beta - E_\alpha \pm 2i\epsilon} \right)^\star + \frac{T_{\beta\alpha}{}^{\pm}}{E_\alpha - E_\beta \pm 2i\epsilon} = -\int d\gamma \left(\frac{T_{\gamma\beta}{}^{\pm}}{E_\beta - E_\gamma \pm i\epsilon} \right)^\star \frac{T_{\gamma\alpha}{}^{\pm}}{E_\alpha - E_\gamma \pm i\epsilon} .$$

The 2ϵs in the denominators on the left-hand side can be replaced with ϵs, since the only important thing is that these are positive infinitesimals. We see then that $\delta(\beta - \alpha) + T_{\beta\alpha}{}^{\pm}/(E_\alpha - E_\beta \pm i\epsilon)$ is unitary. With (3.1.17), this is just the statement that the $\Psi_\alpha{}^{\pm}$ form two orthonormal sets of state-vectors. The unitarity of the S-matrix can be proved in a similar fashion by multiplying (3.2.9) with $\delta(E_\beta - E_\alpha)$ instead of $(E_\alpha - E_\beta \pm 2i\epsilon)^{-1}$.

3.3 Symmetries of the S-Matrix

In this section we will consider both what is meant by the invariance of the S-matrix under various symmetries, and what are the conditions on the Hamiltonian that will ensure such invariance properties.

Lorentz Invariance

For any proper orthochronous Lorentz transformation $x \to \Lambda x + a$, we may define a unitary operator $U(\Lambda, a)$ by specifying that it acts as in Eq. (3.1.1) on either the 'in' *or* the 'out' states. When we say that a theory is Lorentz-invariant, we mean that the same operator $U(\Lambda, a)$ acts as in (3.1.1) on *both* 'in' *and* 'out' states. Since the operator $U(\Lambda, a)$ is unitary, we may write

$$S_{\beta\alpha} = \left(\Psi_\beta{}^{-}, \Psi_\alpha{}^{+} \right) = \left(U(\Lambda, a)\Psi_\beta{}^{-}, U(\Lambda, a)\Psi_\alpha{}^{+} \right)$$

so using (3.1.1), we obtain the Lorentz invariance (actually, covariance) property of the S-matrix: for arbitrary Lorentz transformations $\Lambda^\mu{}_\nu$ and translations a^μ,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

$$
\begin{aligned}
&S_{p_1',\sigma_1',n_1';p_2',\sigma_2',n_2';\cdots,\ p_1,\sigma_1,n_1;\ p_2,\sigma_2,n_2;\cdots} \\
&= \exp\Big(ia_\mu \Lambda^\mu{}_\nu (p'^\nu_1 + p'^\nu_2 + \ldots - p_1^\nu - p_2^\nu - \ldots)\Big) \\
&\quad \times \sqrt{\frac{(\Lambda p_1)^0 (\Lambda p_2)^0 \cdots (\Lambda p_1')^0 (\Lambda p_2')^0 \cdots}{p_1^0 p_2^0 \cdots \quad p_1'^0 p_2'^0 \cdots}} \\
&\quad \times \sum_{\bar\sigma_1 \bar\sigma_2 \cdots} D^{(j_1)}_{\bar\sigma_1 \sigma_1}\Big(W(\Lambda, p_1)\Big) D^{(j_2)}_{\bar\sigma_2 \sigma_2}\Big(W(\Lambda, p_2)\Big) \cdots \\
&\quad \times \sum_{\bar\sigma_1' \bar\sigma_2' \cdots} D^{(j_1')*}_{\bar\sigma_1' \sigma_1'}\Big(W(\Lambda, p_1')\Big) D^{(j_2')*}_{\bar\sigma_2' \sigma_2'}\Big(W(\Lambda, p_2')\Big) \cdots \\
&\quad \times S_{\Lambda p_1',\bar\sigma_1',n_1';\Lambda p_2',\bar\sigma_2',n_2';\cdots,\ \Lambda p_1,\bar\sigma_1,n_1;\Lambda p_2,\bar\sigma_2,n_2;\cdots}\,. \qquad (3.3.1)
\end{aligned}
$$

(Primes are used to distinguish final from initial particles; bars are used to distinguish summation variables.) In particular, since the left-hand side is independent of a^μ, so must be the right-hand side, and so the S-matrix vanishes unless the four-momentum is conserved. We can therefore write the part of the S-matrix that represents actual interactions among the particles in the form:

$$S_{\beta\alpha} - \delta(\beta - \alpha) = -2\pi i\, M_{\beta\alpha} \delta^4(p_\beta - p_\alpha). \qquad (3.3.2)$$

(However, as we will see in the next chapter, the amplitude $M_{\beta\alpha}$ itself contains terms that involve further delta function factors.)

Eq. (3.3.1) should be regarded as a definition of what we mean by the Lorentz invariance of the S-matrix, rather than a theorem, because it is only for certain special choices of Hamiltonian that there exists a unitary operator that acts as in (3.1.1) on both 'in' and 'out' states. We need to formulate conditions on the Hamiltonian that would ensure the Lorentz invariance of the S-matrix. For this purpose, it will be convenient to work with the operator S defined by Eq. (3.2.4):

$$S_{\beta\alpha} = (\Phi_\beta, S\Phi_\alpha).$$

As we have defined the free-particle states Φ_α in Chapter 2, they furnish a representation of the inhomogeneous Lorentz group, so we can always define a unitary operator $U_0(\Lambda, a)$ that induces the transformation (3.1.1) on these states:

$$
\begin{aligned}
&U_0(\Lambda, a)\Phi_{p_1,\sigma_1,n_1;p_2,\sigma_2,n_2;\cdots} = \exp\Big(-ia_\mu \Lambda^\mu{}_\nu (p_1^\nu + p_2^\nu + \ldots)\Big) \\
&\times \sqrt{\frac{(\Lambda p_1)^0 (\Lambda p_2)^0 \cdots}{p_1^0 p_2^0 \cdots}} \sum_{\sigma_1' \sigma_2' \cdots} D^{(j_1)}_{\sigma_1' \sigma_1}\Big(W(\Lambda, p_1)\Big) D^{(j_2)}_{\sigma_2' \sigma_2}\Big(W(\Lambda, p_2)\Big) \cdots \\
&\times \Phi_{\Lambda p_1,\sigma_1',n_1;\Lambda p_2,\sigma_2',n_2;\cdots}\,.
\end{aligned}
$$

Eq. (3.3.1) will thus hold if this unitary operator commutes with the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

S-operator:

$$U_0(\Lambda, a)^{-1} S\, U_0(\Lambda, a) = S\,.$$

This condition can also be expressed in terms of infinitesimal Lorentz transformations. Just as in Section 2.4, there will exist a set of Hermitian operators, a momentum $\mathbf{P}_0$, an angular momentum $\mathbf{J}_0$, and a boost generator $\mathbf{K}_0$, that together with H_0 generate the infinitesimal version of inhomogeneous Lorentz transformations when acting on free-particle states. Eq. (3.3.1) is equivalent to the statement that the S-matrix is unaffected by such transformations, or in other words, that the S-operator commutes with these generators:

$$[H_0, S] = [\mathbf{P}_0, S] = [\mathbf{J}_0, S] = [\mathbf{K}_0, S] = 0. \tag{3.3.3}$$

Because the operators H_0, $\mathbf{P}_0$, $\mathbf{J}_0$, and $\mathbf{K}_0$ generate infinitesimal inhomogeneous Lorentz transformations on the Φ_α, they automatically satisfy the commutation relations (2.4.18)–(2.4.24):

$$[J_0^i, J_0^j] = i\,\epsilon_{ijk} J_0^k\,, \tag{3.3.4}$$

$$[J_0^i, K_0^j] = i\,\epsilon_{ijk} K_0^k\,, \tag{3.3.5}$$

$$[K_0^i, K_0^j] = -i\,\epsilon_{ijk} J_0^k\,, \tag{3.3.6}$$

$$[J_0^i, P_0^j] = i\,\epsilon_{ijk} P_0^k\,, \tag{3.3.7}$$

$$[K_0^i, P_0^j] = -iH_0\delta_{ij}\,, \tag{3.3.8}$$

$$[J_0^i, H_0] = [P_0^i, H_0] = [P_0^i, P_0^j] = 0\,, \tag{3.3.9}$$

$$[K_0^i, H_0] = -i\,P_0^i\,, \tag{3.3.10}$$

where i, j, k, etc. run over the values 1, 2, and 3, and ϵ_{ijk} is the totally antisymmetric quantity with $\epsilon_{123} = +1$.

In the same way, we may define a set of 'exact generators,' operators $\mathbf{P}$, $\mathbf{J}$, $\mathbf{K}$ that together with the full Hamiltonian H generate the transformations (3.1.1) on, say, the 'in' states. (As already mentioned, what is not obvious is that the same operators generate the same transformations on the 'out' states.) The group structure tells us that these exact generators satisfy the same commutation relations:

$$[J^i, J^j] = i\,\epsilon_{ijk} J^k\,, \tag{3.3.11}$$

$$[J^i, K^j] = i\,\epsilon_{ijk} K^k\,, \tag{3.3.12}$$

$$[K^i, K^j] = -i\,\epsilon_{ijk} J^k\,, \tag{3.3.13}$$

$$[J^i, P^j] = i\,\epsilon_{ijk} P^k\,, \tag{3.3.14}$$

$$[K^i, P^j] = -iH\delta_{ij}\,, \tag{3.3.15}$$

$$[J^i, H] = [P^i, H] = [P^i, P^j] = 0\,, \tag{3.3.16}$$

$$[K^i, H] = -i\,P^i. \tag{3.3.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

In virtually all known field theories, the effect of interactions is to add an interaction term V to the Hamiltonian, while leaving the momentum and angular momentum unchanged:

$$H = H_0 + V \ , \qquad \mathbf{P} = \mathbf{P}_0 \ , \qquad \mathbf{J} = \mathbf{J}_0 \ . \tag{3.3.18}$$

(The only known exceptions are theories with topologically twisted fields, such as those with magnetic monopoles, where the angular momentum of states depends on the interactions.) Eq. (3.3.18) implies that the commutation relations (3.3.11), (3.3.14), and (3.3.16) are satisfied provided that the interaction commutes with the free-particle momentum and angular-momentum operators

$$[V, \mathbf{P}_0] = [V, \mathbf{J}_0] = 0. \tag{3.3.19}$$

It is easy to see from the Lippmann–Schwinger equation (3.1.16) or equivalently from (3.1.13) that the operators that generate translations and rotations when acting on the 'in' (and 'out') states are indeed simply $\mathbf{P}_0$ and $\mathbf{J}_0$. Also we easily see that $\mathbf{P}_0$ and $\mathbf{J}_0$ commute with the operator $U(t, t_0)$ defined by Eq.(3.2.6), and hence with the S-operator $U(\infty, -\infty)$. Further, we already know that the S-operator commutes with H_0, because there are energy-conservation delta functions in both terms in (3.2.7). This leaves just the boost generator $\mathbf{K}_0$ which we need to show commutes with the S-operator.

On the other hand, it is *not* possible to set the boost generator $\mathbf{K}$ equal to its free-particle counterpart $\mathbf{K}_0$, because then Eqs. (3.3.15) and (3.3.8) would give $H = H_0$, which is certainly not true in the presence of interactions. Thus when we add an interaction V to H_0, we must also add a correction $\mathbf{W}$ to the boost generator:

$$\mathbf{K} = \mathbf{K}_0 + \mathbf{W}. \tag{3.3.20}$$

Of the remaining commutation relations, let us concentrate on Eq. (3.3.17), which may now be put in the form

$$[\mathbf{K}_0, V] = -[\mathbf{W}, H]. \tag{3.3.21}$$

By itself, the condition (3.3.21) is empty, because for any V we could always define $\mathbf{W}$ by giving its matrix elements between H-eigenstates Ψ_α and Ψ_β as $-(\Psi_\beta, [\mathbf{K}_0, V]\Psi_\alpha)/(E_\beta - E_\alpha)$. Recall that the crucial point in the Lorentz invariance of a theory is not that there should exist a set of exact generators satisfying Eqs. (3.3.11)–(3.3.17), but rather that these operators should act the same way on 'in' and 'out' states; merely finding an operator $\mathbf{K}$ that satisfies Eq. (3.3.21) is not enough. Eq. (3.3.21) does become significant if we add the requirement that matrix elements of $\mathbf{W}$ should be smooth functions of the energies, and in particular should not have singularities of the form $(E_\beta - E_\alpha)^{-1}$. We shall now show that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Eq. (3.3.21), together with an appropriate smoothness condition on $\mathbf{W}$, does imply the remaining Lorentz invariance condition $[\mathbf{K}_0, S] = 0$.

To prove this, let us consider the commutator of $\mathbf{K}_0$ with the operator $U(t, t_0)$ defined by Eq. (3.2.6) for finite t and t_0. Using Eq. (3.3.10) and the fact that $\mathbf{P}_0$ commutes with H_0 yields:

$$[\mathbf{K}_0, \exp(iH_0 t)] = t\mathbf{P}_0 \exp(iH_0 t)$$

while Eq. (3.3.21) (which is equivalent to Eq. (3.3.17)) yields

$$[\mathbf{K}, \exp(iHt)] = t\mathbf{P} \exp(iHt) = t\mathbf{P}_0 \exp(iHt) .$$

The momentum operators then cancel in the commutator of $\mathbf{K}_0$ with U, and we find:

$$[\mathbf{K}_0, U(\tau, \tau_0)] = -\mathbf{W}(\tau)U(\tau, \tau_0) + U(\tau, \tau_0)\mathbf{W}(\tau_0) , \tag{3.3.22}$$

where

$$\mathbf{W}(t) \equiv \exp(iH_0 t)\mathbf{W} \exp(-iH_0 t). \tag{3.3.23}$$

If the matrix elements of $\mathbf{W}$ between H_0-eigenstates are sufficiently smooth functions of energy, then matrix elements of $\mathbf{W}(t)$ between smooth superpositions of energy eigenstates vanish for $t \to \pm\infty$, so Eq. (3.3.22) gives in effect:

$$0 = [\mathbf{K}_0, U(\infty, -\infty)] = [\mathbf{K}_0, S], \tag{3.3.24}$$

as was to be shown. This is the essential result: Eq. (3.3.21), together with the smoothness condition on matrix elements of $\mathbf{W}$ that ensures that $\mathbf{W}(t)$ vanishes for $t \to \pm\infty$, provides a sufficient condition for the Lorentz invariance of the S-matrix. This smoothness condition is a natural one, because it is much like the condition on matrix elements of V that is needed to make $V(t)$ vanish for $t \to \pm\infty$, as required in order to justify the very idea of an S-matrix.

We can also use Eq.(3.3.22) with $\tau = 0$ and $\tau_0 = \mp\infty$ to show that

$$\mathbf{K}\Omega(\mp\infty) = \Omega(\mp\infty)\mathbf{K}_0 , \tag{3.3.25}$$

where $\Omega(\mp\infty)$ is according to (3.1.13) the operator that converts a free-particle state Φ_α into the corresponding 'in' or 'out' state $\Psi_\alpha{}^{\pm}$. Also, it follows trivially from Eqs. (3.3.18) and (3.3.19) that the same is true for the momentum and angular momentum:

$$\mathbf{P}\Omega(\mp\infty) = \Omega(\mp\infty)\mathbf{P}_0 , \tag{3.3.26}$$

$$\mathbf{J}\Omega(\mp\infty) = \Omega(\mp\infty)\mathbf{J}_0 . \tag{3.3.27}$$

Finally, since all Φ_α and $\Psi_\alpha{}^{\pm}$ are eigenstates of H_0 and H respectively with the same eigenvalue E_α, we have

$$H\Omega(\mp\infty) = \Omega(\mp\infty)H_0 . \tag{3.3.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Eqs. (3.3.25)–(3.3.28) show that with our assumptions, 'in' and 'out' states do transform under inhomogeneous Lorentz transformations just like the free-particle states. Also, since these are similarity transformations, we now see that the exact generators **K**, **P**, **J**, and H satisfy the same commutation relations as $\mathbf{K}_0$, $\mathbf{P}_0$, $\mathbf{J}_0$, and H_0. This is why it turned out to be unnecessary in proving the Lorentz invariance of the S-matrix to use the other commutation relations (3.3.12), (3.3.13), and (3.3.15) that involve **K**.

Internal Symmetries

There are various symmetries, like the symmetry in nuclear physics under interchange of neutrons and protons, or the 'charge-conjugation' symmetry between particles and antiparticles, that have nothing directly to do with Lorentz invariance, and further appear the same in all inertial frames. Such a symmetry transformation T acts on the Hilbert space of physical states as a unitary operator $U(T)$ that induces linear transformations on the indices labelling particle species

$$U(T)\Psi_{p_1\sigma_1 n_1;p_2\sigma_2 n_2;\cdots} = \sum_{\bar{n}_1\bar{n}_2\cdots} \mathscr{D}_{\bar{n}_1 n_1}(T)\,\mathscr{D}_{\bar{n}_2 n_2}(T)\cdots \times \Psi_{p_1\sigma_1\bar{n}_1;p_2\sigma_2\bar{n}_2;\cdots}\,. \tag{3.3.29}$$

In accordance with the general discussion in Chapter 2, the $U(T)$ must satisfy the group multiplication rule

$$U(\bar{T})\,U(T) = U(\bar{T}T)\,, \tag{3.3.30}$$

where $\bar{T}T$ is the transformation obtained by first performing the transformation T, then some other transformation $\bar{T}$. Acting on Eq. (3.3.29) with $U(\bar{T})$, we see that the matrices $\mathscr{D}$ satisfy the same rule

$$\mathscr{D}(\bar{T})\mathscr{D}(T) = \mathscr{D}(\bar{T}T). \tag{3.3.31}$$

Also, taking the scalar product of the states obtained by acting with $U(T)$ on two different 'in' states or two different 'out' states, and using the normalization condition (3.1.2), we see that $\mathscr{D}(T)$ must be unitary

$$\mathscr{D}^\dagger(T) = \mathscr{D}^{-1}(T). \tag{3.3.32}$$

Finally, taking the scalar product of the states obtained by acting with $U(T)$ on one 'out' state and one 'in' state shows that $\mathscr{D}$ commutes with the S-matrix, in the sense that

$$\sum_{\bar{N}_1\bar{N}_2\cdots}\sum_{\bar{N}'_1\bar{N}'_2\cdots} \mathscr{D}^*_{\bar{N}'_1 n'_1}(T)\mathscr{D}^*_{\bar{N}'_2 n'_2}(T)\cdots\ \mathscr{D}_{\bar{N}_1 n_1}(T)\,\mathscr{D}_{\bar{N}_2 n_2}(T)\cdots \times S_{p'_1\sigma'_1\bar{N}'_1;p'_2\sigma'_2\bar{N}'_2;\cdots\,,\,p_1\sigma_1\bar{N}_1;p_2\sigma_2\bar{N}_2;\cdots} = S_{p'_1\sigma'_1 n'_1;p'_2\sigma'_2 n'_2;\cdots\,,\,p_1\sigma_1 n_1;p_2\sigma_2 n_2;\cdots}\,. \tag{3.3.33}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Again, this is a definition of what we mean by a theory being invariant under the internal symmetry T, because to derive Eq. (3.3.33) we still need to show that the *same* unitary operator $U(T)$ will induce the transformation (3.3.29) on both 'in' and 'out' states. This will be the case if there is an 'unperturbed' transformation operator $U_0(T)$ that induces these transformations on free-particle states,

$$U_0(T)\Phi_{p_1\sigma_1 n_1;p_2\sigma_2 n_2;\cdots} = \sum_{\bar{N}_1\bar{N}_2\cdots} \mathscr{D}_{\bar{N}_1 n_1}(T)\,\mathscr{D}_{\bar{N}_2 n_2}(T)\cdots\Phi_{p_1\sigma_1\bar{N}_1;p_2\sigma_2\bar{N}_2;\cdots} \tag{3.3.34}$$

and that commutes with both the free-particle and interaction parts of the Hamiltonian

$$U_0^{-1}(T)H_0U_0(T) = H_0\,, \tag{3.3.35}$$

$$U_0^{-1}(T)VU_0(T) = V. \tag{3.3.36}$$

From either the Lippmann–Schwinger equation (3.1.17) or from (3.1.13), we see that the operator $U_0(T)$ will induce the transformations (3.3.29) on 'in' and 'out' states as well as free-particle states, so that we can derive Eq. (3.3.29) taking $U(T)$ as $U_0(T)$.

A special case of great physical importance is that of a one-parameter Lie group, where T is a function of a single parameter θ, with

$$T(\bar{\theta})T(\theta) = T(\bar{\theta}+\theta). \tag{3.3.37}$$

As shown in Section 2.2, in this case the corresponding Hilbert-space operators must take the form

$$U(T(\theta)) = \exp(iQ\theta) \tag{3.3.38}$$

with Q a Hermitian operator. Likewise the matrices $\mathscr{D}(T)$ take the form

$$\mathscr{D}_{n'n}(T(\theta)) = \delta_{n'n}\exp(iq_n\theta)\,, \tag{3.3.39}$$

where q_n are a set of real species-dependent numbers. Here Eq. (3.3.33) simply tells us that the qs are conserved: $S_{\beta\alpha}$ vanishes unless

$$q_{n_1'} + q_{n_2'} + \cdots = q_{n_1} + q_{n_2} + \cdots\ . \tag{3.3.40}$$

The classic example of such a conservation law is that of conservation of electric charge. Also, all known processes conserve baryon number (the number of baryons, such as protons, neutrons, and hyperons, minus the number of their antiparticles) and lepton number (the number of leptons, such as electrons, muons, τ particles, and neutrinos, minus the number of their antiparticles) but as we shall see in Volume II, these conservation laws are believed to be only very good approximations. There are other conservation laws of this type that are definitely only approximate, such

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

as the conservation of the quantity known as strangeness, which was introduced to explain the relatively long life of a class of particles discovered by Rochester and Butler[5] in cosmic rays in 1947. For instance, the mesons now called* K^+ and K^0 are assigned strangeness $+1$ and the hyperons Λ^0, Σ^+, Σ^0, Σ^- are assigned strangeness -1, while the more familiar protons, neutrons, and π mesons (or pions) are taken to have strangeness zero. The conservation of strangeness in strong interactions explains why strange particles are always produced in association with one another, as in reactions like $\pi^+ + n \to K^+ + \Lambda^0$, while the relatively slow decays of strange particles into non-strange ones such as $\Lambda^0 \to p + \pi^-$ and $K^+ \to \pi^+ + \pi^0$ show that the interactions that do not conserve strangeness are very weak.

The classic example of a 'non-Abelian' symmetry whose generators do not commute with one another is isotopic spin symmetry, which was suggested[6] in 1937 on the basis of an experiment[7] that showed the existence of a strong proton–proton force similar to that between protons and neutrons. Mathematically, the group is $SU(2)$, like the covering group of the group of three-dimensional rotations; its generators are denoted t_i, with $i = 1, 2, 3$, and satisfy a commutation relation like (2.4.18):

$$[t_i, t_j] = i\epsilon_{ijk}t_k.$$

To the extent that isotopic spin symmetry is respected, it requires particles to form degenerate multiplets labelled with an integer or half-integer T and with $2T + 1$ components distinguished by their t_3 values, just like the degenerate spin multiplets required by rotational invariance. These include the nucleons p and n with $T = \frac{1}{2}$ and $t_3 = \frac{1}{2}, -\frac{1}{2}$; the pions π^+, π^0, and π^- with $T = 1$ and $t_3 = +1, 0, -1$; and the Λ^0 hyperon with $T = 0$ and $t_3 = 0$. These examples illustrate the relation between electric charge Q, the third component of isotopic spin t_3, the baryon number B, and the strangeness S:

$$Q = t_3 + (B + S)/2.$$

This relation was originally inferred from observed selection rules, but it was interpreted[8] by Gell-Mann and Ne'eman in 1960 to be a consequence of the embedding of both the isospin $\vec{T}$ and the 'hypercharge' $Y \equiv B + S$ in the Lie algebra of a larger but more badly broken non-Abelian internal symmetry, based on the non-Abelian group $SU(3)$. As we will see in Volume II, today both isospin and $SU(3)$ symmetry are understood as incidental consequences of the small masses of the two or three lightest quarks in the modern theory of strong interactions, quantum chromodynamics.

* Superscripts denote electric charges in units of the absolute value of the electronic charge. A 'hyperon' is any particle carrying non-zero strangeness and unit baryon number.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

The implications of isotopic spin symmetry for reactions among strongly interacting particles can be worked out by the same familiar methods that were invented for deriving the implications of rotational invariance. In particular, for a two-body reaction $A + B \to C + D$, Eq. (3.3.33) requires that the S-matrix may be put in the form (suppressing all but isospin labels):

$$S_{t_{C3}t_{D3},t_{A3}t_{B3}} = \sum_{T,t_3} C_{T_C T_D}(T t_3; t_{C3} t_{D3}) C_{T_A T_B}(T t_3; t_{A3} t_{B3}) S_T \ ,$$

where $C_{j_1 j_2}(j\sigma; \sigma_1 \sigma_2)$ is the usual Clebsch–Gordan coefficient[9] for forming a spin j with three-component σ from spins j_1 and j_2 with three-components σ_1 and σ_2, respectively; and S_T is a 'reduced' S-matrix depending on T and on all the suppressed momentum and spin variables, but not on the isospin three-components $t_{A3}, t_{B3}, t_{C3}, t_{D3}$. Of course, this, like all of the consequences of isotopic spin invariance, is only approximate, because this symmetry is not respected by electromagnetic (and other) interactions, as shown for instance by the fact that different members of the same isospin multiplet like p and n have different electric charges and slightly different masses.

Parity

To the extent that the symmetry under the transformation $\mathbf{x} \to -\mathbf{x}$ is really valid, there must exist a unitary operator P under which both 'in' and 'out' states transform as a direct product of single-particle states:

$$\mathsf{P}\Psi^{\pm}_{p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots} = \eta_{n_1}\eta_{n_2}\cdots\Psi^{\pm}_{\mathscr{P}p_1\sigma_1 n_1; \mathscr{P}p_2\sigma_2 n_2; \cdots} \tag{3.3.41}$$

where η_n is the intrinsic parity of particles of species n, and $\mathscr{P}$ reverses the space components of p^μ. (This is for massive particles; the modification for massless particles is obvious.) The parity conservation condition for the S-matrix is then:

$$\begin{aligned} S_{p'_1\sigma'_1 n'_1; p'_2\sigma'_2 n'_2; \cdots \ , \ p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots} &= \eta^*_{n'_1}\eta^*_{n'_2}\cdots\eta_{n_1}\eta_{n_2}\cdots \\ &\times S_{\mathscr{P}p'_1\sigma'_1 n'_1; \mathscr{P}p'_2\sigma'_2 n'_2; \cdots \ , \ \mathscr{P}p_1\sigma_1 n_1; \mathscr{P}p_2\sigma_2 n_2; \cdots} \ . \end{aligned} \tag{3.3.42}$$

Just as for internal symmetries, an operator P satisfying Eq. (3.3.41) will actually exist if the operator P_0 which is defined to act this way on free-particle states commutes with V as well as H_0.

The phases η_n may be inferred either from dynamical models or from experiment, but neither can provide a unique determination of the ηs. This is because we are always free to redefine P by combining it with any conserved internal symmetry operator. For instance, if P is conserved,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

then so is

$$\mathsf{P}' \equiv \mathsf{P}\exp(i\alpha B + i\beta L + i\gamma Q),$$

where B, L, and Q are respectively baryon number, lepton number, and electric charge, and α, β, and γ are arbitrary real phases; hence either P or P' could be called the parity operator. The neutron, proton, and electron have different combinations of values for B, L, and Q, so by judicious choice of the phases α, β, and γ we can define the intrinsic parities of all three particles to be $+1$. However, once we have done this the intrinsic parities of other particles like the charged pion (which can be emitted in a transition $n \to p + \pi^-$) are no longer arbitrary. Also, the intrinsic parity of any particle like the neutral pion π^0 which carries no conserved quantum numbers is always meaningful.

The foregoing remarks help to clarify the question of whether intrinsic parities must always have the values ± 1. It is easy to say that space inversion P has the group multiplication law $\mathsf{P}^2 = 1$; however, the parity operator that is conserved may not be this one, but rather may differ from it by a phase transformation of some sort. In any case, whether or not $\mathsf{P}^2 = 1$, the operator P^2 behaves just like an internal symmetry transformation:

$$\mathsf{P}^2\Psi^{\pm}_{p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots} = \eta^2_{n_1}\eta^2_{n_2}\cdots\Psi^{\pm}_{p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots}.$$

If this internal symmetry is part of a continuous symmetry group of phase transformations, such as the group of multiplication by the phases $\exp(i\alpha B + i\beta L + i\gamma Q)$ with arbitrary values of α, β, and γ, then its inverse square root must also be a member of this group, say I_P, with $I_P^2\mathsf{P}^2 = 1$ and $[I_P, \mathsf{P}] = 0$. (For instance, if $\mathsf{P}^2 = \exp(i\alpha B + \cdots)$, then take $I_P = \exp(-\frac{1}{2}i\alpha B + \cdots)$.) We can then define a new parity operator $\mathsf{P}' \equiv \mathsf{P}I_P$ with $\mathsf{P}'^2 = 1$. This is conserved to the same extent as P, so there is no reason why we should not call this *the* parity operator, in which case the intrinsic parities can only take the values ± 1.

The only sort of theory in which it is not necessarily possible to define parity so that all intrinsic parities have the values ± 1 is one in which there is some discrete internal symmetry which is not a member of any continuous symmetry group of phase transformations.[10] For instance, it is a consequence of angular-momentum conservation that the total number F of all particles of half-integer spin can only change by even numbers, so the internal symmetry operator $(-1)^F$ is conserved. All known particles of half-integer spin have odd values of the sum $B + L$ of baryon number and lepton number, so as far as we know, $(-1)^F = (-1)^{B+L}$. If this is true, then $(-1)^F$ is part of a continuous symmetry group, consisting of the operators $\exp(i\alpha(B + L))$ with arbitrary real α, and has an inverse square root $\exp(-i\alpha(B + L)/2)$. In this case, if $\mathsf{P}^2 = (-1)^F$ then P can be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

redefined so that all intrinsic parities are ± 1. However, if there were to be discovered a particle of half-integer spin and an even value of $B+L$ (such as a so-called Majorana neutrino, with $j = \frac{1}{2}$ and $B + L = 0$), then it would be possible to have $\mathsf{P}^2 = (-1)^F$ without our being able to redefine the parity operator itself to have eigenvalues ± 1. In this case, of course, we would have $\mathsf{P}^4 = 1$, so all particles would have intrinsic parities either ± 1 or (like the Majorana neutrino) $\pm i$.

It follows from Eq. (3.3.42) that if the product of intrinsic parities in the final state is equal to the product of intrinsic parities in the initial state, or equal to minus this product, then the S-matrix must be respectively even or odd overall in the three-momenta. For instance, it was observed[11] in 1951 that a pion can be absorbed by a deuteron from the $\ell = 0$ ground state of the $\pi^- d$ atom, in the reaction $\pi^- + d \to n + n$. (As discussed in Section 3.7, the orbital angular-momentum quantum number ℓ can be used in relativistic physics in the same way as in non-relativistic wave mechanics.) The initial state has total angular-momentum $j = 1$ (the pion and deuteron having spins zero and one, respectively), so the final state must have orbital angular-momentum $\ell = 1$ and total neutron spin $s = 1$. (The other possibilities $\ell = 1$, $s = 0$; $\ell = 0$, $s = 1$; and $\ell = 2$, $s = 1$, which are allowed by angular-momentum conservation, are forbidden by the requirement that the final state be antisymmetric in the two neutrons.) Because the final state has $\ell = 1$, the matrix element is odd under reversal of the direction of all three-momenta, so we can conclude that the intrinsic parities of the particles in this reaction must be related by:

$$\eta_d \eta_{\pi^-} = -\eta_n^2.$$

The deuteron is known to be a bound state of a proton and neutron with even orbital angular-momentum (chiefly $\ell = 0$), and as we have seen we can take the neutron and proton to have the same intrinsic parity, so $\eta_d = \eta_n^2$, and we can conclude that $\eta_{\pi^-} = -1$; that is, the negative pion is a *pseudoscalar* particle. The π^+ and π^0 have also been found to have negative parity, as would be expected from the symmetry (isospin invariance) among these three particles.

The negative parity of the pion has some striking consequences. A spin zero particle that decays into three pions must have intrinsic parity $\eta_\pi^3 = -1$, because in the Lorentz frame in which the decaying particle is at rest, rotational invariance only allows the matrix element to depend on scalar products of the pion momenta with each other, all of which are even under reversal of all momenta. (The triple scalar product $\mathbf{p}_1 \cdot (\mathbf{p}_2 \times \mathbf{p}_3)$ formed from the three pion momenta vanishes because $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 0$.) For the same reason, a spin zero particle that decays into two pions must have intrinsic parity $\eta_\pi^2 = +1$. In particular, among the strange particles discovered in the late 1940s there seemed to be two different particles of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

zero spin (inferred from the angular distribution of their decay products): one, the τ, was identified by its decay into three pions, and hence was assigned a parity -1, while the other, the θ, was identified by its decay into two pions and was assigned a parity $+1$. The trouble with all this was that as the τ and θ were studied in greater detail, they seemed increasingly to have identical masses and lifetimes. After many suggested solutions of this puzzle, Lee and Yang in 1956 finally cut the Gordian knot, and proposed that the τ and θ are the same particle, (now known as the $K^{\pm}$) and that parity is simply not conserved in the weak interactions that lead to its decay.[12]

As we shall see in detail in the next section of this chapter, the rate for a physical process $\alpha \to \beta$ (with $\alpha \neq \beta$) is proportional to $|S_{\beta\alpha}|^2$, with proportionality factors that are invariant under reversal of all three-momenta. As long as the states α and β contain definite numbers of particles of each type, the phase factors in Eq. (3.3.42) have no effect on $|S_{\beta\alpha}|^2$, so Eq. (3.3.42) would imply that the rate for $\alpha \to \beta$ is invariant under the reversal of direction of all three-momenta. As we have seen, this is a trivial consequence of rotational invariance for the decays of a K meson into two or three pions, but it is a non-trivial restriction on rates in more complicated processes. For example, following theoretical suggestions by Lee and Yang, Wu together with a group at the National Bureau of Standards measured the angular distribution of the electron in the final state of the beta decay $\mathrm{Co}^{60} \to \mathrm{Ni}^{60} + e^- + \bar{\nu}$ with a polarized cobalt source.[13] (No attempt was made in this experiment to measure the momentum of the antineutrino or nickel nucleus.) The electrons were found to be preferentially emitted in a direction opposite to that of the spin of the decaying nucleus, which would, of course, be impossible if the decay rate were invariant under a reversal of all three-momenta. A similar result was found in the decay of a positive muon (polarized in its production in the process $\pi^+ \to \mu^+ + \nu$) into a positron, neutrino, and antineutrino.[14] In this way, it became clear that parity is indeed not conserved in the weak interactions responsible for these decays. Nevertheless, for reasons discussed in Section 12.5, parity *is* conserved in the strong and electromagnetic interactions, and therefore continues to play an important part in theoretical physics.

Time-Reversal

We saw in Section 2.6 that the time-reversal operator T acting on a one-particle state $\Psi_{p,\sigma,n}$ gives a state $\Psi_{\mathscr{P}p,-\sigma,n}$ with reversed spin and momentum, times a phase $\zeta_n(-1)^{j-\sigma}$. A multi-particle state transforms as usual as a direct product of one-particle states, except that since this

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

is a time-reversal transformation, we expect 'in' and 'out' states to be interchanged:

$$\mathsf{T}\Psi^{\pm}_{p_1\sigma_1 n_1;\,p_2\sigma_2 n_2\cdots} = \zeta_{n_1}(-1)^{j_1-\sigma_1}\zeta_{n_2}(-1)^{j_2-\sigma_2}\cdots\Psi^{\mp}_{\mathscr{P}p_1\,-\sigma_1 n_1;\,\mathscr{P}p_2\,-\sigma_2 n_2;\cdots}\,. \tag{3.3.43}$$

(Again, this is for massive particles, with obvious modifications being required for massless particles.) It will be convenient to abbreviate this assumption as

$$\mathsf{T}\Psi^{\pm}_{\alpha} = \Psi^{\mp}_{\mathscr{T}\alpha}\,, \tag{3.3.44}$$

where $\mathscr{T}$ indicates a reversal of sign of three-momenta and spins as well as multiplication by the phase factors shown in Eq. (3.3.43). Because T is antiunitary, we have

$$(\Psi^{-}_{\beta},\Psi^{+}_{\alpha}) = (\mathsf{T}\Psi^{+}_{\alpha},\mathsf{T}\Psi^{-}_{\beta})\,, \tag{3.3.45}$$

so the time-reversal invariance condition for the S-matrix is

$$S_{\beta,\alpha} = S_{\mathscr{T}\alpha,\mathscr{T}\beta} \tag{3.3.46}$$

or in more detail

$$\begin{aligned} &S_{p_1'\sigma_1' n_1';\,p_2'\sigma_2' n_2';\cdots\,,\;p_1\sigma_1 n_1;\,p_2\sigma_2 n_2;\cdots} \\ &= \zeta_{n_1'}(-1)^{j_1'-\sigma_1'}\zeta_{n_2'}(-1)^{j_2'-\sigma_2'}\cdots\zeta^{*}_{n_1}(-1)^{j_1-\sigma_1}\zeta^{*}_{n_2}(-1)^{j_2-\sigma_2}\cdots \\ &\quad\times S_{\mathscr{P}p_1\,-\sigma_1\,n_1;\,\mathscr{P}p_2\,-\sigma_2\,n_2;\cdots\,,\;\mathscr{P}p_1'\,-\sigma_1'\,n_1';\,\mathscr{P}p_2'\,-\sigma_2'\,n_2';\cdots}\,. \end{aligned} \tag{3.3.47}$$

Note that in addition to the reversal of momenta and spins, the role of initial and final states is interchanged, as would be expected for a symmetry involving the reversal of time.

The S-matrix will satisfy this transformation rule if the operator T_0 that induces time-reversal transformations on free-particle states

$$\mathsf{T}_0\Phi_{\alpha} \equiv \Phi_{\mathscr{T}\alpha} \tag{3.3.48}$$

commutes not only with the free-particle Hamiltonian (which is automatic) but also with the interaction:

$$\mathsf{T}_0^{-1}H_0\mathsf{T}_0 = H_0\,, \tag{3.3.49}$$

$$\mathsf{T}_0^{-1}V\mathsf{T}_0 = V. \tag{3.3.50}$$

In this case we can take $\mathsf{T} = \mathsf{T}_0$, and use either (3.1.13) or (3.1.16) to show that time-reversal transformations do act as stated in Eq. (3.3.44). For instance, operating on the Lippmann–Schwinger equation (3.1.16) with T and using Eqs. (3.3.48)–(3.3.50), we have

$$\mathsf{T}\Psi^{\pm}_{\alpha} = \Phi_{\mathscr{T}\alpha} + [E_{\alpha} - H_0 \mp i\epsilon]^{-1}V\mathsf{T}\Psi^{\pm}_{\alpha}\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

with the sign of $\pm i\epsilon$ reversed because T is antiunitary. This is just the Lippmann–Schwinger equation for $\Psi^{\mp}_{\mathscr{T}\alpha}$, thus justifying Eq. (3.3.44). Similarly, because T is antiunitary it changes the sign of the i in the exponent of $\Omega(t)$, so that

$$\mathsf{T}\Omega(-\infty)\Phi_\alpha = \Omega(\infty)\Phi_{\mathscr{T}\alpha}\,,$$

again leading to Eq. (3.3.44).

In contrast with the case of parity conservation, the time-reversal invariance condition (3.3.46) does *not* in general tell us that the rate for the process $\alpha \to \beta$ is the same as for the process $\mathscr{T}\alpha \to \mathscr{T}\beta$. However, something like this is true in cases where the S-matrix takes the form

$$S_{\beta\alpha} = S^{(0)}_{\beta\alpha} + S^{(1)}_{\beta\alpha}\,, \tag{3.3.51}$$

where $S^{(1)}$ is small, while $S^{(0)}$ happens to have matrix element zero for some particular process of interest, though it generally has much larger matrix elements than $S^{(1)}$. (For instance, the process might be nuclear beta decay, $N \to N' + e^- + \bar{\nu}$, with $S^{(0)}$ the S-matrix produced by the strong nuclear and electromagnetic interactions alone, and $S^{(1)}$ the correction to the S-matrix produced by the weak interactions. Section 3.5 shows how the use of the 'distorted-wave Born approximation' leads to an S-matrix of the form (3.3.51) in cases of this type. In some cases $S^{(0)}$ is simply the unit operator.) To first order in $S^{(1)}$, the unitarity condition for the S-operator reads

$$1 = S^\dagger S = S^{(0)\dagger}S^{(0)} + S^{(0)\dagger}S^{(1)} + S^{(1)\dagger}S^{(0)}.$$

Using the zeroth-order relation $S^{(0)\dagger}S^{(0)} = 1$, this gives a reality condition for $S^{(1)}$:

$$S^{(1)} = -S^{(0)}S^{(1)\dagger}S^{(0)}. \tag{3.3.52}$$

If $S^{(1)}$ as well as $S^{(0)}$ satisfies the time-reversal condition (3.3.46), then this can be put in the form

$$S^{(1)}_{\beta\alpha} = -\int d\gamma \int d\gamma'\; S^{(0)}_{\beta\gamma'}\, S^{(1)*}_{\mathscr{T}\gamma'\mathscr{T}\gamma}\, S^{(0)}_{\gamma\alpha}. \tag{3.3.53}$$

Since $S^{(0)}$ is unitary, the rates for the processes $\alpha \to \beta$ and $\mathscr{T}\alpha \to \mathscr{T}\beta$ are thus the same if summed over sets $\mathscr{I}$ and $\mathscr{F}$ of final and initial states that are complete with respect to $S^{(0)}$. (By being 'complete' here is meant that if $S^{(0)}_{\alpha'\alpha}$ is non-zero, and either α or α' are in $\mathscr{I}$, then both states are in $\mathscr{I}$; and similarly for $\mathscr{F}$.) In the simplest case we have 'complete' sets $\mathscr{I}$ and $\mathscr{F}$ consisting of just one state each; that is, both the initial and the final states are eigenvectors of $S^{(0)}$ with eigenvalues $e^{2i\delta_\alpha}$ and $e^{2i\delta_\beta}$, respectively. (The δ_α and δ_β are called 'phase shifts'; they are real because

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

$S^{(0)}$ is unitary.) In this case, Eq. (3.3.53) becomes simply:

$$S^{(1)}_{\beta\alpha} = -e^{2i(\delta_\alpha+\delta_\beta)}S^{(1)*}_{\mathcal{T}\beta\,\mathcal{T}\alpha}, \tag{3.3.54}$$

and it is clear that the absolute value of the S-matrix for the process $\alpha \to \beta$ is the same as for the process $\mathcal{T}\alpha \to \mathcal{T}\beta$. This is the case for instance in nuclear beta decay (in the approximation in which we ignore the relatively weak Coulomb interaction between the electron and nucleus in the final state), because both the initial and final states are eigenstates of the strong interaction S-matrix (with $\delta_\alpha = \delta_\beta = 0$). Thus if time-reversal invariance is respected, the differential rate for a beta decay process should be unchanged if we reverse both the momenta and the spin z-components σ of all particles. This prediction was *not* contradicted in the 1956 experiments[13,14] that discovered the non-conservation of parity; for instance, time-reversal invariance is consistent with the observation that electrons from the decay $Co^{60} \to Ni^{60} + e^- + \bar{\nu}$ are preferentially emitted in a direction opposite to that of the Co^{60} spin. As described below, indirect evidence against time-reversal invariance did emerge in 1964, but it remains a useful approximate symmetry in weak as well as strong and electromagnetic interactions.

In some cases we can use a basis of states for which $\mathcal{T}\alpha = \alpha$ and $\mathcal{T}\beta = \beta$, for which Eq. (3.3.54) reads

$$S^{(1)}_{\beta\alpha} = -e^{2i(\delta_\alpha+\delta_\beta)}S^{(1)*}_{\beta\alpha}, \tag{3.3.55}$$

which just says that $iS^{(1)}_{\beta,\alpha}$ has the phase $\delta_\alpha + \delta_\beta$ mod π. This is known as *Watson's theorem*.[15] The phases in Eqs. (3.3.54) or (3.3.55) may be measured in processes where there is interference between different final states. For instance, in the decay of the spin 1/2 hyperon Λ into a nucleon and a pion, the final state can only have orbital angular-momentum $\ell = 0$ or $\ell = 1$; the angular distribution of the pion relative to the Λ spin involves the interference between these states, and hence according to Watson's theorem depends on the difference $\delta_s - \delta_p$ of their phase shifts.

PT

Although the 1957 experiments on parity violation did not rule out time-reversal invariance, they did show immediately that the product PT is not conserved. If conserved this operator would have to be antiunitary for the same reasons as for T, so in processes like nuclear beta decay its consequences would take the form of relations like Eq. (3.3.54):

$$S^{(1)}_{\beta\alpha} = -e^{2i(\delta_\alpha+\delta_\beta)}S^{(1)*}_{\mathcal{P}\mathcal{T}\beta\,\mathcal{P}\mathcal{T}\alpha},$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

where $\mathscr{P}\mathscr{T}$ reverses the signs of all spin z-components but *not* any momenta. Neglecting the final-state Coulomb interaction, it would then follow that there could be no preference for the electron in the decay $Co^{60} \to Ni^{60} + e^- + \bar{\nu}$ to be emitted in the same or the opposite direction to the Co^{60} spin, in contradiction with what was observed.

C, CP, *and* CPT

As already mentioned, there is an internal symmetry transformation, known as charge-conjugation, which interchanges particles and antiparticles. Formally, this entails the existence of a unitary operator C, whose effect on multi-particle states is:

$$\mathsf{C}\Psi^{\pm}_{p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots} = \xi_{n_1}\xi_{n_2}\cdots\Psi^{\pm}_{p_1\sigma_1 n_1^c; p_2\sigma_2 n_2^c; \cdots}, \tag{3.3.56}$$

where n^c is the antiparticle of particle type n, and ξ_n is yet another phase. If this is true for both 'in' and 'out' states then the S-matrix satisfies the invariance conditions

$$\begin{aligned} &S_{p_1'\sigma_1' n_1'; p_2'\sigma_2' n_2'; \cdots, \; p_1\sigma_1 n_1; p_2\sigma_2 n_2; \cdots} \\ &= \xi^*_{n_1'}\xi^*_{n_2'}\cdots\xi_{n_1}\xi_{n_2}\cdots S_{p_1'\sigma_1' n_1^{c\prime}; p_2'\sigma_2' n_2^{c\prime}; \cdots, \; p_1\sigma_1 n_1^c; p_2\sigma_2 n_2^c; \cdots}. \end{aligned} \tag{3.3.57}$$

As with other internal symmetries, the S-matrix will satisfy this condition if the operator C_0 that is defined to act as stated in Eq. (3.3.56) on free-particle states commutes with the interaction V as well as H_0; in this case, we take $\mathsf{C} = \mathsf{C}_0$.

The phases ξ_n are called charge-conjugation parities. Just as for the ordinary parities η_n, the ξ_n are in general not uniquely defined, because for any operator C that is defined to satisfy Eq. (3.3.56), we can find another such operator with different ξ_n by multiplying C with any internal symmetry phase transformation, such as $\exp(i\alpha B + i\beta L + i\gamma Q)$; the only particles whose charge-conjugation parities are individually measurable are those completely neutral particles like the photon or the neutral pion that carry no conserved quantum numbers and are their own antiparticles. In reactions involving only completely neutral particles, Eq. (3.3.57) tells us that the product of the charge-conjugation parities in the initial and final states must be equal; for instance, as we shall see the photon is required by quantum electrodynamics to have charge-conjugation parity $\eta_\gamma = -1$, so the observation of the neutral pion decay $\pi^0 \to 2\gamma$ requires that $\eta_{\pi^0} = +1$; it then follows that the process $\pi^0 \to 3\gamma$ should be forbidden, as is in fact known to be the case. For these two particles, the charge-conjugation parities are real, either $+1$ or -1. Just as for ordinary parity, this will always be the case if all internal phase transformation symmetries are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

members of continuous groups of phase transformations, because then we can redefine C by multiplying by the inverse square root of the internal symmetry equal to C^2, with the result that the new C satisfies $C^2 = 1$.

For general reactions, Eq. (3.3.57) requires that the rate for a process equals the rate for the same process with particles replaced with their corresponding antiparticles. This was not directly contradicted by the 1957 experiments on parity non-conservation (it will be a long time before anyone is able to study the beta decay of anticobalt), but these experiments showed that C is not conserved in the *theory* of weak interactions as modified by Lee and Yang[12] to take account of parity non-conservation. (As we shall see below, the observed violation of TP conservation would imply a violation of C conservation in *any* field theory of weak interactions, not just in the particular theory considered by Lee and Yang.) It is understood today that C as well as P is not conserved in the weak interactions responsible for processes like beta decay and the decay of the pion and muon, though both C and P are conserved in the strong and electromagnetic interactions.

Although the early experiments on parity non-conservation indicated that neither C nor P are conserved in the weak interactions, they left open the possibility that their product CP is universally conserved. For some years it was expected (though not with complete confidence) that CP would be found to be generally conserved. This had particularly important consequences for the properties of the neutral K mesons. In 1954 Gell-Mann and Pais[16] had pointed out that because the K^0 meson is not its own antiparticle (the K^0 carries a non-zero value for the approximately conserved quantity known as strangeness) the particles with definite decay rates would be not K^0 or $\overline{K^0}$, but the linear combinations $K^0 \pm \overline{K^0}$. This was originally explained in terms of C conservation, but with C not conserved in the weak interactions, the argument may be equally well be based on CP conservation. If we arbitrarily define the phases in the CP operator and in the K^0 and $\overline{K^0}$ states so that

$$\mathrm{CP}\Psi_{K^0} = \Psi_{\overline{K^0}}$$

and

$$\mathrm{CP}\Psi_{\overline{K^0}} = \Psi_{K^0}$$

then we can define self-charge-conjugate one-particle states

$$\Psi_{K_1^0} \equiv \frac{1}{\sqrt{2}}[\Psi_{K^0} + \Psi_{\overline{K^0}}]$$

and

$$\Psi_{K_2^0} \equiv \frac{1}{\sqrt{2}}[\Psi_{K^0} - \Psi_{\overline{K^0}}]\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

which have CP eigenvalues $+1$ and -1, respectively. The fastest available decay mode of these particles is into two-pion states, but CP conservation would allow this only** for the K_1, not the K_2. The K_2^0 would thus be expected to decay only by slower modes, into three pions or into a pion, muon or electron, and neutrino. Nevertheless, it was found by Fitch and Cronin in 1964 that the long-lived neutral K-meson does have a small probability for decaying into two pions.[17] The conclusion was that CP is not exactly conserved in the weak interactions, although it seems more nearly conserved than C or P individually.

As we shall see in Chapter 5, there are good reasons to believe that although neither C nor CP is strictly conserved, the product CPT is exactly conserved in all interactions, at least in any quantum field theory. It is CPT that provides a precise correspondence between particles and antiparticles, and in particular it is the fact that CPT commutes with the Hamiltonian that tells us that stable particles and antiparticles have exactly the same mass. Because CPT is antiunitary, it relates the S-matrix for an arbitrary process to the S-matrix for the *inverse* process with all spin three-components reversed and particles replaced with antiparticles. However, in cases where the S-matrix can be divided into a weak term $S^{(1)}$ that produces a given reaction and a strong term $S^{(0)}$ that acts in the initial and final states, we can use the same arguments that were used above in studying the implications of T conservation to show that the rate for any process is equal to the rate for the same process with particles replaced with antiparticles and spin three-components reversed, provided that we sum over sets of initial and final states that are complete with respect to $S^{(0)}$. In particular, although the partial rates for decay of the particle into a pair of final states β_1, β_2 with $S^{(0)}_{\beta_1\beta_2} \neq 0$ may differ from the partial rates for the decay of the antiparticle into the corresponding final states $\mathscr{CPT}\beta_1$ and $\mathscr{CPT}\beta_2$, we shall see in Section 3.5 that (without any approximations) the total decay rate of any particle is equal to that of its antiparticle.

We can now understand why the 1957 experiments on parity violation could be interpreted in the context of the existing theory of weak interactions as evidence that C conservation as well as P conservation are badly violated but CP is not. These theories were field theories, and therefore automatically conserved CPT. Since the experiments showed that PT conservation but not T conservation is badly violated in nuclear beta decay, *any* theory that was consistent with these experiments and in which CPT is conserved would have to also incorporate C but not CP non-conservation.

** The neutral K-mesons have spin zero, so the two-pion final state has $\ell = 0$, and hence $P = +1$. Further, $C = +1$ for two π^0s because the π^0 has $C = +1$, and also for an $\ell = 0$ π^+–π^- state because C interchanges the two pions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Similarly, the observation in 1964 of small violations of CP conservation in the weak interactions together with the assumed invariance of all interactions under CPT allowed the immediate inference that the weak interactions also do not exactly conserve T. This has since been verified[18] by more detailed studies of the K^0–$\overline{K^0}$ system, but it has so far been impossible to find other direct evidence of the failure of invariance under time-reversal.

3.4 Rates and Cross-Sections

The S-matrix $S_{\beta\alpha}$ is the probability amplitude for the transition $\alpha \to \beta$, but what does this have to do with the transition rates and cross-sections measured by experimentalists? In particular, (3.3.2) shows that $S_{\beta\alpha}$ has a factor $\delta^4(p_\beta - p_\alpha)$, which ensures the conservation of the total energy and momentum, so what are we to make of the factor $[\delta^4(p_\beta - p_\alpha)]^2$ in the transition probability $|S_{\beta\alpha}|^2$? The proper way to approach these problems is by studying the way that experiments are actually done, using wave packets to represent particles localized far from each other before a collision, and then following the time-history of these superpositions of multi-particle states. In what follows we will instead give a quick and easy derivation of the main results, actually more a mnemonic than a derivation, with the excuse that (as far as I know) no interesting open problems in physics hinge on getting the fine points right regarding these matters.

We consider our whole system of physical particles to be enclosed in a large box with a macroscopic volume V. For instance, we can take this box as a cube, but with points on opposite sides identified, so that the single-valuedness of the spatial wave function requires the momenta to be quantized

$$\mathbf{p} = \frac{2\pi}{L}(n_1, n_2, n_3)\,, \tag{3.4.1}$$

where the n_i are integers, and $L^3 = V$. Then all three-dimensional delta-functions become

$$\delta^3_V(\mathbf{p}' - \mathbf{p}) \equiv \frac{1}{(2\pi)^3}\int_V d^3x\, e^{i(\mathbf{p}-\mathbf{p}')\cdot\mathbf{x}} = \frac{V}{(2\pi)^3}\delta_{\mathbf{p}',\mathbf{p}}\,, \tag{3.4.2}$$

where $\delta_{\mathbf{p}',\mathbf{p}}$ is an ordinary Kronecker delta symbol, equal to one if the subscripts are equal and zero otherwise. The normalization condition (3.1.2) thus implies that the states we have been using have scalar products in a box which are not just sums of products of Kronecker deltas, but also contain a factor $[V/(2\pi)^3]^N$, where N is the number of particles in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

the state. To calculate transition probabilities we should use states of unit norm, so let us introduce states normalized approximately for our box

$$\Psi_\alpha^{\text{Box}} \equiv \left[(2\pi)^3/V\right]^{N_\alpha/2}\Psi_\alpha \tag{3.4.3}$$

with norm

$$\left(\Psi_\beta^{\text{Box}},\ \Psi_\alpha^{\text{Box}}\right) = \delta_{\beta\alpha} \tag{3.4.4}$$

where $\delta_{\beta\alpha}$ is a product of Kronecker deltas, one for each three-momentum, spin, and species label, plus terms with particles permuted. Correspondingly, the S-matrix may be written

$$S_{\beta\alpha} = \left[V/(2\pi)^3\right]^{(N_\beta+N_\alpha)/2} S_{\beta\alpha}^{\text{Box}}\ , \tag{3.4.5}$$

where $S_{\beta\alpha}^{\text{Box}}$ is calculated using the states (3.4.3).

Of course, if we just leave our particles in the box forever, then every possible transition will occur again and again. To calculate a meaningful transition probability we also have to put our system in a 'time box'. We suppose that the interaction is turned on for only a time T. One immediate consequence is that the energy-conservation delta function is replaced with

$$\delta_T(E_\alpha - E_\beta) = \frac{1}{2\pi}\int_{-T/2}^{T/2} \exp\left(i(E_\alpha - E_\beta)t\right)dt\ . \tag{3.4.6}$$

The probability that a multi-particle system, which is in a state α before the interaction is turned on, is found in a state β after the interaction is turned off, is

$$P(\alpha \to \beta) = \left|S_{\beta\alpha}^{\text{Box}}\right|^2 = \left[(2\pi)^3/V\right]^{(N_\alpha+N_\beta)}\left|S_{\beta\alpha}\right|^2\ . \tag{3.4.7}$$

This is the probability for a transition into one specific box state β. The number of one-particle box states in a momentum-space volume d^3p is $Vd^3p/(2\pi)^3$, because this is the number of triplets of integers n_1, n_2, n_3 for which the momentum (3.4.1) lies in the momentum-space volume d^3p around $\mathbf{p}$. We shall define the final-state interval $d\beta$ as a product of $d^3\mathbf{p}$ for each final particle, so the total number of states in this range is

$$d\mathcal{N}_\beta = \left[V/(2\pi)^3\right]^{N_\beta} d\beta\ . \tag{3.4.8}$$

Hence the total probability for the system to wind up in a range $d\beta$ of final states is

$$dP(\alpha \to \beta) = P(\alpha \to \beta)d\mathcal{N}_\beta = \left[(2\pi)^3/V\right]^{N_\alpha} |S_{\beta\alpha}|^2\, d\beta\ . \tag{3.4.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

We will restrict our attention throughout this section to final states β that are not only different (however slightly) from the initial state α, but that also satisfy the more stringent condition, that no subset of the particles in the state β (other than the whole state itself) has precisely the same four-momentum as some corresponding subset of the particles in the state α. (In the language to be introduced in the next chapter, this means that we are considering only the connected part of the S-matrix.) For such states, we may define a delta function-free matrix element $M_{\beta\alpha}$:

$$S_{\beta\alpha} \equiv -2i\pi\delta_V^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)\delta_T(E_\beta - E_\alpha)M_{\beta\alpha} \,. \tag{3.4.10}$$

Our introduction of the box allows us to interpret the squares of delta functions in $|S_{\beta\alpha}|^2$ for $\beta \neq \alpha$ as

$$\left[\delta_V^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)\right]^2 = \delta_V^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)\delta_V^3(0) = \delta_V^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)V/(2\pi)^3 \,,$$

$$\left[\delta_T(E_\beta - E_\alpha)\right]^2 = \delta_T(E_\beta - E_\alpha)\delta_T(0) = \delta_T(E_\beta - E_\alpha)T/2\pi \,,$$

so Eq. (3.4.9) gives a differential transition probability

$$\begin{aligned} dP(\alpha \to \beta) = (2\pi)^2\left[(2\pi)^3/V\right]^{N_\alpha - 1}(T/2\pi)|M_{\beta\alpha}|^2 \\ \times \ \delta_V^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)\delta_T(E_\beta - E_\alpha)d\beta \,. \end{aligned}$$

If we let V and T be very large, the delta function product here may be interpreted as an ordinary four-dimensional delta function $\delta^4(p_\beta - p_\alpha)$. In this limit, the transition probability is simply proportional to the time T during which the interaction is acting, with a coefficient that may be interpreted as a differential transition *rate*:

$$\begin{aligned} d\Gamma(\alpha \to \beta) &\equiv dP(\alpha \to \beta)/T \\ &= (2\pi)^{3N_\alpha - 2}V^{1-N_\alpha}|M_{\beta\alpha}|^2\delta^4(p_\beta - p_\alpha)d\beta \,, \end{aligned} \tag{3.4.11}$$

where now

$$S_{\beta\alpha} \equiv -2\pi i\delta^4(p_\beta - p_\alpha)M_{\beta\alpha} \,. \tag{3.4.12}$$

This is the master formula which is used to interpret calculations of S-matrix elements in terms of predictions for actual experiments. We will come back to the interpretation of the factor $\delta^4(p_\alpha - p_\beta)d\beta$ later in this section.

There are two cases of special importance:

$N_\alpha = 1$:

Here the volume V cancels in Eq. (3.4.11), which gives the transition rate for a single-particle state α to decay into a general multi-particle state β as

$$d\Gamma(\alpha \to \beta) = 2\pi|M_{\beta\alpha}|^2\delta^4(p_\beta - p_\alpha)d\beta \,. \tag{3.4.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Of course, this makes sense only if the time T during which the interaction acts is much less than the mean lifetime τ_α of the particle α, so we cannot pass to the limit $T \to \infty$ in $\delta_T(E_\alpha - E_\beta)$. There is an unremovable width $\triangle E \simeq 1/T \gtrsim 1/\tau_\alpha$ in this delta function, so Eq. (3.4.13) is only useful if the total decay rate $1/\tau_\alpha$ is much less than any of the characteristic energies of the process.

$N_\alpha = 2$:

Here the rate (3.4.11) is proportional to $1/V$, or in other words, to the density of either particle at the position of the other one. Experimentalists generally report not the transition rate per density, but the *rate per flux*, also known as the *cross-section*. The flux of either particle at the position of the other particle is defined as the product of the density $1/V$ and the relative velocity u_α:

$$\Phi_\alpha = u_\alpha / V \ . \tag{3.4.14}$$

(A general definition of u_α is given below; for the moment we will content ourselves with specifying that if either particle is at rest then u_α is defined as the velocity of the other.) Thus the differential cross-section is

$$d\sigma(\alpha \to \beta) \equiv d\Gamma(\alpha \to \beta)/\Phi_\alpha = (2\pi)^4 u_\alpha^{-1} |M_{\beta\alpha}|^2 \delta^4(p_\beta - p_\alpha) d\beta \ . \tag{3.4.15}$$

Even though the cases $N_\alpha = 1$ and $N_\alpha = 2$ are the most important, transition rates for $N_\alpha \geq 3$ are all measurable in principle, and some of them are very important in chemistry, astrophysics, etc. (For instance, in one of the main reactions that release energy in the sun, two protons and an electron turn into a deuteron and a neutrino.) Section 3.6 presents an application of the master transition rate formula (3.4.11) for general numbers N_α of initial particles.

We next take up the question of the Lorentz transformation properties of rates and cross-sections, which will help us to give a more general definition of the relative velocity u_α in Eq. (3.4.15). The Lorentz transformation rule (3.3.1) for the S-matrix is complicated by the momentum-dependent matrices associated with each particle's spin. To avoid this complication, consider the absolute-value squared of (3.3.1) (after factoring out the Lorentz-invariant delta function in Eq. (3.4.12)), and sum over all spins. The unitarity of the matrices $D^{(j)}_{\bar{\sigma}\sigma}(W)$ (or their analogs for zero mass) then shows that, apart from the energy factors in (3.3.1), the sum is Lorentz-invariant. That is, the quantity

$$\sum_{\text{spins}} |M_{\beta\alpha}|^2 \prod_\beta E \prod_\alpha E \equiv R_{\beta\alpha} \tag{3.4.16}$$

is a scalar function of the four-momenta of the particles in states α and β. (By $\prod_\alpha E$ and $\prod_\beta E$ is meant the product of all the single-particle energies $p^0 = \sqrt{\mathbf{p}^2 + m^2}$ for the particles in the states α and β.)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

We can now write the spin-summed single-particle decay rate (3.4.13) as

$$\sum_{\text{spins}} d\Gamma(\alpha \to \beta) = 2\pi E_\alpha^{-1} R_{\beta\alpha} \delta^4(p_\beta - p_\alpha) d\beta / \prod_\beta E \,.$$

The factor $d\beta/\prod_\beta E$ may be recognized as the product of the Lorentz-invariant momentum-space volume elements (2.5.15), so it is Lorentz-invariant. So also are $R_{\beta\alpha}$ and $\delta^4(p_\beta - p_\alpha)$, leaving just the non-invariant factor $1/E_\alpha$, where E_α is the energy of the single initial particle. Our conclusion then is that the decay rate has the same Lorentz transformation property as $1/E_\alpha$. This is, of course, just the usual special-relativistic time dilation — the faster the particle, the slower it decays.

Similarly, our result (3.4.15) for the spin-summed cross-section may be written as

$$\sum_{\text{spins}} d\sigma(\alpha \to \beta) = (2\pi)^4 u_\alpha^{-1} E_1^{-1} E_2^{-1} R_{\beta\alpha} \delta^4(p_\alpha - p_\beta) d\beta / \prod_\beta E \,,$$

where E_1 and E_2 are the energies of the two particles in the initial state α. It is conventional to define the cross-section to be (when summed over spins) a Lorentz-invariant function of four-momenta. The factors $R_{\beta\alpha}$, $\delta^4(p_\beta - p_\alpha)$, and $d\beta/\prod_\beta E$ are already Lorentz-invariant, so this means that we must define the relative velocity u_α in arbitrary inertial frames so that $u_\alpha E_1 E_2$ is a scalar. We also mentioned earlier that in the Lorentz frame in which one particle (say, particle 1) is at rest, u_α is the velocity of the other particle. This uniquely determines u_α to have the value in general Lorentz frames*

$$u_\alpha = \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} \Big/ E_1 E_2 \tag{3.4.17}$$

where p_1, p_2 and m_1, m_2 are the four-momenta and mass of the two particles in the initial state α.

As a bonus, we note that in the 'center-of-mass' frame, where the total three-momentum vanishes, we have

$$p_1 = (\mathbf{p}, E_1) \,, \qquad p_2 = (-\mathbf{p}, E_2)$$

and here Eq. (3.4.17) gives

$$u_\alpha = \frac{|\mathbf{p}|(E_1 + E_2)}{E_1 E_2} = \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right| \tag{3.4.18}$$

* Eq. (3.4.17) makes it obvious that $E_1 E_2 u_\alpha$ is a scalar. Also, when particle 1 is at rest, we have $\mathbf{p}_1 = 0$, $E_1 = m_1$, so $p_1 \cdot p_2 = -m_1 E_2$, and so Eq. (3.4.17) gives

$$u_\alpha = \sqrt{E_2^2 - m_2^2}/E_2 = |\mathbf{p}_2|/E_2 \,,$$

which is just the velocity of particle 2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

as might have been expected for a relative velocity. However, in this frame u_α is not really a physical velocity; in particular, Eq. (3.4.18) shows that for extremely relativistic particles, it can take values as large as 2.

We now turn to the interpretation of the so-called phase-space factor $\delta^4(p_\beta - p_\alpha)d\beta$, which appears in the general formula (3.4.11) for transition rates, and also in Eqs. (3.4.13) and (3.4.15) for decay rates and cross-sections. We here specialize to the case of the 'center-of-mass' Lorentz frame, where the total three-momentum of the initial state vanishes

$$\mathbf{p}_\alpha = 0\,. \tag{3.4.19}$$

(For $N_\alpha = 1$, this is just the case of a particle decaying at rest.) If the final state consists of particles with momenta $\mathbf{p}'_1, \mathbf{p}'_2, \cdots$, then

$$\delta^4(p_\beta - p_\alpha)d\beta = \delta^3(\mathbf{p}'_1 + \mathbf{p}'_2 + \cdots)\delta(E'_1 + E'_2 + \cdots - E)d^3p'_1 d^3p'_2 \cdots\,, \tag{3.4.20}$$

where $E \equiv E_\alpha$ is the total energy of the initial state. Any one of the $\mathbf{p}'_k$ integrals, say over $\mathbf{p}'_1$, can be done trivially by just dropping the momentum delta function

$$\delta^4(p_\beta - p_\alpha)d\beta \to \delta(E'_1 + E'_2 + \cdots - E)d^3p'_2 \cdots \tag{3.4.21}$$

with the understanding that wherever $\mathbf{p}'_1$ appears (as in E'_1) it must be replaced with

$$\mathbf{p}'_1 = -\mathbf{p}'_2 - \mathbf{p}'_3 - \cdots\,. \tag{3.4.22}$$

We can similarly use the remaining delta function to eliminate any *one* of the remaining integrals.

In the simplest case, there are just two particles in the final state. Here (3.4.21) gives

$$\delta^4(p_\beta - p_\alpha)d\beta \to \delta(E'_1 + E'_2 - E)d^3\mathbf{p}'_2\,.$$

In more detail, this is

$$\delta^4(p_\beta - p_\alpha)d\beta \to \delta\left(\sqrt{|\mathbf{p}_1{}'|^2 + m_1'^2} + \sqrt{|\mathbf{p}_1{}'|^2 + m_2'^2} - E\right)|\mathbf{p}_1{}'|^2 d|\mathbf{p}_1{}'|d\Omega\,, \tag{3.4.23}$$

where

$$\mathbf{p}'_2 = -\mathbf{p}'_1$$

and $d\Omega \equiv \sin\theta\, d\theta\, d\phi$ is the solid-angle differential for $\mathbf{p}_1{}'$. This can be simplified by using the standard formula

$$\delta(f(x)) = \delta(x - x_0)/\,|f'(x_0)|\,,$$

where $f(x)$ is an arbitrary real function with a single simple zero at $x = x_0$. In our case, the argument $E'_1 + E'_2 - E$ of the delta function in Eq. (3.4.23)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

has a unique zero at $|\mathbf{p}_1'| = k'$, where

$$k' = \sqrt{(E^2 - m_1'^2 - m_2'^2)^2 - 4m_1'^2 m_2'^2}/2E \,, \tag{3.4.24}$$

$$E_1' = \sqrt{k'^2 + m_1'^2} = \frac{E^2 - m_2'^2 + m_1'^2}{2E} \,, \tag{3.4.25}$$

$$E_2' = \sqrt{k'^2 + m_2'^2} = \frac{E^2 - m_1'^2 + m_2'^2}{2E} \,, \tag{3.4.26}$$

with derivative

$$\left[\frac{d}{d|\mathbf{p}_1'|}\left(\sqrt{|\mathbf{p}_1'|^2 + m_1'^2} + \sqrt{|\mathbf{p}_1'|^2 + m_2'^2} - E\right)\right]_{|\mathbf{p}_1'|=k'}$$
$$= \frac{k'}{E_1'} + \frac{k'}{E_2'} = \frac{k'E}{E_1'E_2'} \,. \tag{3.4.27}$$

We can thus drop the delta function and the differential $d|\mathbf{p}_1'|$ in Eq. (3.4.23), by dividing by (3.4.27),

$$\delta^4(p_\beta - p_\alpha)d\beta \to \frac{k'E_1'E_2'}{E}d\Omega \tag{3.4.28}$$

with the understanding that k', E_1', and E_2' are given everywhere by Eqs. (3.4.24)–(3.4.26). In particular, the differential rate (3.4.13) for decay of a one-particle state of zero momentum and energy E into two particles is

$$\frac{d\Gamma(\alpha \to \beta)}{d\Omega} = \frac{2\pi k' E_1' E_2'}{E}|M_{\beta\alpha}|^2 \tag{3.4.29}$$

and the differential cross-section for the two-body scattering process $1\,2 \to 1'2'$ is given by Eq. (3.4.15) as

$$\frac{d\sigma(\alpha \to \beta)}{d\Omega} = \frac{(2\pi)^4 k' E_1' E_2'}{E u_\alpha}\,|M_{\beta\alpha}|^2 = \frac{(2\pi)^4 k' E_1' E_2' E_1 E_2}{E^2 k}\,|M_{\beta\alpha}|^2 \,, \tag{3.4.30}$$

where $k \equiv |\mathbf{p}_1| = |\mathbf{p}_2|$.

The above case $N_\beta = 2$ is particularly simple, but there is one nice result for $N_\beta = 3$ that is also worth recording. For $N_\beta = 3$, Eq. (3.4.21) gives

$$\delta^4(p_\beta - p_\alpha)d\beta \to d^3\mathbf{p}_2'\, d^3\mathbf{p}_3'$$
$$\times \delta\left(\sqrt{(\mathbf{p}_2' + \mathbf{p}_3')^2 + m_1'^2} + \sqrt{\mathbf{p}_2'^2 + m_2'^2} + \sqrt{\mathbf{p}_3'^2 + m_3'^2} - E\right) \,.$$

We write the momentum-space volume as

$$d^3p_2'\, d^3p_3' = |\mathbf{p}_2'|^2\, d|\mathbf{p}_2'|\, |\mathbf{p}_3'|^2\, d|\mathbf{p}_3'|\, d\Omega_3\, d\phi_{23}\, d\cos\theta_{23} \,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

where $d\Omega_3$ is the differential element of solid angle for $\mathbf{p}'_3$, and θ_{23} and ϕ_{23} are the polar and azimuthal angles of $\mathbf{p}'_2$ relative to the $\mathbf{p}'_3$ direction. The orientation of the plane spanned by $\mathbf{p}'_2$ and $\mathbf{p}'_3$ is specified by ϕ_{23} and the direction of $\mathbf{p}'_3$, with the remaining angle θ_{23} fixed by the energy-conservation condition

$$\sqrt{\mathbf{p}'^2_2 + 2|\mathbf{p}'_2|\,|\mathbf{p}'_3|\cos\theta_{23} + |\mathbf{p}'_3|^2 + m'^2_1} + \sqrt{|\mathbf{p}'_2|^2 + m'^2_2} + \sqrt{|\mathbf{p}'^2_3| + m'^2_3} = E\,.$$

The derivative of the argument of the delta function with respect to $\cos\theta_{23}$ is

$$\frac{\partial E'_1}{\partial \cos\theta_{23}} = \frac{|\mathbf{p}'_2|\,|\mathbf{p}'_3|}{E'_1}\,,$$

so we can do the integral over $\cos\theta_{23}$ by just dropping the delta function and dividing by this derivative

$$\delta^4(p_\beta - p_\alpha)d\beta \to |\mathbf{p}'_2|\,d\,|\mathbf{p}'_2|\;|\mathbf{p}'_3|\,d\,|\mathbf{p}'_3|\;E'_1\,d\Omega_3\,d\phi_{23}\,.$$

Replacing momenta with energies, this is finally

$$\delta^4(p_\beta - p_\alpha)d\beta \to E'_1\,E'_2\,E'_3\,dE'_2 dE'_3\,d\Omega_3\,d\phi_{23}\,. \tag{3.4.31}$$

But recall that the quantity (3.4.16), obtained by summing $|M_{\beta\alpha}|^2$ over spins and multiplying with the product of energies, is a scalar function of four-momenta. If we approximate this scalar as a constant, then Eq. (3.4.31) tells us that for a fixed initial state, the distribution of events plotted in the E'_2, E'_3 plane is uniform. Any departure from a uniform distribution of events in this plot thus provides a useful clue to the dynamics of the decay process, including possible centrifugal barriers or resonant intermediate states. This is known as a Dalitz plot,[19] because of its use by Dalitz in 1953 to analyze the decay $K^+ \to \pi^+ + \pi^+ + \pi^-$.

3.5 Perturbation Theory

The technique that has historically been most useful in calculating the S-matrix is perturbation theory, an expansion in powers of the interaction term V in the Hamiltonian $H = H_0 + V$. Eqs. (3.2.7) and (3.1.18) give the S-matrix as

$$S_{\beta\alpha} = \delta(\beta - \alpha) - 2i\pi\delta(E_\beta - E_\alpha)T_{\beta\alpha}{}^+$$
$$T_{\beta\alpha}{}^+ = (\Phi_\beta, V\Psi_\alpha{}^+),$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

where $\Psi_\alpha{}^+$ satisfies the Lippmann–Schwinger equation (3.1.17):

$$\Psi_\alpha{}^+ = \Phi_\alpha + \int d\gamma \, \frac{T_{\gamma\alpha}{}^+ \Phi_\gamma}{E_\alpha - E_\gamma + i\epsilon}.$$

Operating on this equation with V and taking the scalar product with Φ_β yields an integral equation for T^+

$$T_{\beta\alpha}{}^+ = V_{\beta\alpha} + \int d\gamma \, \frac{V_{\beta\gamma} T_{\gamma\alpha}{}^+}{E_\alpha - E_\gamma + i\epsilon}, \tag{3.5.1}$$

where

$$V_{\beta\alpha} \equiv (\Phi_\beta, V\Phi_\alpha). \tag{3.5.2}$$

The perturbation series for $T_{\beta\alpha}{}^+$ is obtained by iteration from Eq. (3.5.1)

$$T_{\beta\alpha}{}^+ = V_{\beta\alpha} + \int d\gamma \, \frac{V_{\beta\gamma} V_{\gamma\alpha}}{E_\alpha - E_\gamma + i\epsilon} + \int d\gamma d\gamma' \, \frac{V_{\beta\gamma} V_{\gamma\gamma'} V_{\gamma'\alpha}}{(E_\alpha - E_\gamma + i\epsilon)(E_\alpha - E_{\gamma'} + i\epsilon)} + \cdots. \tag{3.5.3}$$

The method of calculation based on Eq. (3.5.3), which dominated calculations of the S-matrix in the 1930s, is today known as *old-fashioned perturbation theory*. Its obvious drawback is that the energy denominators obscure the underlying Lorentz invariance of the S-matrix. It still has some uses, however, in clarifying the way that singularities of the S-matrix arise from various intermediate states. For the most part in this book, we will rely on a rewritten version of Eq. (3.5.3), known as *time-dependent perturbation theory*, which has the virtue of making Lorentz invariance much more transparent, while somewhat obscuring the contribution of individual intermediate states.

The easiest way to derive the time-ordered perturbation expansion is to use Eq. (3.2.5), which gives the S-operator as

$$S = U(\infty, -\infty),$$

where

$$U(\tau, \tau_0) \equiv \exp(iH_0\tau) \exp(-iH(\tau - \tau_0)) \exp(-iH_0\tau_0).$$

Differentiating this formula for $U(\tau, \tau_0)$ with respect to τ gives the differential equation

$$i\frac{d}{d\tau} U(\tau, \tau_0) = V(\tau) U(\tau, \tau_0), \tag{3.5.4}$$

where

$$V(t) \equiv \exp(iH_0 t) V \exp(-iH_0 t). \tag{3.5.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

(Operators with this sort of time-dependence are said to be defined in the *interaction picture*, to distinguish their time-dependence from the time-dependence $O_H(t) = \exp(iHt)O_H \exp(-iHt)$ required in the Heisenberg picture of quantum mechanics.) Eq. (3.5.4) as well as the initial condition $U(\tau_0, \tau_0) = 1$ is obviously satisfied by the solution of the integral equation

$$U(\tau, \tau_0) = 1 - i \int_{\tau_0}^{\tau} dt\, V(t) U(t, \tau_0). \tag{3.5.6}$$

By iteration of this integral equation, we obtain an expansion for $U(\tau, \tau_0)$ in powers of V

$$\begin{aligned} U(\tau, \tau_0) = 1 - i \int_{\tau_0}^{\tau} dt_1\, V(t_1) + (-i)^2 \int_{\tau_0}^{\tau} dt_1 \int_{\tau_0}^{t_1} dt_2\, V(t_1) V(t_2) \\ + (-i)^3 \int_{\tau_0}^{\tau} dt_1 \int_{\tau_0}^{t_1} dt_2 \int_{\tau_0}^{t_2} dt_3\, V(t_1) V(t_2) V(t_3) + \cdots . \end{aligned} \tag{3.5.7}$$

Setting $\tau = \infty$ and $\tau_0 = -\infty$ then gives the perturbation expansion for the S-operator:

$$\begin{aligned} S = 1 - i \int_{-\infty}^{\infty} dt_1 V(t_1) + (-i)^2 \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{t_1} dt_2\, V(t_1) V(t_2) \\ + (-i)^3 \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{t_1} dt_2 \int_{-\infty}^{t_2} dt_3\, V(t_1) V(t_2) V(t_3) + \cdots . \end{aligned} \tag{3.5.8}$$

This can also be derived directly from the old-fashioned perturbation expansion (3.5.3), by using the Fourier representation of the energy factors in Eq. (3.5.3):

$$(E_\alpha - E_\gamma + i\epsilon)^{-1} = -i \int_0^{\infty} d\tau \exp(i(E_\alpha - E_\gamma)\tau) \tag{3.5.9}$$

with the understanding that such integrals are to be evaluated by inserting a convergence factor $e^{-\epsilon\tau}$ in the integrand, with $\epsilon \to 0+$.

There is a way of rewriting Eq. (3.5.8) that proves very useful in carrying out manifestly Lorentz-invariant calculations. Define the *time-ordered product* of any time-dependent operators as the product with factors arranged so that the one with the latest time-argument is placed leftmost, the next-latest next to the leftmost, and so on. For instance,

$$\begin{aligned} T\{V(t)\} &= V(t)\,, \\ T\{V(t_1)V(t_2)\} &= \theta(t_1 - t_2)V(t_1)V(t_2) + \theta(t_2 - t_1)V(t_2)V(t_1), \end{aligned}$$

and so on, where $\theta(\tau)$ is the step function, equal to $+1$ for $\tau > 0$ and to zero for $\tau < 0$. The time-ordered product of n Vs is a sum over all $n!$ permutations of the Vs, each of which gives the same integral over all

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

$t_1 \cdots t_n$, so Eq. (3.5.8) may be written

$$S = 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} dt_1 dt_2 \ldots dt_n \, T\Big\{V(t_1) \cdots V(t_n)\Big\}. \tag{3.5.10}$$

This is sometimes known as the *Dyson series*.[20] This series can be summed if the $V(t)$ at different times all commute; the sum is then

$$S = \exp\left(-i \int_{-\infty}^{\infty} dt \, V(t)\right) .$$

Of course, this is not usually the case; in general (3.5.10) does not even converge, and is at best an asymptotic expansion in whatever coupling constant factors appear in V. However Eq. (3.5.10) is sometimes written in the general case as

$$S = T \exp\left(-i \int_{-\infty}^{\infty} dt \, V(t)\right)$$

with T indicating here that the expression is to be evaluated by time-ordering each term in the series expansion for the exponential.

We can now readily find one large class of theories for which the S-matrix is manifestly Lorentz-invariant. Since the elements of the S-matrix are the matrix elements of the S-operator between free-particle states Φ_α, Φ_β, etc., what we want is that the S-operator should commute with the operator $U_0(\Lambda, a)$ that produces Lorentz transformations on these free-particle states. Equivalently, the S-operator must commute with the generators of $U_0(\Lambda, a)$: H_0, $\mathbf{P}_0$ $\mathbf{J}_0$, and $\mathbf{K}_0$. To satisfy this requirement, let's try the hypothesis that $V(t)$ is an integral over three-space

$$V(t) = \int d^3x \, \mathscr{H}(\mathbf{x}, t) \tag{3.5.11}$$

with $\mathscr{H}(x)$ a scalar in the sense that

$$U_0(\Lambda, a)\mathscr{H}(x)U_0^{-1}(\Lambda, a) = \mathscr{H}(\Lambda x + a). \tag{3.5.12}$$

(By equating the coefficients of a^0 for infinitesimal transformations it can be checked that $\mathscr{H}(x)$ has a time-dependence consistent with Eq. (3.5.5).) Then S may be written as a sum of four-dimensional integrals

$$S = 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int d^4x_1 \cdots d^4x_n \, T\Big\{\mathscr{H}(x_1) \cdots \mathscr{H}(x_n)\Big\}. \tag{3.5.13}$$

Everything is now manifestly Lorentz invariant, except for the time-ordering of the operator product.

Now, the time-ordering of two spacetime points x_1, x_2 is Lorentz-invariant unless $x_1 - x_2$ is space-like, i.e., unless $(x_1 - x_2)^2 > 0$, so the time-ordering in Eq. (3.5.13) introduces no special Lorentz frame if

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

(though not only if) the $\mathscr{H}(x)$ all commute at space-like or light-like* separations:

$$[\mathscr{H}(x), \mathscr{H}(x')] = 0 \quad \text{for} \quad (x - x')^2 \geq 0. \tag{3.5.14}$$

We can use the results of Section 3.3 to give a formal non-perturbative proof that an interaction (3.5.11) satisfying Eqs. (3.5.12) and (3.5.14) does lead to an S-matrix with the correct Lorentz transformation properties. For an infinitesimal boost, Eq. (3.5.12) gives

$$-i[\mathbf{K}_0, \mathscr{H}(\mathbf{x}, t)] = t\nabla\mathscr{H}(\mathbf{x}, t) + \mathbf{x}\frac{\partial}{\partial t}\mathscr{H}(\mathbf{x}, t)\,, \tag{3.5.15}$$

so integrating over $\mathbf{x}$ and setting $t = 0$,

$$[\mathbf{K}_0, V] = \left[\mathbf{K}_0, \int d^3x\, \mathscr{H}(\mathbf{x}, 0)\right] = [H_0, \mathbf{W}]\,, \tag{3.5.16}$$

where

$$\mathbf{W} \equiv -\int d^3x\, \mathbf{x}\, \mathscr{H}(\mathbf{x}, 0). \tag{3.5.17}$$

If (as is usually the case) the matrix elements of $\mathscr{H}(\mathbf{x}, 0)$ between eigenstates of H_0 are smooth functions of the energy eigenvalues, then the same is true of V, as is necessary for the validity of scattering theory, and also true of $\mathbf{W}$, which is necessary in the proof of Lorentz invariance. The other condition for Lorentz invariance, the commutation relation (3.3.21), is also valid if and only if

$$0 = [\mathbf{W}, V] = \int d^3x \int d^3y\, \mathbf{x}\, [\mathscr{H}(\mathbf{x}, 0), \mathscr{H}(\mathbf{y}, 0)]\,. \tag{3.5.18}$$

This condition would follow from the 'causality' condition (3.5.14), but provides a somewhat less restrictive sufficient condition for Lorentz invariance of the S-matrix.

Theories of this class are not the only ones that are Lorentz invariant, but the most general Lorentz invariant theories are not very different. In particular, there is always a commutation condition something like (3.5.14) that needs to be satisfied. This condition has no counterpart for non-relativistic systems, for which time-ordering is always Galilean-invariant. *It is this condition that makes the combination of Lorentz invariance and quantum mechanics so restrictive.*

* * *

* We write the condition on x and x' here as $(x - x')^2 \geq 0$ instead of $(x - x')^2 > 0$, because as we shall see in Chapter 6, Lorentz invariance can be disturbed by troublesome singularities at $x = x'$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

The methods described so far in this section are useful when the interaction operator V is sufficiently small. There is also a modified version of this approximation, known as the *distorted-wave Born approximation*, that is useful when the interaction contains two terms

$$V = V_s + V_w \tag{3.5.19}$$

with V_w weak but V_s strong. We can define $\Psi_{s\alpha}{}^{\pm}$ as what the 'in' and 'out' states would be if V_s were the whole interaction

$$\Psi_{s\alpha}{}^{\pm} = \Phi_\alpha + (E_\alpha - H_0 \pm i\epsilon)^{-1} V_s \Psi_{s\alpha}{}^{\pm}. \tag{3.5.20}$$

We can then write (3.1.16) as

$$\begin{aligned} T_{\beta\alpha}{}^{+} &= (\Phi_\beta, V\Psi_\alpha{}^{+}) \\ &= \left(\left[\Psi_{s\beta}{}^{-} - (E_\beta - H_0 - i\epsilon)^{-1} V_s \Psi_{s\beta}{}^{-}\right], (V_s + V_w)\Psi_\alpha{}^{+}\right) \\ &= (\Psi_{s\beta}{}^{-}, V_w \Psi_\alpha{}^{+}) \\ &\quad + \left(\Psi_{s\beta}{}^{-}, \left[V_s - V_s(E_\beta - H_0 + i\epsilon)^{-1}(V_s + V_w)\right] \Psi_\alpha{}^{+}\right) \end{aligned}$$

and so

$$T_{\beta\alpha}{}^{+} = (\Psi_{s\beta}{}^{-}, V_w \Psi_\alpha{}^{+}) + (\Psi_{s\beta}{}^{-}, V_s \Phi_\alpha). \tag{3.5.21}$$

The second term on the right-hand side is just what $T_{\beta\alpha}{}^{+}$ would be in the presence of strong interactions alone

$$T_{s\ \beta\alpha}{}^{+} \equiv (\Phi_\beta, V_s \Psi_{s\alpha}{}^{+}) = (\Psi_{s\beta}{}^{-}, V_s \Phi_\alpha). \tag{3.5.22}$$

(For a proof of Eq. (3.5.22), just drop V_w everywhere in the derivation of Eq. (3.5.21).) Eq. (3.5.21) is most useful when this second term vanishes: that is, when the process $\alpha \to \beta$ cannot be produced by the strong interactions alone. (For example, in nuclear beta decay we need a weak nuclear force to turn neutrons into protons, even though we cannot ignore the presence of the strong nuclear force acting in the initial and final nuclear states.) For such processes, the matrix element (3.5.22) vanishes, so Eq. (3.5.21) reads

$$T_{\beta\alpha}{}^{+} = (\Psi_{s\beta}{}^{-},\ V_w \Psi_\alpha{}^{+}). \tag{3.5.23}$$

So far, this is all exact. However, this way of rewriting the T-matrix becomes worthwhile when V_w is so weak that we may neglect its effect on the state $\Psi_\alpha{}^{+}$ in Eq. (3.5.23), and hence replace $\Psi_\alpha{}^{+}$ with the state $\Psi_{s\alpha}{}^{+}$, which takes account only of the strong interaction V_s. In this approximation, Eq. (3.5.23) becomes

$$T_{\beta\alpha}{}^{+} \simeq (\Psi_{s\beta}{}^{-},\ V_w \Psi_{s\alpha}{}^{+}). \tag{3.5.24}$$

This is valid to first order in V_w, but to all orders in V_s. This approximation is ubiquitous in physics; for instance, the S-matrix element for nuclear

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

beta or gamma decay is calculated using Eq. (3.5.24) with V_s the strong nuclear interaction and V_w respectively either the weak nuclear interaction or the electromagnetic interaction, and with $\Psi_{s\beta}{}^{-}$ and $\Psi_{s\alpha}{}^{+}$ the final and initial nuclear states.

3.6 Implications of Unitarity

The unitarity of the S-matrix imposes an interesting and useful condition relating the amplitude $M_{\alpha\alpha}$ for forward scattering in an arbitrary multi-particle state α to the total rate for all reactions in that state. Recall that in the general case, where the state β may or may not be the same as the state α, the S-matrix may be written as in (3.3.2):

$$S_{\beta\alpha} = \delta(\beta - \alpha) - 2\pi i \delta^4(p_\beta - p_\alpha) M_{\beta\alpha}.$$

The unitarity condition then gives

$$\delta(\gamma - \alpha) = \int d\beta\, S^*_{\beta\gamma} S_{\beta\alpha} = \delta(\gamma - \alpha) - 2\pi i \delta^4(p_\gamma - p_\alpha) M_{\gamma\alpha}$$
$$+ 2\pi i \delta^4(p_\gamma - p_\alpha) M^*_{\alpha\gamma} + 4\pi^2 \int d\beta\, \delta^4(p_\beta - p_\gamma) \delta^4(p_\beta - p_\alpha) M^*_{\beta\gamma} M_{\beta\alpha}\,.$$

Cancelling the term $\delta(\gamma - \alpha)$ and a factor $2\pi\delta^4(p_\gamma - p_\alpha)$, we find that for $p_\gamma = p_\alpha$

$$0 = -iM_{\gamma\alpha} + iM^*_{\alpha\gamma} + 2\pi \int d\beta\, \delta^4(p_\beta - p_\alpha) M^*_{\beta\gamma} M_{\beta\alpha}\,. \tag{3.6.1}$$

This is most useful in the special case $\alpha = \gamma$, where it reads

$$\text{Im}\, M_{\alpha\alpha} = -\pi \int d\beta\, \delta^4(p_\beta - p_\alpha) |M_{\beta\alpha}|^2. \tag{3.6.2}$$

Using Eq. (3.4.11), this can be expressed as a formula for the total rate for all reactions produced by an initial state α in a volume V

$$\Gamma_\alpha \equiv \int d\beta\, \frac{d\Gamma(\alpha \to \beta)}{d\beta}$$
$$= (2\pi)^{3N_\alpha - 2} V^{1-N_\alpha} \int d\beta\, \delta^4(p_\beta - p_\alpha) |M_{\beta\alpha}|^2$$
$$= -\frac{1}{\pi} (2\pi)^{3N_\alpha - 2} V^{1-N_\alpha}\, \text{Im}\, M_{\alpha\alpha}. \tag{3.6.3}$$

In particular, where α is a two-particle state, this can be written

$$\text{Im}\, M_{\alpha\alpha} = -u_\alpha \sigma_\alpha / 16\pi^3\,, \tag{3.6.4}$$

where u_α is the relative velocity (3.4.17) in state α, and σ_α is the *total*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

cross-section in this state, given by (3.4.15) as

$$\sigma_\alpha \equiv \int d\beta \; d\sigma(\alpha \to \beta)/d\beta = (2\pi)^4 u_\alpha^{-1} \int d\beta \, |M_{\beta\alpha}|^2 \delta^4(p_\beta - p_\alpha). \quad (3.6.5)$$

This is usually expressed in a slightly different way, in terms of a *scattering amplitude* $f(\alpha \to \beta)$. Eq. (3.4.30) shows that the differential cross-section for *two-body* scattering in the center-of-mass frame is

$$\frac{d\sigma(\alpha \to \beta)}{d\Omega} = \frac{(2\pi)^4 k' E_1' E_2' E_1 E_2}{k E^2} |M_{\beta\alpha}|^2 , \quad (3.6.6)$$

where k' and k are the magnitudes of the momenta in the final and initial states. We therefore define the scattering amplitude as*

$$f(\alpha \to \beta) \equiv -\frac{4\pi^2}{E} \sqrt{\frac{k' E_1' E_2' E_1 E_2}{k}} \, M_{\beta\alpha} , \quad (3.6.7)$$

so that the differential cross-section is simply

$$\frac{d\sigma(\alpha \to \beta)}{d\Omega} = |f(\alpha \to \beta)|^2. \quad (3.6.8)$$

In particular, for *elastic* two-body scattering, we have

$$f(\alpha \to \beta) \equiv -\frac{4\pi^2 E_1 E_2}{E} M_{\beta\alpha} . \quad (3.6.9)$$

Using (3.4.18) for the relative velocity u_α, the unitarity prediction (3.6.3) now reads

$$\operatorname{Im} f(\alpha \to \alpha) = \frac{k}{4\pi} \sigma_\alpha. \quad (3.6.10)$$

This form of the unitarity condition (3.6.3) is known as the *optical theorem*.[22]

There is a pretty consequence of the optical theorem that tells much about the pattern of scattering at high energy. The scattering amplitude f may be expected to be a smooth function of angle, so there must be some solid angle $\Delta\Omega$ within which $|f|^2$ has nearly the same value (say, within a factor 2) as in the forward direction. The total cross-section is then bounded by

$$\sigma_\alpha \geq \int |f|^2 d\Omega \geq \frac{1}{2} |f(\alpha \to \alpha)|^2 \Delta\Omega \geq \frac{1}{2} |\operatorname{Im} f(\alpha \to \alpha)|^2 \Delta\Omega .$$

* The phase of f is conventional, and is motivated by the wave mechanical interpretation[21] of f as the coefficient of the outgoing wave in the solution of the time-independent Schrödinger equation. The normalization of f used here is slightly unconventional for inelastic scattering; usually f is defined so that a ratio of final and initial velocities appears in the formula for the differential cross-section.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Using Eq. (3.6.10) then yields an upper bound on $\Delta\Omega$

$$\Delta\Omega \leq 32\pi^2 \Big/ k^2\sigma_\alpha \ . \tag{3.6.11}$$

As we shall see in the next section, total cross-sections are usually expected to approach constants or grow slowly at high energy, so Eq. (3.6.11) shows that the solid angle around the forward direction within which the differential cross-section is roughly constant shrinks at least as fast as $1/k^2$ for $k \to \infty$. This increasingly narrow peak in the forward direction at high energies is known as a *diffraction peak.*

Returning now to the general case of reactions involving arbitrary numbers of particles, we can use Eq. (3.6.2) together with CPT invariance to say something about the relations for total interaction rates of particles and antiparticles. Because CPT is antiunitary, its conservation does not in general imply any simple relation between a process $\alpha \to \beta$ and the process with particles replaced with their antiparticles. Instead, it provides a relation between a process and the *inverse* process involving antiparticles: we can use the same arguments that allowed us to deduce (3.3.46) from time-reversal invariance to show that CPT invariance requires the S-matrix to satisfy the condition

$$S_{\beta,\alpha} = S_{\mathscr{CPT}\alpha\,,\,\mathscr{CPT}\beta} \ , \tag{3.6.12}$$

where $\mathscr{CPT}$ indicates that we must reverse all spin z-components, change all particles into their corresponding antiparticles, and multiply the matrix element by various phase factors for the particles in the initial state and by their complex conjugates for the particles in the final state. Since CPT invariance also requires particles to have the same masses as their corresponding antiparticles, the same relation holds for the coefficient of $\delta^4(p_\alpha - p_\beta)$ in $S_{\beta\alpha}$:

$$M_{\beta\,,\,\alpha} = M_{\mathscr{CPT}\alpha\,,\,\mathscr{CPT}\beta} \ . \tag{3.6.13}$$

In particular, when the initial and final states are the same the phase factors all cancel, and Eq. (3.6.13) says that

$$\begin{aligned} &M_{p_1\sigma_1 n_1\,;\,p_2\sigma_2 n_2\,;\,\cdots\,,\,p_1\sigma_1 n_1\,;\,p_2\sigma_2 n_2\,;\,\cdots} \\ &\quad = M_{p_1-\sigma_1 n_1^c\,;\,p_2-\sigma_2 n_2^c\,;\,\cdots\,,\,p_1-\sigma_1 n_1^c\,;\,p_2-\sigma_2 n_2^c\,;\,\cdots} \ , \end{aligned} \tag{3.6.14}$$

where a superscript c on n indicates the antiparticle of n. The generalized optical theorem (3.6.2) then tells us that *the total reaction rate from an initial state consisting of some set of particles is the same as for an initial state consisting of the corresponding antiparticles with spins reversed*:

$$\Gamma_{p_1\sigma_1 n_1\,;\,p_2\sigma_2 n_2\,;\,\cdots} = \Gamma_{p_1-\sigma_1 n_1^c\,;\,p_2-\sigma_2 n_2^c\,;\,\cdots} \ . \tag{3.6.15}$$

In particular, applying this to one-particle states, we see that the decay rate of any particle equals the decay rate of the antiparticle with reversed

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

spin. Rotational invariance does not allow particle decay rates to depend on the spin z-component of the decaying particle, so a special case of the general result (3.6.15) is that unstable particles and their corresponding antiparticles have precisely the same lifetimes.

* * *

The same argument that led from the unitarity condition $S^\dagger S = 1$ to our result (3.6.2) also allows us to use the other unitarity relation $SS^\dagger = 1$ to derive the result

$$\text{Im}\, M_{\alpha\alpha} = -\pi \int d\beta\, \delta^4(p_\beta - p_\alpha)|M_{\alpha\beta}|^2. \tag{3.6.16}$$

Putting this together with Eq. (3.6.2) then yields the reciprocity relation

$$\int d\beta\, \delta^4(p_\beta - p_\alpha)|M_{\beta\alpha}|^2 = \int d\beta\, \delta^4(p_\beta - p_\alpha)|M_{\alpha\beta}|^2 \tag{3.6.17}$$

or in other words

$$\int d\beta\, c_\alpha \frac{d\Gamma(\alpha \to \beta)}{d\beta} = \int d\beta\, c_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha}\,, \tag{3.6.18}$$

where $c_\alpha \equiv [V/(2\pi)^3]^{N_\alpha}$. This result can be used in deriving some of the most important results of kinetic theory.[23] If $P_\alpha d\alpha$ is the probability of finding the system in a volume $d\alpha$ of the space of multi-particle states Φ_α, then the rate of decrease in P_α due to transitions to all other states is $P_\alpha \int d\beta\, d\Gamma(\alpha \to \beta)/d\beta$, while the rate of increase of P_α due to transitions from all other states is $\int d\beta\, P_\beta\, d\Gamma(\beta \to \alpha)/d\alpha$; the rate of change of P_α is then

$$\frac{dP_\alpha}{dt} = \int d\beta\, P_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha} - P_\alpha \int d\beta \frac{d\Gamma(\alpha \to \beta)}{d\beta}\,. \tag{3.6.19}$$

It follows immediately that $\int P_\alpha d\alpha$ is time-independent. (Just interchange the labelling of the integration variables in the integral of the second term in Eq. (3.6.19).) On the other hand, the rate of change of the entropy $-\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha)$ is

$$-\frac{d}{dt}\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha) = -\int d\alpha \int d\beta \left(\ln(P_\alpha/c_\alpha) + 1\right) \times \left[P_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha} - P_\alpha \frac{d\Gamma(\alpha \to \beta)}{d\beta}\right].$$

Interchanging the labelling of the integration variables in the second term, this may be written

$$-\frac{d}{dt}\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha) = \int d\alpha \int d\beta\, P_\beta \ln\left(\frac{P_\beta c_\alpha}{P_\alpha c_\beta}\right) \frac{d\Gamma(\beta \to \alpha)}{d\alpha}\,.$$

Now, for any positive quantities x and y, the function $y \ln(y/x)$ satisfies

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

the inequality**

$$y \ln\left(\frac{y}{x}\right) \geq y - x .$$

The rate of change of the entropy is thus bounded by

$$-\frac{d}{dt}\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha) \geq \int d\alpha \int d\beta \left[\frac{P_\beta}{c_\beta} - \frac{P_\alpha}{c_\alpha}\right] c_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha}$$

or interchanging variables of integration in the second term

$$-\frac{d}{dt}\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha) \geq \int d\alpha \int d\beta\, \frac{P_\beta}{c_\beta}\left[c_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha} - c_\alpha \frac{d\Gamma(\alpha \to \beta)}{d\beta}\right] .$$

But the unitarity relation (3.6.18) (with α and β interchanged) tells us that the integral over α on the right-hand side of this inequality vanishes, so we may conclude that the entropy always increases:

$$-\frac{d}{dt}\int d\alpha\, P_\alpha \ln(P_\alpha/c_\alpha) \geq 0 . \tag{3.6.20}$$

This is the 'Boltzmann H-theorem.' This theorem is often derived in statistical mechanics textbooks either by using the Born approximation, for which $|M_{\beta\alpha}|^2$ is symmetric in α and β so that $c_\beta d\Gamma(\beta \to \alpha)/d\alpha = c_\alpha d\Gamma(\alpha \to \beta)/d\beta$, or by assuming time-reversal invariance, which would tell us that $|M_{\beta\alpha}|^2$ is unchanged if we interchange α and β and also reverse all momenta and spins. Of course, neither the Born approximation nor time-reversal invariance are exact, so it is a good thing that the unitarity result (3.6.18) is all we need in order to derive the H-theorem.

The increase of the entropy stops when the probability P_α becomes a function only of conserved quantities such as the total energy and charge, times the factor c_α. In this case the conservation laws require $d\Gamma(\beta \to \alpha)/d\alpha$ to vanish unless $P_\alpha/c_\alpha = P_\beta/c_\beta$, so we can replace P_β with $P_\alpha c_\beta/c_\alpha$ in the first term of Eq. (3.6.19). Using Eq. (3.6.18) again then shows that in this case, P_α is time-independent. Here again, we need only the unitarity relation (3.6.18), not the Born approximation or time-reversal invariance.

3.7 Partial-Wave Expansions*

It is often convenient to work with the S-matrix in a basis of free-particle states in which all variables are discrete, except for the total momentum

** The difference between the left- and right-hand sides approaches the positive quantity $(x-y)^2/2y$ for $x \to y$, and has a derivative with respect to x that is positive- or negative-definite for all $x > y$ or $x < y$, respectively.

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

and energy. This is possible because the components of the momenta $\mathbf{p}_1, \cdots, \mathbf{p}_n$ in an n-particle state of definite total momentum $\mathbf{p}$ and total energy E form a $(3n-4)$-dimensional *compact* space; for instance, for $n = 2$ particles in the center-of-mass frame with $\mathbf{p} = 0$, this space is a two-dimensional spherical surface. Any function on such a compact space may be expanded in a series of generalized 'partial waves', such as the spherical harmonics that are commonly used in representing functions on the two-sphere. We may thus define a basis for these n-particle states that apart from the continuous variables $\mathbf{p}$ and E is discrete: we label the free-particle states in such a basis as $\Phi_{E\,\mathbf{p}\,N}$, with the index N incorporating all spin and species labels as well as whatever indices are used to label the generalized partial waves. These states may conveniently be chosen to be normalized so that their scalar products are:

$$(\Phi_{E'\,\mathbf{p}'\,N'}, \Phi_{E\,\mathbf{p}\,N}) = \delta(E'-E)\delta^3(\mathbf{p}'-\mathbf{p})\delta_{N',N}. \tag{3.7.1}$$

The S-operator then has matrix elements in this basis of the form

$$(\Phi_{E'\,\mathbf{p}'N'},\ S\Phi_{E\,\mathbf{p}\,N}) = \delta(E'-E)\delta^3(\mathbf{p}'-\mathbf{p})S_{N',N}(E,\mathbf{p}), \tag{3.7.2}$$

where $S_{N',N}$ is a unitary matrix. Similarly, the T-operator, whose free-particle matrix elements $(\Phi_\beta, T\Phi_\alpha)$ are defined to be the quantities $T_{\beta\alpha}{}^+$ defined by Eq. (3.1.18), may be expressed in our new basis (in accordance with Eq. (3.4.12)) as

$$(\Phi_{E\,\mathbf{p}'\,N'},\ T\Phi_{E\,\mathbf{p}\,N}) = \delta^3(\mathbf{p}'-\mathbf{p})M_{N',N}(E,\mathbf{p}) \tag{3.7.3}$$

and the relation (3.2.7) is now an ordinary matrix equation:

$$S_{N',N}(E,\mathbf{p}) = \delta_{N',N} - 2i\pi M_{N',N}(E,\mathbf{p}). \tag{3.7.4}$$

We shall use this general formalism in the following section; for the present we will concentrate on reactions in which the initial state involves just two particles.

For example, consider a state consisting of two non-identical particles of species n_1, n_2 with non-zero masses M_1, M_2 and arbitrary spins s_1, s_2. In this case, the states may be labelled by their total momentum $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$, the energy E, the species labels n_1, n_2, the spin z-components σ_1, σ_2, and a pair of integers ℓ, m (with $|m| \leq \ell$) that specify the dependence of the state on the directions of, say, $\mathbf{p}_1$. Alternatively, we can form a convenient discrete basis by using Clebsch–Gordan coefficients[9] to combine the two spins to give a total spin s with z-component μ, and then using Clebsch–Gordan coefficients again to combine this with the orbital angular momentum ℓ with three-component m to form a total angular momentum j with three-component σ. This gives a basis of states $\Phi_{E\,\mathbf{p}\,j\sigma\ell sn}$ (with n a 'channel index' labelling the two particle species' n_1, n_2), defined by their scalar products with the states of definite individual

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

momenta and spin three-components:

$$(\Phi_{\mathbf{p}_1\sigma_1\mathbf{p}_2\sigma_2 n'}\ ,\ \Phi_{E\,\mathbf{p}\,j\sigma\ell s n}) \equiv (|\mathbf{p}_1|E_1E_2/E)^{-1/2}\delta^3(\mathbf{p}-\mathbf{p}_1-\mathbf{p}_2)$$
$$\times\,\delta\left(E-\sqrt{\mathbf{p}_1^2+M_1^2}-\sqrt{\mathbf{p}_2^2+M_2^2}\right)\delta_{n',n}$$
$$\times\sum_{m,\mu}C_{s_1s_2}(s,\mu;\sigma_1,\sigma_2)\,C_{\ell s}(j,\sigma;m,\mu)Y_\ell^m(\hat{\mathbf{p}}_1)\,. \tag{3.7.5}$$

where Y_ℓ^m are the usual spherical harmonics.[24] The factor $(|\mathbf{p}_1|E_1E_2/E)^{-1/2}$ is inserted so that in the center-of-mass frame these states will be properly normalized:

$$(\Phi_{E'\mathbf{p}'j'\sigma'\ell's'n'},\ \Phi_{E0j\sigma\ell sn}) = \delta^3(\mathbf{p}')\delta(E'-E)\delta_{j',j}\delta_{\sigma',\sigma}\delta_{\ell',\ell}\delta_{s',s}\delta_{n',n}\,. \tag{3.7.6}$$

For identical particles to avoid double counting we must integrate only over half of the two-particle momentum space, so an extra factor $\sqrt{2}$ should appear in the scalar product (3.7.6).

In the center-of-mass frame the matrix elements of any momentum-conserving and rotationally-invariant operator O must take the form:

$$(\Phi_{E\,\mathbf{p}'j'\sigma'\ell's'n'}\ ,\ O\ \Phi_{E0j\sigma\ell sn}) = \delta^3(\mathbf{p}')O^j_{\ell's'n',\ell sn}(E)\delta_{jj'}\delta_{\sigma\sigma'}. \tag{3.7.7}$$

(The fact that this is diagonal in j and σ follows from the commutation of O with $\mathbf{J}^2$ and J_3, and the further fact that the coefficient of $\delta_{\sigma\sigma'}$ is independent of σ follows from the commutation of O with $J_1 \pm iJ_2$. This is a special case of a general result known as the Wigner–Eckart theorem.[25]) Applying this to the operator M whose matrix elements are the quantities $M_{\beta\alpha}$, it follows that the scattering amplitude (3.6.7) in the center-of-mass system takes the form

$$f(\mathbf{k}\,\sigma_1,-\mathbf{k}\,\sigma_2,n\ \to\ \mathbf{k}'\,\sigma_1',-\mathbf{k}'\,\sigma_2',n')$$
$$\equiv -4\pi^2\sqrt{\frac{k'E_1'E_2'E_1E_2}{E^2k}}\,M_{\mathbf{k}'\sigma_1'-\mathbf{k}'\sigma_2'n',\mathbf{k}\sigma_1-\mathbf{k}\sigma_2 n}$$
$$= -\frac{4\pi^2}{k}\sum_{j\sigma\ell'm's'\mu'\ell ms\mu} C_{s_1s_2}(s,\mu;\sigma_1,\sigma_2)C_{\ell s}(j,\sigma;m,\mu)$$
$$\times\,C_{s_1's_2'}(s',\mu';\sigma_1',\sigma_2')C_{\ell's'}(j,\sigma;m',\mu')$$
$$\times\,Y_{\ell'}^{m'}(\hat{\mathbf{k}}')Y_\ell^{m*}(\hat{\mathbf{k}})M^j_{\ell's'n',\ell sn}(E). \tag{3.7.8}$$

The differential scattering cross-section is $|f|^2$. We will take the direction of the initial momentum $\mathbf{k}$ to be along the three-direction, in which case

$$Y_\ell^m(\hat{\mathbf{k}}) = \delta_{m0}\sqrt{\frac{2\ell+1}{4\pi}}. \tag{3.7.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Integrating $|f|^2$ over the direction of the final momentum $\mathbf{k}'$ and respectively summing and averaging over final and initial spin three-components gives the total cross-section** for transitions from channel n to n':

$$\sigma(n \to n'; E) = \frac{\pi}{k^2(2s_1+1)(2s_2+1)} \sum_{j\ell s\ell' s'} (2j+1) \times \left|\delta_{\ell'\ell}\delta_{s's}\delta_{n'n} - S^j_{\ell' s' n', \ell s n}(E)\right|^2 . \tag{3.7.10}$$

Summing (3.7.10) over all two-body channels gives the total cross-section for all elastic or inelastic two-body reactions:

$$\sum_{n'} \sigma(n \to n'; E) = \frac{\pi}{k^2(2s_1+1)(2s_2+1)} \sum_{j\ell s} (2j+1) \times \left[(1 - S^j(E))^\dagger (1 - S^j(E))\right]_{\ell s n, \ell s n} . \tag{3.7.11}$$

For comparison, Eqs. (3.7.8), (3.7.9), (3.7.4), and the Clebsch–Gordan sum rules** give the spin-averaged forward scattering amplitude as

$$f(n; E) = \frac{i}{2k(2s_1+1)(2s_2+1)} \sum_{j\ell s} (2j+1)[1 - S^j_{\ell s n, \ell s n}]$$

The optical theorem (3.6.10) then gives the total cross-section as

$$\sigma_{\text{total}}(n; E) = \frac{2\pi}{k^2(2s_1+1)(2s_2+1)} \sum_{j\ell s} (2j+1) \operatorname{Re}\left[1 - S^j(E)\right]_{\ell s n, \ell s n} . \tag{3.7.12}$$

If only two-body channels can be reached from the channel n at energy E, then the matrix $S^j(E)$ (or at least some submatrix that includes channel n) is unitary, and thus

$$\left[(1 - S^j(E))^\dagger (1 - S^j(E))\right]_{\ell s n, \ell s n} = 2 \operatorname{Re}\left[1 - S^j(E)\right]_{\ell s n, \ell s n} , \tag{3.7.13}$$

so (3.7.12) and (3.7.11) are equal. On the other hand, if channels involving three or more particles are open, then the difference of (3.7.12) and (3.7.11)

** In deriving this result, we use standard sum rules[9] for the Clebsch–Gordan coefficients: first

$$\sum_{\sigma_1, \sigma_2} C_{s_1, s_2}(s, \mu; \sigma_1, \sigma_2) C_{s_1, s_2}(\bar{s}, \bar{\mu}; \sigma_1, \sigma_2) = \delta_{s\bar{s}}\delta_{\mu\bar{\mu}}$$

and the same with primes; then

$$\sum_{m\tilde{\sigma}} C_{\ell s}(j, \sigma; m, \tilde{\sigma}) C_{\ell s}(\bar{j}, \bar{\sigma}; m, \tilde{\sigma}) = \delta_{j\bar{j}}\delta_{\sigma\bar{\sigma}}$$

and finally

$$\sum_{\sigma\mu} C_{\ell s}(j, \sigma; 0, \mu) C_{\bar{\ell} s}(j, \sigma; 0, \mu) = \frac{2j+1}{2\ell+1}\delta_{\bar{\ell}\ell} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

gives the total cross-section for producing extra particles:

$$\sigma_{\text{production}}(n;E) = \frac{\pi}{k^2(2s_1+1)(2s_2+1)} \sum_{j\ell s} (2j+1) \times \left[1 - S^j(E)^\dagger S^j(E)\right]_{\ell s n, \ell s n}, \tag{3.7.14}$$

and this must be positive.

The partial wave expansion is particularly useful when applied to processes where the relevant part of the S-matrix is diagonal. This is the case, for instance, if the initial channel n contains just two spinless particles, and no other channels are open at this energy, as in $\pi^+ - \pi^+$ or $\pi^+ - \pi^0$ scattering at energies below the threshold for producing extra pions (provided one ignores weak and electromagnetic interactions). For a pair of spinless particles we have $j = \ell$, and angular momentum conservation keeps the S-matrix diagonal. It is also possible for the S-matrix to be diagonal in certain processes involving particles with spin; for instance in pion–nucleon scattering we can have $j = \ell + \frac{1}{2}$ or $j = \ell - \frac{1}{2}$, but for a given j these two states have opposite parity, so they cannot be connected by non-zero S-matrix elements. In any case, if for some n and E the S-matrix elements $S_{N',j\ell sn}(E,0)$ all vanish unless N' is the two-body state j,ℓ,s,n, then unitarity requires that

$$S^j_{\ell' s' n', \ell s n}(E) = \exp\left[2i\delta_{j\ell sn}(E)\right] \delta_{\ell'\ell}\delta_{s's}\delta_{n'n}, \tag{3.7.15}$$

where $\delta_{j\ell sn}(E)$ is a real phase, commonly known as the *phase shift*. This formula is also often used where the two-body part of the S-matrix is diagonal but channels containing three or more particles are also open; in such cases the phase shift must have a positive imaginary part, to keep (3.7.14) positive. For real phase shifts, the elastic and total cross-sections are then given by Eq. (3.7.10) or Eq. (3.7.12) as:

$$\begin{aligned}\sigma(n \to n;E) &= \sigma_{\text{total}}(n;E) \\ &= \frac{4\pi}{k^2(2s_1+1)(2s_2+1)} \sum_{j\ell s} (2j+1) \sin^2 \delta_{j\ell sn}(E).\end{aligned} \tag{3.7.16}$$

This familiar result is usually derived in non-relativistic quantum mechanics by studying the coordinate space wave function for a particle in a potential. The derivation given here is offered both to show that the partial wave expansion applies for elastic scattering even at relativistic velocities, and also to emphasize that it depends on no particular dynamical assumptions, only on unitarity and invariance principles.

It is also often useful to introduce phase shifts in dealing with problems where several channels are open, forming a few irreducible representations of some internal symmetry group. The classic example of such an internal symmetry is isotopic spin symmetry, for which the channel index n includes

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

a specification of the isospins T_1, T_2 of the two particles together with their three-components t_1, t_2; the states in channel n may be expressed as linear combinations of the tth components of irreducible representations T, with coefficients given by the familiar Clebsch–Gordan coefficients $C_{T_1T_2}(T,t : t_1,t_2)$. Suppose that for the channels and energy of interest the S-matrix is diagonal in ℓ and s as well as j, T, and t. Unitarity and isospin symmetry then allow us to write the S-matrix as

$$S^j_{\ell' s' T' t',\ell s T t} = \exp[2i\delta_{j\ell s T}(E)]\delta_{\ell'\ell}\delta_{s's}\delta_{T'T}\delta_{t't}\,, \tag{3.7.17}$$

with $\delta_{j\ell s T}(E)$ a real phase shift, t-independent according to the Wigner–Eckart theorem. The partial cross-sections can again be calculated from Eq. (3.7.10), and the total cross-section is given by Eq. (3.7.12) as

$$\sigma_{\text{total}}(t_1,t_2;E) = \frac{4\pi}{k^2(2s_1+1)(2s_2+1)} \times \sum_{j\ell sTt}(2j+1)C_{T_1,T_2}(T,t;t_1,t_2)^2 \sin^2\delta_{j\ell sT}(E). \tag{3.7.18}$$

For instance, in pion–pion scattering we have phase shifts $\delta_{\ell\ell 0T}(E)$ with $T = 0$ or $T = 2$ for each even ℓ and $T = 1$ for each odd ℓ, while for pion–nucleon scattering we have phase shifts $\delta_{j\,j\pm\frac{1}{2}\,\frac{1}{2}\,T}$ with $T = \frac{1}{2}$ or $T = \frac{3}{2}$.

We can gain some useful insight about the threshold behavior of the scattering amplitudes and phase shifts from considerations of analyticity that are nearly independent of any dynamical assumptions. Unless there are special circumstances that would produce singularities in momentum space, we would expect the matrix element $M_{\mathbf{k}'\sigma_1'-\mathbf{k}'\sigma_2'n',\,\mathbf{k}\sigma_1-\mathbf{k}\sigma_2 n}$ to be an analytic function† of the three-momenta $\mathbf{k}$ and $\mathbf{k}'$ near $k = 0$ or $k' = 0$ or (for elastic scattering) $k = k' = 0$. Turning to the partial wave expansion (3.7.8) for M, we note that $k^\ell Y_\ell^m(\hat{\mathbf{k}})$ is a simple polynomial function of the three-vector $\mathbf{k}$, so in order for $M_{\mathbf{k}'\sigma_1'-\mathbf{k}'\sigma_2'n',\,\mathbf{k}\sigma_1-\mathbf{k}\sigma_2 n}$ to be an analytic function of the three-momenta $\mathbf{k}$ and $\mathbf{k}'$ near $k = 0$ or $k' = 0$, the coefficients $M^j_{\ell' s' n',\ell s n}$ or equivalently $\delta_{\ell'\ell}\delta_{s's}\delta_{n'n} - S^j_{\ell' s' n',\ell s n}$ must go as $k^{\ell+\frac{1}{2}}k'^{\ell'+\frac{1}{2}}$ when k and/or k' go to zero. Hence for small k and/or k', it is only the lowest partial waves in the initial and/or final state that contribute appreciably to the scattering amplitude. We have three possible cases:

† For instance, in the Born approximation (3.2.8), M is proportional to the Fourier transform of the coordinate space matrix elements of the interaction, and hence is analytic at zero momenta as long as these matrix elements fall off sufficiently rapidly at large separations. The chief exception is for scattering involving long-distance forces, such as the Coulomb force.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Exothermic reactions

Here k' approaches a finite value as $k \to 0$, and in this limit $\delta_{\ell'\ell}\delta_{s's}\delta_{n'n} - S^j_{\ell' s' n', \ell s n}$ goes as $k^{\ell+\frac{1}{2}}$. The cross-section (3.7.11) in this case goes as $k^{2\ell-1}$, where ℓ is here the *lowest* orbital angular momentum that can lead to the reaction. In the most usual case $\ell = 0$, so the reaction cross-section goes as $1/k$. (This is the case, for instance, in the absorption of slow neutrons by complex nuclei, or for the annihilation of electron-positron pairs into photons at low energy, aside from the higher-order effects of Coulomb forces.) The reaction rate is the cross-section times the flux, which goes like k, so the rate for an exothermic reaction behaves like a constant for $k \to 0$. However, it is the cross-section rather than the reaction rate that determines the probability of absorption when a beam crosses a given thickness of target material, and the factor $1/k$ makes this probability very high for slow neutrons in an absorbing material like boron.

Endothermic Reactions

Here the reaction is forbidden until k reaches a finite threshold value, where $k' = 0$. Just above this threshold $\delta_{\ell'\ell}\delta_{s's}\delta_{n'n} - S^j_{\ell' s' n', \ell s n}$ goes as $(k')^{\ell'+\frac{1}{2}}$. The cross-section (3.7.11) in this case goes as $(k')^{2\ell'+1}$, where ℓ' is here the *lowest* orbital angular momentum that can be produced at threshold. In the most usual case $\ell' = 0$, so the reaction cross-section rises above threshold like k', and hence like $\sqrt{E - E_{\text{threshold}}}$. (This is the case, for instance, in the associated production of strange particles, or the production of electron–positron pairs in the scattering of photons.)

Elastic Reactions

Here $k = k'$, so k and k' go to zero together. (This is the case where $n' = n$, or where n' consists of particles in the same isotopic spin multiplets as are those in n.) In elastic scattering the partial waves with $\ell = \ell' = 0$ are always present, so in the limit $k \to 0$ the scattering amplitude (3.7.8) becomes a constant:

$$f(\mathbf{k}, \sigma_1, -\mathbf{k}, \sigma_2, n \to \mathbf{k}', \sigma_1', -\mathbf{k}', \sigma_2', n') \to \sum_{s\sigma} C_{s_1 s_2}(s, \sigma; \sigma_1, \sigma_2) C_{s_1' s_2'}(s, \sigma; \sigma_1', \sigma_2')\, a_s(n \to n')\,, \tag{3.7.19}$$

where a is a constant, known as the *scattering length*, defined by the limit

$$S^s_{0sn',0sn} \to \delta_{n',n} + 2ika_s(n \to n') \tag{3.7.20}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

for $k = k' \to 0$. Summing $4\pi|f|^2$ over final spins and averaging over initial spins gives the total cross-section for the transition $n \to n'$ at $k = k' = 0$:

$$\sigma(n \to n'; k = 0) = \frac{4\pi}{(2s_1 + 1)(2s_2 + 1)} \sum_s (2s + 1) a_s^2(n \to n') . \quad (3.7.21)$$

The classic instance of the use of this formula is in neutron–proton scattering, where there are two scattering lengths, with the spin singlet length a_0 considerably larger than the spin triplet length a_1.

The partial wave expansion can also be used to make a crude guess about the behavior of cross-sections at high energy. With decreasing wavelengths, we may expect scattering to be described more or less classically: a particle of momentum k and orbital angular momentum ℓ would have an impact parameter ℓ/k, and will therefore strike a disk of radius R if $\ell \leq kR$. This can be interpreted as a statement about S-matrix elements:

$$S^j_{\ell' s' n, \ell s n} \to \begin{cases} 0 & \ell \ll kR_n \\ 1 & \ell \gg kR_n \end{cases} , \quad (3.7.22)$$

where R_n is some sort of interaction radius for channel n. For a given $\ell \gg s$, there are $2s + 1$ values of j, all close enough to ℓ to approximate $2j + 1 \simeq 2\ell + 1$, so the sum over j and s in Eq. (3.7.12) merely gives a factor of order

$$\sum_{js} (2j + 1) = (2\ell + 1) \sum_s (2s + 1) = (2\ell + 1)(2s_1 + 1)(2s_2 + 1).$$

The total cross-section is then given for $k \gg 1/R_n$ by Eq. (3.7.12) as

$$\sigma_{\text{total}}(n; E) \to \frac{2\pi}{k^2} \sum_{\ell \leq kR_n} (2\ell + 1) \to 2\pi R_n^2 . \quad (3.7.23)$$

In exactly the same way, Eq. (3.7.10) gives the elastic scattering cross-section

$$\sigma(n \to n; E) \to \pi R_n^2 . \quad (3.7.24)$$

The difference between Eqs. (3.7.23) and (3.7.24) gives an inelastic cross-section πR_n^2, which is what we would expect for collisions with an opaque disk of radius R_n. (The somewhat surprising elastic scattering cross-section πR_n^2 may be attributed to diffraction by the disk.) On the other hand, if we assume along with Eq. (3.7.22) that $S^j_{\ell' s' n, \ell s n}$ is complex only for impact parameters ℓ/k within a small range of width $\Delta_n \ll R_n$ around $\ell/k = R_n$, then using the inequality $|\text{Im}(1 - S^j_{\ell' s' n, \ell s n})| \leq 2$, the same analysis gives a bound on the real part of the forward scattering amplitude

$$|\text{Re} f(n; E)| \leq 2k R_n \Delta_n \ll |\text{Im} f(n; E)| . \quad (3.7.25)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

The smallness of the real part of the forward scattering amplitude at high energy is confirmed by experiment.

So far we have not said anything about whether the interaction radius R_n itself may depend on energy. As a very crude guess, we may take R_n as the distance at which the factor $\exp(-\mu r)$ in the Yukawa potential (1.2.74) takes a value proportional to some unknown power of E, in which case R_n goes as $\log E$ for $E \to \infty$, and the cross-sections go as $(\log E)^2$. As it happens, it has been rigorously shown[26] on the basis of very general assumptions that the total cross-section can grow no faster than $(\log E)^2$ for $E \to \infty$, and in fact the observed proton–proton total cross-section rises something like $(\log E)^2$ at high energies, so this rough picture of high-energy scattering does seem to have some correspondence with reality.

3.8 Resonances*

It often happens that the particles participating in a multi-particle collision can form an intermediate state consisting of a *single* unstable particle R, that eventually decays into the particles observed as the final state. If the total decay rate of R is small the cross-section exhibits a rapid variation (usually a peak), known as a *resonance*, at the energy of the intermediate state R.

We shall see that the behavior of the cross-sections near a resonance is pretty much prescribed by the unitarity condition alone, which is a good thing since there are a number of very different mechanisms that can produce a nearly stable state:

(a) The simplest possibility is that the Hamiltonian can be decomposed into two terms, a 'strong' Hamiltonian $H_0 + V_s$, which has the particle R as an eigenstate, plus a weak perturbation V_w, which allows R to decay into various states, including the initial and final states α, β of our collision process. For instance there is a neutral particle, the Z^0, with $j = 1$ and mass 91 GeV, that would be stable in the absence of the electroweak interactions. These interactions allow the Z^0 to decay into electron–positron pairs, muon–antimuon pairs, etc., but with a total decay rate that is much less than the Z^0 mass. In 1989 the Z^0 particle was seen as a resonance** in electron–positron collisions at CERN and Stanford,

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** Incidentally, this example shows that a resonant state only needs to decay *relatively* slowly; the Z^0 lifetime is 2.6×10^{-25} seconds, which is not long enough for a Z^0 travelling near the speed of light to cross an atomic nucleus. What is important is that the decay rate is 36 times smaller than the rate $\hbar/M_Z$ of oscillation of the Z^0 wave function in its rest frame.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

in the reactions $e^+ + e^- \to Z^0 \to e^+ + e^-$, $e^+ + e^- \to Z^0 \to \mu^+ + \mu^-$, etc.

(b) In some cases a particle is long-lived because there is a potential barrier that nearly prevents its constituents from escaping. The classic example is nuclear alpha decay: it may be energetically possible for a nucleus to emit an alpha particle (a He^4 nucleus) but the strong electrostatic repulsion between the alpha particle and the nucleus creates a barrier region around the daughter nucleus, which the alpha particle is classically forbidden to enter. The decay then can proceed only by quantum mechanical barrier penetration, and is exponentially slow. Such an unstable state shows up as a resonance in the scattering of the alpha particle on the daughter nucleus. For instance, the lowest-energy state of the Be^8 nucleus is unstable against decay into two alpha particles, and is seen as a resonance in He^4–He^4 scattering. (In addition to Coulomb barriers, there are also centrifugal barriers that help lengthen the life of alpha-, beta-, and gamma-unstable nuclei of high spin.)

(c) It is possible for complicated systems to be nearly stable for statistical reasons, without the presence of any potential barriers or weak interactions. For instance, an excited state of a heavy nucleus may be able to decay only if, through a statistical fluctuation, a large part of its energy is concentrated on a single neutron. This state will then show up as a resonance in the scattering of a neutron on the daughter nucleus.

These mechanisms for producing long-lived states are so different that it is truly fortunate that most of the properties of resonances follow from unitarity alone, without regard to the dynamical mechanism that produces the resonance.

First, let's consider the energy dependence of the matrix element for a reaction near a resonance. A wave packet $\int d\alpha\, g(\alpha)\Psi_\alpha{}^+ \exp(-iE_\alpha t)$ of 'in' states has a time-dependence given by (3.1.19)

$$\int d\alpha\, g(\alpha)\Psi_\alpha^+ e^{-iE_\alpha t} = \int d\alpha\, g(\alpha)\Phi_\alpha e^{-iE_\alpha t} + \int d\beta\, \Phi_\beta \int d\alpha\, \frac{e^{-iE_\alpha t}\, g(\alpha)\, T_{\beta\alpha}{}^+}{E_\alpha - E_\beta + i\epsilon}\,.$$

As mentioned in Section 3.1, a pole in the function $T_{\beta\alpha}{}^+$ in the lower-half complex E_α plane would make a contribution to the second term that decays exponentially as $t \to \infty$. Specifically, a pole at $E_\alpha = E_R - i\Gamma/2$ yields a term in the *amplitude* that behaves like $\exp(-iE_R t - \Gamma t/2)$, so it corresponds to a state whose *probability* decays like $\exp(-\Gamma t)$. We conclude then that a long-lived state of energy E_R with a slow decay rate

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Γ produces a term in the scattering amplitude that varies as

$$T_{\beta\alpha}{}^{+} \sim (E_\alpha - E_R + i\Gamma/2)^{-1} + \text{ constant} . \tag{3.8.1}$$

To go further, it will be convenient to adopt as a basis the orthonormal discrete multi-particle states $\Phi_{\mathbf{p}EN}$ discussed in the previous section; $\mathbf{p}$ and E are the total momentum and energy, and N is an index that takes only discrete (though infinitely many) values. In this basis, the S-matrix may be written

$$S_{\mathbf{p}'E'N',\,\mathbf{p}EN} = \delta^3(\mathbf{p}' - \mathbf{p})\delta(E' - E)S_{N'N}(\mathbf{p}, E) . \tag{3.8.2}$$

Near a resonance, we expect the center-of-mass frame amplitude $S(0, E) \equiv \mathcal{S}(E)$ to have the form

$$\mathcal{S}_{N'N}(E) \equiv S_{N'N}(0, E) = \mathcal{S}_{0N'N} + \frac{\mathcal{R}_{N'N}}{E - E_R + i\Gamma/2} , \tag{3.8.3}$$

where $\mathcal{S}_0$ and $\mathcal{R}$ are approximately constant at least over the relatively small range of energies $|E - E_R| \lesssim \Gamma$.

In this basis, the unitarity of the S-matrix is an ordinary matrix equation

$$\mathcal{S}(E)^\dagger \mathcal{S}(E) = 1 . \tag{3.8.4}$$

Applied to Eq. (3.8.3), this tells us that the non-resonant background S-matrix is unitary

$$\mathcal{S}_0{}^\dagger \mathcal{S}_0 = 1 , \tag{3.8.5}$$

and also that the residue matrix $\mathcal{R}$ satisfies the two conditions

$$\mathcal{S}_0{}^\dagger \mathcal{R} + \mathcal{R}^\dagger \mathcal{S}_0 = 0 , \tag{3.8.6}$$

$$-\frac{i}{2}\Gamma \mathcal{S}_0{}^\dagger \mathcal{R} + \frac{i}{2}\Gamma \mathcal{R}^\dagger \mathcal{S}_0 + \mathcal{R}^\dagger \mathcal{R} = 0 . \tag{3.8.7}$$

These conditions can be put in a more transparent form by setting

$$\mathcal{R} \equiv -i\Gamma \mathcal{A} \mathcal{S}_0 . \tag{3.8.8}$$

The unitarity conditions on the matrix $\mathcal{A}$ are then simply

$$\mathcal{A}^\dagger = \mathcal{A} , \qquad \mathcal{A}^2 = \mathcal{A} . \tag{3.8.9}$$

Any such Hermitian idempotent matrix is called a *projection matrix*. Such matrices can always be expressed as a sum of dyads of orthonormal vectors $u^{(r)}$:

$$\mathcal{A}_{N'N} = \sum_r u_{N'}^{(r)} u_N^{(r)*} , \qquad \sum_N u_N^{(r)*} u_N^{(s)} = \delta_{rs} . \tag{3.8.10}$$

The discrete part of the S-matrix is then

$$\mathcal{S}_{N'N}(E) = \sum_{N''} \left[\delta_{N'N''} - i\frac{\Gamma}{E - E_R + i\Gamma/2} \sum_r u_{N'}^{(r)} u_{N''}^{(r)*} \right] \mathcal{S}_{0N''N} . \tag{3.8.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

Each term in the sum over r can be thought of as arising from a different resonant state, all these states having the same values for E_R and Γ.

What has this to do with rates and cross-sections? For simplicity let's now ignore the non-resonant background scattering, setting $\mathcal{S}_{0N'N}$ equal to $\delta_{N'N}$; we will come back to the more general case a little later. Then for the two-body discrete center-of-mass states described in the previous section, Eq. (3.8.11) reads:

$$\mathcal{S}_{j'\sigma'\ell's'n',\,j\sigma\ell sn}(E) = \delta_{j'j}\delta_{\sigma'\sigma}\delta_{\ell'\ell}\delta_{s's}\delta_{n'n} - i\frac{\Gamma}{E - E_R + i\Gamma/2}\sum_r u^{(r)}_{j'\sigma'\ell's'n'}\, u^{(r)*}_{j\sigma\ell sn}\,. \tag{3.8.12}$$

In all cases the label r will include an index σ_R giving the z-component of the total angular momentum of the resonant state; for a resonant state of total angular momentum j_R, σ_R takes $2j_R + 1$ values. If there is no other degeneracy, then r just labels the value of σ_R, and

$$u^{(\sigma_R)}_{j\sigma\ell sn} = \delta_{j_R,j}\delta_{\sigma_R,\sigma}u_{\ell sn}\,, \tag{3.8.13}$$

where $u_{\ell sn}$ are a set of complex amplitudes that (because of the Wigner–Eckart theorem) are independent of σ. Now Eq. (3.8.12) gives the amplitude S^j defined by Eq. (3.7.7) as

$$S^j_{\ell's'n',\ell sn}(E) = \delta_{\ell',\ell}\delta_{s's}\delta_{n'n} - i\delta_{j,j_R}\frac{\Gamma}{E - E_R + i\Gamma/2}\,u_{\ell's'n'}\,u^*_{\ell sn}\,. \tag{3.8.14}$$

Also, Eq. (3.8.10) now reads

$$\sum_{\ell sn}|u_{\ell sn}|^2 + \cdots = 1 \tag{3.8.15}$$

with the dots representing the positive contribution of any states containing three or more particles. As we shall see, the quantities $|u_{\ell sn}|^2$ have the interpretation of branching ratios for the decay of the resonant state into the various accessible two-body states.

Eq. (3.7.12) now gives the total cross-section for all reactions in channel n:

$$\sigma_{\text{total}}(n;E) = \frac{\pi(2j_R+1)}{k^2(2s_1+1)(2s_2+1)}\frac{\Gamma\Gamma_n}{(E-E_R)^2 + \Gamma^2/4}\,, \tag{3.8.16}$$

where

$$\Gamma_n \equiv \Gamma\sum_{\ell s}|u_{\ell sn}|^2\,. \tag{3.8.17}$$

This is a version of the celebrated Breit–Wigner single-level formula.[27] We can also use these results to calculate the cross-section for resonant scattering from an initial two-body channel n to a final two-body channel

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

n'. Using Eq. (3.8.14) in Eq. (3.7.10) gives

$$\sigma(n \to n'; E) = \frac{\pi(2j_R + 1)}{k^2(2s_1 + 1)(2s_2 + 1)} \frac{\Gamma_n \Gamma_{n'}}{(E - E_R)^2 + \Gamma^2/4} . \tag{3.8.18}$$

This shows that the probabilities that the resonant state will decay into any one of the final two-body channels n' are proportional to the $\Gamma_{n'}$. According to Eq. (3.8.15), the sum of the Γ_n (including contributions from final states containing three or more particles) is just equal to the total decay rate Γ, so we can conclude that Γ_n is just the rate for the decay of the resonant state into channel n.

We see in Eqs. (3.8.16) and (3.8.18) the characteristic resonant peak at energy E_R, with a width (the full width at half maximum) equal to the decay rate Γ. (The individual Γ_n are often called partial widths.) Since $\Gamma_n \leq \Gamma$, the total cross-section at the peak of the resonance is roughly bounded by one square wavelength, $(2\pi/k)^2$. This rule, that cross-sections at a single resonance are roughly bounded by a square wavelength, is universally applicable even in classical physics (where energy conservation plays the role played here by unitarity), as for instance in the resonant interaction of sound waves with bubbles in the sea, or gravitational waves with gravitational wave antennae. (In the latter case, the branching ratio for oscillations in any laboratory mass to lose their energy through gravitational radiation is tiny, so the cross-section even at a resonance peak is vastly less then a square wavelength.[28])

Incidentally, it often happens that a resonance is detected, but energy measurements are insufficiently precise to resolve its width. In this case, what is measured experimentally is the integral of the cross-section over the resonant peak. For the total cross-section (3.8.16), this is

$$\int \sigma_{\text{total}}(n; E) dE = \frac{2\pi^2(2j_R + 1)\Gamma_n}{k_R^2(2s_1 + 1)(2s_2 + 1)} . \tag{3.8.19}$$

Such experiments can reveal only the partial width for decay of the resonant state into the initial particles, not the total width or branching ratios.

This formalism can also be applied when the resonant states with a given spin z-component form a multiplet related by some symmetry group. For instance, to the extent that isospin symmetry is respected, for a resonance of total isospin T_R the index r labelling resonant states includes a specification not only of the angular-momentum z-component σ_R, but also of an isospin three-component t_R, taking values $-T_R, -T_R + 1, \cdots T_R$. In this case there is no change in the above results for the total and partial cross-sections, because each two-body channel n has definite values t_1, t_2 of the isospin z-components of the two particles, and hence can only couple

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

to the resonant state with the single value $t_1 + t_2$ for t_R. The partial widths Γ_n here depend on t_1 and t_2 only through factors $C_{T_1,T_2}(T_R, t_R; t_1, t_2)^2$.

The presence of a resonance shows itself in a characteristic behavior of phase shifts near the resonance. Returning to the general formula (3.8.11) (but still taking $\mathcal{S}_0 = 1$), we see from Eq. (3.8.10) that for each individual resonant state r, there is an eigenvector $u_N^{(r)}$ of $\mathcal{S}_{N'N}(E)$ with eigenvalue

$$\exp(2i\delta^{(r)}(E)) = 1 - i\frac{\Gamma}{E - E_R + i\Gamma/2}$$

or, in other words,

$$\tan\delta^{(r)}(E) = -\frac{\Gamma/2}{E - E_R} . \tag{3.8.20}$$

We see that over an energy range of order Γ centered around the resonant energy, the the 'eigenphase' $\delta^{(r)}(E)$ jumps from a value $\nu\pi$ (with ν a positive or negative integer) below the resonance to $(\nu + 1)\pi$ above it. However, in order to use this result to say something about reaction rates, we need to know the eigenvectors $u_N^{(r)}$, which, in general, have components with arbitrary numbers of particles with various momenta, spin, and species.

These results are much more useful in those special cases when the particles in a particular channel N are forbidden (usually by conservation laws) from making a transition to any other channel. With this assumption, it is not difficult to include the effects of a non-resonant background scattering matrix $\mathcal{S}_0$ in the general result (3.8.11). In order for $\mathcal{S}_{N'N}$ to vanish for some particular N and all $N' \neq N$, it is necessary that the same is true of $\mathcal{S}_{0N'N}$, and also true of $u_{N'}^{(r)}$ for any r for which $u_N^{(r)} \neq 0$. The unitarity requirement (3.8.5) then requires that for this N

$$\mathcal{S}_{0N'N} = \exp(2i\delta_{0N})\delta_{N'N}$$

and Eq. (3.8.10) requires that

$$u_N^{(r)*} u_N^{(s)} = \delta_{rs} ,$$

so that there can be only one term r in Eq. (3.8.11) for which $u_N^{(r)} \neq 0$. In this case, Eq. (3.8.11) gives

$$\begin{aligned}\mathcal{S}_{N'N}(E) &= \delta_{N'N}\left[1 - \frac{i\Gamma}{E - E_R + i\Gamma/2}\right]\exp(2i\delta_{0N}) \\ &\equiv \delta_{N'N}\exp(2i\delta_N(E))\end{aligned}$$

with total phase shift

$$\delta_N(E) = \delta_{0N} - \arctan\left(\frac{\Gamma/2}{E - E_R}\right) . \tag{3.8.21}$$

We see that over an energy range of order Γ centered around the resonant energy E_R, the phase shift $\delta_N(E)$ jumps from a value δ_{0N} below the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

resonance to $\delta_{0N} + \pi$ above it. For instance, as we saw in the previous section, these assumptions are satisfied in various two-body reactions such as pion–pion and pion–nucleon scattering at energies below the threshold for producing extra pions, with N incorporating the total and orbital angular-momenta j, ℓ (with $j = \ell$ for pion–pion scattering) and the total angular-momentum z-component σ, as well as the total isospin T and its three-component t. The Wigner–Eckart theorem allows the phase shifts to depend only on j, ℓ, and T, not on t or σ. There are famous resonances in these channels: in pion–pion scattering there is a resonance at 770 MeV called ρ with $j = \ell = 1$, $T = 1$, and $\Gamma = 150$ MeV; in pion–nucleon scattering there is a resonance at 1232 MeV called Δ with $j = \frac{3}{2}$, $\ell = 1$, $T = \frac{3}{2}$ and $\Gamma = 110$ to 120 MeV.

Inspection of Eq. (3.7.12) or Eq. (3.7.18) shows that the total cross-section reaches a peak when the resonant phase shift passes through $\pi/2$ (or odd-integer multiples of $\pi/2$.) The non-resonant phase shifts are typically rather small, so as we saw earlier, σ_{total} will exhibit a sharp peak when the phase shift δ_ℓ goes through $\pi/2$, at an energy close to E_R. However, it sometimes happens that the non-resonant background phase shift δ_{0N} is near $\pi/2$, in which case the cross-section will exhibit a sharp *dip* as the phase shift rises through π near E_R, due to destructive interference between the resonance and the non-resonant background amplitude. Such dips were first observed by Ramsauer and Townsend[29] in 1922, in the scattering of electrons by noble gas atoms.

Problems

1. Consider a theory with a separable interaction; that is,

$$(\Phi_\beta, V\Phi_\alpha) = g\, u_\beta\, u_\alpha^* ,$$

where g is a real coupling constant, and u_α is a set of complex quantities with

$$\sum_\alpha |u_\alpha|^2 = 1.$$

Use the Lippmann–Schwinger equation (3.1.16) to find explicit solutions for the 'in' and 'out' states and the S-matrix.

2. Suppose that a resonance of spin one is discovered in e^+–e^- scattering at a total energy of 150 GeV and with a cross-section (in the center-of-mass frame, averaged over initial spins, and summed over final spins) for elastic e^+–e^- scattering at the peak of the resonance equal to 10^{-34} cm^2. What is the branching ratio for the decay of the resonant state R by the mode $R \to e^- + e^+$? What is the total

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

cross-section for e^+–e^- scattering at the peak of the resonance? (In answering both questions, ignore the non-resonant background scattering.)

3. Express the differential cross-section for two-body scattering in the *laboratory* frame, in which one of the two particles is initially at rest, in terms of kinematic variables and the matrix element $M_{\beta\alpha}$. (Derive the result directly, without using the results derived in this chapter for the differential cross-section in the center-of-mass frame.)

4. Derive the perturbation expansion (3.5.8) directly from the expansion (3.5.3) of old-fashioned perturbation theory.

5. We can define 'standing wave' states $\Psi_\alpha{}^0$ by a modified version of the Lippmann–Schwinger equation

$$\Psi_\alpha{}^0 = \Phi_\alpha + \frac{\mathscr{P}}{E_\alpha - H_0} V \Psi_\alpha{}^0 .$$

Show that the matrix $K_{\beta\alpha} \equiv \pi\delta(E_\beta - E_\alpha)(\Phi_\beta, V\Psi_\alpha{}^0)$ is Hermitian. Show how to express the S-matrix in terms of the K-matrix.

6. Express the differential cross-section for elastic π^+–proton and π^-–proton scattering in terms of the phase shifts for states of definite total angular momentum, parity, and isospin.

7. Show that the states $\Phi_{E\,\mathbf{p}\,j\sigma sn}$ defined by Eq. (3.7.5) are correctly normalized to have the scalar products (3.7.6).

References

1. For more details, see M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, New York, 1964); R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd edn (Springer-Verlag, New York, 1982).

1*a*. B. Lippmann and J. Schwinger, *Phys. Rev.* **79**, 469 (1950).

2. J. A. Wheeler, *Phys. Rev.* **52**, 1107 (1937); W. Heisenberg, *Z. Phys.* **120**, 513, 673 (1943).

3. M. Born, *Z. Phys.*. **37**, 863 (1926); **38**, 803 (1926).

4. C. Møller, *Kgl. Danske Videnskab. Mat. Fys. Medd.* **23**, No. 1 (1945); **22**, No. 19 (1946).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

5. G. D. Rochester and C. C. Butler, *Nature* **160**, 855 (1947). For a historical review, see G. D. Rochester, in *Pions to Quarks – Particle Physics in the 1950s*, ed. by L. M. Brown, M. Dresden, and L. Hoddeson (Cambridge University Press, Cambridge, UK, 1989).

6. G. Breit, E. U. Condon, and R. S. Present, *Phys. Rev.* **50**, 825 (1036); B. Cassen and E. U. Condon, *Phys. Rev.* **50**, 846 (1936); G. Breit and E. Feenberg, *Phys. Rev.* **50**, 850 (1936).

7. M. A. Tuve, N. Heydenberg, and L. R. Hafstad, *Phys. Rev.* **50**, 806 (1936).

8. M. Gell-Mann, Cal. Tech. Synchotron Laboratory Report CTSL-20 (1961); *Phys. Rev.* **125**, 1067 (1962); Y. Ne'eman *Nucl. Phys.* **26**, 222 (1961).

9. See e.g. A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, 1957): Chapter 3 (where $C_{j_1 j_2}(jm; m_1 m_2)$ is denoted $(j_1\, j_2\, j\, m | j_1\, m_1\, j_2\, m_2)$); M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957): Chapter III (where $C_{j_1 j_2}(jm; m_1 m_2)$ is denoted $C(j_1\, j_2\, j\, ;\, m_1\, m_2\, m)$).

10. G. Feinberg and S. Weinberg, *Nuovo Cimento* Serie X, **14**, 571 (1959).

11. W. Chinowsky and J. Steinberger, *Phys. Rev.* **95**, 1561 (1954); also see B. Ferretti, *Report of an International Conference on Fundamental Particles and Low Temperatures, Cambridge, 1946* (The Physical Society, London, 1947).

12. T. D. Lee and C. N. Yang, *Phys. Rev.* **104**, 254 (1956).

13. C. S. Wu *et al.*, *Phys. Rev.* **105**, 1413 (1957).

14. R. Garwin, L. Lederman, and M. Weinrich, *Phys. Rev.* **105**, 1415 (1957); J. I. Friedman and V. L. Telegdi, *Phys. Rev.* **105**, 1681 (1957).

15. K. M. Watson, *Phys. Rev.* **88**, 1163 (1952).

16. M. Gell-Mann and A. Pais, *Phys. Rev.* **97**, 1387 (1955); also see A. Pais and O. Piccioni, *Phys. Rev.* **100**, 1487 (1955).

17. J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Phys. Rev. Letters* **13**, 138 (1964).

18. K. R. Schubert *et al.*, *Phys. Lett.* **31B**, 662 (1970). This reference analyzes neutral kaon data without the assumption of CPT invariance, and finds that the part of the CP-violating amplitude that conserves

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

CPT and violates T has both real and imaginary parts five standard deviations from zero, while the part that conserves T and violates CPT is within a standard deviation of zero.

19. R. H. Dalitz, *Phil. Mag.* **44**, 1068 (1953); also see E. Fabri, *Nuovo Cimento* **11**, 479 (1954).

20. F. J. Dyson, *Phys. Rev.* **75**, 486, 1736 (1949).

21. See, e.g., L. I. Schiff, *Quantum Mechanics*, 1st edn (McGraw-Hill, New York, 1949): Section 19.

22. This was first proved in classical electrodynamics. See, e. g., H. A. Kramers, *Atti Congr. Intern. Fisici, Como, 1927*; reprinted in H. A. Kramers, *Collected Scientific Papers* (North-Holland, Amsterdam, 1956). For the proof in quantum mechanics, see E. Feenberg, *Phys. Rev.* **40**, 40 (1932); N. Bohr, R. E. Peierls, and G. Placzek, *Nature* **144**, 200 (1939).

23. A general version of this argument was given in the late 1960s in unpublished work of C. N. Yang and C. P. Yang. Also see A. Aharony, in *Modern Developments in Thermodynamics* (Wiley, New York, 1973): pp. 95–114, and references therein.

24. See, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957): Chapter 2; M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957): Appendix III; L. D. Landau and E. M. Lifshitz, *Quantum Mechanics — Non Relativistic Theory*, 3rd edn (Pergamon Press, Oxford, 1977): Section 28.

25. E. P. Wigner, *Gruppentheorie* (Friedrich Vieweg und Sohn, Braunschweig, 1931); C. Eckart, *Rev. Mod. Phys.* **2**, 305 (1930).

26. M. Froissart, *Phys. Rev.* **123**, 1053 (1961).

27. G. Breit and E. P. Wigner, *Phys. Rev.* **49**, 519 (1936).

28. See, e.g., S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Section 10.7.

29. R. Kollath, *Phys. Zeit.* **31**, 985 (1931).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.005

4

The Cluster Decomposition Principle

Up to this point we have not had much to say about the detailed structure of the Hamiltonian operator H. This operator can be defined by giving all its matrix elements between states with arbitrary numbers of particles. Equivalently, as we shall show here, any such operator may be expressed as a function of certain operators that create and destroy single particles. We saw in Chapter 1 that such creation and annihilation operators were first encountered in the canonical quantization of the electromagnetic field and other fields in the early days of quantum mechanics. They provided a natural formalism for theories in which massive particles as well as photons can be produced and destroyed, beginning in the early 1930s with Fermi's theory of beta decay.

However, there is a deeper reason for constructing the Hamiltonian out of creation and annihilation operators, which goes beyond the need to quantize any pre-existing field theory like electrodynamics, and has nothing to do with whether particles can actually be produced or destroyed. The great advantage of this formalism is that if we express the Hamiltonian as a sum of products of creation and annihilation operators, with suitable non-singular coefficients, then the S-matrix will automatically satisfy a crucial physical requirement, the cluster decomposition principle,[1] which says in effect that distant experiments yield uncorrelated results. Indeed, it is for this reason that the formalism of creation and annihilation operators is widely used in non-relativistic quantum statistical mechanics, where the number of particles is typically fixed. In relativistic quantum theories, the cluster decomposition principle plays a crucial part in making field theory inevitable. There have been many attempts to formulate a relativistically invariant theory that would not be a local field theory, and it is indeed possible to construct theories that are not field theories and yet yield a Lorentz-invariant S-matrix for two-particle scattering,[2] but such efforts have always run into trouble in sectors with more than two particles: either the three-particle S-matrix is not Lorentz-invariant, or else it violates the cluster decomposition principle.

In this chapter we will first discuss the basis of states containing ar-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

bitrary numbers of bosons and fermions, then define the creation and annihilation operators, and finally show how their use facilitates the construction of Hamiltonians that yield S-matrices satisfying the cluster decomposition condition.

4.1 Bosons and Fermions

The Hilbert space of physical states is spanned by states containing 0, 1, 2, $\cdots$ free particles. These can be free-particle states, or 'in' states, or 'out' states; for definiteness we shall deal here with the free-particle states $\Phi_{\mathbf{p}_1 \sigma_1 n_1, \mathbf{p}_2 \sigma_2 n_2, \cdots}$, but all our results will apply equally to 'in' or 'out' states. As usual, σ labels spin z-components (or helicities, for massless particles) and n labels particle species.

We must now go into a matter that has been passed over in Chapter 3; the symmetry properties of these states. As far as we know, all particles are either *bosons* or *fermions*, the difference being that a state is unchanged by the interchange of two identical bosons, and changes sign under the interchange of two identical fermions. That is

$$\Phi_{\cdots \mathbf{p}\sigma n \cdots \mathbf{p}'\sigma' n \cdots} = \pm \Phi_{\cdots \mathbf{p}'\sigma' n \cdots \mathbf{p}\sigma n \cdots} \tag{4.1.1}$$

with an upper or lower sign if n is a boson or a fermion, respectively, and dots representing other particles that may be present in the state. (Equivalently, this could be stated as a condition on the 'wave functions,' the coefficients of these multi-particle basis vectors in physically allowable state-vectors.) These two cases are often referred to as Bose or Fermi 'statistics'. We will see in the next chapter that Bose and Fermi statistics are only possible for particles that have integer or half-integer spins, respectively, but we shall not need this information in the present chapter. In this section we shall offer a non-rigorous argument that all particles must be either bosons or fermions, and then set up normalization conditions for multi-boson or multi-fermion states.

First note that if two particles with momenta and spins $\mathbf{p}, \sigma$ and $\mathbf{p}', \sigma'$ belong to identical species n, then the state-vectors $\Phi_{\cdots \mathbf{p}\sigma n \cdots \mathbf{p}'\sigma' n \cdots}$ and $\Phi_{\cdots \mathbf{p}'\sigma' n \cdots \mathbf{p}\sigma n \cdots}$ represent the same physical state; if this were not the case then the particles would be distinguished by their order in the labelling of the state-vector, and the first listed would not be identical with the second. Since the two state-vectors are physically indistinguishable, they must belong to the same ray, and so

$$\Phi_{\cdots \mathbf{p}\sigma n \cdots \mathbf{p}'\sigma' n \cdots} = \alpha_n \Phi_{\cdots \mathbf{p}'\sigma' n \cdots \mathbf{p}\sigma n \cdots}, \tag{4.1.2}$$

where α_n is a complex number of unit absolute value. We may regard this as part of the definition of what we mean by identical particles.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

The crux of the matter is to decide on what the phase factor α_n may depend. *If* it depends only on the species index n, then we are nearly done. Interchanging the two particles in Eq. (4.1.2) again, we find

$$\Phi_{\cdots \mathbf{p}\sigma n \cdots \mathbf{p}'\sigma' n \cdots} = \alpha_n^2 \Phi_{\cdots \mathbf{p}\sigma n \cdots \mathbf{p}'\sigma' n \cdots}$$

so that $\alpha_n^2 = 1$, yielding Eq. (4.1.1) as the only two possibilities.

On what else could α_n depend? It might depend on the numbers and species of the other particles in the state (indicated by dots in Eqs. (4.1.1) and (4.1.2)), but this would lead to the uncomfortable result that the symmetry of state-vectors under interchange of particles here on earth may depend on the presence of particles elsewhere in the universe. This is the sort of thing that is ruled out by the cluster decomposition principle, to be discussed later in this chapter. The phase α_n cannot have any non-trivial dependence on the spins of the two particles that are interchanged, because then these spin-dependent phase factors would have to furnish a representation of the rotation group, and there are no non-trivial representations of the three-dimensional rotation group that are one-dimensional – that is, by phase factors. The phase α_n might conceivably depend on the momenta of the two particles that are interchanged, but Lorentz invariance would require α_n to depend only on the scalar $p_1^\mu p_{2\mu}$; this is symmetric under interchange of particles 1 and 2, and therefore such dependence would not change the argument leading to the conclusion that $\alpha_n^2 = 1$.

The logical gap in the above argument is that (although our notation hides the fact) the states $\Phi_{\mathbf{p}_1 \sigma_1 n, \mathbf{p}_2 \sigma_2 n, \cdots}$ may carry a phase factor that depends on the path through momentum space by which the momenta of the particles are brought to the values $\mathbf{p}_1$, $\mathbf{p}_2$, etc. In this case the interchange of two particles twice might change the state by a phase factor, so that $\alpha_n^2 \neq 1$. We will see in Section 9.7 that this is a real possibility in two-dimensional space, but not for three or more spatial dimensions.

What about interchanges of particles belonging to different species? If we like, we can avoid this question by simply agreeing from the beginning to label the state-vector by listing all photon momenta and helicities first, then all electron momenta and spin z-components, and so on through the table of elementary particle types. Alternatively, we can allow the particle labels to appear in any order, and *define* the state-vectors with particle labels in an arbitrary order as equal to the state-vector with particle labels in some standard order times phase factors, whose dependence on the interchange of particles of different species can be anything we like. In order to deal with symmetries like isospin invariance that relate particles of different species, it is convenient to adopt a convention that generalizes Eq. (4.1.1): the state-vector will be taken to be symmetric under inter-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

change of any bosons with each other, or any bosons with any fermions, and antisymmetric with respect to interchange of any two fermions with each other, in all cases, whether the particles are of the same species or not.*

The normalization of these states must be defined in consistency with these symmetry conditions. To save writing, we will use a label q to denote all the quantum numbers of a single particle: its momentum, $\mathbf{p}$, spin z-component (or, for massless particles, helicity) σ, and species n. The N-particle states are thus labelled $\Phi q_1 \dots q_N$ (with $N = 0$ for the vacuum state Φ_0.) For $N = 0$ and $N = 1$ the question of symmetry does not arise: here we have

$$(\Phi_0, \Phi_0) = 1 \tag{4.1.3}$$

and

$$(\Phi_{q'}, \Phi_q) = \delta(q' - q)\,, \tag{4.1.4}$$

where $\delta(q'-q)$ is a product of all the delta functions and Kronecker deltas for the particle's quantum numbers,

$$\delta(q' - q) \equiv \delta^3(\mathbf{p}' - \mathbf{p})\,\delta_{\sigma'\sigma}\delta_{n'n}. \tag{4.1.5}$$

On the other hand, for $N = 2$ the states $\Phi_{q_1'q_2'}$ and $\Phi_{q_2'q_1'}$ are physically the same, so here we must take

$$\left(\Phi_{q_1'q_2'}\,,\ \Phi_{q_1q_2}\right) = \delta(q_1' - q_1)\,\delta(q_2' - q_2) \pm \delta(q_2' - q_1)\,\delta(q_1' - q_2) \tag{4.1.6}$$

the sign $\pm$ being $-$ if both particles are fermions and $+$ otherwise. This obviously is consistent with the above stated symmetry properties of the states. More generally,

$$\left(\Phi_{q_1'q_2'\dots q_M'}, \Phi_{q_1q_2\dots q_N}\right) = \delta_{NM} \sum_{\mathscr{P}} \delta_{\mathscr{P}} \prod_i \delta(q_i - q'_{\mathscr{P}i}). \tag{4.1.7}$$

The sum here is over all permutations $\mathscr{P}$ of the integers $1, 2, \cdots, N$. (For instance, in the first term in Eq. (4.1.6), $\mathscr{P}$ is the identity, $\mathscr{P}1 = 1$, $\mathscr{P}2 = 2$, while in the second term $\mathscr{P}1 = 2$, $\mathscr{P}2 = 1$.) Also, $\delta_{\mathscr{P}}$ is a sign factor equal to -1 if $\mathscr{P}$ involves an odd permutation of fermions (an odd number of fermion interchanges) and $+1$ otherwise. It is easy to see that Eq. (4.1.7) has the desired symmetry or antisymmetry properties under interchange of the q_i, and also under interchange of the q_j'.

* In fact, by the same reasoning, the symmetry or antisymmetry of the state-vector under interchange of particles of the same species but different helicities or spin z-components is purely conventional, because we could have agreed from the beginning to list first the momenta of photons of helicity $+1$, then the momenta of all photons of helicity -1, then the momenta of all electrons of spin z-component $+\frac{1}{2}$, and so on. We adopt the *convention* that the state-vector is symmetric or antisymmetric under interchange of identical bosons or fermions of different helicities or spin z-components in order to facilitate the use of rotational invariance.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

4.2 Creation and Annihilation Operators

Creation and annihilation operators may be defined in terms of their effect on the normalized multi-particle states discussed in the previous section. The *creation operator* $a^\dagger(q)$ (or in more detail, $a^\dagger(\mathbf{p},\sigma,n)$) is defined as the operator that simply adds a particle with quantum numbers q at the front of the list of particles in the state

$$a^\dagger(q)\Phi_{q_1q_2\cdots q_N} \equiv \Phi_{qq_1q_2\cdots q_N} . \tag{4.2.1}$$

In particular, the N-particle state can be obtained by acting on the vacuum with N creation operators

$$a^\dagger(q_1)a^\dagger(q_2)\cdots a^\dagger(q_N)\Phi_0 = \Phi_{q_1\cdots q_N} . \tag{4.2.2}$$

It is conventional for this operator to be called $a^\dagger(q)$; *its* adjoint, which is then called $a(q)$, may be calculated from Eq. (4.1.7). As we shall now show, $a(q)$ removes a particle from any state on which it acts, and is therefore known as an *annihilation operator*. In particular, when the particles $q\,q_1\cdots q_N$ are either all bosons or all fermions, we have

$$a(q)\Phi_{q_1q_2\cdots q_N} = \sum_{r=1}^{N}(\pm)^{r+1}\delta(q-q_r)\Phi_{q_1\cdots q_{r-1}q_{r+1}\cdots q_N} , \tag{4.2.3}$$

with a $+1$ or -1 sign for bosons or fermions, respectively . (Here is the proof. We want to calculate the scalar product of $a(q)\Phi_{q_1q_2\cdots q_N}$ with an arbitrary state $\Phi_{q_1'\cdots q_M'}$. Using Eq. (4.2.1), this is

$$\left(\Phi_{q_1'\cdots q_M'},\ a(q)\Phi_{q_1\cdots q_N}\right) \equiv \left(a^\dagger(q)\Phi_{q_1'\cdots q_M'},\ \Phi_{q_1\cdots q_N}\right) = \left(\Phi_{qq_1'\cdots q_M'},\ \Phi_{q_1\cdots q_N}\right) .$$

We now use Eq. (4.1.7). The sum over permutations $\mathscr{P}$ of $1,2,\cdots,N$ can be written as a sum over the integer r that is permuted into the first place, i.e. $\mathscr{P}r = 1$, and over mappings $\bar{\mathscr{P}}$ of the remaining integers $1,\cdots,r-1,\,r+1,\cdots,N$ into $1,\cdots,N-1$. Furthermore, the sign factor is

$$\delta_{\mathscr{P}} = (\pm)^{r-1}\delta_{\bar{\mathscr{P}}}$$

with upper and lower signs for bosons and fermions, respectively. Hence, using Eq. (4.1.7) twice,

$$\begin{aligned}\left(\Phi_{q_1'\cdots q_M'},\ a(q)\Phi_{q_1\cdots q_N}\right) &= \delta_{N,M+1}\\ &\times \sum_{r=1}^{N}\sum_{\bar{\mathscr{P}}}(\pm)^{r-1}\delta_{\bar{\mathscr{P}}}\delta(q-q_r)\prod_{i=1}^{M}\delta(q_i'-q_{\mathscr{P}i})\\ &= \delta_{N,M+1}\sum_{r=1}^{N}(\pm)^{r-1}\delta(q-q_r)\left(\Phi_{q_1'\cdots q_M'},\ \Phi_{q_1\cdots q_{r-1}q_{r+1}\cdots q_N}\right) .\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

Both sides of Eq. (4.2.3) thus have the same matrix element with any state $\Phi_{q_1'\cdots q_M'}$, and are therefore equal, as was to be shown.) As a special case of Eq. (4.2.3), we note that for both bosons and fermions, $a(q)$ annihilates the vacuum

$$a(q)\Phi_0 = 0\,. \tag{4.2.4}$$

As defined here, the creation and annihilation operators satisfy an important commutation or anticommutation relation. Applying the operator $a(q')$ to Eq. (4.2.1) and using Eq. (4.2.3) gives

$$a(q')a^\dagger(q)\Phi_{q_1\cdots q_N} = \delta(q'-q)\Phi_{q_1\cdots q_N} + \sum_{r=1}^{N}(\pm)^{r+2}\delta(q'-q_r)\Phi_{qq_1\cdots q_{r-1}q_{r+1}\cdots q_N}\,.$$

(The sign in the second term is $(\pm)^{r+2}$ because q_r is in the $(r+1)$-th place in $\Phi_{qq_1\cdots q_N}$.) On the other hand, applying the operator $a^\dagger(q)$ to Eq. (4.2.3) gives

$$a^\dagger(q)a(q')\Phi_{q_1\cdots q_N} = \sum_{r=1}^{N}(\pm)^{r+1}\delta(q'-q_r)\Phi_{qq_1\cdots q_{r-1}q_{r+1}\cdots q_N}\,.$$

Subtracting or adding, we have then

$$\left[a(q')a^\dagger(q) \mp a^\dagger(q)a(q')\right]\ \Phi_{q_1\cdots q_N} = \delta(q'-q)\Phi_{q_1\cdots q_N}\,.$$

But this holds for all states $\Phi_{q_1\cdots q_N}$ (and may easily be seen to hold also for states containing both bosons and fermions) and therefore implies the operator relation

$$a(q')a^\dagger(q) \mp a^\dagger(q)a(q') = \delta(q'-q)\,. \tag{4.2.5}$$

In addition, Eq. (4.2.2) gives immediately

$$a^\dagger(q')a^\dagger(q) \mp a^\dagger(q)a^\dagger(q') = 0 \tag{4.2.6}$$

and so also

$$a(q')a(q) \mp a(q)a(q') = 0\,. \tag{4.2.7}$$

As always, the top and bottom signs apply for bosons and fermions, respectively. According to the conventions discussed in the previous section, the creation and/or annihilation operators for particles of two different species commute if either particle is a boson, and anticommute if both are fermions.

The above discussion could have been presented in reverse order (and in most textbooks usually is). That is, we could have started with the commutation or anticommutation relations Eqs. (4.2.5)–(4.2.7), derived from the canonical quantization of some given field theory. Multi-particle

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

states would have then been defined by Eq. (4.2.2), and their scalar products Eq. (4.1.7) derived from the commutation or anticommutation relations. In fact, as discussed in Chapter 1, such a treatment would be much closer to the way that this formalism developed historically. We have followed an unhistorical approach here because we want to free ourselves from any dependence on pre-existing field theories, and rather wish to understand why field theories are the way they are.

We will now prove the fundamental theorem quoted at the beginning of this chapter: *any* operator $\mathcal{O}$ may be expressed as a sum of products of creation and annihilation operators

$$\begin{aligned}\mathcal{O} = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty}\int dq'_1\cdots dq'_N dq_1\cdots dq_M \\ \times a^\dagger(q'_1)\cdots a^\dagger(q'_N)a(q_M)\cdots a(q_1) \\ \times C_{NM}(q'_1\cdots q'_N q_1\cdots q_M)\,. \end{aligned} \tag{4.2.8}$$

That is, we want to show that the C_{NM} coefficients can be chosen to give the matrix elements of this expression any desired values. We do this by mathematical induction. First, it is trivial that by choosing C_{00} properly, we can give $(\Phi_0, \mathcal{O}\Phi_0)$ any desired value, irrespective of the values of C_{NM} with $N > 0$ and/or $M > 0$. We need only use Eq. (4.2.4) to see that Eq. (4.2.8) has the vacuum expectation value

$$(\Phi_0, \mathcal{O}\Phi_0) = C_{00}\,.$$

Now suppose that the same is true for all matrix elements of $\mathcal{O}$ between N- and M-particle states, with $N < L, M \leq K$ or $N \leq L, M < K$; that is, that these matrix elements have been given some desired values by an appropriate choice of the corresponding coefficients C_{NM}. To see that the same is then also true of matrix elements of $\mathcal{O}$ between any L- and K-particle states, use Eq. (4.2.8) to evaluate

$$\begin{aligned}\left(\Phi_{q'_1\cdots q'_L}, \mathcal{O}\Phi_{q_1\cdots q_K}\right) = L!K!\,C_{LK}(q'_1\cdots q'_L q_1\cdots q_K) \\ +\ \text{terms involving } C_{NM} \text{ with } N < L, M \leq K \text{ or } N \leq L, M < K\,.\end{aligned}$$

Whatever values have already been given to C_{NM} with $N < L, M \leq K$ or $N \leq L, M < K$, there is clearly some choice of C_{LK} which gives this matrix element any desired value.

Of course, an operator need not be expressed in the form (4.2.8), with all creation operators to the left of all annihilation operators. (This is often called the 'normal' order of the operators.) However, if the formula for some operator has the creation and annihilation operators in some other order, we can always bring the creation operators to the left of the annihilation operators by repeated use of the commutation or

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

anticommutation relations, picking up new terms from the delta function in Eq. (4.2.5).

For instance, consider any sort of additive operator F (like momentum, charge, etc.) for which

$$F\Phi_{q_1\cdots q_N} = (f(q_1) + \cdots + f(q_N))\,\Phi_{q_1\cdots q_N}\,. \tag{4.2.9}$$

Such an operator can be written as in Eq. (4.2.8), but using only the term with $N = M = 1$:

$$F = \int dq\, a^\dagger(q) a(q) f(q)\,. \tag{4.2.10}$$

In particular, the free-particle Hamiltonian is always

$$H_0 = \int dq\, a^\dagger(q) a(q) E(q) \tag{4.2.11}$$

where $E(q)$ is the single-particle energy

$$E(\mathbf{p}, \sigma, n) = \sqrt{\mathbf{p}^2 + m_n^2}\,.$$

We will need the transformation properties of the creation and annihilation operators for various symmetries. First, let's consider inhomogeneous proper orthochronous Lorentz transformations. Recall that the N-particle states have the Lorentz transformation property

$$\begin{aligned} U_0(\Lambda,\alpha)\Phi_{\mathbf{p}_1\sigma_1 n_1,\mathbf{p}_2\sigma_2 n_2,\cdots} &= e^{-i(\Lambda p_1)\cdot\alpha}\, e^{-i(\Lambda p_2)\cdot\alpha} \\ &\cdots \sqrt{\frac{(\Lambda p_1)^0 (\Lambda p_2)^0 \cdots}{p_1^0 p_2^0 \cdots}} \\ &\quad \times \sum_{\bar\sigma_1\bar\sigma_2\cdots} D^{(j_1)}_{\bar\sigma_1\sigma_1}\Big(W(\Lambda,p_1)\Big) D^{(j_2)}_{\bar\sigma_2\sigma_2}\Big(W(\Lambda,p_2)\Big) \\ &\cdots \Phi_{\mathbf{p}_{1\Lambda}\bar\sigma_1 n_1,\mathbf{p}_{2\Lambda}\bar\sigma_2 n_2,\cdots}\,. \end{aligned}$$

Here $\mathbf{p}_\Lambda$ is the three-vector part of Λp, $D^{(j)}_{\bar\sigma\sigma}(R)$ is the same unitary spin-j representation of the three-dimensional rotation group as used in Section 2.5, and $W(\Lambda, p)$ is the particular rotation

$$W(\Lambda, p) \equiv L^{-1}(\Lambda p)\Lambda L(p),$$

where $L(p)$ is the standard 'boost' that takes a particle of mass m from rest to four-momentum p^μ. (Of course, m and j depend on the species label n. This is all for $m \neq 0$; we will return to the massless particle case in the following chapter.) Now, these states can be expressed as in Eq. (4.2.2)

$$\Phi_{\mathbf{p}_1\sigma_1 n_1,\mathbf{p}_2\sigma_2 n_2,\cdots} = a^\dagger(\mathbf{p}_1\sigma_1 n_1) a^\dagger(\mathbf{p}_2\sigma_2 n_2)\cdots\Phi_0\,,$$

where Φ_0 is the Lorentz-invariant vacuum state

$$U_0(\Lambda,\alpha)\Phi_0 = \Phi_0\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

In order that the state (4.2.2) should transform properly, it is necessary and sufficient that the creation operator have the transformation rule

$$U_0(\Lambda,\alpha)a^\dagger(\mathbf{p}\sigma n)U_0^{-1}(\Lambda,\alpha) = e^{-i(\Lambda p)\cdot\alpha}\sqrt{(\Lambda p)^0/p^0} \times \sum_{\bar\sigma} D^{(j)}_{\bar\sigma\sigma}\Big(W(\Lambda,p)\Big)a^\dagger(\mathbf{p}_\Lambda\,\bar\sigma\, n)\,. \quad (4.2.12)$$

In the same way, the operators C, P, and T, that induce charge-conjugation, space inversion, and time-reversal transformations on free particle states* transform the creation operators as:

$$\mathsf{C}a^\dagger(\mathbf{p}\,\sigma\,n)\mathsf{C}^{-1} = \xi_n\, a^\dagger(\mathbf{p}\,\sigma\,n^c)\,, \quad (4.2.13)$$

$$\mathsf{P}a^\dagger(\mathbf{p}\,\sigma\,n)\mathsf{P}^{-1} = \eta_n\, a^\dagger(-\mathbf{p}\,\sigma\,n)\,, \quad (4.2.14)$$

$$\mathsf{T}a^\dagger(\mathbf{p}\,\sigma\,n)\mathsf{T}^{-1} = \zeta_n(-1)^{j-\sigma}\, a^\dagger(-\mathbf{p}\,-\sigma\,n)\,, \quad (4.2.15)$$

As mentioned in the previous section, although we have been dealing with operators that create and annihilate particles in free-particle states, the whole formalism can be applied to 'in' and 'out' states, in which case we would introduce operators $a_{\rm in}$ and $a_{\rm out}$ defined in the same way by their action on these states. These operators satisfy a Lorentz transformation rule just like Eq. (4.2.12), but with the true Lorentz transformation operator $U(\Lambda,\alpha)$ instead of the free-particle operator $U_0(\Lambda,\alpha)$.

4.3 Cluster Decomposition and Connected Amplitudes

It is one of the fundamental principles of physics (indeed, of all science) that experiments that are sufficiently separated in space have unrelated results. The probabilities for various collisions measured at Fermilab should not depend on what sort of experiments are being done at CERN at the same time. If this principle were not valid, then we could never make any predictions about any experiment without knowing everything about the universe.

In S-matrix theory, the cluster decomposition principle states that if multi-particle processes $\alpha_1 \to \beta_1, \alpha_2 \to \beta_2, \cdots, \alpha_{\mathscr{N}} \to \beta_{\mathscr{N}}$ are studied in $\mathscr{N}$ very distant laboratories, then the S-matrix element for the overall process

* We omit the subscript '0' on these operators, because in virtually all cases where C, P, and/or T are conserved, the operators that induce these transformations on 'in' and 'out' states are the same as those defined by their action on free-particle states. This is not the case for continuous Lorentz transformations, for which it is necessary to distinguish between the operators $U(\Lambda,a)$ and $U_0(\Lambda,a)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

factorizes. That is,*

$$S_{\beta_1+\beta_2+\cdots+\beta_{n'},\,\alpha_1+\alpha_2+\cdots+\alpha_n} \;\to\; S_{\beta_1\alpha_1}\, S_{\beta_2\alpha_2}\cdots S_{\beta_{n'}\alpha_n} \tag{4.3.1}$$

if for all $i \neq j$, *all* of the particles in states α_i and β_i are at a great spatial distance from *all* of the particles in states α_j and β_j. This factorization of S-matrix elements will ensure a factorization of the corresponding transition probabilities, corresponding to uncorrelated experimental results.

There is a combinatoric trick that allows us to rewrite Eq. (4.3.1) in a more transparent way. Suppose we define the *connected* part of the S-matrix, $S^{\rm C}_{\beta\alpha}$, by the formula**

$$S_{\beta\alpha} = \sum_{\rm PART} (\pm)\, S^{\rm C}_{\beta_1\alpha_1} S^{\rm C}_{\beta_2\alpha_2}\cdots . \tag{4.3.2}$$

Here the sum is over all different ways of partitioning the particles in the state α into clusters $\alpha_1, \alpha_2, \cdots$, and likewise a sum over all ways of partitioning the particles in the state β into clusters $\beta_1, \beta_2, \cdots$, not counting as different those that merely arrange particles within a given cluster or permute whole clusters. The sign is $+$ or $-$ according to whether the rearrangements $\alpha \to \alpha_1\alpha_2\cdots$ and $\beta \to \beta_1\beta_2\cdots$ involve altogether an even or an odd number of fermion interchanges, respectively. The term 'connected' is used because of the interpretation of $S^{\rm C}_{\beta\alpha}$ in terms of diagrams representing different contributions in perturbation theory, to be discussed in the next section.

This is a recursive definition. For each α and β, the sum on the right-hand side of Eq. (4.3.2) consists of a term $S^{\rm C}_{\beta\alpha}$, plus a sum Σ' over products of two or more $S^{\rm C}$-matrix elements, with a total number of particles in each of the states α_j and β_j that is *less* than the number of particles in

* We are here returning to the notation used in Chapter 3; Greek letters α or β stand for a collection of particles, including for each particle a specification of its momentum, spin, and species. Also, $\alpha_1 + \alpha_2 + \cdots + \alpha_{n'}$ is the state formed by combining all the particles in the states $\alpha_1, \alpha_2, \cdots$, and $\alpha_{n'}$, and likewise for $\beta_1 + \beta_2 + \cdots + \beta_{n'}$.

** This decomposition has been used in classical statistical mechanics by Ursell, Mayer, and others, and in quantum statistical mechanics by Lee and Yang and others.[3] It has also been used to calculate many-body ground state energies by Goldstone[4] and Hugenholtz.[5] In all of these applications the purpose of isolating the connected parts of Green's functions, partition functions, resolvents, etc., is to deal with objects with a simple volume dependence. This is essentially our purpose too, because as we shall see, the crucial property of the connected parts of the S-matrix is that they are proportional to a single momentum-conservation delta function, and in a box the delta function becomes a Kronecker delta times the volume. The cluster decomposition is also the same formal device as that used in the theory of noise[6] to decompose the correlation function of several random variables into its 'cumulants'; if the random variable receives contributions from a large number N of independent fluctuations, then each cumulant is proportional to N.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

the states α and β

$$S_{\beta\alpha} = S^{\rm C}_{\beta\alpha} + \sum_{\rm PART}{}' (\pm)\, S^{\rm C}_{\beta_1\alpha_1} S^{\rm C}_{\beta_2\alpha_2} \cdots .$$

Suppose that the $S^{\rm C}$-matrix elements in this sum have already been chosen in such a way that Eq. (4.3.2) is satisfied for states β, α containing together fewer than, say, N particles. Then no matter what values are found in this way for the S-matrix elements appearing in the sum Σ', we can always choose the remaining term $S^{\rm C}_{\beta\alpha}$ so that Eq. (4.3.2) is also satisfied for states α, β containing a total of N particles.† Thus Eq. (4.3.2) contains no information in itself; it is merely a definition of $S^{\rm C}$.

If the states α and β each consist of just a single particle, say with quantum numbers q and q' respectively, then the only term on the right-hand-side of Eq. (4.3.2) is just $S^{\rm C}_{\beta\alpha}$ itself, so for one-particle states

$$S^{\rm C}_{q'q} \equiv S_{q'q} = \delta(q' - q)\,. \tag{4.3.3}$$

(Apart from possible degeneracies, the fact that $S_{q'q}$ is proportional to $\delta(q' - q)$ follows from conservation laws. The absence of any proportionality factor in Eq. (4.3.3) is based on a suitable choice of the relative phase of 'in' and 'out' states.) We are here assuming that single-particle states are stable, so that there are no transitions between single-particle states and any others, such as the vacuum.

For transitions between two-particle states, Eq. (4.3.2) reads

$$S_{q_1'q_2',q_1q_2} = S^{\rm C}_{q_1'q_2',q_1q_2} + \delta(q_1' - q_1)\delta(q_2' - q_2) \pm \delta(q_1' - q_2)\delta(q_2' - q_1)\,. \tag{4.3.4}$$

(We are here using Eq. (4.3.3).) The sign $\pm$ is $-$ if both particles are fermions, and otherwise $+$. We recognize that the two delta function terms just add up to the norm (4.1.6), so here $S^{\rm C}_{\beta\alpha}$ is just $(S-1)_{\beta\alpha}$. But the general case is more complicated.

For transitions between three-particle or four-particle states, Eq. (4.3.2) reads

$$\begin{aligned} S_{q_1'q_2'q_3',q_1q_2q_3} &= S^{\rm C}_{q_1'q_2'q_3',q_1q_2q_3} \\ &+ \delta(q_1' - q_1) S^{\rm C}_{q_2'q_3',q_2q_3} \pm \text{permutations} \\ &+ \delta(q_1' - q_1)\delta(q_2' - q_2)\delta(q_3' - q_3) \pm \text{permutations} \end{aligned} \tag{4.3.5}$$

† A technicality should be mentioned here. This argument works only if we neglect the possibility that for one or more of the connected S-matrix elements in Eq. (4.3.2), the states α_j and β_j both contain no particles at all. We must therefore define the connected vacuum–vacuum element $S^{\rm C}_{0,0}$ to be zero. We do not use Eq. (4.3.2) for the vacuum–vacuum S-matrix $S_{0,0}$, which in the absence of time-varying external fields is simply defined to be unity, $S_{0,0} = 1$. We will have more to say about the vacuum–vacuum amplitude in the presence of external fields in Volume II.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

and

$$\begin{aligned} S_{q_1'q_2'q_3'q_4',q_1q_2q_3q_4} &= S^{\mathrm{C}}_{q_1'q_2'q_3'q_4',q_1q_2q_3q_4} \\ &+ S^{\mathrm{C}}_{q_1'q_2',q_1q_2} S^{\mathrm{C}}_{q_3'q_4',q_3q_4} \pm \text{permutations} \\ &+ \delta(q_1'-q_1) S^{\mathrm{C}}_{q_2'q_3'q_4',q_2q_3q_4} \pm \text{permutations} \\ &+ \delta(q_1'-q_1)\delta(q_2'-q_2) S^{\mathrm{C}}_{q_3'q_4',q_3q_4} \pm \text{permutations} \\ &+ \delta(q_1'-q_1)\delta(q_2'-q_2)\delta(q_3'-q_3)\delta(q_4'-q_4) \pm \text{permutations} . \end{aligned} \tag{4.3.6}$$

(Taking account of all permutations, there are a total of $1+9+6=16$ terms in Eq. (4.3.5) and $1+18+16+72+24=131$ terms in Eq. (4.3.6). If we had not assumed that one-particle states are stable, there would be even more terms.) As explained previously, the definition of $S^{\mathrm{C}}_{\beta\alpha}$ is recursive: we use Eq. (4.3.4) to define $S^{\mathrm{C}}_{\beta\alpha}$ for two-particle states, then use this definition in Eq. (4.3.5) when we define $S^{\mathrm{C}}_{\beta\alpha}$ for three-particle states, then use both of these definitions in Eq. (4.3.6) to obtain the definition of $S^{\mathrm{C}}_{\beta\alpha}$ for four-particle states, and so on.

The point of this definition of the connected part of the S-matrix is that the cluster decomposition principle is equivalent to the requirement that $S^{\mathrm{C}}_{\beta\alpha}$ must vanish when any one or more of the particles in the states β and/or α are far away in space from the others.‡ To see this, suppose that the particles in the states β and α are grouped into clusters $\beta_1, \beta_2, \cdots$ and $\alpha_1, \alpha_2, \cdots$, and that all particles in the set $\alpha_i + \beta_i$ are far from all particles in the set $\alpha_j + \beta_j$ for any $j \neq i$. Then if $S^{\mathrm{C}}_{\beta'\alpha'}$ vanishes if any particles in β' or α' are far from the others, it vanishes if any particles in these states are in different clusters, so the definition (4.3.2) yields

$$S_{\beta\alpha} \to \sum{}^{(1)} (\pm) S^{\mathrm{C}}_{\beta_{11}\alpha_{11}} S^{\mathrm{C}}_{\beta_{12}\alpha_{12}} \cdots \times \sum{}^{(2)} (\pm) S^{\mathrm{C}}_{\beta_{21}\alpha_{21}} S^{\mathrm{C}}_{\beta_{22}\alpha_{22}} \cdots \times \cdots , \tag{4.3.7}$$

where $\Sigma^{(j)}$ is a sum over all different ways of partitioning the clusters β_j and α_j into subclusters $\beta_{j1}, \beta_{j2}, \cdots$ and $\alpha_{j1}, \alpha_{j2}, \cdots$. But referring back to Eq. (4.3.2), this is just the desired factorization property (4.3.1).

For instance, suppose that in the four-particle reaction $1234 \to 1'2'3'4'$, we let particles $1, 2, 1'$, and $2'$ be very far from $3, 4, 3'$, and $4'$. Then if $S^{\mathrm{C}}_{\beta\alpha}$ vanishes when any particles in β and/or α are far from the others, the only terms in Eq. (4.3.6) that survive (in an even more abbreviated notation) are

‡ In order to give a meaning to 'far', we will have to Fourier transform S^{C}, so that each three-momentum label $\mathbf{p}$ is replaced with a spatial coordinate three-vector $\mathbf{x}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

$$S_{1'2'3'4',1234} \to S^{C}_{1'2',12}S^{C}_{3'4',34} + (\delta_{1'1}\delta_{2'2} \pm \delta_{1'2}\delta_{2'1})S^{C}_{3'4',34} + (\delta_{3'3}\delta_{4'4} \pm \delta_{3'4}\delta_{4'3})S^{C}_{1'2',12} + (\delta_{1'1}\delta_{2'2} \pm \delta_{1'2}\delta_{2'1})(\delta_{3'3}\delta_{4'4} \pm \delta_{3'4}\delta_{4'3}) .$$

Comparison with Eq. (4.3.4) shows that this is just the required factorization condition (4.3.1)

$$S_{1'2'3'4',1234} \to S_{1'2',12}S_{3'4',34} .$$

We have formulated the cluster decomposition principle in coordinate space, as the condition that $S^{C}_{\beta\alpha}$ vanishes if any particles in the states β or α are far from any others. It is convenient for us to reexpress this in momentum space. The coordinate space matrix elements are defined as a Fourier transform

$$S^{C}_{\mathbf{x}'_1\mathbf{x}'_2\cdots,\mathbf{x}_1\mathbf{x}_2\cdots} \equiv \int d^3\mathbf{p}'_1 d^3\mathbf{p}'_2 \cdots d^3\mathbf{p}_1 d^3\mathbf{p}_2 \cdots \; S^{C}_{\mathbf{p}'_1\mathbf{p}'_2\cdots\mathbf{p}_1\mathbf{p}_2\cdots} \times e^{i\mathbf{p}'_1\cdot\mathbf{x}'_1} e^{i\mathbf{p}'_2\cdot\mathbf{x}'_2}\cdots e^{-i\mathbf{p}_1\cdot\mathbf{x}_1} e^{-i\mathbf{p}_2\cdot\mathbf{x}_2\cdots} . \quad (4.3.8)$$

(We are here temporarily dropping spin and species labels, which just go along with the momentum or coordinate labels.) If $|S^{C}_{\mathbf{p}'_1\mathbf{p}'_2\cdots\mathbf{p}_1\mathbf{p}_2\cdots}|$ were sufficiently well behaved (to be specific, if it were Lebesgue integrable) then according to the Riemann–Lebesgue theorem[7] the integral (4.3.8) would vanish when any combination of spatial coordinates goes to infinity. Now, this is certainly too strong a requirement. Translational invariance tells us that the connected part of the S-matrix, like the S-matrix itself, can only depend on differences of coordinate vectors, and therefore does not change at all if all of the x_i and x'_j vary together, with their differences held constant. This requires that the elements of S^{C} in a momentum basis must, like those of S, be proportional to a three-dimensional delta function that ensures momentum conservation (and makes $|S^{C}_{\mathbf{p}'_1\mathbf{p}'_2\cdots\mathbf{p}_1\mathbf{p}_2\cdots}|$ *not* Lebesgue integrable), as well as the energy-conservation delta function required by scattering theory. That is, we can write

$$S^{C}_{\mathbf{p}'_1\mathbf{p}'_2\cdots,\mathbf{p}_1\mathbf{p}_2\cdots} = \delta^3(\mathbf{p}'_1 + \mathbf{p}'_2 + \cdots - \mathbf{p}_1 - \mathbf{p}_2 - \cdots) \times \delta(E'_1 + E'_2 + \cdots - E_1 - E_2 - \cdots)C_{\mathbf{p}'_1\mathbf{p}'_2\cdots,\mathbf{p}_1\mathbf{p}_2\cdots} . \quad (4.3.9)$$

This is no problem: the cluster decomposition principle only requires that Eq. (4.3.8) vanish when the *differences* among some of the $\mathbf{x}_i$ and/or $\mathbf{x}'_i$ become large. However, if C itself in Eq. (4.3.9) contained additional delta functions of linear combinations of the three-momenta, then this principle would not be satisfied. For instance, suppose that there were a delta function in C that required that the sum of the $\mathbf{p}'_i$ and $-\mathbf{p}_j$ for some subset of the particles vanished. Then Eq. (4.3.8) would not vary if all of the $\mathbf{x}'_i$ and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

$\mathbf{x}_j$ for the particles in that subset moved together (with constant differences) away from all the other $\mathbf{x}'_k$ and $\mathbf{x}_\ell$, in contradiction to the cluster decomposition principle. Loosely speaking then, the cluster decomposition principle simply says that *the connected part of the S-matrix, unlike the S-matrix itself, contains just a single momentum-conservation delta function.*

In order to put this a bit more precisely, we can say that the coefficient function $C_{\mathbf{p}'_1\mathbf{p}'_2\cdots,\mathbf{p}_1\mathbf{p}_2\cdots}$ in Eq. (4.3.9) is a smooth function of its momentum labels. But how smooth? It would be most straightforward if we could simply require that $C_{\mathbf{p}'_1\mathbf{p}'_2\cdots,\mathbf{p}_1\mathbf{p}_2\cdots}$ be analytic in all of the momenta at $\mathbf{p}'_1 = \mathbf{p}'_2 = \cdots = \mathbf{p}_1 = \mathbf{p}_2 = \cdots = 0$. This requirement would indeed guarantee that $S^{\rm C}_{\mathbf{x}'_1\mathbf{x}'_2\cdots,\mathbf{x}_1\mathbf{x}_2\cdots}$ vanishes exponentially fast when any of the $\mathbf{x}$ and $\mathbf{x}'$ is very distant from any of the other $\mathbf{x}$ and $\mathbf{x}'$. However, an exponential fall-off of $S^{\rm C}$ is not an essential part of the cluster decomposition principle, and, in fact, the requirement of analyticity is not met in all theories. Most notably, in theories with massless particles, $S^{\rm C}$ can have poles at certain values of the $\mathbf{p}$ and $\mathbf{p}'$. For instance, as we will see in Chapter 10, if a massless particle can be emitted in the transition $1 \to 3$ and absorbed in the transition $2 \to 4$, then $S^{\rm C}_{34,12}$ will have a term proportional to $1/(p_1 - p_3)^2$. After Fourier transforming, such poles yield terms in $S^{\rm C}_{\mathbf{x}'_1\mathbf{x}'_2\cdots,\mathbf{x}_1\mathbf{x}_2\cdots}$ that fall off only as negative powers of coordinate differences.[1] There is no need to formulate the cluster decomposition principle so stringently that such behavior is ruled out. Thus the 'smoothness' condition on $S^{\rm C}$ should be understood to allow various poles and branch-cuts at certain values of the $\mathbf{p}$ and $\mathbf{p}'$, but not singularities as severe as delta functions.

4.4 Structure of the Interaction

We now ask, what sort of Hamiltonian will yield an S-matrix that satisfies the cluster decomposition principle? It is here that the formalism of creation and annihilation operators comes into its own. The answer is contained in the theorem that the S-matrix satisfies the cluster decomposition principle if (and as far as I know, only if) the Hamiltonian can be expressed as in Eq. (4.2.8):

$$\begin{aligned} H = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \int & d\,q'_1 \cdots d\,q'_N\, d\,q_1 \cdots d\,q_M \\ & \times a^\dagger(q'_1) \cdots a^\dagger(q'_N) a(q_M) \cdots a(q_1) \\ & \times h_{NM}(q'_1 \cdots q'_N\,,\; q_1 \cdots q_M) \end{aligned} \tag{4.4.1}$$

with coefficient functions h_{NM} that contain just a *single* three-dimensional momentum-conservation delta function (returning here briefly to a more

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

explicit notation)

$$h_{NM}(\mathbf{p}'_1\sigma'_1n'_1\cdots\mathbf{p}'_N\sigma'_Nn'_N\ ,\ \mathbf{p}_1\sigma_1n_1\cdots\mathbf{p}_M\sigma_Mn_M)$$
$$=\delta^3(\mathbf{p}'_1+\cdots+\mathbf{p}'_N-\mathbf{p}_1-\cdots-\mathbf{p}_M)$$
$$\times\tilde{h}_{NM}(\mathbf{p}'_1\sigma'_1n'_1\cdots\mathbf{p}'_N\sigma'_Nn'_N\ ,\ \mathbf{p}_1\sigma_1n_1\cdots\mathbf{p}_M\sigma_Mn_M)\ , \qquad (4.4.2)$$

where $\tilde{h}_{NM}$ contains no delta function factors. Note that Eq. (4.4.1) by itself has no content — we saw in Section 4.2 that *any* operator can be put in this form. It is only Eq. (4.4.1) combined with the requirement that h_{NM} has only the single delta function shown in Eq. (4.4.2) that guarantees that the S-matrix satisfies the cluster decomposition principle.

The validity of this theorem in perturbation theory will become obvious when we develop the Feynman diagram formalism in Chapter 6. The trusting reader may prefer to skip the rest of the present chapter, and move on to consider the implications of this theorem in Chapter 5. However, the proof has some instructive features, and will help to clarify in what sense the field theory of the next chapter is inevitable.

To prove this theorem, we make use of perturbation theory in its time-dependent form. (One of the advantages of time-dependent perturbation theory is that it makes the combinatorics underlying the cluster decomposition principle much more transparent; if E is a sum of one-particle energies then e^{-iEt} is a product of functions of the individual energies, while $[E-E_\alpha+i\epsilon]^{-1}$ is not.) The S-matrix is given by Eq. (3.5.10) as*

$$S_{\beta\alpha}=\sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{\infty}dt_1\cdots dt_n\Big(\Phi_\beta,T\Big\{V(t_1)\cdots V(t_n)\Big\}\Phi_\alpha\Big)\ , \qquad (4.4.3)$$

where the Hamiltonian is split into a free-particle part H_0 and an interaction V, and

$$V(t)\equiv\exp(iH_0t)V\exp(-iH_0t)\ . \qquad (4.4.4)$$

Now, the states Φ_α and Φ_β may be expressed as in Eq. (4.2.2) as products of creation operators acting on the vacuum Φ_0, and $V(t)$ is itself a sum of products of creation and annihilation operators, so each term in the sum (4.4.3) may be written as a sum of vacuum expectation values of products of creation and annihilation operators. By using the commutation or anticommutation relations (4.2.5) we may move each annihilation operator in turn to the right past all the creation operators. For each annihilation operator moved to the right past a creation operator we have two terms, as shown by writing Eq. (4.2.5) in the form

$$a(q')a^\dagger(q)=\pm a^\dagger(q)a(q')+\delta(q'-q)\ .$$

* We are now adopting the convention that for $n=0$, the time-ordered product in Eq. (4.4.3) is taken as the unit operator, so the $n=0$ term in the sum just yields the term $\delta(\beta-\alpha)$ in $S_{\beta\alpha}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

Moving other creation operators past the annihilation operator in the first term generates yet more terms. But Eq. (4.2.4) shows that any annihilation operator that moves all the way to the right and acts on Φ_0 gives zero, so in the end all we have left is the delta functions. In this way, the vacuum expectation value of a product of creation and annihilation operators is given by a sum of different terms, each term equal to a product of delta functions and $\pm$ signs from the commutators or anticommutators. It follows that each term in Eq. (4.4.3) may be expressed as a sum of terms, each term equal to a product of delta functions and $\pm$ signs from the commutators or anticommutators and whatever factors are contributed by $V(t)$, integrated over all the times and integrated and summed over the momenta, spins, and species in the arguments of the delta functions.

Each of the terms generated in this way may be symbolized by a diagram. (This is not yet the full Feynman diagram formalism, because we are not yet going to associate numerical quantities with the ingredients in the diagrams; we are using the diagrams here only as a way of keeping track of three-momentum delta functions.) Draw n points, called *vertices*, one for each $V(t)$ operator. For each delta function produced when an annihilation operator in one of these $V(t)$ operators moves past a creation operator in the initial state Φ_α, draw a line coming into the diagram from below that ends at the corresponding vertex. For each delta function produced when an annihilation operator in the adjoint of the final state Φ_β moves past a creation operator in one of the $V(t)$, draw a line from the corresponding vertex upwards out of the diagram. For each delta function produced when an annihilation operator in one $V(t)$ moves past a creation operator in another $V(t)$ draw a line between the two corresponding vertices. Finally, for each delta function produced when an annihilation operator in the adjoint of the final state moves past a creation operator in the initial state, draw a line from bottom to top, right through the diagrams. Each of the delta functions associated with one of these lines enforces the equality of the momentum arguments of the pair of creation and annihilation operators represented by the line. There is also at least one delta function contributed by each of the vertices, which enforces the conservation of the total three-momentum at the vertex.

Such a diagram may be connected (every point connected to every other by a set of lines) and if not connected, it breaks up into a number of connected pieces. The $V(t)$ operator associated with a vertex in one connected component effectively commutes with the $V(t)$ associated with any vertex in any other connected component, because for this diagram, we are not including any terms in which an annihilation operator in one vertex destroys a particle that is produced by a creation operator in the other vertex — if we did, the two vertices would be in the same connected

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

component. Thus the matrix element in Eq. (4.4.3) can be expressed as a sum over *products* of contributions, one from each connected component:

$$\left(\Phi_\beta, T\left\{V(t_1)\cdots V(t_n)\right\}\Phi_\alpha\right) = \sum_{\text{clusterings}} (\pm) \prod_{j=1}^{v} \left(\Phi_{\beta_j}, T\left\{V(t_{j1})\cdots V(t_{jn_j})\right\}\Phi_{\alpha_j}\right)_C . \quad (4.4.5)$$

Here the sum is over all ways of splitting up the incoming and outgoing particles and $V(t)$ operators into v clusters (including a sum over v from 1 to n) with the n_j operators $V(t_{j_1})\cdots V(t_{jn_j})$ and the subsets of initial particles α_j and final particles β_j all in the jth cluster. Of course, this means that

$$n = n_1 + \cdots + n_v$$

and also the set α is the union of all the particles in the subsets $\alpha_1, \alpha_2, \cdots \alpha_v$, and likewise for the final state. Some of the clusters in Eq. (4.4.5) may contain no vertices at all, i.e., $n_j = 0$; for these factors, we must take the matrix element factor in Eq. (4.4.5) to vanish unless β_j and α_j are both one-particle states (in which case it is just a delta function $\delta(\alpha_j - \beta_j)$), because the only connected diagrams without vertices consist of a single line running through the diagram from bottom to top. Most important, the subscript C in Eq. (4.4.5) means that we exclude any contributions corresponding to disconnected diagrams, that is, any contributions in which any $V(t)$ operator or any initial or final particle is not connected to every other by a sequence of particle creations and annihilations.

Now let us use Eq. (4.4.5) in the sum (4.4.3). Every time variable is integrated from $-\infty$ to $+\infty$, so it makes no difference which of the $t_1, \cdots t_n$ are sorted out into each cluster. The sum over clusterings therefore yields a factor $n!/n_1!n_2!\cdots n_v!$, equal to the number of ways of sorting out n vertices into v clusters, each containing $n_1, n_2, \cdots$ vertices:

$$\int_{-\infty}^{\infty} dt_1 \cdots dt_n \left(\Phi_\beta,\ T\left\{V(t_1)\cdots V(t_n)\right\}\Phi_\alpha\right)$$
$$= \sum_{\text{PART}} (\pm) \sum_{\substack{n_1 \cdots n_v \\ n_1+\cdots+n_v=n}} \frac{n!}{n_1!n_2!\cdots n_v!} \prod_{j=1}^{v} \int_{-\infty}^{\infty} dt_{j_1} \cdots dt_{jn_j}$$
$$\times \left(\Phi_{\beta_j}, T\left\{V(t_{j_1})\cdots V(t_{jn_j})\right\}\Phi_{\alpha_j}\right)_C .$$

The first sum here is over all ways of partitioning the particles in the initial and final states into clusters $\alpha_1 \cdots \alpha_v$ and $\beta_1 \cdots \beta_v$ (including a sum over the number v of clusters). The factor $n!$ here cancels the $1/n!$ in Eq. (4.4.3), and the factor $(-i)^n$ in the perturbation series for (4.4.5) can be written as a product $(-i)^{n_1} \cdots (-i)^{n_v}$, so instead of summing over n and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

then summing separately over $n_1, \cdots n_v$ constrained by $n_1 + \cdots + n_v = n$, we can simply sum independently over each $n_1, \cdots n_v$. This gives finally

$$S_{\beta\alpha} = \sum_{\text{PART}} (\pm) \prod_{j=1}^{v} \sum_{n_j=0}^{\infty} \frac{(-i)^{n_j}}{n_j!} \int_{-\infty}^{\infty} dt_{j_1} \cdots dt_{jn_j} \times \left(\Phi_{\beta_j},\ T\left\{ V(t_{j_1}) \cdots V(t_{jn_j}) \right\} \Phi_{\alpha_j} \right)_{\text{C}} .$$

Comparing this with the definition (4.3.2) of the connected matrix elements $S^{\text{C}}_{\beta\alpha}$, we see that these matrix elements are just given by the factors in the product here

$$S^{\text{C}}_{\beta\alpha} = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} dt_1 \cdots dt_n \left(\Phi_\beta, T\left\{ V(t_1) \cdots V(t_n) \right\} \Phi_\alpha \right)_{\text{C}} . \qquad (4.4.6)$$

(The subscript j is dropped on all the ts and ns, as these are now mere integration and summation variables.) We see that $S^{\text{C}}_{\beta\alpha}$ is calculated by a very simple prescription: *$S^{\text{C}}_{\beta\alpha}$ is the sum of all contributions to the S-matrix that are connected, in the sense that we drop all terms in which any initial or final particle or any operator $V(t)$ is not connected to all the others by a sequence of particle creations and annihilations.* This justifies the adjective 'connected' for S^{C}.

As we have seen, momentum is conserved at each vertex and along every line, so the connected parts of the S-matrix individually conserve momentum: $S^{\text{C}}_{\beta\alpha}$ contains a factor $\delta^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)$. What we want to prove is that $S^{\text{C}}_{\beta\alpha}$ contains no other delta functions.

We now make the assumption that the coefficient fractions h_{NM} in the expansion (4.4.1) of the Hamiltonian in terms of creation and annihilation operators are proportional to a *single* three-dimensional delta function, that ensures momenta conservation. This is automatically true for the free-particle Hamiltonian H_0, so it is also then true separately for the interaction V. Returning to the graphical interpretation of the matrix elements that we have been using, this means that each vertex contributes one three-dimensional delta function. (The other delta functions in matrix elements $V_{\gamma\delta}$ simply keep the momentum of any particle that is not created or annihilated at the corresponding vertex unchanged.) Now, most of these delta functions simply go to fix the momentum of intermediate particles. The only momenta that are left unfixed by such delta functions are those that circulate in loops of internal lines. (Any line which if cut leaves the diagram disconnected carries a momentum that is fixed by momentum conservation as some linear combination of the momenta of the lines coming into or going out of the diagram. If the diagram has L lines that can all be cut at the same time without the diagram becoming disconnected, then we say it has L independent loops, and there are L

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

momenta that are not fixed by momentum conservation.) With V vertices, I internal lines, and L loops, there are V delta functions, of which $I - L$ go to fix internal momenta, leaving $V - I + L$ delta functions relating the momenta of incoming or outgoing particles. But a well-known topological identity** tells that for any graph consisting of C connected pieces, the numbers of vertices, internal lines, and loops are related by

$$V - I + L = C \,. \tag{4.4.7}$$

Hence for a connected matrix element like $S^{C}_{\beta\alpha}$, which arises from graphs with $C = 1$, we find just a single three-dimensional delta function $\delta^3(\mathbf{p}_\beta - \mathbf{p}_\alpha)$, as was to be proved.

It was not important in the above argument that the time variables were integrated from $-\infty$ to $+\infty$. Thus exactly the same arguments can be used to show that if the coefficients $h_{N,M}$ in the Hamiltonian contain just single delta functions, then $U(t, t_0)$ can also be decomposed into connected parts, each containing a single momentum-conservation delta function factor. On the other hand, the connected part of the S-matrix also contains an energy-conservation delta function, and when we come to Feynman diagrams in Chapter 6 we shall see that $S^{C}_{\beta\alpha}$ contains only a single energy-conservation delta function, $\delta(E_\beta - E_\alpha)$, while $U(t, t_0)$ contains no energy-conservation delta functions at all.

It should be emphasized that the requirement that h_{NM} in Eq. (4.4.1) should have only a single three-dimensional momentum conservation delta function factor is very far from trivial, and has far-reaching implications. For instance, assume that V has non-vanishing matrix elements between two-particle states. Then Eq. (4.4.1) must contain a term with $N = M = 2$, and coefficient

$$v_{2,2}(\mathbf{p}'_1\mathbf{p}'_2, \mathbf{p}_1\,\mathbf{p}_2) = V_{\mathbf{p}'_1\mathbf{p}'_2,\mathbf{p}_1\,\mathbf{p}_2} \,. \tag{4.4.8}$$

(We are here temporarily dropping spin and species labels.) But then the matrix element of the interaction between three-particle states is

$$\begin{aligned} V_{\mathbf{p}'_1\mathbf{p}'_2\mathbf{p}'_3,\mathbf{p}_1\,\mathbf{p}_2\,\mathbf{p}_3} &= v_{3,3}(\mathbf{p}'_1\mathbf{p}'_2\mathbf{p}'_3, \mathbf{p}_1\,\mathbf{p}_2\,\mathbf{p}_3) \\ &\quad + v_{2,2}(\mathbf{p}'_1\mathbf{p}'_2, \mathbf{p}_1\mathbf{p}_2)\,\delta^3(\mathbf{p}'_3 - \mathbf{p}_3) \pm \text{permutations.} \end{aligned} \tag{4.4.9}$$

** A graph consisting of a single vertex has $V = 1$, $L = 0$, and $C = 1$. If we add $V - 1$ vertices with just enough internal lines to keep the graph connected, we have $I = V - 1$, $L = 0$, and $C = 1$. Any additional internal lines attached (without new vertices) to the same connected graph produce an equal number of loops, so $I = V + L - 1$ and $C = 1$. If a disconnected graph consists of C such connected parts, the sums of I, V, and L in each connected part will than satisfy $\sum I = \sum V + \sum L - C$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

As mentioned at the beginning of this chapter, we might try to make a relativistic quantum theory that is not a field theory by choosing $v_{2,2}$ so that the two–body S-matrix is Lorentz-invariant, and adjusting the rest of the Hamiltonian so that there is no scattering in states containing three or more particles. We would then have to take $v_{3,3}$ to cancel the other terms in Eq. (4.4.9)

$$v_{3,3}(\mathbf{p}'_1\mathbf{p}'_2\mathbf{p}'_3, \mathbf{p}_1\,\mathbf{p}_2\,\mathbf{p}_3) = -v_{2,2}(\mathbf{p}'_1\mathbf{p}'_2, \mathbf{p}_1\mathbf{p}_2)\,\delta^3(\mathbf{p}'_3 - \mathbf{p}_3) \mp \text{ permutations.} \tag{4.4.10}$$

However, this would mean that each term in $v_{3,3}$ contains *two* delta function factors (recall that $v_{2,2}(\mathbf{p}'_1\mathbf{p}'_2, \mathbf{p}_1\mathbf{p}_2)$ has a factor $\delta^3(\mathbf{p}'_1 + \mathbf{p}'_2 - \mathbf{p}_1 - \mathbf{p}_2)$) and this would violate the cluster decomposition principle. Thus in a theory satisfying the cluster decomposition principle, the existence of scattering processes involving two particles makes processes involving three or more particles inevitable.

* * *

When we set out to solve three-body problems in quantum theories that satisfy the cluster decomposition principle, the term $v_{3,3}$ in Eq. (4.4.9) gives no particular trouble, but the extra delta function in the other terms makes the Lippmann–Schwinger equation difficult to solve directly. The problem is that these delta functions make the kernel $[E_\alpha - E_\beta + i\epsilon]^{-1}V_{\beta\alpha}$ of this equation not square-integrable, even after we factor out an overall momentum conservation delta function. In consequence, it cannot be approximated by a finite matrix, even one of very large rank. To solve problems involving three or more particles, it is necessary to replace the Lippmann–Schwinger equation with one that has a connected right-hand side. Such equations have been developed for the scattering of three or more particles,[8,9] and in non-relativistic scattering problems they can be solved recursively, but they have not turned out to be useful in relativistic theories and so will not be described in detail here.

However, recasting the Lippmann–Schwinger equation in this manner is useful in another way. Our arguments in this section have so far relied on perturbation theory. I do not know of any non-perturbative proof of the main theorem of this section, but it has been shown[9] that these reformulated non-perturbative dynamical equations are *consistent* with the requirement that $U^C(t, t_0)$ (and hence S^C) should also contain only a single momentum-conservation delta function, as required by the cluster decomposition principle, provided that the Hamiltonian satisfies our condition that the coefficient functions $h_{N,M}$ each contain only a single momentum-conservation delta function.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

Problems

1. Define generating functionals for the S-matrix and its connected part:

$$F[v] \equiv 1 + \sum_{N=1}^{\infty} \sum_{M=1}^{\infty} \frac{1}{N!M!} \int v^*(q_1') \cdots v^*(q_N') v(q_1) \cdots v(q_M) \times S_{q_1' \cdots q_N', q_1 \cdots q_M} \, dq_1' \cdots dq_N' \, dq_1 \cdots dq_M$$

$$F^C[v] \equiv \sum_{N=1}^{\infty} \sum_{M=1}^{\infty} \frac{1}{N!M!} \int v^*(q_1') \cdots v^*(q_N') v(q_1) \cdots v(q_M) \times S^C_{q_1' \cdots q_N', q_1 \cdots q_M} \, dq_1' \cdots dq_N' \, dq_1 \cdots dq_M.$$

   Derive a formula relating $F[v]$ and $F^C[v]$. (You may consider the purely bosonic case.)

2. Consider an interaction

$$V = g \int d^3\mathbf{p}_1 \, d^3\mathbf{p}_2 \, d^3\mathbf{p}_3 \, d^3\mathbf{p}_4 \, \delta^3(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_3 - \mathbf{p}_4) \times a^\dagger(\mathbf{p}_1) \, a^\dagger(\mathbf{p}_2) \, a(\mathbf{p}_3) \, a(\mathbf{p}_4) \, ,$$

   where g is a real constant and $a(\mathbf{p})$ is the annihilation operator of a spinless boson of mass $M > 0$. Use perturbation theory to calculate the S-matrix element for scattering of these particles in the center-of-mass frame to order g. What is the corresponding differential cross-section?

3. A *coherent state* Φ_λ is defined to be an eigenstate of the annihilation operators $a(q)$ with eigenvalues $\lambda(q)$. Construct such a state as a superposition of the multi-particle states $\Phi_{q_1 q_2 \cdots q_N}$.

References

1. The cluster decomposition principle seems to have been first stated explicitly in quantum field theory by E. H. Wichmann and J. H. Crichton, *Phys. Rev.* **132**, 2788 (1963).

2. See, e.g., B. Bakamijian and L. H. Thomas, *Phys. Rev.* **92**, 1300 (1953).

3. For references, see T. D. Lee and C. N. Yang, *Phys. Rev.* **113**, 1165 (1959).

4. J. Goldstone, *Proc. Roy. Soc. London* **A239**, 267 (1957)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

5. N. M. Hugenholtz, *Physica* **23**, 481 (1957).

6. See, e.g., R. Kubo, *J. Math. Phys.* **4**, 174 (1963).

7. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, Oxford, 1937): Section 1.8.

8. L. D. Faddeev. *Zh. Ekxper. i Teor. Fiz.* **39**. 1459 (1961) (translation: *Soviet Phys — JETP* **12**, 1014 (1961)); *Dokl. Akad. Nauk. SSSR* **138**, 565 (1961) and **145**, 30 (1962) (translations *Soviet Physics — Doklady* **6**, 384 (1961) and **7**, 600 (1963)).

9. S. Weinberg, *Phys. Rev.* **133**, B232 (1964)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.006

5

Quantum Fields and Antiparticles

We now have all the pieces needed to motivate the introduction of quantum fields.[1] In the course of this construction, we shall encounter some of the most remarkable and universal consequences of the union of relativity with quantum mechanics: the connection between spin and statistics, the existence of antiparticles, and various relationships between particles and antiparticles, including the celebrated CPT theorem.

5.1 Free Fields

We have seen in Chapter 3 that the S-matrix will be Lorentz-invariant if the interaction can be written as

$$V(t) = \int d^3x \, \mathscr{H}(\mathbf{x}, t) \,, \tag{5.1.1}$$

where $\mathscr{H}$ is a scalar, in the sense that

$$U_0(\Lambda, a)\mathscr{H}(x)U_0^{-1}(\Lambda, a) = \mathscr{H}(\Lambda x + a) \,, \tag{5.1.2}$$

and satisfies the additional condition:

$$[\mathscr{H}(x), \mathscr{H}(x')] = 0 \quad \text{for} \quad (x - x')^2 \geq 0 \,. \tag{5.1.3}$$

As we shall see, there are more general possibilities, but none of them are very different from this. (For the present we are leaving it as an open question whether Λ here is restricted to a proper orthochronous Lorentz transformation, or can also include space inversions.) In order to facilitate also satisfying the cluster decomposition principle we are going to construct $\mathscr{H}(x)$ out of creation and annihilation operators, but here we face a problem: as shown by Eq. (4.2.12), under Lorentz transformations each such operator is multiplied by a matrix that depends on the momentum carried by that operator. How can we couple such operators together to make a scalar? The solution is to build $\mathscr{H}(x)$ out of *fields* — both

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

annihilation fields $\psi_\ell^+(x)$ and creation fields $\psi_\ell^-(x)$:

$$\psi_\ell^+(x) = \sum_{\sigma n} \int d^3p\, u_\ell(x;\mathbf{p},\sigma,n) a(\mathbf{p},\sigma,n)\,, \tag{5.1.4}$$

$$\psi_\ell^-(x) = \sum_{\sigma n} \int d^3p\, v_\ell(x;\mathbf{p},\sigma,n) a^\dagger(\mathbf{p},\sigma,n). \tag{5.1.5}$$

with coefficients* $u_\ell(x;\mathbf{p},\sigma,n)$ and $v_\ell(x;\mathbf{p},\sigma,n)$ chosen so that under Lorentz transformations each field is multiplied with a position-independent matrix:

$$U_0(\Lambda,a)\psi_\ell^+(x)U_0^{-1}(\Lambda,a) = \sum_{\bar\ell} D_{\ell\bar\ell}(\Lambda^{-1})\psi_{\bar\ell}^+(\Lambda x + a)\,, \tag{5.1.6}$$

$$U_0(\Lambda,a)\psi_\ell^-(x)U_0^{-1}(\Lambda,a) = \sum_{\bar\ell} D_{\ell\bar\ell}(\Lambda^{-1})\psi_{\bar\ell}^-(\Lambda x + a)\,. \tag{5.1.7}$$

(We might, in principle, have different transformation matrices $D^\pm$ for the annihilation and creation fields, but as we shall see, it is always possible to choose the fields so that these matrices are the same.) By applying a second Lorentz transformation $\bar\Lambda$, we find that

$$D(\Lambda^{-1})D(\bar\Lambda^{-1}) = D((\bar\Lambda\Lambda)^{-1})\,,$$

so taking $\Lambda_1 = (\Lambda)^{-1}$ and $\Lambda_2 = (\bar\Lambda)^{-1}$, we see that the D-matrices furnish a *representation* of the homogeneous Lorentz group:

$$D(\Lambda_1)D(\Lambda_2) = D(\Lambda_1\Lambda_2)\,. \tag{5.1.8}$$

There are many such representations, including the scalar $D(\Lambda) = 1$, the vector $D(\Lambda)^\mu{}_\nu = \Lambda^\mu{}_\nu$, and a host of tensor and spinor representations. These particular representations are irreducible, in the sense that it is not possible by a choice of basis to reduce all $D(\Lambda)$ to the same block-diagonal form, with two or more blocks. However, we do not require at this point that $D(\Lambda)$ be irreducible; in general it is a block-diagonal matrix with an arbitrary array of irreducible representations in the blocks. That is, the index ℓ here includes a label that runs over the types of particle described and the irreducible representations in the different blocks, as well as another that runs over the components of the individual irreducible representations. Later we will separate these fields into irreducible fields that each describe only a single particle species (and its antiparticle) and transform irreducibly under the Lorentz group.

* A reminder: the labels n and σ run over all different particle species and spin z-components, respectively.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Once we have learned how to construct fields satisfying the Lorentz transformation rules (5.1.6) and (5.1.7), we will be able to construct the interaction density as

$$\mathscr{H}(x) = \sum_{NM} \sum_{\ell'_1 \cdots \ell'_N} \sum_{\ell_1 \cdots \ell_M} g_{\ell'_1 \cdots \ell'_N,\, \ell_1 \cdots \ell_M} \times \psi^-_{\ell'_1}(x) \cdots \psi^-_{\ell'_N}(x)\, \psi^+_{\ell_1}(x) \cdots \psi^+_{\ell_M}(x) \tag{5.1.9}$$

and this will be a scalar in the sense of Eq. (5.1.2) if the constant coefficients $g_{\ell'_1 \cdots \ell'_N,\, \ell_1 \cdots \ell_M}$ are chosen to be Lorentz covariant, in the sense that for all Λ:

$$\sum_{\ell'_1 \cdots \ell'_N} \sum_{\ell_1 \cdots \ell_M} D_{\ell'_1 \bar{\ell}'_1}(\Lambda^{-1}) \cdots D_{\ell'_N \bar{\ell}'_N}(\Lambda^{-1}) D_{\ell_1 \bar{\ell}_1}(\Lambda^{-1}) \cdots D_{\ell_M \bar{\ell}_M}(\Lambda^{-1}) \times g_{\ell'_1 \cdots \ell'_N \ell_1 \cdots \ell_M} = g_{\bar{\ell}'_1 \cdots \bar{\ell}'_N \bar{\ell}_1 \cdots \bar{\ell}_M}. \tag{5.1.10}$$

(Note that we do not include derivatives here, because we regard the derivatives of components of these fields as just additional sorts of field components.) The task of finding coefficients $g_{\ell'_1 \cdots \ell'_N,\, \ell_1 \cdots \ell_M}$ that satisfy Eq. (5.1.10) is no different in principle (and not much more difficult in practice) than that of using Clebsch–Gordan coefficients to couple together various representations of the three-dimensional rotation group to form rotational scalars. Later we will be able to combine creation and annihilation fields so that this density also commutes with itself at space-like separations.

Now, what shall we take as the coefficient functions $u_\ell(x; \mathbf{p}, \sigma, n)$ and $v_\ell(x; \mathbf{p}, \sigma, n)$? Eq. (4.2.12) and its adjoint give the transformation rules** for the annihilation and creation operators

$$U_0(\Lambda, b) a(\mathbf{p}, \sigma, n) U_0^{-1}(\Lambda, b) = \exp\Big(i(\Lambda p) \cdot b\Big) \sqrt{(\Lambda p)^0 / p^0} \times \sum_{\bar{\sigma}} D^{(j_n)}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda, p)\Big) a(\mathbf{p}_\Lambda, \bar{\sigma}, n)\,, \tag{5.1.11}$$

$$U_0(\Lambda, b) a^\dagger(\mathbf{p}, \sigma, n) U_0^{-1}(\Lambda, b) = \exp\Big(-i(\Lambda p) \cdot b\Big) \sqrt{(\Lambda p)^0 / p^0} \times \sum_{\bar{\sigma}} D^{(j_n)*}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda, p)\Big) a^\dagger(\mathbf{p}_\Lambda, \bar{\sigma}, n) \tag{5.1.12}$$

where j_n is the spin of particles of species n, and $\mathbf{p}_\Lambda$ is the three-vector part of Λp. (We have used the unitarity of the rotation matrices $D^{(j_n)}_{\sigma\bar{\sigma}}$ to put both Eqs. (5.1.11) and (5.1.12) in the form shown here.) Also, as we saw in Section 2.5 the volume element d^3p/p^0 is Lorentz-invariant, so we

** This is for massive particles. The case of zero mass will be taken up in Section 5.9.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

can replace d^3p in Eqs. (5.1.4) and (5.1.5) with $d^3(\Lambda p)p^0/(\Lambda p)^0$. Putting this all together, we find

$$U_0(\Lambda,b)\psi_\ell^+(x)U_0^{-1}(\Lambda,b) = \sum_{\sigma\bar{\sigma}n}\int d^3(\Lambda p)\, u_\ell(x;\mathbf{p},\sigma,n) \times \exp\Big(i(\Lambda p)\cdot b\Big)D^{(j_n)}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda,p)\Big)\sqrt{p^0/(\Lambda p)^0}\, a(\mathbf{p}_\Lambda,\bar{\sigma},n)$$

and

$$U_0(\Lambda,b)\psi_\ell^-(x)U_0^{-1}(\Lambda,b) = \sum_{\sigma\bar{\sigma}n}\int d^3(\Lambda p)\, v_\ell(x;\mathbf{p},\sigma,n) \times \exp\Big(-i(\Lambda p)\cdot b\Big)D^{(j_n)*}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda,p)\Big)\sqrt{p^0/(\Lambda p)^0}\, a^\dagger(\mathbf{p}_\Lambda,\bar{\sigma},n)\,.$$

We see that in order for the fields to satisfy the Lorentz transformation rules (5.1.6) and (5.1.7), it is necessary and sufficient that

$$\sum_{\bar{\ell}} D_{\ell\bar{\ell}}(\Lambda^{-1})u_{\bar{\ell}}(\Lambda x+b;\mathbf{p}_\Lambda,\sigma,n) = \sqrt{p^0/(\Lambda p)^0} \times \sum_{\bar{\sigma}} D^{(j_n)}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda,p)\Big)\exp\Big(+i(\Lambda p)\cdot b\Big)u_\ell(x;\mathbf{p},\bar{\sigma},n)$$

and

$$\sum_{\bar{\ell}} D_{\ell\bar{\ell}}(\Lambda^{-1})v_{\bar{\ell}}(\Lambda x+b;\mathbf{p}_\Lambda,\sigma,n) = \sqrt{p^0/(\Lambda p)^0} \times \sum_{\bar{\sigma}} D^{(j_n)*}_{\sigma\bar{\sigma}}\Big(W^{-1}(\Lambda,p)\Big)\exp\Big(-i(\Lambda p)\cdot b\Big)v_\ell(x;\mathbf{p},\bar{\sigma},n)$$

or somewhat more conveniently

$$\sum_{\bar{\sigma}} u_{\bar{\ell}}(\Lambda x+b;\mathbf{p}_\Lambda,\bar{\sigma},n)D^{(j_n)}_{\bar{\sigma}\sigma}\Big(W(\Lambda,p)\Big) = \sqrt{p^0/(\Lambda p)^0} \times \sum_{\ell} D_{\bar{\ell}\ell}(\Lambda)\exp\Big(i(\Lambda p)\cdot b\Big)u_\ell(x;\mathbf{p},\sigma,n) \tag{5.1.13}$$

and

$$\sum_{\bar{\sigma}} v_{\bar{\ell}}(\Lambda x+b;\mathbf{p}_\Lambda,\bar{\sigma},n)D^{(j_n)*}_{\bar{\sigma}\sigma}\Big(W(\Lambda,p)\Big) = \sqrt{p^0/(\Lambda p)^0} \times \sum_{\ell} D_{\bar{\ell}\ell}(\Lambda)\exp\Big(-i(\Lambda p)\cdot b\Big)v_\ell(x;\mathbf{p},\sigma,n)\,. \tag{5.1.14}$$

These are the fundamental requirements that will allow us to calculate the u_ℓ and v_ℓ coefficient functions in terms of a finite number of free parameters.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

We will use Eqs. (5.1.13) and (5.1.14) in three steps, considering in turn three different types of proper orthochronous Lorentz transformation:

Translations

First we consider Eqs. (5.1.13) and (5.1.14) with $\Lambda = 1$ and b arbitrary. We see immediately that $u_\ell(x;\mathbf{p},\sigma,n)$ and $v_\ell(x;\mathbf{p},\sigma,n)$ must take the form

$$u_\ell(x;\mathbf{p},\sigma,n) = (2\pi)^{-3/2} e^{ip\cdot x} u_\ell(\mathbf{p},\sigma,n)\,, \tag{5.1.15}$$

$$v_\ell(x;\mathbf{p},\sigma,n) = (2\pi)^{-3/2} e^{-ip\cdot x} v_\ell(\mathbf{p},\sigma,n)\,, \tag{5.1.16}$$

so the fields are Fourier transforms:

$$\psi^+_\ell(x) = \sum_{\sigma,n} (2\pi)^{-3/2} \int d^3p\; u_\ell(\mathbf{p},\sigma,n) e^{ip\cdot x} a(\mathbf{p},\sigma,n)\,, \tag{5.1.17}$$

and

$$\psi^-_\ell(x) = \sum_{\sigma,n} (2\pi)^{-3/2} \int d^3p\; v_\ell(\mathbf{p},\sigma,n) e^{-ip\cdot x} a^\dagger(\mathbf{p},\sigma,n)\,. \tag{5.1.18}$$

(The factors $(2\pi)^{-3/2}$ could be absorbed into the definition of u_ℓ and v_ℓ, but it is conventional to show them explicitly in these Fourier integrals.) Using Eqs. (5.1.15) and (5.1.16), we see that Eqs. (5.1.13) and (5.1.14) are satisfied if and only if

$$\sum_{\bar\sigma} u_{\bar\ell}(\mathbf{p}_\Lambda,\bar\sigma,n) D^{(j_n)}_{\bar\sigma\sigma}\Big(W(\Lambda,p)\Big) = \sqrt{\frac{p^0}{(\Lambda p)^0}} \sum_\ell D_{\bar\ell\ell}(\Lambda) u_\ell(\mathbf{p},\sigma,n) \tag{5.1.19}$$

and

$$\sum_{\bar\sigma} v_{\bar\ell}(\mathbf{p}_\Lambda,\bar\sigma,n) D^{(j_n)*}_{\bar\sigma\sigma}\Big(W(\Lambda,p)\Big) = \sqrt{\frac{p^0}{(\Lambda p)^0}} \sum_\ell D_{\bar\ell\ell}(\Lambda) v_\ell(\mathbf{p},\sigma,n)\,. \tag{5.1.20}$$

for arbitrary homogeneous Lorentz transformations Λ.

Boosts

Next take $\mathbf{p} = 0$ in Eqs. (5.1.19) and (5.1.20), and let Λ be the standard boost $L(q)$ that takes a particle of mass m from rest to some four-momentum q^μ. Then $L(p) = 1$, and

$$W(\Lambda,p) \equiv L^{-1}(\Lambda p)\Lambda L(p) = L^{-1}(q)L(q) = 1\,.$$

Hence in this special case, Eqs. (5.1.19) and (5.1.20) give

$$u_{\bar\ell}(\mathbf{q},\sigma,n) = (m/q^0)^{1/2} \sum_\ell D_{\bar\ell\ell}(L(q))\; u_\ell(0,\sigma,n) \tag{5.1.21}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and

$$v_{\bar{\ell}}(\mathbf{q},\sigma,n) = (m/q^0)^{1/2} \sum_{\ell} D_{\bar{\ell}\ell}(L(q))\, v_{\ell}(0,\sigma,n)\,. \tag{5.1.22}$$

In other words, if we know the quantities $u_\ell(0,\sigma,n)$ and $v_\ell(0,\sigma,n)$ for zero momentum, then for a given representation $D(\Lambda)$ of the homogeneous Lorentz group, we know the functions $u_\ell(\mathbf{p},\sigma,n)$ and $v_\ell(\mathbf{p},\sigma,n)$ for all $\mathbf{p}$. (Explicit formulas for the matrices $D_{\bar{\ell}\ell}(L(q))$ will be given for arbitrary representations of the homogeneous Lorentz group in Section 5.7.)

Rotations

Next, take $\mathbf{p} = 0$, but this time let Λ be a Lorentz transformation with $\mathbf{p}_\Lambda = 0$; that is, take Λ as a rotation R. Here obviously $W(\Lambda,p) = R$, and so Eqs. (5.1.19) and (5.1.20) read

$$\sum_{\bar{\sigma}} u_{\bar{\ell}}(0,\bar{\sigma},n)\, D^{(j_n)}_{\bar{\sigma}\sigma}(R) = \sum_{\ell} D_{\bar{\ell}\ell}(R) u_\ell(0,\sigma,n) \tag{5.1.23}$$

and

$$\sum_{\bar{\sigma}} v_{\bar{\ell}}(0,\bar{\sigma},n)\, D^{(j_n)*}_{\bar{\sigma}\sigma}(R) = \sum_{\ell} D_{\bar{\ell}\ell}(R) v_\ell(0,\sigma,n)\,, \tag{5.1.24}$$

or equivalently

$$\sum_{\bar{\sigma}} u_{\bar{\ell}}(0,\bar{\sigma},n) \mathbf{J}^{(j_n)}_{\bar{\sigma}\sigma} = \sum_{\ell} \mathscr{J}_{\bar{\ell}\ell} u_\ell(0,\sigma,n) \tag{5.1.25}$$

and

$$\sum_{\bar{\sigma}} v_{\bar{\ell}}(0,\bar{\sigma},n) \mathbf{J}^{(j_n)*}_{\bar{\sigma}\sigma} = -\sum_{\ell} \mathscr{J}_{\bar{\ell}\ell} v_\ell(0,\sigma,n), \tag{5.1.26}$$

where $\mathbf{J}^{(j)}$ and $\mathscr{J}$ are the angular-momentum matrices in the representations $D^{(j)}(R)$ and $D(R)$, respectively. Any representation $D(\Lambda)$ of the homogeneous Lorentz group obviously yields a representation of the rotation group when Λ is restricted to rotations R; Eqs. (5.1.25) and (5.1.26) tell us that if the field $\psi^{\pm}_{\ell}(x)$ is to describe particles of some particular spin j, then this representation $D(R)$ must contain among its irreducible components the spin-j representation $D^{(j)}(R)$, with the coefficients $u_\ell(0,\sigma,n)$ and $v_\ell(0,\sigma,n)$ simply describing how the spin-j representation of the rotation group is embedded in $D(R)$. We shall see in Section 5.6 that each *irreducible* representation of the proper orthochronous Lorentz group contains any given irreducible representation of the rotation group at most once, so that if the fields $\psi^{+}_{\ell}(x)$ and $\psi^{-}_{\ell}(x)$ transform irreducibly, then they are unique up to overall scale. More generally, the number of free

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

parameters in the annihilation or creation fields (including their overall scales) is equal to the number of irreducible representations in the field.

It is straightforward to show that coefficient functions $u_\ell(\mathbf{p},\sigma,n)$ and $v_\ell(\mathbf{p},\sigma,n)$ given by Eqs. (5.1.21) and (5.1.22), with $u_\ell(0,\sigma,n)$ and $v_\ell(0,\sigma,n)$ satisfying Eqs. (5.1.23) and (5.1.24), will automatically satisfy the more general requirements (5.1.19) and (5.1.20). This is left as an exercise for the reader.

Let us now return to the cluster decomposition principle. Inserting Eqs. (5.1.17) and (5.1.18) in Eq. (5.1.9) and integrating over $\mathbf{x}$, the interaction Hamiltonian is

$$\begin{aligned} V = \sum_{NM} \int d^3\mathbf{p}'_1 \cdots d^3\mathbf{p}'_N \, d^3\mathbf{p}_1 \cdots d^3\mathbf{p}_M \sum_{\sigma'_1 \cdots \sigma'_N} \sum_{\sigma_1 \cdots \sigma_M} \sum_{n'_1 \cdots n'_N} \sum_{n_1 \cdots n_M} \\ \times a^\dagger(\mathbf{p}'_1 \sigma'_1 n'_1) \cdots a^\dagger(\mathbf{p}'_N \sigma'_N n'_N) \, a(\mathbf{p}_M \sigma_M n_M) \cdots a(\mathbf{p}_1 \sigma_1 n_1) \\ \times \mathscr{V}_{NM}(\mathbf{p}'_1 \sigma'_1 n'_1 \cdots \mathbf{p}'_N \sigma'_N n'_N \, , \, \mathbf{p}_1 \sigma_1 n_1 \cdots \mathbf{p}_M \sigma_M n_M) \end{aligned} \tag{5.1.27}$$

with coefficient functions given by

$$\begin{aligned} \mathscr{V}_{NM}(\mathbf{p}'_1 \sigma'_1 n'_1 \cdots, \mathbf{p}_1 \sigma_1 n_1 \cdots) = \delta^3(\mathbf{p}'_1 + \cdots - \mathbf{p}_1 - \cdots) \\ \times \tilde{\mathscr{V}}_{NM}(\mathbf{p}'_1 \sigma'_1 n'_1 \cdots, \mathbf{p}_1 \sigma_1 n_1 \cdots) \, , \end{aligned} \tag{5.1.28}$$

where

$$\begin{aligned} \tilde{\mathscr{V}}_{NM}(\mathbf{p}'_1 \sigma'_1 n'_1 \cdots \mathbf{p}'_N \sigma'_N n'_N \, , \, \mathbf{p}_1 \sigma_1 n_1 \cdots \mathbf{p}_M \sigma_M n_M) = (2\pi)^{3-3N/2-3M/2} \\ \times \sum_{\ell'_1 \cdots \ell'_N} \sum_{\ell_1 \cdots \ell_M} g_{\ell'_1 \cdots \ell'_N \ell_1 \cdots \ell_M} \, v_{\ell'_1}(\mathbf{p}'_1 \sigma'_1 n'_1) \cdots v_{\ell'_N}(\mathbf{p}'_N \sigma'_N n'_N) \\ \times u_{\ell_1}(\mathbf{p}_1 \sigma_1 n_1) \cdots u_{\ell_M}(\mathbf{p}_M \sigma_M n_M) \, . \end{aligned} \tag{5.1.29}$$

This interaction is manifestly of the form that will guarantee that the S-matrix satisfies the cluster decomposition principle: $\mathscr{V}_{NM}$ has a single delta function factor, with a coefficient $\tilde{\mathscr{V}}_{NM}$ that (at least for a finite number of field types) has at most branch point singularities at zero particle momenta. In fact, we could turn this argument around; any operator can be written as in Eq. (5.1.27), and the cluster decomposition principle requires that the coefficient $\mathscr{V}_{NM}$ may be written as in Eq. (5.1.28) as the product of a single momentum-conservation delta function times a smooth coefficient function. Any sufficiently smooth function (but *not* one containing additional delta functions) can be expressed as in Eq. (5.1.29).[†] *The cluster decomposition principle together with Lorentz invariance thus makes it natural that the interaction density should be constructed out of the annihilation and creation fields.*

[†] For general functions the indices ℓ and ℓ' may have to run over an infinite range. The reasons for restricting ℓ and ℓ' to a finite range have to do with the principle of renormalizability, discussed in Chapter 12.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

If all we needed were to construct a scalar interaction density that satisfied the cluster decomposition principle, then we could combine annihilation and creation operators in arbitrary polynomials (5.1.9), with coupling coefficients $g_{\ell'_1 \cdots \ell'_N, \ell_1 \cdots \ell_M}$ subject only to the invariance condition (5.1.10) (and a suitable reality condition). However, for the Lorentz invariance of the S-matrix it is necessary also that the interaction density satisfy the commutation condition (5.1.3). This condition is not satisfied for arbitrary functions of the creation and annihilation fields because

$$[\psi^+_\ell(x), \psi^-_{\bar{\ell}}(y)]_\mp = (2\pi)^{-3} \sum_{\sigma n} \int d^3p \, u_\ell(\mathbf{p}, \sigma, n) v_{\bar{\ell}}(\mathbf{p}, \sigma, n) e^{ip\cdot(x-y)} \quad (5.1.30)$$

(with the sign $\mp$ indicating a commutator or anticommutator if the particles destroyed and created by the components ψ^+_ℓ and $\psi^-_{\bar{\ell}}$ are bosons or fermions, respectively,) and in general this does not vanish even for $x - y$ space-like. It is obviously not possible to avoid this problem by making the interaction density out of creation or annihilation fields alone, for then the interaction could not be Hermitian. The only way out of this difficulty is to combine annihilation and creation fields in linear combinations:

$$\psi_\ell(x) \equiv \kappa_\ell \psi^+_\ell(x) + \lambda_\ell \psi^-_{\bar{\ell}}(x) \,, \quad (5.1.31)$$

with the constants κ and λ and any other arbitrary constants in the fields adjusted so that for $x - y$ space-like

$$[\psi_\ell(x), \psi_{\ell'}(y)]_\mp = [\psi_\ell(x), \psi^\dagger_{\ell'}(y)]_\mp = 0. \quad (5.1.32)$$

We will see in subsequent sections of this chapter how to do this for various irreducibly transforming fields. (By including explicit constants κ and λ in Eq. (5.1.31) we are leaving ourselves free to choose the overall scale of the annihilation and creation fields in any way that seems convenient.) The Hamiltonian density $\mathscr{H}(x)$ will satisfy the commutation condition (5.1.3) if it is constructed out of such fields and their adjoints, with an even number of any field components that destroy and create fermions.

The condition (5.1.32) is often described as a *causality* condition, because if $x - y$ is space-like then no signal can reach y from x, so that a measurement of ψ_ℓ at point x should not be able to interfere with a measurement of $\psi_{\ell'}$ or $\psi^\dagger_{\ell'}$ at point y. Such considerations of causality are plausible for the electromagnetic field, any one of whose components may be measured at a given spacetime point, as shown in a classic paper of Bohr and Rosenfeld.[2] However, we will be dealing here with fields like the Dirac field of the electron that do not seem in any sense measurable. The point of view taken here is that Eq. (5.1.32) is needed for the Lorentz invariance of the S-matrix, without any ancillary assumptions about measurability or causality.

There is an obstacle to the construction of fields (5.1.31) satisfying

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

(5.1.32). It may be that the particles that are destroyed and created by these fields carry non-zero values of one or more conserved quantum numbers like the electric charge. For instance, if particles of species n carry a value $q(n)$ for the electric charge Q, then

$$[Q, a(\mathbf{p}, \sigma, n)] = -q(n)a(\mathbf{p}, \sigma, n) ,$$

$$[Q, a^\dagger(\mathbf{p}, \sigma, n)] = +q(n)a^\dagger(\mathbf{p}, \sigma, n) .$$

In order that $\mathscr{H}(x)$ should commute with the charge operator Q (or some other symmetry generator) it is necessary that it be formed out of fields that have simple commutation relations with Q:

$$[Q, \psi_\ell(x)] = -q_\ell \psi_\ell(x) \tag{5.1.33}$$

for then we can make $\mathscr{H}(x)$ commute with Q by constructing it as a sum of products of fields $\psi_{\ell_1}\psi_{\ell_2}\cdots$ and adjoints $\psi^\dagger_{m_1}\psi^\dagger_{m_2}\cdots$ such that

$$q_{\ell_1} + q_{\ell_2} + \cdots - q_{m_1} - q_{m_2} - \cdots = 0.$$

Now, Eq. (5.1.33) is satisfied for one particular component $\psi^+_\ell(x)$ of the annihilation field if and only if all particle species n that are destroyed by the field carry the same charge $q(n) = q_\ell$, and it is satisfied for one particular component $\psi^-_\ell(x)$ of the creation field if and only if all particle species $\bar{n}$ that are created by the field carry the charge $q(\bar{n}) = -q_\ell$. We see that in order for such a theory to conserve quantum numbers like electric charge, there must be a doubling of particle species carrying non-zero values of such quantum numbers: if a particular component of the annihilation field destroys a particle of species n, then the same component of the creation field must create particles of a species $\bar{n}$, known as the *antiparticles* of the particles of species n, which have opposite values of all conserved quantum numbers. *This is the reason for antiparticles.*

If the representation $D(\Lambda)$ is not irreducible, then we can adopt a basis for the fields in which $D(\Lambda)$ breaks up into blocks along the main diagonal, such that fields that belong to different blocks do not transform into each other under Lorentz transformations. Also, Lorentz transformations have no effect on the particle species. Therefore, instead of considering one big field, including many irreducible components and many particle species, we shall from now on restrict our attention to fields that destroy only a single type of particle (dropping the label n) and create only the corresponding antiparticle, and that transform irreducibly under the Lorentz group (which as mentioned above may or may not be supposed to include space inversion), with the understanding that, in general, we shall have to consider many different such fields, some perhaps formed as the derivatives of other fields. In the following sections we are going to finish the determination of the coefficient functions $u_\ell(\mathbf{p}, \sigma)$ and $v_\ell(\mathbf{p}, \sigma)$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

fix the relative values of the constants κ and λ, and deduce the relations between the properties of particles and antiparticles first for fields that belong to the simplest irreducible representations of the Lorentz group, the scalar, vector, and Dirac spinor representations. After that we will repeat the analysis for a completely general irreducible representation.

A word about field equations. Inspection of Eqs. (5.1.31), (5.1.17), and (5.1.18) shows that all the components of a field of definite mass m satisfy the Klein–Gordon equation:

$$(\Box - m^2)\psi_\ell(x) = 0\,. \tag{5.1.34}$$

Some fields satisfy other field equations as well, depending on whether or not there are more field components than independent particle states. Traditionally in quantum field theory one begins with such field equations, or with the Lagrangian from which they are derived, and then uses them to derive the expansion of the fields in terms of one-particle annihilation and creation operators. In the approach followed here, we start with the particles, and derive the fields according to the dictates of Lorentz invariance, with the field equations arising almost incidentally as a byproduct of this construction.

* * *

A technicality must be mentioned here. According to the theorem proved in Section 4.4, the condition that guarantees that a theory will satisfy the cluster decomposition principle is that the interaction can be expressed as a sum of products of creation and annihilation operators, with all creation operators to the left of all annihilation operators, and with coefficients that contain only a single momentum-conservation delta function. For this reason, we should write the interaction in the 'normal ordered' form

$$V = \int d^3x \;:\mathscr{F}(\psi(x),\psi^\dagger(x)):\tag{5.1.35}$$

the colons indicating that the enclosed expression is to be rewritten (ignoring non-vanishing commutators or anticommutators, but including minus signs for permutations of fermionic operators) so that all creation operators stand to the left of all annihilation operators. By using the commutation or anticommutation relations of the fields, any such normal-ordered function of the fields can just as well be written as a sum of ordinary products of the fields with c-number coefficients. Rewriting $:\mathscr{F}:$ in this way makes it obvious that despite the normal ordering, $:\mathscr{F}(\psi(x),\psi^\dagger(x)):$ will commute with $:\mathscr{F}(\psi(y),\psi^\dagger(y)):$ when $x-y$ is space-like, if it is constructed out of fields that satisfy Eq. (5.1.32), with even numbers of any fermionic field components.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

5.2 Causal Scalar Fields

We first consider one-component annihilation and creation fields $\phi^+(x)$ and $\phi^-(x)$ that transform as the simplest of all representations of the Lorentz group, the scalar, with $D(\Lambda) = 1$. Restricted to rotations, this is just the scalar representation of the rotation group, for which $\mathcal{J} = 0$, so Eqs. (5.1.25) and (5.1.26) have no solutions except for $j = 0$, in which case $\sigma, \bar{\sigma}$ take only the value zero. Thus a scalar field can only describe particles of zero spin. Assuming also for the moment that the field describes only a single species of particle, with no distinct antiparticle (and dropping the species label n as well as the spin label σ and the field label ℓ), the quantities $u_\ell(0\sigma n)$ and $v_\ell(0\sigma n)$ are here just the numbers $u(0)$ and $v(0)$. It is conventional to adjust the overall scales of the annihilation and creation fields so that these constants both have the values $(2m)^{-1/2}$. Eqs. (5.1.21) and (5.1.22) then give simply

$$u(\mathbf{p}) = (2p^0)^{-1/2} \tag{5.2.1}$$

and

$$v(\mathbf{p}) = (2p^0)^{-1/2} . \tag{5.2.2}$$

The fields (5.1.17) and (5.1.18) are then, in the scalar case,

$$\phi^+(x) = \int d^3p\, (2\pi)^{-3/2}(2p^0)^{-1/2} a(\mathbf{p}) e^{ip\cdot x} \tag{5.2.3}$$

and

$$\phi^-(x) = \int d^3p\, (2\pi)^{-3/2}(2p^0)^{-1/2} a^\dagger(\mathbf{p}) e^{-ip\cdot x} = \phi^{+\dagger}(x). \tag{5.2.4}$$

A Hamiltonian density $\mathcal{H}(x)$ that is formed as a polynomial in $\phi^+(x)$ and $\phi^-(x)$ will automatically satisfy the requirement (5.1.9), that it transform as a scalar. It remains to satisfy the other condition for the Lorentz invariance of the S-matrix, that $\mathcal{H}(x)$ commute with $\mathcal{H}(y)$ at space-like separations $x - y$. If $\mathcal{H}(x)$ were a polynomial in $\phi^+(x)$ alone, there would be no problem. All annihilation operators commute or anticommute, so $\phi^+(x)$ either commutes or anticommutes with $\phi^+(y)$ for all x and y, according to whether the particle is a boson or fermion, respectively:

$$[\phi^+(x), \phi^+(y)]_\mp = 0 . \tag{5.2.5}$$

Hence any $\mathcal{H}(x)$ formed as a polynomial in $\phi^+(x)$ (or, for fermions, any such even polynomial) will commute with $\mathcal{H}(y)$ for all x and y. The problem, of course, is that, in order to be Hermitian, $\mathcal{H}(x)$ must involve $\phi^{+\dagger}(x) = \phi^-(x)$ as well as $\phi^+(x)$, and $\phi^+(x)$ does not commute or anticommute with $\phi^-(y)$ for general space-like separations. Using the commutation (for bosons) or anticommutation (for fermions) relations

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

(4.2.5), we have

$$[\phi^+(x), \phi^-(y)]_\mp = \int \frac{d^3p\, d^3p'}{(2\pi)^3(2p^0 \cdot 2p'^0)^{1/2}} e^{ip\cdot x} e^{-ip'\cdot y} \delta^3(\mathbf{p} - \mathbf{p}'),$$

which collapses to the single integral

$$[\phi^+(x), \phi^-(y)]_\mp = \Delta_+(x - y), \tag{5.2.6}$$

where Δ_+ is a standard function:

$$\Delta_+(x) \equiv \frac{1}{(2\pi)^3} \int \frac{d^3p}{2p^0} e^{ip\cdot x}. \tag{5.2.7}$$

This is manifestly Lorentz-invariant, and therefore for space-like x it can depend only on the invariant square $x^2 > 0$. We can thus evaluate $\Delta_+(x)$ for space-like x by choosing the coordinate system so that

$$x^0 = 0, \quad |\mathbf{x}| = \sqrt{x^2}.$$

Eq. (5.2.7) then gives

$$\begin{aligned}\Delta_+(x) &= \frac{1}{(2\pi)^3} \int \frac{d^3p}{2\sqrt{\mathbf{p}^2 + m^2}} e^{i\mathbf{p}\cdot\mathbf{x}} \\ &= \frac{4\pi}{(2\pi)^3} \int_0^\infty \frac{p^2 dp}{2\sqrt{\mathbf{p}^2 + m^2}} \frac{\sin(p\sqrt{x^2})}{p\sqrt{x^2}}.\end{aligned}$$

Changing the variable of integration to $u \equiv p/m$, this is

$$\Delta_+(x) = \frac{m}{4\pi^2\sqrt{x^2}} \int_0^\infty \frac{u\,du}{\sqrt{u^2 + 1}} \sin(m\sqrt{x^2}u) \tag{5.2.8}$$

or, in terms of a standard Hankel function,

$$\Delta_+(x) = \frac{m}{4\pi^2\sqrt{x^2}} K_1\left(m\sqrt{x^2}\right). \tag{5.2.9}$$

This isn't zero, so what are we to do with it? Note that even though $\Delta_+(x)$ is not zero, for $x^2 > 0$ it is even in x^μ. Instead of using only $\phi^+(x)$, suppose we try to construct $\mathscr{H}(x)$ out of a linear combination

$$\phi(x) \equiv \kappa\phi^+(x) + \lambda\phi^-(x).$$

Using Eq. (5.2.6), we have then for $x - y$ space-like

$$\begin{aligned}[\phi(x), \phi^\dagger(y)]_\mp &= |\kappa|^2[\phi^+(x), \phi^-(y)]_\mp + |\lambda|^2[\phi^-(x), \phi^+(y)]_\mp \\ &= (|\kappa|^2 \mp |\lambda|^2)\Delta_+(x - y)\end{aligned}$$

$$\begin{aligned}[\phi(x), \phi(y)]_\mp &= \kappa\lambda([\phi^+(x), \phi^-(y)]_\mp + [\phi^-(x), \phi^+(y)]_\mp) \\ &= \kappa\lambda(1 \mp 1)\Delta_+(x - y).\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Both of these will vanish if and only if the particle is a *boson* (i.e., it is the top sign that applies) and κ and λ are equal in magnitude

$$|\kappa| = |\lambda| .$$

We can change the relative phase of κ and λ by redefining the phases of the states so that $a(\mathbf{p}) \to e^{i\alpha}a(\mathbf{p})$, $a^\dagger(\mathbf{p}) \to e^{-i\alpha}a^\dagger(\mathbf{p})$, and hence $\kappa \to \kappa e^{i\alpha}$, $\lambda \to \lambda e^{-i\alpha}$. Taking $\alpha = \frac{1}{2}\text{Arg}(\lambda/\kappa)$, we can in this way make κ and λ equal in phase, and hence equal.

Redefining $\phi(x)$ to absorb the overall factor $\kappa = \lambda$, we have then

$$\phi(x) = \phi^+(x) + \phi^{+\dagger}(x) = \phi^\dagger(x) . \tag{5.2.10}$$

The interaction density $\mathscr{H}(x)$ will commute with $\mathscr{H}(y)$ at space-like separations $x - y$ if formed as a normal-ordered polynomial in the self-adjoint scalar field $\phi(x)$.

Even though the choice of the relative phase of the two terms in Eq. (5.2.10) is a matter of convention, it is a convention that once adopted must be used wherever a scalar field for this particle appears in the interaction Hamiltonian density. For instance, suppose that the interaction density involved not only the field (5.2.10), but also another scalar field for the same particle

$$\tilde{\phi}(x) = e^{i\alpha}\phi^+(x) + e^{-i\alpha}\phi^{+\dagger}(x)$$

with α an arbitrary phase. This $\tilde{\phi}$, like ϕ, would be causal in the sense that $\tilde{\phi}(x)$ commutes with $\tilde{\phi}(y)$ when $x - y$ is space-like, but $\tilde{\phi}(x)$ would not commute with $\phi(y)$ at space-like separations, and therefore we cannot have both of these fields appearing in the same theory.

If the particles that are destroyed and created by $\phi(x)$ carry some conserved quantum number like electric charge, then $\mathscr{H}(x)$ will conserve the quantum number if and only if each term in $\mathscr{H}(x)$ contains equal numbers of operators $a(\mathbf{p})$ and $a(\mathbf{p})^\dagger$. But this is impossible if $\mathscr{H}(x)$ is formed as a polynomial in $\phi(x) = \phi^+(x) + \phi^{+\dagger}(x)$. To put this another way, in order that $\mathscr{H}(x)$ should commute with the charge operator Q (or some other symmetry generator) it is necessary that it be formed out of fields that have simple commutation relations with Q. This is true for $\phi^+(x)$ and its adjoint, for which

$$[Q, \phi^+(x)]_- = -q\phi^+(x) ,$$
$$[Q, \phi^{+\dagger}(x)]_- = +q\phi^{+\dagger}(x) ,$$

but not for the self-adjoint field (5.2.10).

In order to deal with this problem, we must suppose that there are *two* spinless bosons, with the same mass m, but charges $+q$ and $-q$, respectively. Let $\phi^+(x)$ and $\phi^{+c}(x)$ denote the annihilation fields for these

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

two particles, so that*

$$[Q, \phi^+(x)]_- = -q\phi^+(x)\,,$$
$$[Q, \phi^{+c}(x)]_- = +q\phi^{+c}(x)\,.$$

Define $\phi(x)$ as the linear combination

$$\phi(x) = \kappa\phi^+(x) + \lambda\phi^{+c\dagger}(x)\,,$$

which manifestly has the same commutator with Q as $\phi^+(x)$ alone

$$[Q, \phi(x)]_- = -q\phi(x)\,.$$

The commutator or anticommutator of $\phi(x)$ with its adjoint is then, at space-like separation

$$\begin{aligned}[\phi(x), \phi^\dagger(y)]_\mp &= |\kappa|^2[\phi^+(x), \phi^{+\dagger}(y)] + |\lambda|^2[\phi^{+c\dagger}(x), \phi^{+c}(y)]_\mp \\ &= (|\kappa|^2 \mp |\lambda|^2)\Delta_+(x-y)\,,\end{aligned}$$

while $\phi(x)$ and $\phi(y)$ automatically commute or anticommute with each other for all x and y because ϕ^+ and $\phi^{+c\dagger}$ destroy and create different particles. In deriving this result, we have tacitly assumed that the particle and antiparticle have the same mass, so that the commutators or anticommutators involve the same function $\Delta_+(x-y)$. Fermi statistics is again ruled out here, because it is not possible that $\phi(x)$ should anticommute with $\phi^\dagger(y)$ at space-like separations unless $\kappa = \lambda = 0$, in which case the fields simply vanish. So a spinless particle must be a boson.

For Bose statistics, in order that a complex $\phi(x)$ should commute with $\phi^\dagger(y)$ at space-like separations, it is necessary and sufficient that $|\kappa|^2 = |\lambda|^2$, as well as for the particle and antiparticle to have the same mass. By redefining the relative phase of states of these two particles, we can again give κ and λ the same phase, in which case $\kappa = \lambda$. This common factor can again be eliminated by a redefinition of the field ϕ, so that

$$\phi(x) = \phi^+(x) + \phi^{c+\dagger}(x)$$

or in more detail

$$\phi(x) = \int \frac{d^3p}{(2\pi)^{3/2}(2p^0)^{1/2}}\left[a(\mathbf{p})e^{ip\cdot x} + a^{c\dagger}(\mathbf{p})e^{-ip\cdot x}\right]. \tag{5.2.11}$$

This is the essentially unique causal scalar field. This formula can be used both for purely neutral spinless particles that are their own antiparticles (in which case we take $a^c(\mathbf{p}) = a(\mathbf{p})$), and for particles with distinct antiparticles (for which $a^c(\mathbf{p}) \neq a(\mathbf{p})$).

* The label 'c' denotes 'charge conjugate'. It should be kept in mind that a particle that carries no conserved quantum numbers may or may not be its own antiparticle, with $a^c(\mathbf{p}) = a(\mathbf{p})$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

For future use, we note here that the commutator of the complex scalar field with its adjoint is

$$[\phi(x), \phi^\dagger(y)] = \Delta(x-y)\,, \tag{5.2.12}$$

where

$$\Delta(x-y) \equiv \Delta_+(x-y) - \Delta_+(y-x) = \int \frac{d^3p}{2p^0(2\pi)^3}[e^{ip\cdot(x-y)} - e^{-ip\cdot(x-y)}]. \tag{5.2.13}$$

Let's now consider the effect of the various inversion symmetries on this field. First, from the results of Section 4.2, we can readily see that the effect of the space-inversion operator on the annihilation and creation operators is:**

$$\mathsf{P}a(\mathbf{p})\mathsf{P}^{-1} = \eta^* a(-\mathbf{p})\,, \tag{5.2.14}$$

$$\mathsf{P}a^{c\dagger}(\mathbf{p})\mathsf{P}^{-1} = \eta^c a^{c\dagger}(-\mathbf{p})\,, \tag{5.2.15}$$

where η and η^c are the intrinsic parities of the particle and antiparticle, respectively. Applying these results to the annihilation field (5.2.3) and the charge-conjugate of the creation field (5.2.4), and changing the variable of integration from $\mathbf{p}$ to $-\mathbf{p}$, we see that

$$\mathsf{P}\phi^+(x)\mathsf{P}^{-1} = \eta^*\phi^+(\mathscr{P}x) \tag{5.2.16}$$

$$\mathsf{P}\phi^{+c\dagger}(x)\mathsf{P}^{-1} = \eta^c\phi^{+c\dagger}(\mathscr{P}x)\,, \tag{5.2.17}$$

where as before $\mathscr{P}x = (-\mathbf{x}, x^0)$. We see that in general applying the space inversion to the scalar field $\phi(x) = \phi^+(x) + \phi^{+c\dagger}(x)$ would give a different field $\phi_P = \eta^*\phi^+ + \eta^c\phi^{+c\dagger}$. Both fields are separately causal, but if ϕ and $\phi_P^\dagger$ appear in the same interaction then we are in trouble, because in general they do not commute at space-like separations. The only way to preserve Lorentz invariance as well as parity conservation and the hermiticity of the interaction is to require that ϕ_P be proportional to ϕ, and hence that

$$\eta^c = \eta^*. \tag{5.2.18}$$

That is, *the intrinsic parity $\eta\eta^c$ of a state containing a spinless particle and its antiparticle is even.* We have now simply

$$\mathsf{P}\phi(x)\mathsf{P}^{-1} = \eta^*\phi(\mathscr{P}x)\,. \tag{5.2.19}$$

** We are omitting the subscript 0 on inversion operators P, C, and T, because in virtually all cases where these inversions are good symmetries, the same operators induce inversion transformations on 'in' and 'out' states and on free-particle states.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

These results also apply when the spinless particle is its own antiparticle, for which $\eta^c = \eta$, and imply that the intrinsic parity of such a particle is real: $\eta = \pm 1$.

Charge-conjugation can be handled in much the same way. From the results of Section 4.2, we have

$$\mathsf{C}a(\mathbf{p})\mathsf{C}^{-1} = \xi^* a^c(\mathbf{p}) , \tag{5.2.20}$$

$$\mathsf{C}a^{c\dagger}(\mathbf{p})\mathsf{C}^{-1} = \xi^c a^\dagger(\mathbf{p}) , \tag{5.2.21}$$

where ξ and ξ^c are the phases associated with the operation of charge-conjugation on one-particle states. It follows then that

$$\mathsf{C}\phi^+(x)\mathsf{C}^{-1} = \xi^* \phi^{+c}(x) , \tag{5.2.22}$$

$$\mathsf{C}\phi^{+c\dagger}(x)\mathsf{C}^{-1} = \xi^c \phi^{+\dagger}(x) . \tag{5.2.23}$$

In order that $\mathsf{C}\phi(x)\mathsf{C}^{-1}$ should be proportional to the field $\phi^\dagger(x)$ with which it commutes at space-like separations, it is evidently necessary that

$$\xi^c = \xi^* . \tag{5.2.24}$$

Just as for ordinary parity, the intrinsic charge-conjugation parity $\xi\xi^c$ of a state consisting of a spinless particle and its antiparticle is even. We now have simply

$$\mathsf{C}\phi(x)\mathsf{C}^{-1} = \xi^* \phi^\dagger(x) . \tag{5.2.25}$$

Again, these results apply also in the case where the particle is its own antiparticle, where $\xi^c = \xi$. In this case the charge-conjugation parity like the ordinary parity must be real, $\xi = \pm 1$.

Finally we come to time-reversal. From Section 4.2 we have

$$\mathsf{T}a(\mathbf{p})\mathsf{T}^{-1} = \zeta^* a(-\mathbf{p}) , \tag{5.2.26}$$

$$\mathsf{T}a^{c\dagger}(\mathbf{p})\mathsf{T}^{-1} = \zeta^c a^{c\dagger}(-\mathbf{p}) . \tag{5.2.27}$$

Recalling that T is antiunitary, and again changing the variable of integration from $\mathbf{p}$ to $-\mathbf{p}$, we find that

$$\mathsf{T}\phi^+(x)\mathsf{T}^{-1} = \zeta^* \phi^+(-\mathscr{P}x) \tag{5.2.28}$$

$$\mathsf{T}\phi^{+c\dagger}(x)\mathsf{T}^{-1} = \zeta^c \phi^{+c\dagger}(-\mathscr{P}x) . \tag{5.2.29}$$

In order for $\mathsf{T}\phi(x)\mathsf{T}^{-1}$ to be simply related to the field ϕ at the time-reversed point $-\mathscr{P}x$, we must have

$$\zeta^c = \zeta^* \tag{5.2.30}$$

and then

$$\mathsf{T}\phi(x)\mathsf{T}^{-1} = \zeta^* \phi(-\mathscr{P}x) . \tag{5.2.31}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

5.3 Causal Vector Fields

We now take up the next simplest kind of field, which transforms as a four-vector, the simplest non-trivial representation of the homogeneous Lorentz group. There are massive particles, the $W^{\pm}$ and Z^0, that at low energies are described by such fields and that play an increasing role in modern elementary particle physics, so this example is not merely of pedagogical interest. (Also, although we are here considering only massive particles, one approach to quantum electrodynamics is to describe the photon in terms of a massive vector field in the limit of very small mass). For the moment we will suppose that only one species of particle is described by this field (dropping the species label n); then we shall consider the possibility that the field describes both a particle and a distinct antiparticle.

In the four-vector representation of the Lorentz group, the rows and columns of the representation matrices $D(\Lambda)$ are labelled with four-component indices μ, ν, etc., with

$$D(\Lambda)^{\mu}{}_{\nu} \equiv \Lambda^{\mu}{}_{\nu}. \tag{5.3.1}$$

The annihilation and creation parts of the vector field are written:

$$\phi^{+\mu}(x) = \sum_{\sigma} (2\pi)^{-3/2} \int d^3p \, u^{\mu}(\mathbf{p}, \sigma) \, a(\mathbf{p}, \sigma) \, e^{ip \cdot x} \,, \tag{5.3.2}$$

$$\phi^{-\mu}(x) = \sum_{\sigma} (2\pi)^{-3/2} \int d^3p \, v^{\mu}(\mathbf{p}, \sigma) \, a^{\dagger}(\mathbf{p}, \sigma) \, e^{-ip \cdot x} \,. \tag{5.3.3}$$

The coefficient functions $u^{\mu}(\mathbf{p}, \sigma)$ and $v^{\mu}(\mathbf{p}, \sigma)$ for arbitrary momentum are given in terms of those for zero momentum by Eqs. (5.1.21) and (5.1.22), which here read:

$$u^{\mu}(\mathbf{p}, \sigma) = (m/p^0)^{1/2} L(p)^{\mu}{}_{\nu} u^{\nu}(0, \sigma) \,, \tag{5.3.4}$$

$$v^{\mu}(\mathbf{p}, \sigma) = (m/p^0)^{1/2} L(p)^{\mu}{}_{\nu} v^{\nu}(0, \sigma) \,. \tag{5.3.5}$$

(We are using the usual summation convention for spacetime indices μ, ν, etc.) Also, the coefficient functions at zero momentum are subject to the conditions (5.1.25) and (5.1.26):

$$\sum_{\bar{\sigma}} u^{\mu}(0, \bar{\sigma}) \mathbf{J}^{(j)}_{\bar{\sigma}\sigma} = \mathscr{J}^{\mu}{}_{\nu} u^{\nu}(0, \sigma) \tag{5.3.6}$$

and

$$-\sum_{\bar{\sigma}} v^{\mu}(0, \bar{\sigma}) \mathbf{J}^{(j)*}_{\bar{\sigma}\sigma} = \mathscr{J}^{\mu}{}_{\nu} v^{\nu}(0, \sigma) \,. \tag{5.3.7}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

The rotation generators $\mathcal{J}^{\mu}{}_{\nu}$ in the four-vector representation are given by Eq. (5.3.1) as

$$(\mathcal{J}_k)^0{}_0 = (\mathcal{J}_k)^0{}_i = (\mathcal{J}_k)^i{}_0 = 0\,, \tag{5.3.8}$$

$$(\mathcal{J}_k)^i{}_j = -i\epsilon_{ijk}\,, \tag{5.3.9}$$

with i, j, k here running over the values 1, 2, and 3. We note in particular that $\mathcal{J}^2$ takes the form

$$(\mathcal{J}^2)^0{}_0 = (\mathcal{J}^2)^0{}_i = (\mathcal{J}^2)^i{}_0 = 0\,, \tag{5.3.10}$$

$$(\mathcal{J}^2)^i{}_j = 2\delta^i{}_j\,. \tag{5.3.11}$$

From Eqs. (5.3.6) and (5.3.7) it follows then that

$$\sum_{\bar{\sigma}} u^0(0,\bar{\sigma})(\mathbf{J}^{(j)})^2_{\bar{\sigma}\sigma} = 0\,, \tag{5.3.12}$$

$$\sum_{\bar{\sigma}} u^i(0,\bar{\sigma})(\mathbf{J}^{(j)})^2_{\bar{\sigma}\sigma} = 2u^i(0,\sigma) \tag{5.3.13}$$

and

$$\sum_{\bar{\sigma}} v^0(0,\bar{\sigma})(\mathbf{J}^{(j)*})^2_{\bar{\sigma}\sigma} = 0\,, \tag{5.3.14}$$

$$\sum_{\bar{\sigma}} v^i(0,\bar{\sigma})(\mathbf{J}^{(j)*})^2_{\bar{\sigma}\sigma} = 2v^i(0,\sigma)\,. \tag{5.3.15}$$

Also, we recall the familiar result that $(\mathbf{J}^{(j)})^2_{\bar{\sigma}\sigma} = j(j+1)\delta_{\bar{\sigma}\sigma}$. From Eqs. (5.3.12)–(5.3.15) we see that there are just two possibilities for the spin of the particle described by the vector field: either $j = 0$, for which at $\mathbf{p} = 0$ only u^0 and v^0 are non-zero, or else $j = 1$ (so that $j(j+1) = 2$), for which at $\mathbf{p} = 0$ only the space-components u^i and v^i are non-zero. Let us look in a little more detail at each of these two possibilities.

Spin Zero

By an appropriate choice of normalization of the fields, we can take the only non-vanishing component of $u^\mu(0)$ and $v^\mu(0)$ to have the conventional values:

$$u^0(0) = i(m/2)^{1/2}$$
$$v^0(0) = -i(m/2)^{1/2}\,.$$

(The label σ here takes only the single value zero, and is therefore dropped.) Then Eqs. (5.3.4) and (5.3.5) yield for general momenta

$$u^\mu(\mathbf{p}) = ip^\mu(2p^0)^{-1/2} \tag{5.3.16}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and

$$v^{\mu}(\mathbf{p}) = -ip^{\mu}(2p^0)^{-1/2}. \tag{5.3.17}$$

The vector annihilation and creation fields here are nothing but the derivatives of the scalar annihilation and creation fields $\phi^{\pm}$ for a spinless particle that were defined in the previous section:

$$\phi^{+\mu}(x) = \partial^{\mu}\phi^{+}(x) \qquad \phi^{-\mu}(x) = \partial^{\mu}\phi^{-}(x)\,. \tag{5.3.18}$$

It is obvious that the causal vector field for a spinless particle is also just the derivative of the causal scalar field:

$$\phi^{\mu}(x) = \phi^{+\mu}(x) + \phi^{-\mu}(x) = \partial^{\mu}\phi(x)\,. \tag{5.3.19}$$

Hence we need not explore this case any further here.

Spin One

From Eqs. (5.3.6) and (5.3.7) we see immediately that the vectors $u^i(0,0)$ and $v^i(0,0)$ for $\sigma = 0$ are in the 3-direction. By a suitable normalization of the fields, we can take these vectors to have the values

$$u^{\mu}(0,0) = v^{\mu}(0,0) = (2m)^{-1/2}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{5.3.20}$$

with four-vector components listed always in the order 1, 2, 3, 0. To find the other components, we use Eqs. (5.3.6), (5.3.7), and (5.3.9) to calculate the effect of the raising and lowering operators $J_1^{(1)} \pm iJ_2^{(1)}$ on u and v. This gives:

$$u^{\mu}(0,+1) = -v^{\mu}(0,-1) = -\frac{1}{\sqrt{2}}(2m)^{-1/2}\begin{bmatrix} 1 \\ +i \\ 0 \\ 0 \end{bmatrix}, \tag{5.3.21}$$

$$u^{\mu}(0,-1) = -v^{\mu}(0,+1) = \frac{1}{\sqrt{2}}(2m)^{-1/2}\begin{bmatrix} 1 \\ -i \\ 0 \\ 0 \end{bmatrix}. \tag{5.3.22}$$

Applying Eqs. (5.3.4) and (5.3.5) now yields

$$u^{\mu}(\mathbf{p},\sigma) = v^{\mu *}(\mathbf{p},\sigma) = (2p^0)^{-1/2}e^{\mu}(\mathbf{p},\sigma) \tag{5.3.23}$$

where

$$e^{\mu}(\mathbf{p},\sigma) \equiv L^{\mu}{}_{\nu}(\mathbf{p})e^{\nu}(0,\sigma) \tag{5.3.24}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

with

$$e^{\mu}(0,0) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad e^{\mu}(0,+1) = -\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ +i \\ 0 \\ 0 \end{bmatrix}, \quad e^{\mu}(0,-1) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \\ 0 \\ 0 \end{bmatrix}. \tag{5.3.25}$$

The annihilation and creation fields (5.3.2) and (5.3.3) here are

$$\phi^{+\mu}(x) = \phi^{-\mu\dagger}(x) = (2\pi)^{-3/2} \sum_{\sigma} \int \frac{d^3p}{\sqrt{2p^0}} e^{\mu}(\mathbf{p},\sigma)\, a(\mathbf{p},\sigma)\, e^{ip\cdot x}. \tag{5.3.26}$$

The fields $\phi^{+\mu}(x)$ and $\phi^{+\nu}(y)$ of course commute (or anticommute) for all x and y, but $\phi^{+\mu}(x)$ and $\phi^{-\nu}(y)$ do not. Their commutator (for bosons) or anticommutator (for fermions) is

$$[\phi^{+\mu}(x), \phi^{-\nu}(y)]_{\mp} = \int \frac{d^3p}{(2\pi)^3 2p^0} e^{ip\cdot(x-y)}\, \Pi^{\mu\nu}(\mathbf{p}) \tag{5.3.27}$$

where

$$\Pi^{\mu\nu}(\mathbf{p}) \equiv \sum_{\sigma} e^{\mu}(\mathbf{p},\sigma) e^{\nu *}(\mathbf{p},\sigma). \tag{5.3.28}$$

A straightforward calculation using Eq. (5.3.25) shows that $\Pi^{\mu\nu}(0)$ is the projection matrix on the space orthogonal to the time-direction, and Eq. (5.3.24) then shows that $\Pi^{\mu\nu}(\mathbf{p})$ is the projection matrix on the space orthogonal to the four-vector p^{μ}:

$$\Pi^{\mu\nu}(\mathbf{p}) = \eta^{\mu\nu} + p^{\mu}p^{\nu}/m^2. \tag{5.3.29}$$

The commutator (or anticommutator) (5.3.27) may then be written in terms of the Δ_+ function defined in the previous section, as

$$[\phi^{+\mu}(x), \phi^{-\nu}(y)]_{\mp} = \left[\eta^{\mu\nu} - \frac{\partial^{\mu}\partial^{\nu}}{m^2}\right] \Delta_+(x-y). \tag{5.3.30}$$

For our present purposes, the important thing about this expression is that for $x - y$ space-like it does not vanish and is *even* in $x - y$. We can therefore repeat the reasoning of the previous section in seeking to construct a causal field: we form a linear combination of annihilation and creation fields

$$v^{\mu}(x) \equiv \kappa\phi^{+\mu}(x) + \lambda\phi^{-\mu}(x)$$

for which, for $x - y$ spacelike,

$$[v^{\mu}(x), v^{\nu}(y)]_{\mp} = \kappa\lambda\,[1 \mp 1] \left[\eta^{\mu\nu} - \frac{\partial^{\mu}\partial^{\nu}}{m^2}\right] \Delta_+(x-y)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and

$$[v^{\mu}(x), v^{\nu\dagger}(y)]_{\mp} = (|\kappa|^2 \mp |\lambda|^2)\left[\eta^{\mu\nu} - \frac{\partial^{\mu}\partial^{\nu}}{m^2}\right]\Delta_{+}(x-y).$$

In order for both to vanish for space-like $x-y$, it is necessary and sufficient that the spin one particles be *bosons* and that $|\kappa| = |\lambda|$. By a suitable choice of phase of the one-particle states we can give κ and λ the same phase, so that $\kappa = \lambda$, and then drop the common factor κ by redefining the overall normalization of the field. After all this, we find that the causal vector field for a massive particle of spin one is

$$v^{\mu}(x) = \phi^{+\mu}(x) + \phi^{+\mu\dagger}(x)\,. \tag{5.3.31}$$

We note that this is real:

$$v^{\mu}(x) = v^{\mu\dagger}(x)\,. \tag{5.3.32}$$

However, if the particles it describes carry a non-zero value of some conserved quantum number Q, then we cannot construct an interaction that conserves Q out of such a field. Instead, we must suppose that there is another boson of the same mass and spin which carries an opposite value of Q, and construct the causal field as

$$v^{\mu}(x) = \phi^{+\mu}(x) + \phi^{+c\mu\dagger}(x) \tag{5.3.33}$$

or in more detail

$$v^{\mu}(x) = (2\pi)^{-3/2}\sum_{\sigma}\int \frac{d^3p}{\sqrt{2p^0}} \times [e^{\mu}(\mathbf{p},\sigma)a(\mathbf{p},\sigma)e^{ip\cdot x} + e^{\mu *}(\mathbf{p},\sigma)a^{c\dagger}(\mathbf{p},\sigma)e^{-ip\cdot x}]\,, \tag{5.3.34}$$

where the superscript c indicates operators that create the antiparticle that is charge-conjugate to the particle annihilated by $\phi^{+\mu}(x)$. This again is a causal field, but no longer real. We can also use this formula for the case of a purely neutral, spin one particle that is its own antiparticle, by simply setting $a^{c}(\mathbf{p}) = a(\mathbf{p})$. In either case, the commutator of a vector field with its adjoint is

$$[v^{\mu}(x), v^{\nu\dagger}(y)] = \left[\eta^{\mu\nu} - \frac{\partial^{\mu}\partial^{\nu}}{m^2}\right]\Delta(x-y)\,, \tag{5.3.35}$$

where $\Delta(x-y)$ is the function (5.2.13).

The real and complex fields we have constructed for a massive, spin one particle satisfy interesting field equations. First, since p^{μ} in the exponential in Eq. (5.3.26) satisfies $p^2 = -m^2$, the field satisfies the Klein–Gordon equation:

$$(\Box - m^2)v^{\mu}(x) = 0\,, \tag{5.3.36}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

just as for the scalar field. In addition, since Eq. (5.3.24) shows that

$$e^{\mu}(\mathbf{p},\sigma)\,p_{\mu}=0\,, \tag{5.3.37}$$

we now have another equation

$$\partial_{\mu}v^{\mu}(x)=0\,. \tag{5.3.38}$$

In the limit of small mass, Eqs. (5.3.36) and (5.3.38) are just the equations for the potential four-vector of electrodynamics in what is called Lorentz gauge.

However, we cannot obtain electrodynamics from just any theory of massive spin one particles by letting the mass go to zero. The trouble can be seen by considering the rate of production of a spin one particle by an interaction density $\mathscr{H}=J_{\mu}v^{\mu}$, where J_{μ} is an arbitrary four-vector current. Squaring the matrix element and summing over the spin z-components of the spin one particle gives a rate proportional to

$$\sum_{\sigma}|<J_{\mu}>e^{\mu}(\mathbf{p},\sigma)^{*}|^{2}=<J_{\mu}><J_{\nu}>^{*}\,\Pi^{\mu\nu}(\mathbf{p})\,,$$

where $\mathbf{p}$ is the momentum of the emitted spin one particle, and $<J_{\mu}>$ is the matrix element of the current (say, at $x=0$) between the initial and final states of all other particles. The term $p^{\mu}p^{\nu}/m^{2}$ in $\Pi^{\mu\nu}(\mathbf{p})$ will, in general, cause the emission rate to blow up when $m\to 0$. The only way to avert this catastrophe is to suppose that $<J_{\mu}>p^{\mu}$ vanishes, which in coordinate space is just the statement that the current J^{μ} must be *conserved*, in the sense that $\partial_{\mu}J^{\mu}=0$. Indeed, the need for conservation of the current can be seen by simply counting states. A massive spin one particle has three spin states, which can be taken as the states with helicity $+1$, 0, and -1, while any massless, spin one particle like the photon can only have helicities $+1$ and -1: the current conservation condition just ensures that the helicity zero states of the spin one particle are not emitted in the limit of zero mass.

The inversions can be dealt with in much the same way as for the scalar field discussed in the previous section. To evaluate the effect of space inversion, we need a formula for $e^{\mu}(-\mathbf{p},\sigma)$. Using $L^{\mu}{}_{\nu}(-\mathbf{p})=\mathscr{P}^{\mu}{}_{\rho}L^{\rho}{}_{\tau}(\mathbf{p})\mathscr{P}^{\tau}{}_{\nu}$ and Eq. (5.3.24), we have

$$e^{\mu}(-\mathbf{p},\sigma)=-\mathscr{P}^{\mu}{}_{\nu}e^{\nu}(\mathbf{p},\sigma). \tag{5.3.39}$$

Also, to evaluate the effect of time-reversal we need a formula for $(-1)^{1+\sigma}e^{\mu *}(-\mathbf{p},-\sigma)$. Using $(-1)^{1+\sigma}e^{\mu\ *}(0,-\sigma)=-e^{\mu}(0,\sigma)$ and the above formula for $L^{\mu}{}_{\nu}(-\mathbf{p})$, we find

$$(-1)^{1+\sigma}e^{\mu *}(-\mathbf{p},-\sigma)=\mathscr{P}^{\mu}{}_{\nu}\,e^{\nu}(\mathbf{p},\sigma). \tag{5.3.40}$$

Using these results and the transformation properties of the annihilation and creation operators given in Section 4.2, it is straightforward to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

work out the inversion transformation properties of the annihilation and creation fields. Once again we find that, in order for causal fields to be transformed into other fields with which they commute at space-like separations, it is necessary that the intrinsic space inversion, charge-conjugation, and time-reversal phases for spin one particles and their antiparticles be related by

$$\eta^{c} = \eta^{*} , \tag{5.3.41}$$

$$\xi^{c} = \xi^{*} , \tag{5.3.42}$$

$$\zeta^{c} = \zeta^{*} . \tag{5.3.43}$$

(In particular all phases must be real if the spin one particle is its own antiparticle.) With these phase conditions satisfied, our causal vector field (5.3.34) has the inversion transformation properties

$$\mathsf{P}v^{\mu}(x)\mathsf{P}^{-1} = -\eta^{*}\mathscr{P}^{\mu}{}_{\nu}v^{\nu}(\mathscr{P}x) , \tag{5.3.44}$$

$$\mathsf{C}v^{\mu}(x)\mathsf{C}^{-1} = \xi^{*}v^{\mu\dagger}(x) , \tag{5.3.45}$$

$$\mathsf{T}v^{\mu}(x)\mathsf{T}^{-1} = \zeta^{*}\mathscr{P}^{\mu}{}_{\nu}v^{\nu}(-\mathscr{P}x) . \tag{5.3.46}$$

In particular, the minus sign in Eq. (5.3.44) means that a vector field that transforms as a polar vector, with no extra phases or signs accompanying the matrix $\mathscr{P}^{\mu}{}_{\nu}$, describes a spin one particle with intrinsic parity $\eta = -1$.

5.4 The Dirac Formalism

Among all the representations of the homogeneous Lorentz group, there is one that plays a special role in physics. As we saw in Section 1.1, this representation was introduced into the theory of the electron by Dirac,[3] but as so often happens it was already known to mathematicians,[4] because it provides the basis of one of the two broad classes of representations of the rotation or Lorentz groups (actually, of their covering groups — see Section 2.7) in any number of dimensions. From the point of view we are following here, the structure and properties of any quantum field are dictated by the representation of the homogeneous Lorentz group under which it transforms, so it will be natural for us to describe the Dirac formalism as it first appeared in mathematics, rather than as it was introduced by Dirac.

By a representation of the homogeneous Lorentz group, we mean a set of matrices $D(\Lambda)$ satisfying the group multiplication law

$$D(\bar{\Lambda})D(\Lambda) = D(\bar{\Lambda}\Lambda) .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Just as for the unitary operators $U(\Lambda)$, we can study the properties of these matrices by considering the infinitesimal case,

$$\Lambda^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu} , \tag{5.4.1}$$

$$\omega_{\mu\nu} = -\omega_{\nu\mu} , \tag{5.4.2}$$

for which

$$D(\Lambda) = 1 + \frac{i}{2}\,\omega_{\mu\nu}\mathcal{J}^{\mu\nu} \tag{5.4.3}$$

with $\mathcal{J}^{\mu\nu} = -\mathcal{J}^{\nu\mu}$ a set of matrices satisfying the commutation relations (2.4.12):

$$i[\mathcal{J}^{\mu\nu}, \mathcal{J}^{\rho\sigma}] = \eta^{\nu\rho}\mathcal{J}^{\mu\sigma} - \eta^{\mu\rho}\mathcal{J}^{\nu\sigma} - \eta^{\sigma\mu}\mathcal{J}^{\rho\nu} + \eta^{\sigma\nu}\mathcal{J}^{\rho\mu} . \tag{5.4.4}$$

To find such a set of matrices, suppose we first construct matrices γ^{μ} that satisfy the *anti*commutation relations

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2\eta^{\mu\nu} \tag{5.4.5}$$

and tentatively define

$$\mathcal{J}^{\mu\nu} = -\frac{i}{4}[\gamma^{\mu}, \gamma^{\nu}] , \tag{5.4.6}$$

It is elementary, using Eq. (5.4.5), to show that

$$[\mathcal{J}^{\mu\nu}, \gamma^{\rho}] = -i\gamma^{\mu}\eta^{\nu\rho} + i\gamma^{\nu}\eta^{\mu\rho} \tag{5.4.7}$$

and from this we easily see that Eq. (5.4.6) does indeed satisfy the desired commutation relation Eq. (5.4.4). We shall further assume that the matrices γ_{μ} are *irreducible*; that is, that there is no proper subspace that is left invariant by all these matrices. Otherwise we could choose some smaller set of field components, which would transform as in Eqs. (5.4.3) and (5.4.6), with an irreducible set of γ_{μ}s.

Any set of matrices satisfying a relation like Eq. (5.4.5) (or its Euclidean analog, with $\eta_{\mu\nu}$ replaced with a Kronecker delta) is called a *Clifford algebra*. The importance in mathematics of this particular representation of the homogeneous Lorentz group (or, more accurately, its covering group) arises from the fact (shown in Section 5.6) that the most general irreducible representation of the Lorentz group is either a tensor, or a spinor transforming as in Eqs. (5.4.3) and (5.4.6), or a direct product of a spinor and a tensor.

The commutation relation (5.4.7) can be summarized by saying that γ^{ρ} is a *vector*, in the sense that Eq. (5.4.3) satisfies

$$D(\Lambda)\gamma^{\rho}D^{-1}(\Lambda) = \Lambda_{\sigma}{}^{\rho}\gamma^{\sigma} . \tag{5.4.8}$$

In the same sense, the unit matrix is trivially a *scalar*

$$D(\Lambda)\,1\,D^{-1}(\Lambda) = 1 \tag{5.4.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and Eq. (5.4.4) shows that $\mathscr{J}^{\rho\sigma}$ is an antisymmetric *tensor*

$$D(\Lambda)\mathscr{J}^{\rho\sigma}D^{-1}(\Lambda) = \Lambda_\mu{}^\rho \Lambda_\nu{}^\sigma \mathscr{J}^{\mu\nu} . \tag{5.4.10}$$

The matrices γ^μ can be used to construct other totally antisymmetric tensors

$$\mathscr{A}^{\rho\sigma\tau} \equiv \gamma^{[\rho}\gamma^\sigma\gamma^{\tau]} , \tag{5.4.11}$$

$$\mathscr{P}^{\rho\sigma\tau\eta} = \gamma^{[\rho}\gamma^\sigma\gamma^\tau\gamma^{\eta]} . \tag{5.4.12}$$

The brackets here are a standard notation, indicating that we are to sum over all permutations of the indices within the brackets, with a plus or minus sign for even or odd permutations, respectively. For instance, Eq. (5.4.11) is shorthand for

$$\begin{aligned}\mathscr{A}^{\rho\sigma\tau} \equiv\ & \gamma^\rho\gamma^\sigma\gamma^\tau - \gamma^\rho\gamma^\tau\gamma^\sigma - \gamma^\sigma\gamma^\rho\gamma^\tau \\ & + \gamma^\tau\gamma^\rho\gamma^\sigma + \gamma^\sigma\gamma^\tau\gamma^\rho - \gamma^\tau\gamma^\sigma\gamma^\rho .\end{aligned}$$

By repeated use of Eq. (5.4.5) we can write any product of γs as a sum of antisymmetrized products of γs times a product of metric tensors, so the totally antisymmetric tensors form a complete basis for the set of all matrices that can be constructed from the Dirac matrices.

This formalism automatically contains a parity transformation, conventionally taken as

$$\beta \equiv i\gamma^0 . \tag{5.4.13}$$

Applied to the Dirac matrices, this gives

$$\beta\gamma^i\beta^{-1} = -\gamma^i , \qquad \beta\gamma^0\beta^{-1} = +\gamma^0 . \tag{5.4.14}$$

(We here label indices so that μ runs over values $0, 1, 2, \cdots$.) The same similarity transformation, applied to any product of γ-matrices, then yields just a plus or minus sign, according to whether the product contains an even or an odd number of γs with space-like indices, respectively. In particular,

$$\beta\mathscr{J}^{ij}\beta^{-1} = \mathscr{J}^{ij} , \tag{5.4.15}$$

$$\beta\mathscr{J}^{i0}\beta^{-1} = -\mathscr{J}^{i0} . \tag{5.4.16}$$

Everything so far in this section applies in any number of spacetime dimensions and for any 'metric' $\eta_{\mu\nu}$. In four spacetime dimensions, however, there is a special feature, that no totally antisymmetric tensor can have more than four indices, so the sequence of tensors $\mathbf{1}, \gamma^\rho, \mathscr{J}^{\rho\sigma}, \mathscr{A}^{\rho\sigma\tau}, \cdots$ terminates with the tensor (5.4.12). Furthermore, each of these tensors transforms differently under Lorentz and/or parity transformations so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

they are all linearly independent.* The number of linearly independent components of these tensors is one for **1**, four for γ^ρ, six for $\mathscr{J}^{\rho\sigma}$, four for $\mathscr{A}^{\rho\sigma\tau}$, and one for $\mathscr{P}^{\mu\nu\rho\sigma}$, or 16 independent components in all. (The general rule is that a totally antisymmetric tensor with n indices in d dimensions has a number of independent components equal to the binomial coefficient $d!/n!(d-n)!$) There are at most v^2 independent $v \times v$ matrices, so they must have at least $\sqrt{16} = 4$ rows and columns. Dirac matrices of the minimum dimensionality are necessarily irreducible; if reducible, the subspace left invariant by these matrices would furnish a representation of lower dimensionality. We shall therefore take the γ^μ to be 4×4 matrices.

(More generally, in any even number d of spacetime dimensions, one can form antisymmetric tensors with $0, 1, \cdots, d$ indices, which contain altogether a number of independent components equal to

$$\sum_{n=0}^{d} \frac{d!}{n!(d-n)!} = 2^d ,$$

so the γ-matrices must have at least $2^{d/2}$ rows and columns. In spaces or spacetimes with odd dimensionality, the totally antisymmetric tensors of rank n and $d-n$ can be linearly related by the conditions**

$$\gamma^{[\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_r]} \propto \epsilon^{\mu_1\mu_2\cdots\mu_d}\gamma_{[\mu_{r+1}}\gamma_{\mu_{r+2}}\cdots\gamma_{\mu_d]},$$

for $r = 0, 1, 2, ..., d-1$, with $\epsilon^{\mu_1\mu_2\cdots\mu_d}$ totally antisymmetric, and the left-hand side taken as the unit matrix for $r = 0$. Under these conditions there are only 2^{d-1} independent tensors, requiring γ-matrices of dimensionality at least $2^{(d-1)/2}$.)

Returning now to four spacetime dimensions, we shall choose an explicit set of 4×4 γ-matrices. One very convenient choice is

$$\gamma^0 = -i\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{\gamma} = -i\begin{bmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{bmatrix}, \tag{5.4.17}$$

where **1** is the unit 2×2 matrix, and the components of $\boldsymbol{\sigma}$ are the usual

* Alternatively, these matrices can be shown to be linearly independent by noting that they form an orthogonal set, with the scalar product of two matrices defined by the trace of their product. Note that none of these matrices can vanish, because each component of each of these tensors is proportional to a product of different γ-matrices, and such a product has a square equal to plus or minus the product of the corresponding squares, and hence equal to ± 1.

** This constraint does not interfere with the inclusion of space inversion in the Dirac representation of the Lorentz group in odd-dimensional spacetime, because here the tensor $\epsilon^{\mu_1\mu_2\cdots\mu_d}$ is even under inversion of space coordinates. If we don't care about space inversion, we can also construct $2^{(d-1)/2}$-dimensional irreducible representations of the proper orthochronous Lorentz group in even spacetime dimensions by imposing the above condition relating antisymmetrized products of r and $d-r$ Dirac matrices. An example is provided by the submatrices in Eqs. (5.4.19) and (5.4.20) below.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.4.18}$$

(The σ_i are just the 2×2 γ-matrices in three dimensions.) It can be shown[5] that any other irreducible set of γ-matrices are related to these by a similarity transformation. From Eq. (5.4.17), we can easily calculate the Lorentz group generators (5.4.6):

$$\mathscr{J}^{ij} = \frac{1}{2}\epsilon_{ijk}\begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{bmatrix} \tag{5.4.19}$$

$$\mathscr{J}^{i0} = +\frac{i}{2}\begin{bmatrix} \sigma_i & 0 \\ 0 & -\sigma_i \end{bmatrix}. \tag{5.4.20}$$

(Here ϵ_{ijk} is the totally antisymmetric tensor in three dimensions, with $\epsilon_{123} \equiv +1$.) We note that these are block-diagonal, so the Dirac matrices provide a *reducible* representation of the proper orthochronous Lorentz group, the direct sum of two irreducible representation with $\mathscr{J}^{ij} = \pm i\epsilon_{ijk}\mathscr{J}^{k0}$.

It is convenient to write the totally antisymmetric tensors (5.4.11) and (5.4.12) in a somewhat simpler way. The matrix (5.4.12) is totally antisymmetric, and therefore proportional to the pseudotensor $\epsilon^{\rho\sigma\tau\eta}$, defined as a totally antisymmetric quantity with $\epsilon^{0123} = +1$. Setting ρ, σ, τ, η equal to 0,1,2,3, respectively, we see that

$$\mathscr{P}^{\rho\sigma\tau\eta} = 4!\, i\epsilon^{\rho\sigma\tau\eta}\gamma_5, \tag{5.4.21}$$

where

$$\gamma_5 \equiv -i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{5.4.22}$$

The matrix γ_5 is a pseudoscalar in the sense that

$$[\mathscr{J}^{\rho\sigma}, \gamma_5] = 0, \tag{5.4.23}$$

$$\beta\gamma_5\beta^{-1} = -\gamma_5. \tag{5.4.24}$$

Similarly, $\mathscr{A}^{\rho\sigma\tau}$ must be proportional to $\epsilon^{\rho\sigma\tau\eta}$ contracted with some matrix $\mathscr{A}_\eta$, and by setting ρ, σ, τ equal in turn to 0,1,2 or 0,1,3 or 0,2,3 or 1,2,3, we find

$$\mathscr{A}^{\rho\sigma\tau} = 3!\, i\epsilon^{\rho\sigma\tau\eta}\gamma_5\gamma_\eta. \tag{5.4.25}$$

The 16 independent 4×4 matrices can therefore be taken as the components of the scalar 1, the vector γ^ρ, the antisymmetric tensor $\mathscr{J}^{\rho\sigma}$, the 'axial' vector $\gamma_5\gamma_\eta$, and the pseudoscalar γ_5. It is easy to see that the matrix γ_5 has unit square

$$\gamma_5{}^2 = 1 \tag{5.4.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and anticommutes with all γ^μ

$$\{\gamma_5, \gamma^\mu\} = 0\,. \tag{5.4.27}$$

The notation γ_5 is particularly appropriate, because the anticommutation relations (5.4.26) and (5.4.27), together with Eq. (5.4.5) show that $\gamma^0, \gamma^1, \gamma^2, \gamma^3, \gamma_5$ provide a Clifford algebra in five spacetime dimensions. For the particular 4×4 representation (5.4.17) of the γ-matrices, the matrix γ_5 is

$$\gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,. \tag{5.4.28}$$

This representation is convenient because it reduces $\mathcal{J}^{\rho\sigma}$ and γ_5 to block-diagonal form. As we shall see, this makes it particularly useful for dealing with particles in the ultra-relativistic limit, $v \to c$. (It is not, however, the representation described in Section 1.1 that was originally introduced by Dirac, because Dirac was mostly interested in electrons in atoms where $v << c$, and in this case it is more convenient to adopt a representation for which γ^0 rather than γ_5 is diagonal.)

The representation of the homogeneous Lorentz group we have constructed here is not unitary, because the generators $\mathcal{J}^{\rho\sigma}$ are not all represented by Hermitian matrices. In particular, in the representation (5.4.17) we have $\mathcal{J}^{ij}$ Hermitian, but $\mathcal{J}^{i0}$ is anti-Hermitian. Such reality conditions can conveniently be written in a manifestly Lorentz-invariant fashion by introducing the matrix $\beta \equiv i\gamma^0$ of Eq. (5.4.13), which in the representation (5.4.17) takes the form

$$\beta = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\,. \tag{5.4.29}$$

Inspection of Eq. (5.4.17) shows that

$$\beta\gamma^{\mu\dagger}\beta = -\gamma^\mu \tag{5.4.30}$$

and it follows then that

$$\beta\mathcal{J}^{\rho\sigma\dagger}\beta = \mathcal{J}^{\rho\sigma}\,. \tag{5.4.31}$$

Hence, though not unitary, the matrices $D(\Lambda)$ satisfy the pseudounitarity relation

$$\beta\, D(\Lambda)^\dagger \beta = D(\Lambda)^{-1}\,. \tag{5.4.32}$$

Also, γ_5 is Hermitian and anticommutes with β so

$$\beta{\gamma_5}^\dagger\beta = -\gamma_5 \tag{5.4.33}$$

and it follows that

$$\beta(\gamma_5\gamma_\mu)^\dagger\beta = -\gamma_5\gamma_\mu. \tag{5.4.34}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

The Dirac and related matrices also have important symmetry properties. Inspection of Eqs. (5.4.17) and (5.4.18) shows that γ_μ is symmetric for $\mu = 0, 2$ and antisymmetric for $\mu = 1, 3$, so

$$\gamma_\mu^{\mathrm{T}} = -\mathscr{C}\gamma_\mu\mathscr{C}^{-1}\,, \tag{5.4.35}$$

where T denotes a transpose, and

$$\mathscr{C} \equiv \gamma_2\beta = -i\begin{bmatrix} \sigma_2 & 0 \\ 0 & -\sigma_2 \end{bmatrix}. \tag{5.4.36}$$

It follows immediately that

$$\mathscr{J}_{\mu\nu}^{\mathrm{T}} = -\mathscr{C}\mathscr{J}_{\mu\nu}\mathscr{C}^{-1}\,, \tag{5.4.37}$$

$$\gamma_5^{\mathrm{T}} = +\mathscr{C}\gamma_5\mathscr{C}^{-1}\,, \tag{5.4.38}$$

$$(\gamma_5\gamma_\mu)^{\mathrm{T}} = +\mathscr{C}\gamma_5\gamma_\mu\mathscr{C}^{-1}\,. \tag{5.4.39}$$

These signs will prove significant when we consider the charge-conjugation properties of various currents in the next section. Of course, we can combine our results for adjoints and transposes to obtain the complex conjugates of the Dirac and allied matrices:

$$\gamma_\mu^* = \beta\mathscr{C}\gamma_\mu\mathscr{C}^{-1}\beta\,, \tag{5.4.40}$$

$$\mathscr{J}_{\mu\nu}^* = -\beta\mathscr{C}\mathscr{J}_{\mu\nu}\mathscr{C}^{-1}\beta\,, \tag{5.4.41}$$

$$\gamma_5^* = -\beta\mathscr{C}\gamma_5\mathscr{C}^{-1}\beta\,, \tag{5.4.42}$$

$$(\gamma_5\gamma_\mu)^* = -\beta\mathscr{C}\gamma_5\gamma_\mu\mathscr{C}^{-1}\beta. \tag{5.4.43}$$

5.5 Causal Dirac Fields

We now want to construct particle annihilation and antiparticle creation fields that transform under the Lorentz group according to the Dirac representation of this group, discussed in the previous section. In general these take the form given in Eqs. (5.1.17) and (5.1.18):

$$\psi_\ell^+(x) = (2\pi)^{-3/2}\sum_\sigma \int d^3p\; u_\ell(\mathbf{p},\sigma)e^{ip\cdot x}a(\mathbf{p},\sigma) \tag{5.5.1}$$

and

$$\psi_\ell^{-c}(x) = (2\pi)^{-3/2}\sum_\sigma \int d^3p\; v_\ell(\mathbf{p},\sigma)e^{-ip\cdot x}a^{c\dagger}(\mathbf{p},\sigma). \tag{5.5.2}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

with the particle species label omitted here. In order to calculate the coefficient functions $u_\ell(\mathbf{p},\sigma)$ and $v_\ell(\mathbf{p},\sigma)$ appearing in these formulas, we must first use Eqs. (5.1.25) and (5.1.26) to find u_ℓ and v_ℓ for zero momentum, and then apply Eqs. (5.1.21) and (5.1.22) to calculate them for arbitrary momenta, with $D_{\bar{\ell}\ell}(\Lambda)$ in both cases taken as the 4×4 Dirac representation of the homogeneous Lorentz group discussed in the previous section.

Using Eq. (5.4.19), the zero-momentum conditions (5.1.25) and (5.1.26) read*

$$\sum_{\bar{\sigma}} u_{\bar{m}\pm}(0,\bar{\sigma})\mathbf{J}^{(j)}_{\bar{\sigma}\sigma} = \sum_{m} \tfrac{1}{2}\boldsymbol{\sigma}_{\bar{m}m} u_{m\pm}(0,\sigma)$$

and

$$-\sum_{\bar{\sigma}} v_{\bar{m}\pm}(0,\bar{\sigma})\mathbf{J}_{\bar{\sigma}\sigma}{}^{(j)*} = \sum_{m} \tfrac{1}{2}\boldsymbol{\sigma}_{\bar{m}m} v_{m\pm}(0,\sigma)\,.$$

In other words, if we regard $u_{m\pm}(0,\sigma)$ and $v_{m\pm}(0,\sigma)$ as the m,σ elements of matrices $U_\pm$ and $V_\pm$, we have in matrix notation

$$U_\pm \mathbf{J}^{(j)} = \tfrac{1}{2}\boldsymbol{\sigma} U_\pm \tag{5.5.3}$$

and

$$-V_\pm \mathbf{J}^{(j)*} = \tfrac{1}{2}\boldsymbol{\sigma} V_\pm. \tag{5.5.4}$$

Now, the $(2j+1)$-dimensional matrices $\mathbf{J}^{(j)}$ and $-\mathbf{J}^{(j)*}$ and the 2×2 matrices $\frac{1}{2}\boldsymbol{\sigma}$ all provide irreducible representations of the Lie algebra of the rotation group. A general theorem of group theory known as Schur's lemma[6] tells us that when a matrix like $U_\pm$ or $V_\pm$ connects two such representations as in Eqs. (5.5.3) and (5.5.4), the matrix must either vanish (a possibility of no interest here) or else be square and non-singular. Hence the Dirac field can only describe particles of spin $j = \frac{1}{2}$ (so that $2j+1 = 2$) and the matrices $\mathbf{J}^{(1/2)}$ and $-\mathbf{J}^{(1/2)*}$ must be the same as $\frac{1}{2}\boldsymbol{\sigma}$ up to a similarity transformation. In fact, in the standard representation (2.5.21), (2.5.22) of the rotation generators, we have $\mathbf{J}^{(1/2)} = \frac{1}{2}\boldsymbol{\sigma}$ and $-\mathbf{J}^{(1/2)*} = \frac{1}{2}\sigma_2\boldsymbol{\sigma}\sigma_2$. It follows then that $U_\pm$ and $V_\pm\sigma_2$ must commute with $\boldsymbol{\sigma}$, and hence must be proportional to the unit matrix:

$$u_{m,\pm}(0,\sigma) = c_\pm \delta_{m\sigma}\,, \qquad v_{m,\pm}(0,\sigma) = -id_\pm(\sigma_2)_{m\sigma}\,. \tag{5.5.5}$$

* We are here dropping the species label n, and replacing the four-component index ℓ with a pair of indices, one 2-valued index m labelling the rows and columns of the submatrices in Eqs. (5.4.19) and (5.4.20), and a second index taking values $\pm$, labelling the rows and columns of the supermatrix in Eqs. (5.4.19) and (5.4.20).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

In other words

$$u(0, \tfrac{1}{2}) = \begin{bmatrix} c_+ \\ 0 \\ c_- \\ 0 \end{bmatrix}, \qquad u(0, -\tfrac{1}{2}) = \begin{bmatrix} 0 \\ c_+ \\ 0 \\ c_- \end{bmatrix},$$

$$v(0, \tfrac{1}{2}) = \begin{bmatrix} 0 \\ d_+ \\ 0 \\ d_- \end{bmatrix}, \qquad v(0, -\tfrac{1}{2}) = -\begin{bmatrix} d_+ \\ 0 \\ d_- \\ 0 \end{bmatrix}$$

and the spinors at finite momentum are

$$u(\mathbf{p}, \sigma) = \sqrt{m/p^0}\, D\Big(L(p)\Big) u(0, \sigma)\,, \tag{5.5.6}$$

$$v(\mathbf{p}, \sigma) = \sqrt{m/p^0}\, D\Big(L(p)\Big) v(0, \sigma)\,. \tag{5.5.7}$$

It now only remains to say something about the constants $c_\pm$ and $d_\pm$. In general, these are quite arbitrary — we could even choose c_- and d_- or c_+ and d_+ to be zero if we liked, so that the Dirac field would have only *two* non-vanishing components. The only physical principle that could tell us anything about the relative values of the $c_\pm$ or the $d_\pm$ is the conservation of parity. We recall that under a space inversion, the particle annihilation and antiparticle creation operators undergo the transformations:

$$\mathsf{P} a(\mathbf{p}, \sigma) \mathsf{P}^{-1} = \eta^* a(-\mathbf{p}, \sigma) \tag{5.5.8}$$

$$\mathsf{P} a^{c\dagger}(\mathbf{p}, \sigma) \mathsf{P}^{-1} = \eta^c a^{c\dagger}(-\mathbf{p}, \sigma) \tag{5.5.9}$$

and so

$$\mathsf{P} \psi_\ell^+(x) \mathsf{P}^{-1} = \eta^* (2\pi)^{-3/2} \sum_\sigma \int d^3p\; u_\ell(-\mathbf{p}, \sigma) e^{ip\cdot\mathscr{P}x} a(\mathbf{p}, \sigma)\,, \tag{5.5.10}$$

$$\mathsf{P} \psi_\ell^{-c}(x) \mathsf{P}^{-1} = \eta^c (2\pi)^{-3/2} \sum_\sigma \int d^3p\; v_\ell(-\mathbf{p}, \sigma) e^{-ip\cdot\mathscr{P}x} a^{c\dagger}(\mathbf{p}, \sigma). \tag{5.5.11}$$

Also, Eqs. (5.4.16), (5.1.21), and (5.1.22) give

$$u(-\mathbf{p}, \sigma) = \sqrt{m/p^0}\, \beta D(L(\mathbf{p})) \beta u(0, \sigma) \tag{5.5.12}$$

$$v(-\mathbf{p}, \sigma) = \sqrt{m/p^0}\, \beta D(L(\mathbf{p})) \beta v(0, \sigma)\,. \tag{5.5.13}$$

(Since $\beta^2 = 1$, we are no longer making a distinction between β and β^{-1}.) In order that the parity operator should transform the annihilation and creation fields at the point x into something proportional to these fields

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

at $\mathscr{P}x$, it is necessary that $\beta u(0,\sigma)$ and $\beta v(0,\sigma)$ be proportional to $u(0,\sigma)$ and $v(0,\sigma)$, respectively:

$$\beta u(0,\sigma) = b_u u(0,\sigma)\,, \qquad \beta v(0,\sigma) = b_v v(0,\sigma)\,, \tag{5.5.14}$$

where b_u and b_v are sign factors, $b_u^2 = b_v^2 = 1$. In this case, the fields have the simple space-inversion properties:

$$\mathsf{P}\psi^+(x)\mathsf{P}^{-1} = \eta^* b_u \beta \psi^+(\mathscr{P}x)\,, \tag{5.5.15}$$

$$\mathsf{P}\psi^{-c}(x)\mathsf{P}^{-1} = \eta^c b_v \beta \psi^{-c}(\mathscr{P}x). \tag{5.5.16}$$

By adjusting the overall scales of the fields, we can choose the coefficient functions at zero momentum to have the form:

$$u(0,\tfrac{1}{2}) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\b_u\\0\end{bmatrix}, \qquad u(0,-\tfrac{1}{2}) = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\b_u\end{bmatrix}, \tag{5.5.17}$$

$$v(0,\tfrac{1}{2}) = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\b_v\end{bmatrix}, \qquad v(0,-\tfrac{1}{2}) = \frac{-1}{\sqrt{2}}\begin{bmatrix}1\\0\\b_v\\0\end{bmatrix}. \tag{5.5.18}$$

Now let's try to put together the annihilation and creation fields in a linear combination

$$\psi(x) = \kappa\psi^+(x) + \lambda\psi^{-c}(x) \tag{5.5.19}$$

that commutes or anticommutes with itself and its adjoint at space-like separations. A straightforward calculation gives

$$[\psi_\ell(x), \psi^\dagger_{\bar{\ell}}(y)]_\mp = (2\pi)^{-3}\int d^3p\,[|\kappa|^2 N_{\ell\bar{\ell}}(\mathbf{p})e^{ip\cdot(x-y)} \mp |\lambda|^2 M_{\ell\bar{\ell}}(\mathbf{p})e^{-ip\cdot(x-y)}]\,, \tag{5.5.20}$$

where

$$N_{\ell\bar{\ell}}(\mathbf{p}) \equiv \sum_\sigma u_\ell(\mathbf{p},\sigma)u^*_{\bar{\ell}}(\mathbf{p},\sigma)\,, \tag{5.5.21}$$

$$M_{\ell\bar{\ell}}(\mathbf{p}) \equiv \sum_\sigma v_\ell(\mathbf{p},\sigma)v^*_{\bar{\ell}}(\mathbf{p},\sigma)\,. \tag{5.5.22}$$

By using either the eigenvalue conditions (5.5.14) or the explicit formulas (5.5.17) and (5.5.18), we find at zero momentum:

$$N(0) = \frac{1+b_u\beta}{2}\,, \qquad M(0) = \frac{1+b_v\beta}{2}\,. \tag{5.5.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

From Eqs. (5.5.6) and (5.5.7) we have then

$$N(\mathbf{p}) = \frac{m}{2p^0} D\Big(L(p)\Big)[1 + b_u\beta]D^\dagger\Big(L(p)\Big)\,, \tag{5.5.24}$$

$$M(\mathbf{p}) = \frac{m}{2p^0} D\Big(L(p)\Big)[1 + b_v\beta]D^\dagger\Big(L(p)\Big)\,. \tag{5.5.25}$$

The pseudounitarity condition (5.4.32) yields

$$D\Big(L(p)\Big)\beta D^\dagger\Big(L(p)\Big) = \beta$$

and

$$D\Big(L(p)\Big)D^\dagger\Big(L(p)\Big) = D\Big(L(p)\Big)\beta D^{-1}\Big(L(p)\Big)\beta\,.$$

We also recall that $\beta = i\gamma^0$, so by using the Lorentz transformation rule (5.4.8) we have

$$D\Big(L(p)\Big)\beta D^{-1}\Big(L(p)\Big) = iL_\mu{}^0(p)\gamma^\mu = -ip_\mu\gamma^\mu/m. \tag{5.5.26}$$

Putting this together, we find**

$$N(\mathbf{p}) = \frac{1}{2p^0}[-ip^\mu\gamma_\mu + b_u m]\beta\,, \tag{5.5.27}$$

$$M(\mathbf{p}) = \frac{1}{2p^0}[-ip^\mu\gamma_\mu + b_v m]\beta\,. \tag{5.5.28}$$

Using this in Eq. (5.5.20) yields finally

$$\begin{aligned}[\psi_\ell(x), \psi^\dagger_{\tilde{\ell}}(y)]_\mp = \big(&|\kappa|^2[-\gamma^\mu\partial_\mu + b_u m]\beta\Delta_+(x-y)\\ &\mp|\lambda|^2[-\gamma^\mu\partial_\mu + b_v m]\beta\Delta_+(y-x)\big)_{\ell\tilde{\ell}},\end{aligned} \tag{5.5.29}$$

where Δ_+ is the function introduced in Section 5.2:

$$\Delta_+(x) \equiv \int \frac{d^3p}{2p^0(2\pi)^3} e^{ip\cdot x}\,.$$

We saw in Section 5.2 that for $x - y$ space-like $\Delta_+(x-y)$ is an even function of $x - y$ and, of course, this implies that its first derivatives are odd functions of $x - y$. Hence, in order that both the derivative and the non-derivative terms in the commutator or anticommutator should vanish at space-like separations, it is necessary and sufficient that

$$|\kappa|^2 = \mp|\lambda|^2 \tag{5.5.30}$$

** Sometimes an extra factor $\sqrt{p^0/m}$ is included in the Dirac spinors, so that m appears in place of p^0 in the denominators of the spin sums (5.5.27) and (5.5.28). The normalization convention used here has the advantage that it goes smoothly over to the case $m = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and

$$|\kappa|^2 b_u = \pm|\lambda|^2 b_v \,. \tag{5.5.31}$$

Clearly Eq. (5.5.30) is only possible if we choose the bottom sign, $\mp = +$; that is, *the particles described by a Dirac field must be fermions.* It is also then necessary that $|\kappa|^2 = |\lambda|^2$ and $b_u = -b_v$. Just as for scalars we have the freedom to redefine the relative phase of the creation and annihilation operators to make the ratio κ/λ real and positive, in which case $\kappa = \lambda$, and by adjusting the overall scale and phase of the field ψ we may then take

$$\kappa = \lambda = 1. \tag{5.5.32}$$

Finally if we like we can replace ψ with $\gamma_5\psi$, which changes the sign of both b_u and b_v, so we can always take

$$b_u = -b_v = +1. \tag{5.5.33}$$

For future use, we record here that the Dirac field is now

$$\psi_\ell(x) = (2\pi)^{-3/2} \sum_\sigma \int d^3p \, [u_\ell(\mathbf{p},\sigma) e^{ip\cdot x} a(\mathbf{p},\sigma) + v_\ell(\mathbf{p},\sigma) e^{-ip\cdot x} a^{c\dagger}(\mathbf{p},\sigma)] \tag{5.5.34}$$

while the coefficient functions at zero momentum are

$$u(0,\tfrac{1}{2}) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad u(0,-\tfrac{1}{2}) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \tag{5.5.35}$$

$$v(0,\tfrac{1}{2}) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \qquad v(0,-\tfrac{1}{2}) = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \tag{5.5.36}$$

The spin sums are

$$N(\mathbf{p}) = \frac{1}{2p^0}[-ip^\mu\gamma_\mu + m]\beta \,, \tag{5.5.37}$$

$$M(\mathbf{p}) = \frac{1}{2p^0}[-ip^\mu\gamma_\mu - m]\beta, \tag{5.5.38}$$

so the anticommutator is given by Eq. (5.5.20) as

$$[\psi_\ell(x), \psi^\dagger_{\ell'}(y)]_+ = \{[-\gamma^\mu\partial_\mu + m]\beta\}_{\ell\ell'} \, \Delta(x-y) \,. \tag{5.5.39}$$

Now let's return to the requirement that under a space inversion the field $\psi(x)$ must transform into something proportional to $\psi(\mathscr{P}x)$. For this

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

to be possible the phases in Eqs. (5.5.15) and (5.5.16) must be equal, and so the intrinsic parities of particles and their antiparticles must be related by

$$\eta^c = -\eta^* \,. \tag{5.5.40}$$

That is, *the intrinsic parity $\eta\eta^c$ of a state consisting of a spin $\frac{1}{2}$ particle and its antiparticle is odd.* It is for this reason that negative parity mesons like the ρ^0 and J/ψ can be interpreted as s-wave bound states of quark–antiquark pairs. Eqs. (5.5.15) and (5.5.16) now give the transformation of the causal Dirac field under space inversion as

$$\mathsf{P}\psi(x)\mathsf{P}^{-1} = \eta^*\beta\psi(\mathscr{P}x) \,. \tag{5.5.41}$$

Before going on to the other inversions, this is a good place to mention that Eqs. (5.5.14), (5.5.33), and (5.5.26) show that $u(\mathbf{p},\sigma)$ and $v(\mathbf{p},\sigma)$ are eigenvectors of $-ip^\mu\gamma_\mu/m$ with eigenvalues $+1$ and -1, respectively:

$$(ip^\mu\gamma_\mu + m)u(\mathbf{p},\sigma) = 0 \,, \qquad (-ip^\mu\gamma_\mu + m)v(\mathbf{p},\sigma) = 0 \,. \tag{5.5.42}$$

It follows then that the field (5.5.34) satisfies the differential equation

$$(\gamma^\mu\partial_\mu + m)\psi(x) = 0 \,. \tag{5.5.43}$$

This is the celebrated Dirac equation for a free particle of spin $\frac{1}{2}$. From the point of view adopted here, the free-particle Dirac equation is nothing but a Lorentz-invariant record of the convention that we have used in putting together the two irreducible representations of the proper orthochronous Lorentz group to form a field that transforms simply also under space inversion.

In order to work out the charge-conjugation and time-reversal properties of the Dirac field, we will need expressions for the complex-conjugates of the u and v coefficient functions. These functions are real for zero momentum, but to obtain the coefficient functions at finite momentum we have to multiply with the complex matrix $D(L(p))$. From Eq. (5.4.41) we see that for general real $\omega_{\mu\nu}$:

$$[\exp(\tfrac{1}{2}i\mathscr{J}^{\mu\nu}\omega_{\mu\nu})]^* = \beta\mathscr{C}\exp(\tfrac{1}{2}i\mathscr{J}^{\mu\nu}\omega_{\mu\nu})\mathscr{C}^{-1}\beta$$

and so in particular

$$D(L(p))^* = \beta\mathscr{C}D(L(p))\mathscr{C}^{-1}\beta \,.$$

We also note that $\mathscr{C}^{-1}\beta u(0,\sigma) = -v(0,\sigma)$ and $\mathscr{C}^{-1}\beta v(0,\sigma) = -u(0,\sigma)$, so

$$u^*(\mathbf{p},\sigma) = -\beta\mathscr{C}v(\mathbf{p},\sigma) \,, \tag{5.5.44}$$

$$v^*(\mathbf{p},\sigma) = -\beta\mathscr{C}u(\mathbf{p},\sigma) \,. \tag{5.5.45}$$

In order for the field to transform under charge-conjugation into another field with which it commutes at space-like separations, it is necessary

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

again that the charge conjugation parities of the particle and antiparticle be related by

$$\xi^c = \xi^* \,. \tag{5.5.46}$$

In this case, the field transforms as

$$\mathsf{C}\psi(x)\mathsf{C}^{-1} = -\xi^*\beta\mathscr{C}\psi^*(x) \,. \tag{5.5.47}$$

(We are calling the Hermitian adjoint of the field on the right-hand side ψ^* instead of $\psi^\dagger$ to emphasize that this is still a column vector, not a row.)

Although we have been distinguishing particles from their antiparticles, we have not ruled out the possibility that the two are actually identical. Such spin $\frac{1}{2}$ particles are called *Majorana fermions*. Following the same reasoning that led to Eq. (5.5.47), the Dirac field of such a particle must satisfy the reality condition

$$\psi(x) = -\beta\mathscr{C}\psi^*(x) \,. \tag{5.5.48}$$

For Majorana fermions the intrinsic space-inversion parity must be imaginary, $\eta = \pm i$, while the charge-conjugation parity must be real, $\xi = \pm 1$.

There is an important difference between fermions and bosons in the intrinsic charge-conjugation phase of states consisting of a particle and its antiparticle. Such a state may be written

$$\Phi \equiv \sum_{\sigma,\sigma'} \int d^3p \int d^3p' \; \chi(\mathbf{p},\sigma;\mathbf{p}',\sigma') a^\dagger(\mathbf{p},\sigma)\, a^{c\dagger}(\mathbf{p}',\sigma')\Phi_0 \,,$$

where Φ_0 is the vacuum state. Under charge-conjugation, this state is transformed into

$$\mathsf{C}\Phi = \xi\xi^c \sum_{\sigma,\sigma'} \int d^3p \int d^3p' \; \chi(\mathbf{p},\sigma;\mathbf{p}',\sigma') a^{c\dagger}(\mathbf{p},\sigma)\, a^\dagger(\mathbf{p}',\sigma')\Phi_0 \,.$$

Interchanging the variables of integration and summation and using the anticommutation of the creation operators and Eq. (5.5.46), we can rewrite this as

$$\mathsf{C}\Phi = -\sum_{\sigma,\sigma'} \int d^3p \int d^3p' \; \chi(\mathbf{p}',\sigma';\mathbf{p},\sigma) a^\dagger(\mathbf{p},\sigma)\, a^{c\dagger}(\mathbf{p}',\sigma')\Phi_0 \,.$$

That is, *the intrinsic charge-conjugation parity of a state consisting of a particle described by a Dirac field and its antiparticle is odd*, in the sense that if the wave function χ of the state is even or odd under interchange of the momenta and spins of the particle and antiparticle, then the charge-conjugation operator applied to such a state gives a sign -1 or $+1$, respectively. The classic example here is positronium, the bound state of an electron and a positron. The two lowest states are a pair of nearly degenerate s-wave states with total spin $s = 0$ and $s = 1$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

known respectively as para- and ortho-positronium. The wave function for these two states is even under interchange of momenta and odd or even respectively under interchange of spin z-components, so para- and ortho-positronium have $\mathsf{C} = +1$ and $\mathsf{C} = -1$, respectively. These values are dramatically confirmed in the decay modes of positronium: para-positronium decays rapidly into a pair of photons (each of which has $\mathsf{C} = -1$), while ortho-positronium can only decay much more slowly into three or more photons. In the same way, single ρ^0 and ω^0 mesons are produced as resonances in high-energy electron–positron annihilation through a one-photon intermediate state, so they must have $\mathsf{C} = -1$, which is consistent with their interpretation as quark–antiquark bound states with orbital angular momentum zero and total quark spin one.

Now we come to time-reversal. Recall the transformation properties of the particle annihilation and antiparticle creation operators given by Eq. (4.2.15):

$$\mathsf{T}a(\mathbf{p},\sigma)\mathsf{T}^{-1} = \zeta^*(-1)^{\frac{1}{2}-\sigma}a(-\mathbf{p},-\sigma)\,, \tag{5.5.49}$$

$$\mathsf{T}a^{c\dagger}(\mathbf{p},\sigma)\mathsf{T}^{-1} = \zeta^c(-1)^{\frac{1}{2}-\sigma}a^{c\dagger}(-\mathbf{p},-\sigma)\,. \tag{5.5.50}$$

Time-reversal of the field (5.5.34) thus gives

$$\begin{aligned}\mathsf{T}\psi_\ell(x)\mathsf{T}^{-1} = (2\pi)^{-3/2}\sum_\sigma \int d^3p\,(-1)^{\frac{1}{2}-\sigma}\\ \times\,[\zeta^* u^*_\ell(\mathbf{p},\sigma)e^{-ip\cdot x}a(-\mathbf{p},-\sigma) + \zeta^c v^*_\ell(\mathbf{p},\sigma)e^{ip\cdot x}a^{c\dagger}(-\mathbf{p},-\sigma)]\end{aligned}$$

In order to put this back in the form given for ψ, we shall redefine the variables of integration and summation as $-\mathbf{p}$ and $-\sigma$, so we need formulas for $u^*_\ell(-\mathbf{p},-\sigma)$ and $v^*_\ell(-\mathbf{p},-\sigma)$ in terms of $u_\ell(\mathbf{p},\sigma)$ and $v_\ell(\mathbf{p},\sigma)$, respectively. For this purpose, we can use the fact that $\mathscr{J}^{i0}$ anticommutes with β and commutes with γ_5 together with our former result for $D(L(\mathbf{p}))^*$ to write

$$D^*(L(-\mathbf{p})) = \gamma_5\beta D^*(L(\mathbf{p}))\beta\gamma_5 = \gamma_5\mathscr{C}D(L(\mathbf{p}))\mathscr{C}^{-1}\gamma_5\,.$$

Also, Eqs. (5.4.36) and (5.5.35)–(5.5.36) give

$$\gamma_5\mathscr{C}^{-1}u(0,-\sigma) = (-1)^{\frac{1}{2}-\sigma}u(0,\sigma)\,,$$

$$\gamma_5\mathscr{C}^{-1}v(0,-\sigma) = (-1)^{\frac{1}{2}-\sigma}v(0,\sigma)\,,$$

so

$$(-1)^{\frac{1}{2}-\sigma}u^*(-\mathbf{p},-\sigma) = -\gamma_5\mathscr{C}\,u(\mathbf{p},\sigma)\,, \tag{5.5.51}$$

$$(-1)^{\frac{1}{2}-\sigma}v^*(-\mathbf{p},-\sigma) = -\gamma_5\mathscr{C}\,v(\mathbf{p},\sigma)\,. \tag{5.5.52}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

We see then that in order for time-reversal to take the Dirac field into something proportional to itself at the time-reversed point (with which it would anticommute at space-like separations) it is necessary that the intrinsic time-reversal phases be related by

$$\zeta^c = \zeta^* \tag{5.5.53}$$

and in this case

$$\mathsf{T}\psi(x)\mathsf{T}^{-1} = -\zeta^*\gamma_5\mathscr{C}\psi(-\mathscr{P}x). \tag{5.5.54}$$

Now let us consider how to construct scalar interaction densities out of the Dirac fields and their adjoints. As already mentioned the Dirac representation is not unitary, so $\psi^\dagger\psi$ is not a scalar. To deal with this complication it is convenient to define a new sort of adjoint:

$$\bar{\psi} \equiv \psi^\dagger\beta. \tag{5.5.55}$$

Using the pseudounitarity condition (5.4.32), we see that the fermion bilinears constructed with $\bar{\psi}$ have the Lorentz transformation property

$$U_0(\Lambda)[\bar{\psi}(x)M\psi(x)]U_0^{-1}(\Lambda) = \bar{\psi}(\Lambda x)D(\Lambda)\,M\,D^{-1}(\Lambda)\psi(\Lambda x). \tag{5.5.56}$$

Also, under a space inversion

$$\mathsf{P}[\bar{\psi}(x)M\psi(x)]\mathsf{P}^{-1} = \bar{\psi}(\mathscr{P}x)\beta M\beta\psi(\mathscr{P}x). \tag{5.5.57}$$

Taking the matrix M as $\mathbf{1}$, γ^μ, $\mathscr{J}^{\mu\nu}$, $\gamma_5\gamma^\mu$, or γ_5 yields a bilinear $\bar{\psi}M\psi$ that transforms as a scalar, vector, tensor, axial vector, and pseudoscalar, respectively. (The terms 'axial' and 'pseudo' indicate that these have space-inversion properties opposite to those of ordinary vectors and scalars: a pseudoscalar has negative parity, while the space and time components of an axial vector have positive and negative parity, respectively.) These results apply also when the two fermion fields in the bilinear refer to different particle species, except that in this case a space inversion also yields a ratio of the intrinsic parities.

For instance, the original Fermi theory of beta decay involved an interaction density proportional to $\bar{\psi}_p\gamma^\mu\psi_n\,\bar{\psi}_e\gamma_\mu\psi_\nu$. Later it was realized that the most general Lorentz-invariant and parity-conserving non-derivative beta decay interaction takes the form of a linear combination of products like this, with γ_μ replaced with any one of the five covariant types of 4×4 matrices 1, γ^μ, $\mathscr{J}^{\mu\nu}$, $\gamma_5\gamma^\mu$, or γ_5. (As discussed in Chapter 2, we are defining the space-inversion operator so that the proton, neutron, and electron all have parity $+1$. If the neutrino is massless then its parity may also be defined as $+1$, if necessary by replacing the neutrino field with $\gamma_5\psi_\nu$.) When Lee and Yang[7] called parity conservation into question in 1956, they expanded the list of possible non-derivative interactions to include ten terms proportional to $\bar{\psi}_pM\psi_n\,\bar{\psi}_eM\psi_\nu$ and also $\bar{\psi}_pM\psi_n\,\bar{\psi}_eM\gamma_5\psi_\nu$, with M running over the matrices 1, γ^μ, $\mathscr{J}^{\mu\nu}$, $\gamma_5\gamma^\mu$, or γ_5.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

It is also of some interest to study the charge-conjugation properties of these bilinears. Using Eqs. (5.5.47) and (5.4.35)–(5.4.39), we have

$$\begin{aligned} \mathsf{C}(\bar{\psi}M\psi)\mathsf{C}^{-1} &= (\beta\mathscr{C}\psi)^{\mathrm{T}}\beta M(\beta\mathscr{C}\psi^*) = -(\beta\mathscr{C}\psi^*)^{\mathrm{T}}M^{\mathrm{T}}\mathscr{C}\psi \\ &= \bar{\psi}\mathscr{C}^{-1}M^{\mathrm{T}}\mathscr{C}\psi = \pm\bar{\psi}M\psi \end{aligned} \tag{5.5.58}$$

the sign in the last expression being + for the matrices $\mathbf{1}$, $\gamma_5\gamma_\mu$, and γ_5, and − for γ_μ and $\mathscr{J}_{\mu\nu}$. (The minus sign in the first line arises from Fermi statistics. We ignore a c-number anticommutator.) A boson field that interacts with the current $\bar{\psi}M\psi$ must therefore have $\mathsf{C} = +1$ for scalars, pseudoscalars, or axial vectors, and $\mathsf{C} = -1$ for vectors or antisymmetric tensors. This is one way of seeing that the π^0 (which couples to pseudoscalar or axial-vector nucleon currents) has $\mathsf{C} = +1$, while the photon has $\mathsf{C} = -1$.

5.6 General Irreducible Representations of the Homogeneous Lorentz Group*

We shall now generalize from the special cases of vector and Dirac fields to the case of a field that transforms according to a general irreducible representation of the homogeneous Lorentz group. All fields may be constructed as direct sums of these irreducible fields.

A general representation of the proper orthochronous homogeneous Lorentz group (or, more properly, its infinitesimal part) is provided by a set of matrices $\mathscr{J}_{\mu\nu}$ satisfying the same commutation relations (5.4.4) as the generators of the group

$$[\mathscr{J}_{\mu\nu}, \mathscr{J}_{\rho\sigma}] = i\left(\mathscr{J}_{\rho\nu}\eta_{\sigma\mu} + \mathscr{J}_{\mu\rho}\eta_{\nu\sigma} - \mathscr{J}_{\sigma\nu}\eta_{\rho\mu} - \mathscr{J}_{\mu\sigma}\eta_{\nu\rho}\right) , \tag{5.6.1}$$

(Of course, $\mathscr{J}_{\mu\nu} = -\mathscr{J}_{\nu\mu}$, and indices on $\mathscr{J}_{\mu\nu}$ are as usual raised or lowered by contraction with $\eta^{\mu\nu}$ or $\eta_{\mu\nu}$.) To see how to construct such matrices, first divide the six independent components of $\mathscr{J}_{\mu\nu}$ into two three-vectors: an angular momentum matrix

$$\mathscr{J}_1 = \mathscr{J}_{23} , \qquad \mathscr{J}_2 = \mathscr{J}_{31} , \qquad \mathscr{J}_3 = \mathscr{J}_{12} \tag{5.6.2}$$

and a boost

$$\mathscr{K}_1 = \mathscr{J}_{10} , \qquad \mathscr{K}_2 = \mathscr{J}_{20} , \qquad \mathscr{K}_3 = \mathscr{J}_{30} . \tag{5.6.3}$$

Eq. (5.6.1) then reads

$$[\mathscr{J}_i, \mathscr{J}_j] = i\epsilon_{ijk}\mathscr{J}_k , \tag{5.6.4}$$

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

$$[\mathscr{J}_i, \mathscr{K}_j] = i\epsilon_{ijk}\mathscr{K}_k \,, \tag{5.6.5}$$

$$[\mathscr{K}_i, \mathscr{K}_j] = -i\epsilon_{ijk}\mathscr{J}_k \,, \tag{5.6.6}$$

where i, j, k run over the values $1, 2, 3$, and ϵ_{ijk} is the totally antisymmetric quantity with $\epsilon_{123} \equiv +1$. Eq. (5.6.4) just says that the matrices $\mathscr{J}$ generate a representation of the rotation subgroup of the Lorentz group, and Eq. (5.6.5) just represents the fact that $\mathscr{K}$ is a three-vector. The minus sign in the right-hand side of Eq. (5.6.6) arises from the fact that $\eta_{00} = -1$, and plays a crucial role in what follows.

It is very convenient to replace the matrices $\mathscr{J}$ and $\mathscr{K}$ with two decoupled spin three-vectors, writing

$$\mathscr{A} \equiv \frac{1}{2}(\mathscr{J} + i\mathscr{K}) \,, \tag{5.6.7}$$

$$\mathscr{B} \equiv \frac{1}{2}(\mathscr{J} - i\mathscr{K}) \,. \tag{5.6.8}$$

It is easy to see that the commutation relations (5.6.4)–(5.6.6) are equivalent to

$$[\mathscr{A}_i, \mathscr{A}_j] = i\epsilon_{ijk}\ \mathscr{A}_k \,, \tag{5.6.9}$$

$$[\mathscr{B}_i, \mathscr{B}_j] = i\epsilon_{ijk}\ \mathscr{B}_k \,, \tag{5.6.10}$$

$$[\mathscr{A}_i, \mathscr{B}_j] = 0 \,, \tag{5.6.11}$$

We find matrices satisfying Eqs. (5.6.9)–(5.6.11) in the same way that we find matrices representing the spins of a pair of uncoupled particles — as a direct sum. That is, we label the rows and columns of these matrices with a pair of integers and/or half-integers a, b, running over the values

$$a = -A, -A+1, \cdots, +A \,, \tag{5.6.12}$$

$$b = -B, -B+1, \cdots, +B \tag{5.6.13}$$

and take**

$$(\mathscr{A})_{a'b',ab} = \delta_{b'b}\ \mathbf{J}^{(A)}_{a'a} \,, \tag{5.6.14}$$

$$(\mathscr{B})_{a'b',ab} = \delta_{a'a}\ \mathbf{J}^{(B)}_{b'b} \,, \tag{5.6.15}$$

where $\mathbf{J}^{(A)}$ and $\mathbf{J}^{(B)}$ are the standard spin matrices for spins A or B:

$$\left(\mathbf{J}^{(A)}_3\right)_{a'a} = a\delta_{a'a} \,, \tag{5.6.16}$$

$$\left(\mathbf{J}^{(A)}_1 \pm i\mathbf{J}^{(A)}_2\right)_{a'a} = \delta_{a',a\pm1}\sqrt{(A \mp a)(A \pm a + 1)} \,, \tag{5.6.17}$$

** There is an alternative formalism,[8] based on the fact that the spin j representation of the rotation group can be written as the symmetrized direct product of $2j$ spin $1/2$ representations — i.e., as a symmetric $SU(2)$ tensor with $2j$ two-valued indices. We can therefore write fields belonging to the (A, B) representation with $2A$ two-valued $(1/2, 0)$ indices and $2B$ two-valued $(0, 1/2)$ indices, the latter written with dots to distinguish them from the former.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and likewise for $\mathbf{J}^{(B)}$. The representation is labelled by the values of the positive integers and/or half-integers A and B. We see that the (A,B) representation has dimensionality $(2A+1)(2B+1)$.

The finite-dimensional representations of the homogeneous Lorentz group are not unitary, because $\mathscr{A}$ and $\mathscr{B}$ are Hermitian, and therefore $\mathscr{J}$ is Hermitian but $\mathscr{K}$ is *anti-Hermitian.* This is because of the i in Eqs. (5.6.7) and (5.6.8), which is required by the minus sign in (5.6.6), and hence stems from the fact that the homogeneous Lorentz group is not the same as the four-dimensional rotation group $SO(4)$, a compact group, but instead is the non-compact group known as $SO(3,1)$. It is only compact groups that can have finite-dimensional unitary representations (aside from representations in which the non-compact part is represented trivially, by the identity). There is no problem in working with non-unitary representations, because the objects we are now concerned with are fields, not wave functions, and do not need to have a Lorentz-invariant positive norm.

In contrast, the rotation group is represented unitarily, with its generators represented by the Hermitian matrices

$$\mathscr{J} = \mathscr{A} + \mathscr{B}\,, \tag{5.6.18}$$

By the usual rules of vector addition, we can see that a field that transforms according to the (A,B) representation of the homogeneous Lorentz group has components that rotate like objects of spin j, with

$$j = A+B,\ A+B-1, \cdots,\ |A-B|\,.$$

This is enough to identify the (A,B) representations with the perhaps more familiar tensors and spinors. For instance, a $(0,0)$ field is obviously scalar, with only a single $j=0$ component. A $(\frac{1}{2},0)$ or $(0,\frac{1}{2})$ field can only have $j=+\frac{1}{2}$; these are the top (i.e., $\gamma_5=+1$) and bottom ($\gamma_5=-1$) two components of the Dirac spinor. A $(\frac{1}{2},\frac{1}{2})$ field has components with $j=1$ and $j=0$, corresponding to the spatial part $\mathbf{v}$ and time-component v^0 of a four-vector v^μ. More generally, an (A,A) field contains terms with only integer spins $2A, 2A-1, \cdots, 0$, and corresponds to a traceless symmetric tensor of rank $2A$. (Note that the number of independent components of a symmetric tensor of rank $2A$ in four dimensions is

$$\frac{4\cdot 5\cdots(4+2A-1)}{(2A)!} = \frac{(3+2A)!}{6(2A)!}$$

and the tracelessness condition reduces this to

$$\frac{(3+2A)!}{6(2A)!} - \frac{(1+2A)!}{6(2A-2)!} = (2A+1)^2$$

as expected for an (A,A) field.) One more example: a $(1,0)$ or $(0,1)$ field can only have $j=1$, and corresponds to an antisymmetric tensor $F^{\mu\nu}$ that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

satisfies the further irreducibility 'duality' conditions

$$F^{\mu\nu} = \pm \frac{i}{2} \epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}$$

for $(1,0)$ and $(0,1)$ fields, respectively. Of course, it is only in four dimensions that an antisymmetric two-index tensor $F^{\mu\nu}$ can be divided into such 'self-dual' and 'anti-self-dual' parts.

A general tensor of rank N transforms as the direct product of N $(\frac{1}{2},\frac{1}{2})$ four-vector representations. It can therefore be decomposed (by suitable symmetrizations and antisymmetrizations and extracting traces) into irreducible terms (A,B) with $A = N/2, N/2-1, \cdots$ and $B = N/2, N/2-1, \cdots$. In this way, we can construct any irreducible representation (A,B) for which $A+B$ is an integer. The spin representations, for which $A+B$ is half an odd integer, can similarly be constructed from the direct product of these tensor representations and the Dirac representation $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$. For instance, taking the direct product of the vector $(\frac{1}{2},\frac{1}{2})$ representation and the Dirac $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ representation gives a spinor-vector ψ^μ, that transforms according to the reducible representation

$$(\tfrac{1}{2},\tfrac{1}{2})\otimes[(\tfrac{1}{2},0)\oplus(0,\tfrac{1}{2})] = (\tfrac{1}{2},1)\oplus(\tfrac{1}{2},0)\oplus(1,\tfrac{1}{2})\oplus(0,\tfrac{1}{2}).$$

The quantity $\gamma_\mu\psi^\mu$ would transform as an ordinary $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ Dirac field, so we can isolate the $(\frac{1}{2},1)\oplus(1,\frac{1}{2})$ representation† by requiring that $\gamma_\mu\psi^\mu = 0$. This is the *Rarita–Schwinger field*.[9]

So far in this section we have only considered the representations of the proper orthochronous Lorentz group. In any representation of the Lorentz group including space inversion, there must be a matrix β which reverses the signs of tensors with odd numbers of space indices, and in particular

$$\beta\mathscr{J}\beta^{-1} = +\mathscr{J}\,, \qquad \beta\mathscr{K}\beta^{-1} = -\mathscr{K}\,. \tag{5.6.19}$$

In terms of the matrices (5.6.7) and (5.6.8), this is

$$\beta\mathscr{A}\beta^{-1} = \mathscr{B}\,, \qquad \beta\mathscr{B}\beta^{-1} = \mathscr{A}\,. \tag{5.6.20}$$

Thus an irreducible (A,B) representation of the proper orthochronous homogeneous Lorentz group does not provide a representation of the Lorentz group including space inversion unless $A = B$. As we have seen, these (A,A) representations are the scalar, the vector, and the symmetric traceless tensors. For $A \neq B$, the irreducible representations of the Lorentz

† According to Eq. (5.6.18), such a field transforms under ordinary rotations as a direct sum of two $j = 3/2$ and two $j = \frac{1}{2}$ components. The doubling is eliminated by imposing the Dirac equation $[\gamma^\nu\partial_\nu + m]\psi^\mu = 0$, and the remaining $j = \frac{1}{2}$ component is eliminated by requiring that $\partial_\mu\psi^\mu = 0$. With these conditions, the field describes a single particle of spin $j = 3/2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

group including space inversion are the direct sums $(A,B)\oplus(B,A)$, of dimensionality $2(2A+1)(2B+1)$. One of these is the $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ Dirac representation discussed in Section 5.4. The 4×4 matrix (5.4.29) provides the β-matrix for this representation. Another familiar example is the $(1,0)\oplus(0,1)$ representation, which as we have seen is just the antisymmetric tensor of second rank, including both self-dual and anti-self-dual parts.

5.7 General Causal Fields*

We now proceed to construct causal fields that transform according to the general irreducible (A,B) representations described in the previous section. The index ℓ is replaced here with a pair of indices a,b, running over the ranges (5.6.12), (5.6.13), so the fields are now written as

$$\psi_{ab}(x)=(2\pi)^{-3/2}\sum_{\sigma}\int d^3p\ \Big[\kappa\, a(\mathbf{p},\sigma)e^{ip\cdot x}u_{ab}(\mathbf{p},\sigma) +\lambda\, a^{c\dagger}(\mathbf{p},\sigma)e^{-ip\cdot x}v_{ab}(\mathbf{p},\sigma)\Big] \tag{5.7.1}$$

with κ and λ arbitrary constants. We are here leaving open the possibility that this particle is its own antiparticle, in which case $a^c(\mathbf{p},\sigma)=a(\mathbf{p},\sigma)$.

Our first task is to find the zero-momentum coefficient functions $u_{ab}(0,\sigma)$ and $v_{ab}(0,\sigma)$. The fundamental conditions (5.1.25)–(5.1.26) on $u(0,\sigma)$ and $v(0,\sigma)$ read here

$$\sum_{\bar\sigma}u_{\bar a\bar b}(0,\bar\sigma)\mathbf{J}^{(j)}_{\bar\sigma\sigma}=\sum_{a,b}\mathscr{J}_{\bar a\bar b,ab}u_{ab}(0,\sigma)\,,$$

$$-\sum_{\bar\sigma}v_{\bar a\bar b}(0,\bar\sigma)\mathbf{J}^{(j)*}_{\bar\sigma\sigma}=\sum_{a,b}\mathscr{J}_{\bar a\bar b,ab}v_{ab}(0,\sigma)\,,$$

or using Eqs. (5.6.14)–(5.6.15)

$$\sum_{\bar\sigma}u_{\bar a\bar b}(0,\bar\sigma)\mathbf{J}^{(j)}_{\bar\sigma\sigma}=\sum_{a}\mathbf{J}^{(A)}_{\bar a a}u_{a\bar b}(0,\sigma)+\sum_{b}\mathbf{J}^{(B)}_{\bar b b}u_{\bar a b}(0,\sigma)\,, \tag{5.7.2}$$

$$-\sum_{\bar\sigma}v_{\bar a\bar b}(0,\bar\sigma)\mathbf{J}^{(j)*}_{\bar\sigma\sigma}=\sum_{a}\mathbf{J}^{(A)}_{\bar a a}v_{a\bar b}(0,\sigma)+\sum_{b}\mathbf{J}^{(B)}_{\bar b b}v_{\bar a b}(0,\sigma)\,. \tag{5.7.3}$$

But Eq. (5.7.2) is the defining condition for the Clebsch–Gordan coefficients $C_{AB}(j\sigma;ab)$! These coefficients are defined by the requirement that

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

if Ψ_{ab} are states that under an infinitesimal rotation transform as

$$\delta\Psi_{ab} = i\sum_{\bar{a}} \boldsymbol{\theta}\cdot\mathbf{J}^{(A)}_{\bar{a}a}\Psi_{\bar{a}b} + i\sum_{\bar{b}} \boldsymbol{\theta}\cdot\mathbf{J}^{(B)}_{\bar{b}b}\Psi_{a\bar{b}}$$

then, under the same rotation, the state

$$\Psi^{j}{}_{\sigma} \equiv \sum_{ab} C_{AB}(j\sigma;ab)\Psi_{ab}$$

transforms as

$$\delta\Psi^{j}{}_{\sigma} = i\sum_{\bar{\sigma}} \boldsymbol{\theta}\cdot\mathbf{J}^{(j)}_{\bar{\sigma}\sigma}\Psi^{j}{}_{\bar{\sigma}} .$$

Inspection of Eq. (5.7.2) shows that this requirement is satisfied by the coefficients $u_{ab}(0,\sigma)$, and therefore, up to a possible proportionality factor, $u_{ab}(0,\sigma)$ is just $C_{AB}(j\sigma;ab)$. This constant is conventionally chosen so that

$$u_{ab}(0,\sigma) = (2m)^{-1/2}C_{AB}(j\sigma;ab) . \tag{5.7.4}$$

This result is unique because each irreducible (A,B) representation of the homogeneous Lorentz group contains a given spin j representation of the rotation group at most *once*. Similarly, inspection of Eqs. (5.6.16)–(5.6.17) shows that the complex conjugates of the angular momentum matrices are

$$-\mathbf{J}^{(j)*}_{\sigma\,\sigma'} = (-1)^{\sigma-\sigma'}\mathbf{J}^{(j)}_{-\sigma,-\sigma'} . \tag{5.7.5}$$

Therefore if we write Eq. (5.7.3) in terms of $(-1)^{j-\sigma}v_{ab}(\mathbf{p},-\sigma)$, it takes the same form as Eq. (5.7.2). With a suitable adjustment of a constant factor, the unique solution for $v(0,\sigma)$ is

$$v_{ab}(0,\sigma) = (-1)^{j+\sigma}u_{ab}(0,-\sigma) . \tag{5.7.6}$$

We must now perform a boost to calculate the coefficient functions for finite momentum. For a fixed direction $\hat{\mathbf{p}} \equiv \mathbf{p}/|\mathbf{p}|$, we can write the boost (2.5.24) as a function of a parameter θ defined by

$$\cosh\theta = \sqrt{\mathbf{p}^2+m^2}/m , \qquad \sinh\theta = |\mathbf{p}|/m \tag{5.7.7}$$

and write $L^{\mu}{}_{\nu}(\theta)$ in place of $L^{\mu}{}_{\nu}(p)$, where

$$\begin{aligned} L^{i}{}_{k}(\theta) &= \delta_{ik} + (\cosh\theta - 1)\hat{p}_i\hat{p}_k, \\ L^{i}{}_{0}(\theta) &= L^{0}{}_{i}(p) = \hat{p}_i\sinh\theta, \\ L^{0}{}_{0}(\theta) &= \cosh\theta . \end{aligned} \tag{5.7.8}$$

The advantage of this parameterization is that

$$L(\bar{\theta})L(\theta) = L(\bar{\theta}+\theta). \tag{5.7.9}$$

For infinitesimal θ, we have $[L(\theta)]^{\mu}{}_{\nu} \to \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu}$, where $\omega^{i}{}_{0} = \omega^{0}{}_{i} = \hat{p}_i\theta$ and $\omega^{i}{}_{j} = \omega^{0}{}_{0} = 0$. Following the same reasoning that led from Eq.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

(2.2.24) to Eq. (2.2.26), it follows then that

$$D\Big(L(p)\Big) = \exp(-i\hat{\mathbf{p}} \cdot \mathscr{K}\theta) \,. \tag{5.7.10}$$

This is for any representation of the homogeneous Lorentz group; for the irreducible (A, B) representations, Eqs. (5.6.7) and (5.6.8) give

$$i\mathscr{K} = \mathscr{A} - \mathscr{B} \tag{5.7.11}$$

and since $\mathscr{A}$ and $\mathscr{B}$ are commuting matrices

$$D\Big(L(p)\Big) = \exp(-\hat{\mathbf{p}} \cdot \mathscr{A}\theta) \exp(+\hat{\mathbf{p}} \cdot \mathscr{B}\theta) \,. \tag{5.7.12}$$

In more detail, using Eqs. (5.6.14) and (5.6.15)

$$D\Big(L(p)\Big)_{a'b',ab} = \Big(\exp\big(-\hat{\mathbf{p}} \cdot \mathbf{J}^{(A)}\theta\big)\Big)_{a'a} \Big(\exp\big(+\hat{\mathbf{p}} \cdot \mathbf{J}^{(B)}\theta\big)\Big)_{b'b} \,. \tag{5.7.13}$$

Eqs. (5.7.4) and (5.7.6) then give the coefficient functions at finite momentum as

$$u_{ab}(\mathbf{p}, \sigma) = \frac{1}{\sqrt{2p^0}} \sum_{a'b'} \Big(\exp\big(-\hat{\mathbf{p}} \cdot \mathbf{J}^{(A)}\theta\big)\Big)_{aa'} \Big(\exp\big(+\hat{\mathbf{p}} \cdot \mathbf{J}^{(B)}\theta\big)\Big)_{bb'} \times C_{AB}(j\sigma; a'b') \tag{5.7.14}$$

and

$$v_{ab}(\mathbf{p}, \sigma) = (-1)^{j+\sigma} u_{ab}(\mathbf{p}, -\sigma) \,. \tag{5.7.15}$$

These results give the field explicitly for a given transformation type (A, B), so the field (5.1.31) of this type is unique up to the choice of the constant factors κ and λ.

It is very easy in this formalism to construct Lorentz scalar interaction densities. The (A, B) representation of the homogeneous Lorentz group is just the direct product of the $(A, 0)$ and $(0, B)$ representations, so the general Lorentz transformation rules (5.1.6), (5.1.7) read here

$$U_0(\Lambda)\psi_{ab}(x)U_0^{-1}(\Lambda) = \sum_{a'b'} D^{A0}_{a,a'}(\Lambda^{-1}) D^{0B}_{b,b'}(\Lambda^{-1}) \psi_{a'b'}(\Lambda x) \,. \tag{5.7.16}$$

Furthermore, Eqs. (5.6.14) and (5.6.15) show that the matrix generators of the $(A, 0)$ and $(0, B)$ representations are just the spin matrices for spin A and B, respectively. Thus we can construct scalars of the form

$$\sum_{a_1 a_2 \cdots a_n} \sum_{b_1 b_2 \cdots b_n} g_{a_1 a_2 \cdots a_n; b_1 b_2 \cdots b_n} \psi^{(1)}_{a_1 b_1}(x) \, \psi^{(2)}_{a_2 b_2}(x) \cdots \psi^{(n)}_{a_n b_n}(x) \tag{5.7.17}$$

by simply taking $g_{a_1 a_2 \cdots a_n; b_1 b_2 \cdots b_n}$ as the product of a coefficient for coupling spins $A_1, A_2, \cdots A_n$ to make a scalar and a coefficient for coupling spins $B_1, B_2, \cdots B_n$ to make a scalar. (Even though we do not explicitly consider interactions involving derivatives, we will in this way obtain the most general interaction involving n fields, because the derivative of a field of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

type (A, B) can always be decomposed into fields of other types without derivatives.) For instance, the most general Lorentz scalar formed from a product of three fields of transformation types (A_1, B_1), (A_2, B_2), and (A_3, B_3) is

$$g \sum_{a_1 a_2 a_3} \sum_{b_1 b_2 b_3} \begin{pmatrix} A_1 & A_2 & A_3 \\ a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} B_1 & B_2 & B_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \psi^{(1)}_{a_1 b_1} \psi^{(2)}_{a_2 b_2} \psi^{(3)}_{a_3 b_3} \quad (5.7.18)$$

with a single free parameter g. This is the most general three-field interaction. (The brackets in (5.7.18) denote the Wigner "three-j" symbols:[10]

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \equiv \sum_{m_3'} C_{j_1 j_2}(j_3 m_3', m_1 m_2) C_{j_3 j_3}(00, m_3' m_3)$$

which describe the coupling of three spins to make a rotational scalar.)

For the S-matrix to be Lorentz-invariant it is not enough that the interaction density $\mathscr{H}(x)$ be a scalar like (5.7.18); it is also necessary that $\mathscr{H}(x)$ should commute with $\mathscr{H}(y)$ at space-like separations $x - y$. To see how to satisfy this condition, consider the commutator or anticommutator of two fields for the same particle species, a field ψ of type (A, B), and the adjoint $\tilde{\psi}^\dagger$ of a field $\tilde{\psi}$ of type $(\tilde{A}, \tilde{B})$. We find

$$\left[\psi_{ab}(x), \tilde{\psi}^\dagger_{\tilde{a}\tilde{b}}(y)\right]_\mp = (2\pi)^{-3} \int d^3p \, (2p^0)^{-1} \pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) \times \left[\kappa \tilde{\kappa}^* e^{ip\cdot(x-y)} \mp \lambda \tilde{\lambda}^* e^{-ip\cdot(x-y)}\right], \quad (5.7.19)$$

where $\pi(\mathbf{p})$ is the spin sum

$$(2p^0)^{-1} \pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) \equiv \sum_\sigma u_{ab}(\mathbf{p}, \sigma) \tilde{u}^*_{\tilde{a}\tilde{b}}(\mathbf{p}, \sigma) = \sum_\sigma v_{ab}(\mathbf{p}, \sigma) \tilde{v}^*_{\tilde{a}\tilde{b}}(\mathbf{p}, \sigma) \quad (5.7.20)$$

and as usual, the top and bottom signs are for bosons and fermions, respectively. (We allow here for different coefficients $\tilde{\kappa}$ and $\tilde{\lambda}$ in the $\tilde{\psi}$ field.) In more detail

$$\pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) = \sum_{a'b'} \sum_{\tilde{a}'\tilde{b}'} \sum_\sigma C_{AB}\left(j\sigma; a'b'\right) C_{\tilde{A}\tilde{B}}\left(j\sigma; \tilde{a}'\tilde{b}'\right) \times \left(\exp\left(-\hat{\mathbf{p}} \cdot \mathbf{J}^{(A)}\theta\right)\right)_{aa'} \left(\exp\left(\hat{\mathbf{p}} \cdot \mathbf{J}^{(B)}\theta\right)\right)_{bb'} \times \left(\exp\left(-\hat{\mathbf{p}} \cdot \mathbf{J}^{(A)}\theta\right)\right)^*_{\tilde{a}\tilde{a}'} \left(\exp\left(\hat{\mathbf{p}} \cdot \mathbf{J}^{(B)}\theta\right)\right)^*_{\tilde{b}\tilde{b}'}. \quad (5.7.21)$$

The function $\pi(\mathbf{p})$ has been calculated explicitly.[11] What concerns us here is the fact that it turns out to be the mass-shell value of a *polynomial* function P of $\mathbf{p}$ and p^0:

$$\pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) = P_{ab,\tilde{a}\tilde{b}}\left(\mathbf{p}, \sqrt{\mathbf{p}^2 + m^2}\right) \quad (5.7.22)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and that P is even or odd according to whether $2A + 2\tilde{B}$ is an even or odd integer

$$P(-\mathbf{p}, -p^0) = (-)^{2A+2\tilde{B}} P(\mathbf{p}, p^0) . \tag{5.7.23}$$

We shall check this here for just one particular direction of $\mathbf{p}$. Taking $\mathbf{p}$ in the three-direction, (5.7.21) gives

$$\pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) = \sum_{\sigma} C_{AB}(j\sigma; ab)\, C_{\tilde{A}\tilde{B}}(j\sigma; \tilde{a}\tilde{b}) \exp\Big([-a + b - \tilde{a} + \tilde{b}]\,\theta\Big)$$

The Clebsch–Gordan coefficients vanish unless $\sigma = a + b$ and $\sigma = \tilde{a} + \tilde{b}$, so we can replace

$$-a + b - \tilde{a} + \tilde{b} = -2a + \sigma + 2\tilde{b} - \sigma = 2\tilde{b} - 2a .$$

We can write $\exp(\pm\theta)$ as $(p^0 \pm p^3)/m$, so here

$$\pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) = \sum_{\sigma} C_{AB}(j\sigma; ab) C_{\tilde{A}\tilde{B}}(j\sigma, \tilde{a}\tilde{b}) \times \begin{cases} \left[(p^0 + p^3)/m\right]^{2\tilde{b}-2a} & (\tilde{b} \geq a) \\ \left[(p^0 - p^3)/m\right]^{2a-2\tilde{b}} & (a \geq \tilde{b}) \end{cases}$$

where $p^0 \equiv \sqrt{\mathbf{p}^2 + m^2}$. We see that $\pi(\mathbf{p})$ can indeed be written as the mass-shell value of a polynomial $P(\mathbf{p}, p^0)$. Also, $2\tilde{b} - 2a$ equals $2\tilde{B} + 2A$ minus an even integer, so the polynomial satisfies the reflection condition (5.7.23).

Any polynomial in $\mathbf{p}$ and $\sqrt{\mathbf{p}^2 + m^2}$ can be written in a form *linear* in $\sqrt{\mathbf{p}^2 + m^2}$ (by expressing even powers of $\sqrt{\mathbf{p}^2 + m^2}$ in terms of $\mathbf{p}$) so $\pi(\mathbf{p})$ can be written

$$\pi_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) = P_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) + 2\sqrt{\mathbf{p}^2 + m^2}\, Q_{ab,\tilde{a}\tilde{b}}(\mathbf{p}) , \tag{5.7.24}$$

where P and Q are now polynomials in $\mathbf{p}$ alone, with

$$P(-\mathbf{p}) = (-)^{2A+2\tilde{B}} P(\mathbf{p}) \tag{5.7.25}$$

$$Q(-\mathbf{p}) = -(-)^{2A+2\tilde{B}} Q(\mathbf{p}) . \tag{5.7.26}$$

For $x - y$ space-like, we can adopt a Lorentz frame in which $x^0 = y^0$, and write Eq. (5.7.19) as

$$[\psi_{ab}(x), \psi^{\dagger}_{\tilde{a}\tilde{b}}(y)]_{\mp} = [\kappa\tilde{\kappa}^* \mp (-)^{2A+2\tilde{B}} \lambda\tilde{\lambda}^*] P_{ab,\tilde{a}\tilde{b}}(-i\nabla)\Delta_+(\mathbf{x} - \mathbf{y}, 0) + [\kappa\tilde{\kappa}^* \pm (-)^{2A+2\tilde{B}} \lambda\tilde{\lambda}^*] Q_{ab,\tilde{a}\tilde{b}}(-i\nabla)\delta^3(\mathbf{x} - \mathbf{y}) .$$

In order that this should vanish when $\mathbf{x} \neq \mathbf{y}$, we must have

$$\kappa\tilde{\kappa}^* = \pm(-1)^{2A+2\tilde{B}} \lambda\tilde{\lambda}^* . \tag{5.7.27}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Now let us consider the special case where ψ and $\tilde{\psi}$ are the same, so in particular $A = \tilde{A}$ and $B = \tilde{B}$. (It is unavoidable that such commutators or anticommutators will appear in $[\mathscr{H}(x), \mathscr{H}(y)]$, because the hermiticity of the Hamiltonian requires that if $\mathscr{H}(x)$ involves ψ, it also involves $\psi^\dagger$.) In this case, Eqs. (5.7.27) gives

$$|\kappa|^2 = \pm(-)^{2A+2B}|\lambda|^2 .$$

This is possible if and only if

$$\pm(-1)^{2A+2B} = +1 \tag{5.7.28}$$

and

$$|\kappa|^2 = |\lambda|^2 . \tag{5.7.29}$$

Of course, $2A + 2B$ differs from $2j$ by an *even* integer, so Eq. (5.7.28) says that *our particle is a boson or fermion according to whether* $2j$ *is even or odd.* This is the general relation between spin and statistics,[12] of which we have already seen special examples for particles described by scalar, vector, or Dirac fields.

Now let's return to the general case, where the fields ψ and $\tilde{\psi}$ may be different. Using Eq. (5.7.27), and dividing both sides by $|\tilde{\kappa}|^2 = |\tilde{\lambda}|^2$, we have

$$\frac{\kappa}{\tilde{\kappa}} = (-1)^{2B+2\tilde{B}} \frac{\lambda}{\tilde{\lambda}} .$$

It follows that, for any field,

$$\lambda = (-)^{2B} c\ \kappa , \tag{5.7.30}$$

where c is the same factor for all fields of a given particle. Furthermore, Eq. (5.7.29) shows that c is just a phase, $|c| = 1$. We can therefore eliminate c for all fields by a redefinition of the relative phase of the operators $a(\mathbf{p}, \sigma)$ and $a^{c\dagger}(\mathbf{p}, \sigma)$, so that $c = 1$, and hence $\lambda = (-)^{2B}\kappa$. Also, the factor κ for each field type may be eliminated by a redefinition of the over-all scale of the field. We emerge from all this with a formula for the (A, B) field of a given particle, that is unique up to overall scale

$$\psi_{ab}(x) = (2\pi)^{-3/2} \sum_\sigma \int d^3p \left[u_{ab}(\mathbf{p}, \sigma) a(\mathbf{p}, \sigma) e^{ip\cdot x} + (-)^{2B} v_{ab}(\mathbf{p}, \sigma) a^{c\dagger}(\mathbf{p}, \sigma) e^{-ip\cdot x} \right] . \tag{5.7.31}$$

The different fields for a given particle do not really represent possibilities that are physically distinct. For instance, the possible fields for $j = 0$ are those of type (A, A) (because the triangle inequality $|A-B| \leq j \leq A+B$ here requires $A = B$). Starting with a $(0,0)$ scalar field ϕ, we can easily

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

construct such (A, A) fields from the $2A$th derivative

$$\{\partial_{\mu_1} \cdots \partial_{\mu_{2A}}\}\phi\,, \tag{5.7.32}$$

where $\{\}$ here denotes the traceless part; for instance

$$\{\partial_\mu \partial_\nu\} \equiv \frac{\partial^2}{\partial x^\mu \partial x^\nu} - \tfrac{1}{4}\,\eta_{\mu\nu}\,\Box\,.$$

(Recall that a traceless symmetric tensor of rank N transforms according to the $(N/2, N/2)$ representation.) But Eq. (5.7.31) represents the *unique* causal (A, B) field for a given particle of spin j, so the (A, A) fields (5.7.31) for $j = 0$ can be nothing but linear combinations of the $2A$th derivatives (5.7.32) of a scalar field.

More generally, *any* field (A, B) for a given particle of spin j can be expressed as a differential operator of rank $2B$ acting on the field[13] $\varphi_\sigma(x)$ of type $(j, 0)$ (or a differential operator of rank $2A$ acting on the field of type $(0, j)$). To see this, consider the field

$$\{\partial_{\mu_1} \cdots \partial_{\mu_{2B}}\}\ \varphi_\sigma\,. \tag{5.7.33}$$

This transforms as the direct product of the representations (B, B) and $(j, 0)$, and hence by the usual rules of vector addition, it can be decomposed into fields transforming according to all the irreducible representations (A, B) with $|j - B| \leq A \leq j + B$, or equivalently $|A - B| \leq j \leq A + B$. Since Eq. (5.7.31) represents the unique field of type (A, B) for a given particle of spin j, it can be nothing** but the (A, B) field obtained from the derivatives (5.7.33).

Now let us consider the behavior of these fields under inversions, beginning with space inversion. Using the results of Section 4.2, the space-inversion properties of the particle annihilation and antiparticle creation operators are:

$$\mathsf{P}a(\mathbf{p}, \sigma)\mathsf{P}^{-1} = \eta^* a(-\mathbf{p}, \sigma)\,, \tag{5.7.34}$$

$$\mathsf{P}a^{c\dagger}(\mathbf{p}, \sigma)\mathsf{P}^{-1} = \eta^c a^{c\dagger}(-\mathbf{p}, \sigma)\,, \tag{5.7.35}$$

where η and η^c are the intrinsic parities for the particle and antiparticle, respectively. The general causal (A, B) field (5.7.31) thus transforms under

** The only possible flaw in this argument would be if some of the (A, B) fields obtained in this way actually vanished. But in this case, the $(j, 0)$ field φ_σ would satisfy a field equation $\sum_\sigma M_\sigma(\partial/\partial x)\varphi_\sigma(x) = 0$ and hence, for *each* $\bar\sigma$, $\sum_\sigma M_\sigma(ip)u_\sigma(\mathbf{p}, \bar\sigma) = 0$. For the (j, σ) representation the Clebsch–Gordan coefficient $C_{j0}(j\bar\sigma; \sigma 0)$ is just the Kronecker symbol $\delta_{\bar\sigma\sigma}$, so this would require $\sum_\sigma M_\sigma(ip)D_{\sigma\sigma'}(L(p)) = 0$, which is impossible unless all the $M_\sigma(ip)$ vanish, since $D(\Lambda)$ has an inverse $D(\Lambda^{-1})$. The $(j, 0)$ fields $\varphi_\sigma(x)$ thus satisfy no field equation other than the Klein–Gordon equation $(\Box - m^2)\ \varphi_\sigma(x) = 0$, and therefore none of the (A, B) fields obtained from (5.7.33) can vanish.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

the parity operator P into

$$\mathsf{P}\psi^{AB}_{ab}(x)\mathsf{P}^{-1} = (2\pi)^{-3/2}\sum_{\sigma}\int d^3p\,\Big[\eta^* a(-\mathbf{p},\sigma)e^{ip\cdot x}\,u^{AB}_{ab}(\mathbf{p},\sigma) + \eta^{\mathrm{c}}(-)^{2B}a^{\mathrm{c}\dagger}(-\mathbf{p},\sigma)e^{-ip\cdot x}\,v^{AB}_{ab}(\mathbf{p},\sigma)\Big]. \tag{5.7.36}$$

We want to change the integration variable from $\mathbf{p}$ to $-\mathbf{p}$, and for this purpose we will need to evaluate $u_{ab}(-\mathbf{p},\sigma)$ and $v_{ab}(-\mathbf{p},\sigma)$. To do this, we need only glance back at (5.7.14) and (5.7.15), and use the symmetry property of the Clebsch–Gordan coefficient[14]

$$C_{AB}(j\sigma;ab) = (-)^{A+B-j}C_{BA}(j\sigma;ba). \tag{5.7.37}$$

This gives

$$u^{AB}_{ab}(-\mathbf{p},\sigma) = (-)^{A+B-j}\,u^{BA}_{ba}(\mathbf{p},\sigma)\,, \tag{5.7.38}$$

$$v^{AB}_{ab}(-\mathbf{p},\sigma) = (-)^{A+B-j}\,v^{BA}_{ba}(\mathbf{p},\sigma)\,, \tag{5.7.39}$$

so

$$\mathsf{P}\psi^{AB}_{ab}(x)\mathsf{P}^{-1} = (2\pi)^{-3/2}\sum_{\sigma}\int d^3p\,(-1)^{A+B-j} \times \Big[\eta^* a(\mathbf{p},\sigma)e^{ip\cdot\mathscr{P}x}u^{BA}_{ba}(\mathbf{p},\sigma) + \eta^{\mathrm{c}}(-)^{2B}a^{\mathrm{c}\dagger}(\mathbf{p},\sigma)e^{-ip\cdot\mathscr{P}x}v^{BA}_{ba}(\mathbf{p},\sigma)\Big]\,, \tag{5.7.40}$$

where, as before, $\mathscr{P}x \equiv (-\mathbf{x}, x^0)$. This is the causal field ψ^{BA}_{ba} evaluated at $\mathscr{P}x$, except that the coefficients of the annihilation and creation terms may not be the same as called for in Eq. (5.7.31). But these coefficients *must* be the same up to an overall constant factor as in Eq. (5.7.31) because, aside from scale, Eq. (5.7.31) is the unique causal field of any type. Hence the ratio of the coefficients of the two terms in Eq. (5.7.40) must be the same as in Eq. (5.7.31) (but with B replaced with A because this is supposed to be a (B,A) field):

$$\eta^{\mathrm{c}}(-)^{2B}/\eta^* = (-)^{2A}\,. \tag{5.7.41}$$

But $A-B$ differs from the spin j by only an integer, so this gives

$$\eta^{\mathrm{c}} = \eta^*(-)^{2j}\,. \tag{5.7.42}$$

We saw special cases of this result in Sections 5.2, 5.3, and 5.5, where $j=0$, $j=1$, and $j=\frac{1}{2}$, respectively. We now see that the result is general; *the intrinsic parity $\eta^{\mathrm{c}}\eta$ of a particle–antiparticle pair is $+1$ for bosons, and -1 for fermions.* Using Eq. (5.7.42) in Eq. (5.7.40), our final result for space inversion is

$$\mathsf{P}\psi^{AB}_{ab}(x)\mathsf{P}^{-1} = \eta^*(-)^{A+B-j}\psi^{BA}_{ba}(-\mathbf{x},x^0)\,. \tag{5.7.43}$$

Let's see how this applies to the Dirac field. For the top $(\frac{1}{2},0)$ and bottom $(0,\frac{1}{2})$ components of the Dirac field, the sign $(-1)^{A+B-j}$ is just

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

$+1$, so the parity operator simply takes $\mathbf{x}$ into $-\mathbf{x}$; reverses the top and bottom components; and multiplies the field with η^*. The reversal of the top and bottom components of the Dirac field is accomplished by the matrix β in (5.5.41).

Now let us consider charge-conjugation. Its effect on the particle annihilation and antiparticle creation operators is

$$\mathsf{C}a(\mathbf{p},\sigma)\,\mathsf{C}^{-1} = \xi^*\, a^{\mathrm{c}}(\mathbf{p},\sigma), \tag{5.7.44}$$

$$\mathsf{C}a^{\mathrm{c}\dagger}(\mathbf{p},\sigma)\,\mathsf{C}^{-1} = \xi^{\mathrm{c}} a^{\dagger}(\mathbf{p},\sigma)\,, \tag{5.7.45}$$

where ξ and ξ^{c} are the charge-conjugation parities of the particle and antiparticle, respectively. Applying this transformation to the field (5.7.31), we find

$$\mathsf{C}\psi^{AB}_{ab}(x)\mathsf{C}^{-1} = (2\pi)^{-3/2}\sum_{\sigma}\int d^3p\; u^{AB}_{ab}(\mathbf{p},\sigma) \times\left[\xi^* a^{\mathrm{c}}(\mathbf{p},\sigma)e^{ip\cdot x} + \xi^{\mathrm{c}}(-)^{2B}a^{\dagger}(\mathbf{p},-\sigma)(-)^{j-\sigma}e^{-ip\cdot x}\right]. \tag{5.7.46}$$

It is useful to compare this formula for the charge-conjugate of an (A,B) field with the adjoint of the (B,A) field for the same particle:

$$\psi^{BA\dagger}_{ba}(x) = (2\pi)^{-3/2}\sum_{\sigma}\int d^3p\, u^{BA*}_{ba}\;(\mathbf{p},\sigma) \times\left[(-1)^{2A}(-1)^{j-\sigma}a^{\mathrm{c}}(\mathbf{p},-\sigma)e^{ip\cdot x} + a^{\dagger}(\mathbf{p},\sigma)e^{-ip\cdot x}\right]. \tag{5.7.47}$$

To calculate the u^*, we use our previous result

$$\mathbf{J}^{(j)*} = -\mathscr{C}\mathbf{J}^{(j)}\mathscr{C}^{-1}\,, \qquad \mathscr{C}_{\bar{\sigma}\sigma} \propto (-1)^{j-\sigma}\delta_{\bar{\sigma},-\sigma}\,.$$

The Clebsch–Gordan coefficient in Eq. (5.7.14) is real, so

$$u^{BA}_{ba}(\mathbf{p},\sigma)^* = \frac{1}{\sqrt{2p^0}}\sum_{a'b'}\Big(\exp(-\hat{\mathbf{p}}\cdot\mathbf{J}^{(A)}\theta)\Big)_{-a,-a'}\Big(\exp\,(\hat{\mathbf{p}}\cdot\mathbf{J}^{(B)}\theta)\Big)_{-b,-b'} \times(-)^{a'-a}(-)^{b'-b}C_{BA}(j\sigma;b'a')\,.$$

We use the reflection property of the Clebsch–Gordan coefficients[14]

$$C_{BA}(j,-\sigma;-b',-a') = C_{AB}(j\sigma;a'b') \tag{5.7.48}$$

and the fact that those coefficients vanish unless $a'+b'=\sigma$, to write

$$u^{BA}_{-b,-a}(\mathbf{p},-\sigma)^* = (-)^{a+b-\sigma}\; u^{AB}_{ab}(\mathbf{p},\sigma)\,. \tag{5.7.49}$$

The field adjoint (5.7.47) is then (replacing $a\to -a,\ b\to -b,\ \sigma\to-\sigma$)

$$\psi^{BA\dagger}_{-b,-a}(x) = (2\pi)^{-3/2}\sum_{\sigma}\int d^3p\;(-)^{a+b-\sigma}u^{AB}_{ab}(\mathbf{p},\sigma) \times\left[(-)^{2A}(-)^{j+\sigma}\; a^{\mathrm{c}}(\mathbf{p},\sigma)e^{ip\cdot x} + a^{\dagger}(\mathbf{p},\sigma)e^{-ip\cdot x}\right].$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Using the sign relation $(-)^{-2A-j} = (-)^{2B+j}$, this is

$$(-)^{-2A-a-b-j}\ \psi^{BA^\dagger}_{-b,-a}(x) = (2\pi)^{-3/2}\sum_\sigma \int d^3p\ u^{AB}_{ab}(\mathbf{p},\sigma) \times \left[a^c(\mathbf{p},\sigma)e^{ip\cdot x} + (-)^{j-\sigma+2B}a^\dagger(\mathbf{p},-\sigma)e^{-ip\cdot x}\right]. \quad (5.7.50)$$

In order that $\mathsf{C}\ \psi^{AB}_{ab}(x)\ \mathsf{C}^{-1}$ should commute or anticommute with all ordinary fields at space-like separations, it is necessary that it be proportional to $\psi^{BA^\dagger}_{-b,-a}(x)$, because this is the adjoint of the unique causal field of transformation type (B,A). Comparing Eq. (5.7.50) with Eq. (5.7.46), we see that this is only possible if the charge-conjugation parities are related by

$$\xi^* = \xi^c\,, \quad (5.7.51)$$

in which case

$$\mathsf{C}\ \psi^{AB}_{ab}(x)\ \mathsf{C}^{-1} = \xi^*(-)^{-2A-a-b-j}\ \psi^{BA^\dagger}_{-b-a}(x). \quad (5.7.52)$$

We have already encountered the relation (5.7.51) for spins 0, 1, and $\frac{1}{2}$ in Sections 5.2, 5.3, and 5.5, and noted some of its implications for electron–positron and quark–antiquark states in Section 5.5.

In particular, for a particle that is its own antiparticle, Eq. (5.7.52) is satisfied without any charge-conjugation operator on the left-hand side or phase ξ^* on the right:

$$\psi^{AB}_{ab}(x) = (-)^{-2A-a-b-j}\ \psi^{BA\dagger}_{-b-a}(x)\,. \quad (5.7.53)$$

We have already seen an example of this sort of reality condition for Majorana spin $\frac{1}{2}$ particles in Section 5.5.

Finally we come to time-reversal. Applied to particle annihilation and antiparticle creation operators, this gives

$$\mathsf{T}a(\mathbf{p},\sigma)\mathsf{T}^{-1} = \zeta^*(-1)^{j-\sigma}a(-\mathbf{p},-\sigma)\,, \quad (5.7.54)$$

$$\mathsf{T}a^{c\dagger}(\mathbf{p},\sigma)\mathsf{T}^{-1} = \zeta^c(-1)^{j-\sigma}a^{c\dagger}(-\mathbf{p},-\sigma)\,. \quad (5.7.55)$$

The irreducible field (5.7.31) thus has the transformation property

$$\mathsf{T}\psi^{AB}_{ab}(x)\mathsf{T}^{-1} = (2\pi)^{-3/2}\sum_\sigma \int d^3p\, u^{AB*}_{ab}(\mathbf{p},\sigma)(-1)^{j-\sigma} \times \left[\zeta^* a(-\mathbf{p},-\sigma)e^{-ip\cdot x} + \zeta^c(-1)^{2B}a^{c\dagger}(-\mathbf{p},-\sigma)e^{ip\cdot x}\right]. \quad (5.7.56)$$

To calculate the complex conjugate of the coefficient function, we use Eq. (5.7.14) and the standard formula[14]

$$C_{AB}(j,\sigma;a,b) = (-)^{A+B-j}C_{AB}(j,-\sigma;-a,-b) \quad (5.7.57)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and find:

$$u^{AB*}_{ab}(-\mathbf{p},-\sigma) = (-)^{a+b+\sigma+A+B-j} u^{AB}_{-a,-b}(\mathbf{p},\sigma)\,. \tag{5.7.58}$$

Changing the variables of integration and summation in Eq. (5.7.56) to $-\mathbf{p}$ and $-\sigma$, we find that in order for an (A,B) field to be transformed by time-reversal into something proportional to another (A,B) field, it is necessary that

$$\zeta^c = \zeta^*\,, \tag{5.7.59}$$

in which case

$$\mathsf{T}\psi^{AB}_{ab}(x)\mathsf{T}^{-1} = (-)^{a+b+A+B-2j}\,\zeta^*\,\psi^{AB}_{-a,-b}(\mathbf{x},-x^0). \tag{5.7.60}$$

* * *

It should be mentioned that from time to time various difficulties have been reported[15] in the field theory of particles with spin $j \geq 3/2$. Generally, these are encountered in the study of the propagation of a higher spin field in the presence of c-number external field. Depending on the details of the theory, the difficulties encountered include non-causality, inconsistency, unphysical mass states, and violation of unitarity. I will not go into details about these problems here, because it seems to me that they are not relevant to the calculational scheme described in this chapter, for the following reasons:

(1) The fields $\psi_{ab}(x)$ have been constructed here directly from the creation and annihilation operators for physical particles, so no question of inconsistency or unphysical mass states can arise. These are free fields, but by incorporating them into an interaction Hamiltonian density in the interaction picture, we can use perturbation theory to calculate S-matrix elements that automatically satisfy the cluster decomposition principle. As long as the interaction Hamiltonian is Hermitian, there can be no difficulty with unitarity. Lorentz invariance is guaranteed in perturbation theory as long as we add appropriate local but non-covariant terms in the Hamiltonian density; though a rigorous proof is lacking, there is no reason to doubt that this is always possible. Thus any difficulties with higher spin can only arise when we try to go beyond perturbation theory.
(2) As discussed in Section 13.6, the solution of field equations in the presence of a c-number background field (the context where all the problems with higher spin have been found) does go beyond perturbation theory, in that the results correspond to summing an infinite subset of terms in the perturbation series. This partial summation is justified, even for weak external fields, if the fields are sufficiently slowly varying, the smallness of energy denominators making up for the weakness of the fields. But the results obtained in this way depend on all the details of the interaction

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

of the high-spin particle with the external fields: not only the multipole moments of the particle but also possible terms in the interaction that are non-linear in the external fields. The problems reported[15] with higher spin have been encountered only for higher-spin particles that have been arbitrarily assumed to have only very simple interactions with external fields. No one has shown that the problems persist for arbitrary interactions, and as we shall see in Chapter 12, particles of higher spin are expected to have interactions of all possible types allowed by symmetry principles.

(3) In fact, there are good reasons to believe that the problems with higher spin disappear if the interaction with external fields is sufficiently complicated. For one thing, there is no doubt about the existence of higher-spin *particles*, including various stable nuclei and hadronic resonances. If there is any problem with higher spin, it can only be for 'point' particles, that is, those whose interactions with external fields are particularly simple. It should be kept in mind that the requirement of simplicity depends on the choice of which field we choose to represent the higher-spin particle. Remember that any free field types for a given particle can be expressed as a derivative operator acting on any other field type, so in the interaction picture any interaction with external fields may be written in terms of any field types we like, but interactions that are simple when expressed in terms of a field of one type may look complicated when expressed in terms of a field of another type. So the requirement of simplicity does not seem to have any objective content.

(4) Also, both higher-dimensional 'Kaluza–Klein' theories and string theories provide examples of consistent theories of a charged massive particles of spin two interacting with an electromagnetic background field.[16] (It was found that the consistency of the theory depends on the assumption of realistic external fields that satisfy the field equations, a point generally neglected in earlier work.) Reformulating this work in the interaction picture, the spin two particle is represented by a $(1,1)$ free field, but as mentioned above, the interactions may be reexpressed in the interaction picture in terms of any field type (A,B) that contains the $j=2$ representation of the rotation group.

5.8 The CPT Theorem

We have seen that the demands of relativity combined with quantum mechanics require the existence of antiparticles. Not only is it necessary that every particle have an antiparticle (which may for a purely neutral particle be itself); there is a precise relation between the properties of particles and antiparticles, that can be summarized in the statement that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

for an appropriate choice of inversion phases, the product CPT *of all the inversions is conserved.* This is the celebrated CPT theorem.*

As a first step in the proof, let us work out the effect of the product CPT on free fields of various types. For a scalar, vector, or Dirac field the results of Sections 5.2, 5.3, and 5.5 give

$$\mathrm{CPT}\ \phi(x)\ [\mathrm{CPT}]^{-1} = \zeta^*\xi^*\eta^*\phi^\dagger(-x)\,, \tag{5.8.1}$$

$$\mathrm{CPT}\ \phi_\mu(x)\ [\mathrm{CPT}]^{-1} = -\zeta^*\xi^*\eta^*\phi_\mu^\dagger(-x)\,, \tag{5.8.2}$$

$$\mathrm{CPT}\ \psi(x)\ [\mathrm{CPT}]^{-1} = -\zeta^*\xi^*\eta^*\gamma_5\psi^*(-x)\,. \tag{5.8.3}$$

(Of course, the phases ζ, ξ, and η depend on the species of particle described by each field.) We are going to choose the phases so that for all particles

$$\zeta\,\xi\,\eta = 1\,. \tag{5.8.4}$$

Then any tensor $\phi_{\mu_1\ldots\mu_n}$ formed from any set of scalar and vector fields and their derivatives transforms into

$$\mathrm{CPT}\ \phi_{\mu_1\ldots\mu_n}(x)\ [\mathrm{CPT}]^{-1} = (-)^n\phi^\dagger_{\mu_1\ldots\mu_n}(-x)\,. \tag{5.8.5}$$

(Any complex numerical coefficient appearing in these tensors is transformed into its complex conjugate because CPT is antiunitary.) We can easily see that the same transformation rule applies to tensors formed from bilinear combinations of Dirac fields. Applying Eq. (5.8.3) to such a bilinear gives

$$\begin{aligned}\mathrm{CPT}[\bar\psi_1(x)M\psi_2(x)][\mathrm{CPT}]^{-1} &= \psi_1^T(-x)\gamma_5\beta M^*\gamma_5\psi_2^*(-x)\\ &= [\bar\psi_1(-x)\gamma_5 M\gamma_5\psi_2(-x)]^\dagger\,.\end{aligned} \tag{5.8.6}$$

(A minus sign from the anticommutation of β and γ_5 is cancelled by the minus sign from the anticommutation of fermionic operators.) If the bilinear is a tensor of rank n, then M is a product of n modulo 2 Dirac matrices, so $\gamma_5 M\gamma_5 = (-1)^n M$, and the bilinear therefore satisfies Eq. (5.8.5).

A Hermitian scalar interaction density $\mathscr{H}(x)$ must be formed from

* The original proofs of this theorem were by Lüders and Pauli.[17] It has been proved rigorously in axiomatic field theory,[18] by using commutativity assumptions to extend the Lorentz invariance of the theory to the complex Lorentz group, then using complex Lorentz transformations to prove a reflection property of vacuum expectation values of products of fields, and then using this reflection property to infer the existence of an antiunitary operator that induces CPT transformations on the fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

tensors with an *even* total number of spacetime indices, and therefore

$$\mathsf{CPT}\ \mathscr{H}(x)\ [\mathsf{CPT}]^{-1} = \mathscr{H}(-x)\,. \tag{5.8.7}$$

More generally (and somewhat more easily) we can see that the same is true for Hermitian scalars formed from the fields $\psi^{AB}_{ab}(x)$ belonging to one or more of the general irreducible representations of the homogeneous Lorentz group. Putting together our results in the previous section for the effects of inversions on such fields, we find

$$\mathsf{CPT}\ \psi^{AB}_{ab}(x)\ [\mathsf{CPT}]^{-1} = (-1)^{2B}\psi^{AB\dagger}_{ab}(-x)\,. \tag{5.8.8}$$

(For the Dirac field the factor $(-1)^{2B}$ is supplied by the matrix γ_5 in Eq. (5.8.3).) In order to couple together a product $\psi^{A_1B_1}_{a_1b_1}(x)\,\psi^{A_2B_2}_{a_2b_2}(x)\cdots$ to form a scalar $\mathscr{H}(x)$, it is necessary that both $A_1+A_2+\cdots$ and $B_1+B_2+\cdots$ be integers, so $(-1)^{2B_1+2B_2+\cdots}=1$, and so a Hermitian scalar $\mathscr{H}(x)$ will automatically satisfy Eq. (5.8.7).

From Eq. (5.8.7) it follows immediately that CPT commutes with the interaction $V \equiv \int d^3x\,\mathscr{H}(\vec{x},0)$:

$$\mathsf{CPT}\ V\ [\mathsf{CPT}]^{-1} = V\,. \tag{5.8.9}$$

Also, in any theory CPT commutes with the free-particle Hamiltonian H_0. Thus the operator CPT, which has been defined here by its operation on free-particle operators, acts on 'in' and 'out' states in the way described in Section 3.3. The physical consequences of this symmetry principle have already been discussed in Sections 3.3 and 3.6.

5.9 Massless Particle Fields

Up to this point we have dealt only with the fields of massive particles. For some of these fields, such as the scalar and Dirac fields discussed in Sections 5.2 and 5.5, there is no special problem in passing to the limit of zero mass. On the other hand, we saw in Section 5.3 that there *is* a difficulty in taking the zero-mass limit of the vector field for a particle of spin one: at least one of the polarization vectors blows up in this limit. In fact, we shall see in this section that the creation and annihilation operators for physical massless particles of spin $j \geq 1$ cannot be used to construct all of the irreducible (A,B) fields that can be constructed for finite mass. This peculiar limitation on field types will lead us naturally to the introduction of gauge invariance.

Just as we did for massive particles, let us attempt to construct a general free field for a massless particle as a linear combination of the annihilation operators $a(\mathbf{p},\sigma)$ for particles of momentum $\mathbf{p}$ and helicity σ,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

and the corresponding creation operators $a^{c\dagger}(\mathbf{p},\sigma)$ for the antiparticles:*

$$\psi_\ell(x) = (2\pi)^{-3/2}\int d^3p \sum_\sigma \left[\kappa\, a(\mathbf{p},\sigma)u_\ell(\mathbf{p},\sigma)\, e^{ip\cdot x} + \lambda\, a^{c\dagger}(\mathbf{p},\sigma)v_\ell(\mathbf{p},\sigma)\, e^{-ip\cdot x}\right] \tag{5.9.1}$$

where now $p^0 \equiv |\mathbf{p}|$. The creation operators transform just like the one-particle states in Eq. (2.5.42)

$$U(\Lambda)a^\dagger(\mathbf{p},\sigma)U^{-1}(\Lambda) = \sqrt{\frac{(\Lambda p)^0}{p^0}}\, \exp\Big(i\sigma\theta(p,\Lambda)\Big)a^\dagger(\mathbf{p}_\Lambda,\sigma)\,, \tag{5.9.2}$$

$$U(\Lambda)a^{c\dagger}(\mathbf{p},\sigma)U^{-1}(\Lambda) = \sqrt{\frac{(\Lambda p)^0}{p^0}}\, \exp\Big(i\sigma\theta(p,\Lambda)\Big)a^{c\dagger}(\mathbf{p}_\Lambda,\sigma)\,, \tag{5.9.3}$$

and hence also

$$U(\Lambda)a(\mathbf{p},\sigma)U^{-1}(\Lambda) = \sqrt{\frac{(\Lambda p)^0}{p^0}}\, \exp\Big(-i\sigma\theta(p,\Lambda)\Big)a(\mathbf{p}_\Lambda,\sigma)\,, \tag{5.9.4}$$

where $p_\Lambda \equiv \Lambda p$, and θ is the angle defined by Eqs. (2.5.43). Hence if we want the field to transform according to some representation $D(\Lambda)$ of the homogeneous Lorentz group

$$U(\Lambda)\psi_\ell(x)U^{-1}(\Lambda) = \sum_{\bar\ell} D_{\ell\bar\ell}(\Lambda^{-1})\psi_{\bar\ell}(\Lambda x)\,, \tag{5.9.5}$$

then we must take the coefficient functions u and v to satisfy the relations

$$u_{\bar\ell}(\mathbf{p}_\Lambda,\sigma)\, \exp\Big(i\sigma\theta(p,\Lambda)\Big) = \sqrt{\frac{p^0}{(\Lambda p)^0}}\sum_\ell D_{\bar\ell\ell}(\Lambda)u_\ell(\mathbf{p},\sigma)\,, \tag{5.9.6}$$

$$v_{\bar\ell}(\mathbf{p}_\Lambda,\sigma)\, \exp\Big(-i\sigma\theta(p,\Lambda)\Big) = \sqrt{\frac{p^0}{(\Lambda p)^0}}\sum_\ell D_{\bar\ell\ell}(\Lambda)v_\ell(\mathbf{p},\sigma) \tag{5.9.7}$$

in place of Eqs. (5.1.19) and (5.1.20). (Again, $p_\Lambda \equiv \Lambda p$.) As in the massive particle case, we can satisfy these requirements by setting (in place of

* We deal here with only a single species of particle, and drop the species label n. Also, κ and λ are constant coefficients to be determined by the requirement of causality with some convenient choice of normalization of the coefficient functions u_ℓ and v_ℓ.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Eqs. (5.1.21) and (5.1.22))

$$u_{\bar{\ell}}(\mathbf{p},\sigma) = \sqrt{\frac{|\mathbf{k}|}{p^0}} \sum_{\ell} D_{\bar{\ell}\ell}\left(\mathscr{L}(p)\right) u_{\ell}(\mathbf{k},\sigma)\,, \tag{5.9.8}$$

$$v_{\bar{\ell}}(\mathbf{p},\sigma) = \sqrt{\frac{|\mathbf{k}|}{p^0}} \sum_{\ell} D_{\bar{\ell}\ell}\left(\mathscr{L}(p)\right) v_{\ell}(\mathbf{k},\sigma)\,, \tag{5.9.9}$$

where $\mathbf{k}$ is a standard momentum, say $(0,0,k)$, and $\mathscr{L}(p)$ is a standard Lorentz transformation that takes a massless particle from momentum $\mathbf{k}$ to momentum $\mathbf{p}$. Also, in place of Eqs. (5.1.23) and (5.1.24), the coefficient functions at the standard momentum must satisfy

$$u_{\bar{\ell}}(\mathbf{k},\sigma)\, \exp\Big(i\sigma\theta(k,W)\Big) = \sum_{\ell} D_{\bar{\ell}\ell}(W) u_{\ell}(\mathbf{k},\sigma) \tag{5.9.10}$$

$$v_{\bar{\ell}}(\mathbf{k},\sigma)\, \exp\Big(-i\sigma\theta(k,W)\Big) = \sum_{\ell} D_{\bar{\ell}\ell}(W) v_{\ell}(\mathbf{k},\sigma) \tag{5.9.11}$$

where $W^{\mu}{}_{\nu}$ is an arbitrary element of the 'little group' for four-momentum $k = (\mathbf{k}, |\mathbf{k}|)$, i.e., an arbitrary Lorentz transformation that leaves this four-momentum invariant.

We can extract the content of Eqs. (5.9.10) and (5.9.11) by considering separately the two kinds of little-group elements in Eq. (2.5.28). For a rotation $R(\theta)$ by an angle θ around the z-axis, given by Eq. (2.5.27),

$$R^{\mu}{}_{\nu}(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\,,$$

we find from Eqs. (5.9.10) and (5.9.11)

$$u_{\bar{\ell}}(\mathbf{k},\sigma) e^{i\sigma\theta} = \sum_{\ell} D_{\bar{\ell}\ell}\Big(R(\theta)\Big) u_{\ell}(\mathbf{k},\sigma) \tag{5.9.12}$$

$$v_{\bar{\ell}}(\mathbf{k},\sigma) e^{-i\sigma\theta} = \sum_{\ell} D_{\bar{\ell}\ell}\Big(R(\theta)\Big) v_{\ell}(\mathbf{k},\sigma)\ . \tag{5.9.13}$$

For combined rotations and boosts $S(\alpha,\beta)$ in the $x-y$ plane, given by (2.5.26),

$$S^{\mu}{}_{\nu}(\alpha,\beta) = \begin{bmatrix} 1 & 0 & -\alpha & \alpha \\ 0 & 1 & -\beta & \beta \\ \alpha & \beta & 1-\gamma & \gamma \\ \alpha & \beta & -\gamma & 1+\gamma \end{bmatrix}\,,$$

$$\gamma \equiv (\alpha^2+\beta^2)/2\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

Eqs. (5.9.10) and (5.9.11) give

$$u_{\bar{\ell}}(\mathbf{k},\sigma) = \sum_{\ell} D_{\bar{\ell}\ell}\Big(S(\alpha,\beta)\Big) u_{\ell}(\mathbf{k},\sigma) \ , \tag{5.9.14}$$

$$v_{\bar{\ell}}(\mathbf{k},\sigma) = \sum_{\ell} D_{\bar{\ell}\ell}\Big(S(\alpha,\beta)\Big) v_{\ell}(\mathbf{k},\sigma) \ . \tag{5.9.15}$$

Eqs. (5.9.12)–(5.9.15) are the conditions that determine the coefficient functions u and v at the standard momentum $\mathbf{k}$; Eqs. (5.9.8) and (5.9.9) then give them at arbitrary momenta. The equations for v are just the complex conjugates of the equations for u, so with a suitable adjustment of the constants κ and λ we may normalize the coefficient functions so that

$$v_{\ell}(\mathbf{p},\sigma) = u_{\ell}(\mathbf{p},\sigma)^* \ . \tag{5.9.16}$$

The problem is that we cannot find a u_{ℓ} that satisfies Eq. (5.9.14) for general representations of the homogeneous Lorentz group, even for those representations for which it is possible to construct fields for particles of a given helicity in the case $m \neq 0$.

To see what goes wrong here, let's try to construct the four-vector $[(\frac{1}{2},\frac{1}{2})]$ field for a massless particle of helicity ± 1. In the four-vector representation, we have simply

$$D^{\mu}{}_{\nu}(\Lambda) = \Lambda^{\mu}{}_{\nu} \ .$$

It is conventional to write the coefficient function u_{μ} here in terms of a 'polarization vector' e_{μ}:

$$u_{\mu}(\mathbf{p},\sigma) \equiv (2p^0)^{-1/2} e_{\mu}(\mathbf{p},\sigma) \ , \tag{5.9.17}$$

so that Eq. (5.9.8) gives

$$e^{\mu}(\mathbf{p},\sigma) = \mathscr{L}(\mathbf{p})^{\mu}{}_{\nu} e^{\nu}(\mathbf{k},\sigma) \ . \tag{5.9.18}$$

Also, Eqs. (5.9.12) and (5.9.14) read here

$$e^{\mu}(\mathbf{k},\sigma)\, e^{i\sigma\theta} = R(\theta)^{\mu}{}_{\nu} e^{\nu}(\mathbf{k},\sigma) \ , \tag{5.9.19}$$

$$e^{\mu}(\mathbf{k},\sigma) = S(\alpha,\beta)^{\mu}{}_{\nu} e^{\nu}(\mathbf{k},\sigma) \ . \tag{5.9.20}$$

Eq. (5.9.19) requires that (up to a constant which can be absorbed into the coefficients κ and λ),

$$e^{\mu}(\mathbf{k},\pm 1) = (1,\pm i,0,0)/\sqrt{2} \ . \tag{5.9.21}$$

But then Eq. (5.9.20) would require also that $\alpha \pm i\beta = 0$, which is impossible for general real α,β. We therefore cannot satisfy the fundamental

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

requirement (5.9.14) or (5.9.10); instead, we have here

$$D^{\mu}{}_{\nu}\Big(W(\theta,\alpha,\beta)\Big)e^{\nu}(\mathbf{k},\pm 1) = S^{\mu}{}_{\lambda}(\alpha,\beta)R^{\lambda}{}_{\nu}(\theta)e^{\nu}(\mathbf{k},\pm 1)$$
$$= \exp(\pm i\theta)\left\{e^{\mu}(\mathbf{k},\pm 1) + \frac{(\alpha \pm i\beta)}{\sqrt{2}|\mathbf{k}|}k^{\mu}\right\}. \quad (5.9.22)$$

We have thus come to the conclusion that no four-vector field can be constructed from the annihilation and creation operators for a particle of mass zero and helicity ± 1.

Let's temporarily close our eyes to this difficulty, and go ahead anyway, using Eqs. (5.9.18) and (5.9.21) to define a polarization vector for arbitrary momentum, and take the field as

$$a_{\mu}(x) = \int d^3p\,(2\pi)^{-3/2}(2p^0)^{-1/2}$$
$$\times \sum_{\sigma=\pm 1}\left[e_{\mu}(\mathbf{p},\sigma)e^{ip\cdot x}a(\mathbf{p},\sigma) + e_{\mu}(\mathbf{p},\sigma)^{*}e^{-ip\cdot x}a^{c\dagger}(\mathbf{p},\sigma)\right]. \quad (5.9.23)$$

We will come back later to consider how such a field can be used as an ingredient in a physical theory.

The field (5.9.23) of course satisfies

$$\Box a^{\mu}(x) = 0\,. \quad (5.9.24)$$

Other properties of the field follow from those of the polarization vector. (We shall need these properties of the polarization vector later when we come to quantum electrodynamics.). Note that the Lorentz transformation $\mathscr{L}(p)$ that takes a massless particle momentum from $\mathbf{k}$ to $\mathbf{p}$ may be written as a 'boost' $\mathscr{B}(|\mathbf{p}|)$ along the z-axis which takes the particle from energy $|\mathbf{k}|$ to energy $|\mathbf{p}|$, followed by a standardized rotation $R(\hat{\mathbf{p}})$ that takes the z-axis into the direction of $\mathbf{p}$. Since $e^{\nu}(\mathbf{k},\pm 1)$ is a purely spatial vector with only x and y components, it is unaffected by the boost along the z-axis, and so

$$e^{\mu}(\mathbf{p},\pm 1) = R(\hat{\mathbf{p}})^{\mu}{}_{\nu}\,e^{\nu}(\mathbf{k},\pm 1)\,. \quad (5.9.25)$$

In particular, $e^0(\mathbf{k},\pm 1) = 0$ and $\mathbf{k}\cdot\mathbf{e}(\mathbf{k},\pm 1) = 0$ so

$$e^0(\mathbf{p},\pm 1) = 0 \quad (5.9.26)$$

and

$$\mathbf{p}\cdot\mathbf{e}(\mathbf{p},\pm 1) = 0\,. \quad (5.9.27)$$

It follows that

$$a^0(x) = 0 \quad (5.9.28)$$

and

$$\nabla\cdot\mathbf{a}(x) = 0\,. \quad (5.9.29)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

As we shall see in Chapter 9, these are the conditions satisfied by the vacuum vector potential of electrodynamics in what is called Coulomb or radiation gauge.

The fact that a^0 vanishes in all Lorentz frames shows vividly that a^μ cannot be a four-vector. Instead, Eq. (5.9.22) shows that for a general momentum $\mathbf{p}$ and a general Lorentz transformation Λ, in place of Eq. (5.9.6) we have

$$e^\mu(\mathbf{p}_\Lambda, \pm 1) \exp(\pm i\theta(\mathbf{p}, \Lambda)) = D^\mu{}_\nu(\Lambda)\, e^\nu(\mathbf{p}, \pm 1) + p^\mu \Omega_\pm(\mathbf{p}, \Lambda)\,, \tag{5.9.30}$$

so that under a general Lorentz transformation

$$U(\Lambda) a_\mu(x) U^{-1}(\Lambda) = \Lambda^\nu{}_\mu a_\nu(\Lambda x) + \partial_\mu \Omega(x, \Lambda)\,, \tag{5.9.31}$$

where $\Omega(x, \Lambda)$ is a linear combination of annihilation and creation operators, whose precise form will not concern us here. As we will see in more detail in Chapter 8, we will be able to use a field like $a^\mu(x)$ as an ingredient in Lorentz-invariant physical theories if the couplings of $a^\mu(x)$ are not only formally Lorentz-invariant (that is, invariant under formal Lorentz transformations under which $a^\mu \to \Lambda^\mu{}_\nu\, a^\nu$), but are also invariant under the 'gauge' transformations $a_\mu \to a_\mu + \partial_\mu \Omega$. This is accomplished by taking the couplings of a_μ to be of the form $a_\mu j^\mu$, where j^μ is a four-vector current with $\partial_\mu j^\mu = 0$.

Although there is no ordinary four-vector field for massless particles of helicity ± 1, there is no problem in constructing an antisymmetric tensor field for such particles. From Eq. (5.9.22) and the invariance of k^μ under the little group we see immediately that

$$\begin{aligned} &D^\mu{}_\rho\Big(W(\theta, \alpha, \beta)\Big) D^\nu{}_\sigma\Big(W(\theta, \alpha, \beta)\Big) \Big(k^\rho e^\sigma(\mathbf{k}, \pm 1) - k^\sigma e^\rho(\mathbf{k}, \pm 1)\Big) \\ &\quad = e^{\pm i\theta}\Big(k^\mu e^\nu(\mathbf{k}, \pm 1) - k^\nu e^\mu(\mathbf{k}, \pm 1)\Big)\ . \end{aligned} \tag{5.9.32}$$

This shows that the coefficient function that satisfies Eq. (5.9.6) for the antisymmetric tensor representation of the homogeneous Lorentz group is (with an appropriate choice of normalization)

$$u^{\mu\nu}(\mathbf{p}, \pm 1) = i(2\pi)^{-3/2}(2p^0)^{-3/2}\, [p^\mu e^\nu(\mathbf{p}, \pm 1) - p^\nu e^\mu(\mathbf{p}, \pm 1)\ ]\,, \tag{5.9.33}$$

where $e^\mu(\mathbf{p}, \pm 1)$ is given by Eq. (5.9.25). Using this together with Eq. (5.9.23) gives the general antisymmetric tensor field for massless particles of helicity ± 1 in the form

$$f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu\ . \tag{5.9.34}$$

Note that this is a tensor even though a^μ is not a four-vector, because the extra term in Eq. (5.9.31) drops out in Eq. (5.9.34). Note also that Eqs. (5.9.34), (5.9.24), (5.9.28), and (5.9.29) show that $f^{\mu\nu}$ satisfies the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

vacuum Maxwell equations:

$$\partial_\mu f^{\mu\nu} = 0\,, \tag{5.9.35}$$

$$\epsilon^{\rho\sigma\mu\nu}\partial_\sigma f_{\mu\nu} = 0\,. \tag{5.9.36}$$

To calculate the commutation relations for the tensor fields we need sums over helicities of the bilinears $e^\mu e^{\nu *}$. The explicit formula (5.9.21) gives

$$\sum_{\sigma=\pm 1} e^i(\mathbf{k},\sigma)e^j(\mathbf{k},\sigma)^* = \delta_{ij} - \frac{k^i k^j}{|\mathbf{k}|^2}$$

and so, using Eq. (5.9.25),

$$\sum_{\sigma=\pm 1} e^i(\mathbf{p},\sigma)e^j(\mathbf{p},\sigma)^* = \delta_{ij} - \frac{p^i p^j}{|\mathbf{p}|^2}\,. \tag{5.9.37}$$

A straightforward calculation gives then

$$[f_{\mu\nu}(x),\, f_{\rho\sigma}(y)^\dagger] = (2\pi)^{-3}\left[-\eta_{\mu\rho}\partial_\nu\partial_\sigma + \eta_{\nu\rho}\partial_\mu\partial_\sigma + \eta_{\mu\sigma}\partial_\nu\partial_\rho - \eta_{\nu\sigma}\partial_\mu\partial_\rho\right]$$
$$\times \int d^3p(2p^0)^{-1}\left[|\kappa|^2 e^{ip\cdot(x-y)} - |\lambda|^2 e^{-ip\cdot(x-y)}\right]\,. \tag{5.9.38}$$

This clearly vanishes for $x^0 = y^0$ if and only if

$$|\kappa|^2 = |\lambda|^2 \tag{5.9.39}$$

in which case since $f_{\mu\nu}$ is a tensor the commutator also vanishes for all space-like separations. Eq. (5.9.39) also implies that the commutator of the a^μ vanishes at equal times, and as we shall see in Chapter 8 this is enough to yield a Lorentz-invariant S-matrix. The relative phase of the creation and annihilation operators can be adjusted so that $\kappa = \lambda$; the fields are then Hermitian if the particles are their own charge-conjugates, as is the case for the photon.

Why should we want to use fields like $a^\mu(x)$ in constructing theories of massless particles of spin one, rather than being content with fields like $f^{\mu\nu}(x)$ with simple Lorentz transformation properties? The presence of the derivatives in Eq. (5.9.34) means that an interaction density constructed solely from $f_{\mu\nu}$ and its derivatives will have matrix elements that vanish more rapidly for small massless particle energy and momentum than one that uses the vector field a_μ. Interactions in such a theory will have a correspondingly rapid fall-off at large distances, faster than the usual inverse-square law. This is perfectly possible, but gauge-invariant theories that use vector fields for massless spin one particles represent a more

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

general class of theories, including those that are actually realized in nature.

Parallel remarks apply to gravitons, massless particles of helicity ± 2. From the annihilation and creation operators for such particles we can construct a tensor $R_{\mu\nu\rho\sigma}$ with the algebraic properties of the Riemann–Christoffel curvature tensor: antisymmetric within the pairs μ,ν and ρ,σ, and symmetric between the pairs. However, in order to incorporate the usual inverse-square gravitational interactions we need to introduce a field $h_{\mu\nu}$ that transforms as a symmetric tensor, up to gauge transformations of the sort associated in general relativity with general coordinate transformations. Thus in order to construct a theory of massless particles of helicity ± 2 that incorporates long-range interactions, it is necessary for it to have a symmetry something like general covariance. As in the case of electromagnetic gauge invariance, this is achieved by coupling the field to a conserved 'current' $\theta^{\mu\nu}$, now with two spacetime indices, satisfying $\partial_\mu \theta^{\mu\nu} = 0$. The only such conserved tensor is the energy-momentum tensor, aside from possible total derivative terms that do not affect the long-range behavior of the force produced.** The fields of massless particles of spin $j \geq 3$ would have to couple to conserved tensors with three or more spacetime indices, but aside from total derivatives there are none, so *high-spin massless particles cannot produce long-range forces.*

* * *

The problems we have encountered in constructing four-vector fields for helicities ± 1 or symmetric tensor fields for helicity ± 2 are just special cases of a more general limitation. To see this, let's consider how to construct fields for massless particles belonging to arbitrary representations of the homogeneous Lorentz group. As we saw in Section 5.6, any representation $D(\Lambda)$ of the homogeneous Lorentz group can be decomposed into $(2A+1)(2B+1)$-dimensional representations (A,B), for which the generators of the homogeneous Lorentz group are represented by

$$(\mathscr{J}_{ij})_{a'b',ab} = \epsilon_{ijk}\left[(J_k^{(A)})_{a'a}\,\delta_{b'b} + (J_k{}^{(B)})_{b'b}\,\delta_{a'a}\right],$$
$$(\mathscr{J}_{k0})_{a'b',ab} = -i\left[(J_k^{(A)})_{a'a}\,\delta_{b'b} - (J_k{}^{(B)})_{b'b}\,\delta_{a'a}\right],$$

where $\mathbf{J}^{(j)}$ are the angular-momentum matrices for spin j. For θ infinites-

** If $\theta^{\mu_1\cdots\mu_N}$ is a tensor current satisfying $\partial_{\mu_1}\theta^{\mu_1\cdots\mu_N} = 0$, then $\int d^3x\,\theta^{0\mu_2\cdots\mu_N}$ is a conserved quantity that transforms like a tensor of rank $N-1$. The only such conserved tensors are the scalar 'charges' associated with various continuous symmetries, and the energy-momentum four-vector. The conservation of any other four-vector, or any tensor of higher rank, would forbid all but forward collisions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

imal, $D(R(\theta)) = 1 + i\mathcal{J}_{12}\theta$, so Eqs. (5.9.12) and (5.9.13) give

$$\sigma u_{ab}(\mathbf{k},\sigma) = (a+b)u_{ab}(\mathbf{k},\sigma)\,,$$

$$-\sigma v_{ab}(\mathbf{k},\sigma) = (a+b)v_{ab}(\mathbf{k},\sigma)\,,$$

and so $u_{ab}(\mathbf{k},\sigma)$ and $v_{ab}(\mathbf{k},\sigma)$ must vanish unless $\sigma = a+b$ and $\sigma = -a-b$, respectively. Also, letting α and β become infinitesimal in Eq. (5.9.14) gives

$$\begin{aligned}
0 &= (\mathcal{J}_{31} + \mathcal{J}_{01})_{ab,a'b'}\, u_{a'b'}(\mathbf{k},\sigma) \\
&= (J_2^{(A)} + iJ_1^{(A)})_{aa'}u_{a'b}(\mathbf{k},\sigma) + (J_2^{(B)} - iJ_1^{(B)})_{bb'}u_{ab'}(\mathbf{k},\sigma)\,, \\
0 &= (\mathcal{J}_{32} + \mathcal{J}_{02})_{ab,a'b'}\, u_{a'b'}(\mathbf{k},\sigma) \\
&= (-J_1^{(A)} + iJ_2^{(A)})_{aa'}u_{a'b}(\mathbf{k},\sigma) + (-J_1^{(B)} - iJ_2^{(B)})_{bb'}u_{ab'}(\mathbf{k},\sigma)\,,
\end{aligned}$$

or more simply

$$\begin{aligned}
\left(J_1^{(A)} - i\,J_2^{(A)}\right)_{aa'}\, u_{a'b}(\mathbf{k},\sigma) &= 0\,, \\
\left(J_1^{(B)} + i\,J_2^{(B)}\right)_{bb'}\, u_{ab'}(\mathbf{k},\sigma) &= 0\quad.
\end{aligned}$$

These require that $u_{ab}(\mathbf{k},\sigma)$ vanishes unless

$$a = -A,\quad b = +B \tag{5.9.40}$$

and the same is obviously also true of $v_{ab}(\mathbf{k},\sigma)$. Putting this together, we see that a field of type (A,B) can be formed only from the annihilation operators for a massless particle of helicity σ and the creation operators for the antiparticle of helicity $-\sigma$, where

$$\sigma = B - A\,. \tag{5.9.41}$$

For instance, the $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ parts of the Dirac field for a massless particle can only destroy particles of helicity $-\frac{1}{2}$ and $+\frac{1}{2}$ respectively, and create antiparticles of helicity $+\frac{1}{2}$ and $-\frac{1}{2}$, respectively. In the 'two-component' theory of the neutrino, there is only a $(\frac{1}{2},0)$ field and its adjoint, so neutrinos have helicity $-\frac{1}{2}$ and antineutrinos helicity $+\frac{1}{2}$ in this theory.

By the same methods as in Section 5.7, it can be shown that the $(j,0)$ and $(0,j)$ fields for massless particles of spin j (i.e., helicity $\mp j$) commute with each other and their adjoints at space-like separations if the coefficients of the annihilation and creation terms in Eq. (5.9.1) satisfy Eq. (5.9.39). The relative phase of the annihilation and creation operators may then be adjusted so that these coefficients are equal. It is easy to see that the fields for a massless particle of spin j of type $(A, A+j)$ or $(B+j,B)$ are just the $2A$th or $2B$th derivatives of fields of type $(0,j)$ or $(j,0)$, respectively, so these more general fields do not need to be considered separately here.

We can now see why it was impossible to construct a vector field for massless particles of helicity ± 1. A vector field transforms according to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

the $(\frac{1}{2}, \frac{1}{2})$ representation, and hence according to Eq. (5.9.41) can only describe helicity zero. (It *is*, of course, possible to construct a vector field for helicity zero — just take the derivative $\partial_\mu \phi$ of a massless scalar field ϕ.) The simplest covariant massless field for helicity ± 1 has the Lorentz transformation type $(1,0) \oplus (0,1)$; that is, it is an antisymmetric tensor $f_{\mu\nu}$. Similarly, the simplest covariant massless field for helicity ± 2 has the Lorentz transformation type $(2,0) \oplus (0,2)$: a fourth rank tensor which like the Riemann–Christoffel curvature tensor is antisymmetric within each pair of indices and symmetric between the two pairs.

The discussion of the inversions P, C, T given in the previous section can be carried over to the case of zero mass with only obvious modifications.

Problems

1. Show that if the zero-momentum coefficient functions satisfy the conditions (5.1.23) and (5.1.24), then the coefficient functions (5.1.21) and (5.1.22) for arbitrary momentum satisfy the defining conditions Eqs. (5.1.19) and (5.1.20).

2. Consider a free field $\psi_\ell^\mu(x)$ which annihilates and creates a self-charge-conjugate particle of spin $\frac{3}{2}$ and mass $m \neq 0$. Show how to calculate the coefficient functions $u_\ell^\mu(\mathbf{p}, \sigma)$, which multiply the annihilation operators $a(\mathbf{p}, \sigma)$ in this field, in such a way that the field transforms under Lorentz transformations like a Dirac field ψ_ℓ with an extra four-vector index μ. What field equations and algebraic and reality conditions does this field satisfy? Evaluate the matrix $P^{\mu\nu}(p)$, defined (for $p^2 = -m^2$) by

$$\sum_\sigma u_\ell^\mu(\mathbf{p}, \sigma) u_m^{\nu *}(\mathbf{p}, \sigma) \equiv (2p^0)^{-1} P_{\ell m}^{\mu\nu}(p).$$

   What are the commutation relations of this field? How does the field transform under the inversions P, C, T?

3. Consider a free field $h^{\mu\nu}(x)$ satisfying $h^{\mu\nu}(x) = h^{\nu\mu}(x)$ and $h^\mu{}_\mu(x) = 0$, which annihilates and creates a particle of spin two and mass $m \neq 0$. Show how to calculate the coefficient functions $u^{\mu\nu}(\mathbf{p}, \sigma)$, which multiply the annihilation operators $a(\mathbf{p}, \sigma)$ in this field, in such a way that the field transforms under Lorentz transformations like a tensor. What field equations does this field satisfy? Evaluate the function $P^{\mu\nu,\kappa\lambda}(p)$, defined by

$$\sum_\sigma u^{\mu\nu}(\mathbf{p}, \sigma) u^{\kappa\lambda *}(\mathbf{p}, \sigma) \equiv (2p^0)^{-1} P^{\mu\nu,\kappa\lambda}(p).$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

What are the commutation relations of this field? How does the field transform under the inversions P, C, T?

4. Show that the fields for a massless particle of spin j of type $(A, A+j)$ or $(B+j, B)$ are the $2A$th or $2B$th derivatives of fields of type $(0, j)$ or $(j, 0)$, respectively.

5. Work out the transformation properties of fields of transformation type $(j,0)+(0,j)$ for massless particles of helicity $\pm j$ under the inversions P, C, T.

6. Consider a generalized Dirac field ψ that transforms according to the $(j,0)+(0,j)$ representation of the homogeneous Lorentz group. List the tensors that can be formed from products of the components of ψ and $\psi^\dagger$. Check your result against what we found for $j = \frac{1}{2}$.

7. Consider a general field ψ_{ab} describing particles of spin j and mass $m \neq 0$, that transforms according to the (A, B) representation of the homogeneous Lorentz group. Suppose it has an interaction Hamiltonian of the form

$$V = \int d^3x \, [\psi_{ab}(x) J^{ab}(x) + J^{ab\dagger}(x) \psi^\dagger_{ab}(x)] \, ,$$

where J^{ab} is an external c-number current. What is the asymptotic behavior of the matrix element for emitting these particles for energy $E \gg m$ and definite helicity? (Assume that the Fourier transform of the current has values for different a, b that are of the same order of magnitude, and that do not depend strongly on E.)

References

1. The point of view adopted in this chapter was presented in a series of papers: S. Weinberg, *Phys. Rev.* **133**, B1318 (1964); **134**, B882 (1964); **138**, B988 (1965); **181**, 1893 (1969). A similar approach has been followed in unpublished lectures by E. Wichmann.

2. N. Bohr and L. Rosenfeld, *Kgl. Danske Vidensk. Selskab Mat.-Fys. Medd.*, No. 12 (1933) (translation in *Selected Papers of Leon Rosenfeld*, ed. by R. S. Cohen and J. Stachel (Reidel, Dordrecht, 1979)); *Phys. Rev.* **78**, 794 (1950).

3. P. A. M. Dirac, *Proc. Roy. Soc.* (London) **A117**, 610 (1928).

4. E. Cartan, *Bull. Soc. Math. France* **41**, 53 (1913).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

5. See, e.g., J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Cambridge, MA, 1955): Appendix A2; H. Georgi, *Lie Algebras in Particle Physics* (Benjamin–Cummings, Reading, MA, 1982): pp. 15, 198. The original reference is I. Schur, *Sitz. Preuss. Akad.*, p. 406 (1905).

6. See, e.g., H. Georgi, *Lie Algebras in Particle Physics* (Benjamin/Cummings, Reading, MA, 1982): pp. 15, 198. The original reference is I. Schur, *Sitz. Preuss. Akad.*, p. 406 (1905).

7. T. D. Lee and C. N. Yang, *Phys. Rev.* **104**, 254 (1956).

8. See, e.g., B. L. van der Waerden, *Die gruppentheoretische Methode in der Quantenmechanik* (Springer Verlag, Berlin, 1932); G. Ya. Lyubarski, *The Applications of Group Theory in Physics*, translated by S. Dedijer (Pergamon Press, New York, 1960).

9. W. Rarita and J. Schwinger, *Phys. Rev.* **60**, 61 (1941).

10. See, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957): Chapter 3.

11. S. Weinberg, *Phys. Rev.* **181**, 1893 (1969), Section V.

12. M. Fierz, *Helv. Phys. Acta* **12**, 3 (1939); W. Pauli, *Phys. Rev.* **58**, 716 (1940). Non-perturbative proofs in axiomatic field theory were given by G. Lüders and B. Zumino, *Phys. Rev.* **110**, 1450 (1958) and N. Burgoyne, *Nuovo Cimento* **8**, 807 (1958). Also see R. F. Streater and A. S. Wightman, *PCT, Spin & Statistics, and All That* (Benjamin, New York, 1968).

13. Fields in the $(j,0)+(0,j)$ representation were introduced by H. Joos, *Fortschr. Phys.* **10**, 65 (1962); S. Weinberg, *Phys. Rev.* **133**, B1318 (1964).

14. A. R. Edmonds, Ref. 10, or M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957): Chapter III .

15. G. Velo and D. Zwanziger, *Phys. Rev.* **186**, 1337 (1969); **188**, 2218 (1969); A. S. Wightman, in *Proceedings of the Fifth Coral Gables Conference on Symmetry Principles at High Energy*, ed. by T. Gudehus, G. Kaiser, and A. Perlmutter (Gordon and Breach, New York, 1969); B. Schroer, R. Seiler, and J. A. Swieca, *Phys. Rev.* D **2**, 2927 (1970); and other references quoted therein.

16. C. R. Nappi and L. Witten, *Phys. Rev.* D **40**, 1095 (1989); P. C. Argyres and C. R. Nappi, *Phys. Lett.* **B224**, 89 (1989). For the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

derivation of a consistent theory of $j = 3/2$ particles in external fields from a Kaluza–Klein theory, see S. D. Rindani and M. Sivakumar, *J. Phys. G: Nucl. Phys.* **12**, 1335 (1986); *J. Phys. C: Particles & Fields* **49**, 601 (1991).

17. G. Lüders, *Kong. Dansk. Vid. Selskab, Mat.-Fys. Medd.* **28**, 5 (1954); *Ann. Phys.* **2**, 1 (1957); W. Pauli, *Nuovo Cimento* **6**, 204 (1957). When Lüders first considered how the inversions are related, it was still taken for granted that P is conserved, so his theorem stated that C conservation is equivalent to T invariance.

18. R. Jost, *Helv. Phys. Acta* **30**, 409 (1957); F. J. Dyson, *Phys. Rev.* **110**, 579 (1958). Also see Streater and Wightman, Ref. 12.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.007

6

The Feynman Rules

In previous chapters the use of covariant free fields in the construction of the Hamiltonian density has been motivated by the requirement that the S-matrix satisfy Lorentz invariance and cluster decomposition conditions. With the Hamiltonian density constructed in this way, it makes no difference which form of perturbation theory we use to calculate the S-matrix; the results will automatically satisfy these invariance and clustering conditions in each order in the interaction density. Nevertheless, there are obvious practical advantages in using a version of perturbation theory in which the Lorentz invariance and cluster decomposition properties of the S-matrix are kept manifest at every stage in the calculation. This was not true for the perturbation theory used in the 1930s, now known as 'old-fashioned perturbation theory', described at the beginning of Section 3.5. The great achievement of Feynman, Schwinger, and Tomonaga in the late 1940s was to develop perturbative techniques for calculating the S-matrix, in which Lorentz invariance and cluster decomposition properties are transparent throughout. This chapter will outline the diagrammatic calculational technique first described by Feynman at the Poconos Conference in 1948. Feynman was led to these diagrammatic rules in part through his development of a path-integral approach, which will be the subject of Chapter 9. In this chapter, we shall use the approach described by Dyson[1] in 1949, which until the 1970s was the basis of almost all analyses of perturbation theory in quantum field theory, and still provides a particularly transparent introduction to the Feynman rules.

6.1 Derivation of the Rules

Our starting point is a formula for the S-matrix, obtained by putting together the Dyson series (3.5.10) with expression (4.2.2) for the free-particle states:

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

$$S_{\mathbf{p}'_1\sigma'_1n'_1;\mathbf{p}'_2\sigma'_2n'_2;\cdots\,,\,\mathbf{p}_1\sigma_1n_1;\mathbf{p}_2\sigma_2n_2;\cdots}$$
$$= \sum_{N=0}^{\infty} \frac{(-i)^N}{N!} \int d^4x_1 \cdots d^4x_N \Big(\Phi_0, \cdots a(\mathbf{p}'_2\sigma'_2n'_2)a(\mathbf{p}'_1\sigma'_1n'_1)$$
$$\times T\Big\{\mathscr{H}(x_1)\cdots\mathscr{H}(x_N)\Big\}a^\dagger(\mathbf{p}_1\sigma_1n_1)a^\dagger(\mathbf{p}_2\sigma_2n_2)\cdots\Phi_0\Big) \quad . \quad (6.1.1)$$

As a reminder: $\mathbf{p}$, σ, and n label particle momenta, spin, and species; primes denote labels for particles in the final state; Φ_0 is the free-particle vacuum state; a and $a^\dagger$ are annihilation and creation operators; T indicates a time-ordering, which puts the $\mathscr{H}(x)$ in an order in which the arguments x^0 decrease from left to right; and $\mathscr{H}(x)$ is the interaction Hamiltonian density, taken as a polynomial in the fields and their adjoints

$$\mathscr{H}(x) = \sum_i g_i \mathscr{H}_i(x) \,, \quad (6.1.2)$$

each term $\mathscr{H}_i$ being a product of definite numbers of fields and field adjoints of each type. The field of a particle of species n that transforms under a particular representation of the homogeneous Lorentz group (with or without space inversions) is given by

$$\psi_\ell(x) = \sum_\sigma (2\pi)^{-3/2} \int d^3p \, \Big[u_\ell(\mathbf{p},\sigma,n)\; a(\mathbf{p},\sigma,n)\, e^{ip\cdot x}$$
$$+ v_\ell(\mathbf{p},\sigma,n) a^\dagger(\mathbf{p},\sigma,n^c)\, e^{-ip\cdot x}\Big] \quad . \quad (6.1.3)$$

Here n^c denotes the antiparticle of the species n, and $\exp(\pm ip\cdot x)$ is calculated with p^0 set equal to $\sqrt{\mathbf{p}^2 + m_n^2}$. The coefficient functions u_ℓ and v_ℓ depend on the Lorentz transformation properties of the field and the spin of the particle it describes; they were calculated in Chapter 5. (For instance, in the scalar field the u_ℓ for a particle of energy E is simply $(2E)^{-1/2}$, while in a Dirac field u_ℓ and v_ℓ are the normalized Dirac spinors introduced in Section 5.5.) The index ℓ on the field should here be understood to indicate the particle type and the representation of the Lorentz group by which the field transforms, as well as including a running index labelling the components in this representation. There is no need to deal separately with interactions that involve derivatives of fields; from our point of view, the derivative of a field (6.1.3) is just another field described by (6.1.3), with different u_ℓ and v_ℓ. We will here make a distinction between some particle species that we arbitrarily call 'particles', for instance electrons, protons, etc., and those we call 'antiparticles', such as positrons and antiprotons. The field operators that destroy particles and create antiparticles are called simply 'fields'; their adjoints, which destroy antiparticles and create particles, are called 'field adjoints'. Of course,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

some particle species like the photon and π^0 are their own antiparticles; for these the field adjoints are proportional to the fields.

We now proceed to move all annihilation operators to the right in Eq. (6.1.1), repeatedly using for this purpose the commutation or anti-commutation relations:

$$a(\mathbf{p}\,\sigma\,n)a^\dagger(\mathbf{p}'\sigma'n') = \pm a^\dagger(\mathbf{p}'\sigma'n')a(\mathbf{p}\,\sigma\,n) + \delta^3(\mathbf{p}'-\mathbf{p})\delta_{\sigma'\sigma}\delta_{n'n} \tag{6.1.4}$$

$$a(\mathbf{p}\,\sigma\,n)a(\mathbf{p}'\sigma'n') = \pm a(\mathbf{p}'\sigma'n')a(\mathbf{p}\,\sigma\,n) \tag{6.1.5}$$

$$a^\dagger(\mathbf{p}\,\sigma\,n)a^\dagger(\mathbf{p}'\sigma'n') = \pm a^\dagger(\mathbf{p}'\sigma'n')a^\dagger(\mathbf{p}\,\sigma\,n) \tag{6.1.6}$$

(and likewise for antiparticles), the $\pm$ sign on the right being $-$ if both particles n, n' are fermions, and $+$ if either or both are bosons. Whenever an annihilation operator appears on the extreme right (or a creation operator on the extreme left), the corresponding contribution to Eq. (6.1.1) vanishes, because these operators annihilate the vacuum state:

$$a(\mathbf{p}\,\sigma\,n)\,\Phi_0 = 0\,, \tag{6.1.7}$$

$$\Phi_0^\dagger\, a^\dagger(\mathbf{p}\,\sigma\,n) = 0. \tag{6.1.8}$$

The remaining contributions to Eq. (6.1.1) are those arising from the delta function terms on the right-hand side of Eq. (6.1.4), with every creation and annihilation operator in the initial or final states or in the interaction Hamiltonian density paired in this way with some other annihilation or creation operator.

In this way, the contribution to Eq. (6.1.1) of a given order in each of the terms $\mathscr{H}_i$ in the polynomial $\mathscr{H}(\psi(x), \psi^\dagger(x))$ is given by a sum, over all ways of pairing creation and annihilation operators,[2] of the integrals of products of factors, as follows:

(a) Pairing of a final particle having quantum numbers $\mathbf{p}', \sigma', n'$ with a field adjoint $\psi_\ell^\dagger(x)$ in $\mathscr{H}_i(x)$ yields a factor

$$\left[a(\mathbf{p}'\,\sigma'n'),\, \psi_\ell^\dagger(x)\right]_\mp = (2\pi)^{-3/2}e^{-ip'\cdot x}u_\ell^*(\mathbf{p}'\sigma'n')\,. \tag{6.1.9}$$

(b) Pairing of a final antiparticle having quantum numbers $\mathbf{p}', \sigma', n'^c$ with a field $\psi_\ell(x)$ in $\mathscr{H}_i(x)$ yields a factor

$$\left[a(\mathbf{p}'\sigma'\,n'^c), \psi_\ell(x)\right]_\mp = (2\pi)^{-3/2}e^{-ip'\cdot x}v_\ell(\mathbf{p}'\sigma'\,n')\,. \tag{6.1.10}$$

(c) Pairing of an initial particle having quantum numbers $\mathbf{p}, \sigma, n$ with a field $\psi_\ell(x)$ in $\mathscr{H}_i(x)$ yields a factor

$$\left[\psi_\ell(x), a^\dagger(\mathbf{p}\sigma n)\right]_\mp = (2\pi)^{-3/2}e^{ip\cdot x}u_\ell(\mathbf{p}\sigma n)\,. \tag{6.1.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

(d) Pairing of an initial antiparticle having quantum numbers $\mathbf{p}, \sigma, n^c$ with a field adjoint $\psi_\ell^\dagger(x)$ in $\mathscr{H}_i(x)$ yields a factor

$$\left[\psi_\ell^\dagger(x), a^\dagger(\mathbf{p}\ \sigma\ n^c)\right]_\mp = (2\pi)^{-3/2} e^{ip\cdot x} v_\ell^*(\mathbf{p}\ \sigma\ n)\,. \tag{6.1.12}$$

(e) Pairing of a final particle (or antiparticle) having numbers $\mathbf{p}', \sigma', n'$ with an initial particle (or antiparticle) having quantum numbers $\mathbf{p}, \sigma, n$ yields a factor

$$\left[a(\mathbf{p}'\sigma' n'), a^\dagger(\mathbf{p}\ \sigma\ n)\right]_\mp = \delta^3(\mathbf{p}' - \mathbf{p})\delta_{\sigma'\sigma}\delta_{n'n}\,. \tag{6.1.13}$$

(f) Pairing of a field $\psi_\ell(x)$ in $\mathscr{H}_i(x)$ with a field adjoint $\psi_m^\dagger(y)$ in $\mathscr{H}_j(y)$ yields a factor*

$$\begin{aligned}\theta(x-y)\left[\psi_\ell^+(x), \psi_m^{+\dagger}(y)\right]_\mp \pm \theta(y-x)\left[\psi_m^{-\dagger}(y), \psi_\ell^-(x)\right]_\mp \\ \equiv -i\,\Delta_{\ell m}(x,y)\,,\end{aligned} \tag{6.1.14}$$

where ψ^+ and ψ^- are the terms in ψ that destroy particles and create antiparticles, respectively:

$$\psi_\ell^+(x) = (2\pi)^{-3/2} \int d^3p \sum_\sigma u_\ell(\mathbf{p}\ \sigma\ n)\, e^{ip\cdot x} a(\mathbf{p}\ \sigma\ n)\,, \tag{6.1.15}$$

$$\psi_\ell^-(x) = (2\pi)^{-3/2} \int d^3p \sum_\sigma v_\ell(\mathbf{p}\ \sigma\ n)\, e^{-ip\cdot x} a^\dagger(\mathbf{p}\,\sigma\, n^c)\,. \tag{6.1.16}$$

Recall that $\theta(x-y)$ is a step function, equal to $+1$ for $x^0 > y^0$ and zero for $x^0 < y^0$. These step functions appear in Eq. (6.1.14) because of the time-ordering in Eq. (6.1.1); we can encounter a pairing of an annihilation field $\psi^+(x)$ in $\mathscr{H}(x)$ with a creation field $\psi^{+\dagger}(y)$ in $\mathscr{H}(y)$ only if $\mathscr{H}(x)$ was initially to the left of $\mathscr{H}(y)$ in Eq. (6.1.1), i.e., if $x^0 > y^0$; similarly, we encounter a pairing of an annihilation field $\psi^{-\dagger}(y)$ in $\mathscr{H}(y)$ with a creation field $\psi^-(x)$ in $\mathscr{H}(x)$ only if $\mathscr{H}(y)$ was initially to the left of $\mathscr{H}(x)$ in Eq. (6.1.1), i.e., if $y^0 > x^0$. (The $\pm$ sign in the second term in (6.1.14) will be explained a little later.) The quantity (6.1.14) is known as a *propagator*; it is calculated in the following section.

The S-matrix is obtained by multiplying these factors together, along with additional numerical factors to be discussed below, then integrating over $x_1 \cdots x_N$, then summing over all pairings, and then over the numbers of interactions of each type. Before filling in all the details, it will be convenient first to describe a diagrammatic formalism for keeping track of all these pairings.

* If the interaction $\mathscr{H}(x)$ is written in the normal-ordered form, as in Eq. (5.1.33), then there is no pairing of fields and field adjoints in the *same* interaction. Otherwise some sort of regularization is needed to give meaning to $\Delta_{\ell m}(0)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

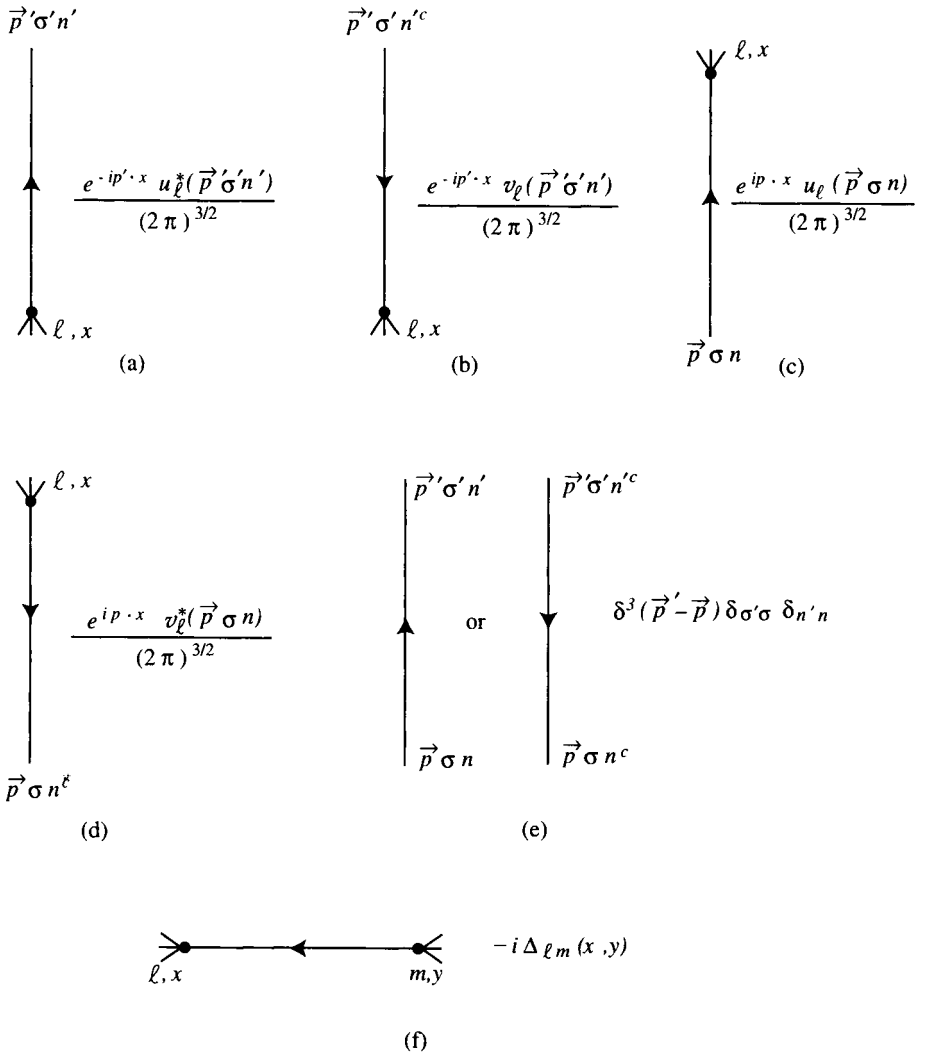

Figure 6.1. Graphical representation of pairings of operators arising in the coordinate-space evaluation of the S-matrix. The expressions on the right are the factors that must be included in the coordinate-space integrand of the S-matrix for each line of the Feynman diagram.

The rules for calculating the S-matrix are conveniently summarized in terms of *Feynman diagrams.* (See Figure 6.1.) The diagrams consist of points called *vertices*, each representing one of the $\mathscr{H}_i(x)$, and *lines*, each representing the pairing of a creation with an annihilation operator. More specifically:

(a) The pairing of a final particle with a field adjoint in one of the $\mathscr{H}(x)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

is represented by a line running from the vertex representing that $\mathscr{H}(x)$ upwards out of the diagram, carrying an arrow pointed upwards.
(b) The pairing of a final antiparticle with a field in one of the $\mathscr{H}(x)$ is also represented by a line running from the vertex representing that $\mathscr{H}(x)$ upwards out of the diagram, but carrying an arrow pointed downwards. (Arrows are omitted throughout for particles like γ, π^0, etc. that are their own antiparticles.)
(c) The pairing of an initial particle with a field in one of the $\mathscr{H}(x)$ is represented by a line running into the diagram from below, ending in the vertex representing that $\mathscr{H}(x)$, carrying an arrow pointed upwards.
(d) The pairing of an initial antiparticle with a field in one of the $\mathscr{H}(x)$ is also represented by a line running into the diagram from below, ending in the vertex representing that $\mathscr{H}(x)$, but carrying an arrow pointed downwards.
(e) The pairing of a final particle or antiparticle with an initial particle or antiparticle is represented by a line running clear through the diagram from bottom to top, not touching any vertex, with an arrow pointed upwards or downwards for particles or antiparticles, respectively.
(f) The pairing of a field in $\mathscr{H}(x)$ with a field adjoint in $\mathscr{H}(y)$ is represented by a line joining the vertices representing $\mathscr{H}(x)$ and $\mathscr{H}(y)$, carrying an arrow pointing from y to x.

Note that arrows always point in the direction a particle is moving, and opposite to the direction an antiparticle is moving. (As mentioned above, arrows should be omitted for particles like photons that are their own antiparticles.) The arrow direction indicated in rule (f) is consistent with this convention because a field adjoint in $\mathscr{H}_j(y)$ can either create a particle destroyed by a field in $\mathscr{H}_i(x)$, or destroy an antiparticle created by a field in $\mathscr{H}_i(x)$. Note also that since every field or field adjoint in $\mathscr{H}_i(x)$ must be paired with something, the total number of lines at a vertex of type i, corresponding to a term $\mathscr{H}_i(x)$ in Eq. (6.1.2), is just equal to the total number of field or field adjoint factors in $\mathscr{H}_i(x)$. Of these lines, the number with arrows pointed into the vertex or out of it equals the number of fields or field adjoints respectively in the corresponding interaction term.

To calculate the contribution to the S-matrix for a given process, of a given order N_i in each of the interaction terms $\mathscr{H}_i(x)$ in Eq. (6.1.2), we must carry out the following steps:

(i) Draw all Feynman diagrams containing N_i vertices of each type i, and containing a line coming into the diagrams from below for each particle or antiparticle in the initial state, and a line going upwards out of the diagram for every particle or antiparticle in the final state, together

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

with any number of internal lines running from one vertex to another, as required to give each vertex the proper number of attached lines. The lines carry arrows as described above, each of which may point upwards or downwards. Each vertex is labelled with an interaction type i and spacetime coordinate x^μ. Each internal or external line is labelled at the end where it runs into a vertex with a field type ℓ (corresponding to the field $\psi_\ell(x)$ or $\psi_\ell^\dagger(x)$ that creates or destroys the particle or antiparticle at that vertex), and each external line where it enters or leaves the diagram is labelled with the quantum numbers $\mathbf{p},\sigma,n$ or $\mathbf{p}',\sigma',n'$ of the initial or final particle (or antiparticle).

(ii) For each vertex of type i, include a factor $-i$ (from the $(-i)^N$ in Eq. (6.1.1)) and a factor g_i (the coupling constant multiplying the product of fields in $\mathscr{H}_i(x)$). For each line running upwards out of the diagram, include a factor (6.1.9) or (6.1.10), depending on whether the arrow is pointing up or down. For each line running from below into the diagram, include a factor (6.1.11) or (6.1.12), again depending on the arrow direction. For each line running straight through the diagram include a factor (6.1.13). For each internal line connecting two vertices include a factor (6.1.14).

(iii) Integrate the product of all these factors over the coordinates $x_1, x_2, \cdots$ of each vertex.

(iv) Add up the results obtained in this way from each Feynman diagram. The complete perturbation series for the S-matrix is obtained by adding up the contributions of each order in each interaction type, up to whatever order our strength permits.

Note that we have not included the factor $1/N!$ from Eq. (6.1.1) in these rules, because the time-ordered product in Eq. (6.1.1) is a sum over the $N!$ permutations of $x_1 x_2 \cdots x_N$, each permutation giving the same contribution to the final result. To put this another way, a Feynman diagram with N vertices is one of $N!$ identical diagrams, which differ only in permutations of the labels on the vertices, and this yields a factor of $N!$ which cancels the $1/N!$ in Eq. (6.1.1). (There are exceptions to this rule, discussed below.) For this reason, henceforth we do *not* include more than one of a set of Feynman diagrams that differ only by relabelling the vertices.

In some cases, there are additional combinatoric factors or signs that must be included in the contribution of individual Feynman diagrams:

(v) Suppose that an interaction $\mathscr{H}_i(x)$ contains (among other fields and field adjoints) M factors of the *same* field. Suppose that each of these fields is paired with a field adjoint in a different interaction (different for each one), or in the initial or final state. The first of these field adjoints

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

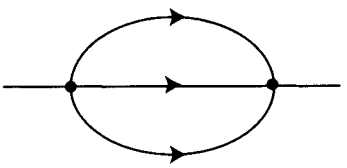

Figure 6.2. Example of a graph requiring extra combinatoric factors in the S-matrix. For an interaction involving, say, three factors of some field (as well as other fields) we usually include a factor $1/3!$ in the interaction Hamiltonian, to cancel factors arising from sums over ways of pairing these fields with their adjoints in other interactions. But in this diagram there are two such factors of $1/3!$ and only $3!$ different pairings, so we are left with an extra factor of $1/3!$.

can be paired with any one of the M identical fields in $\mathscr{H}_i(x)$; the second with any one of the remaining $M-1$ identical fields; and so on, yielding an extra factor of $M!$. To compensate for this, it is conventional to define the coupling constants g_i so that an explicit factor $1/M!$ appears in any $\mathscr{H}_i(x)$ containing M identical fields (or field adjoints.) For instance, the interaction of Mth order in a scalar field $\phi(x)$ would be written $g\phi^M/M!$. (More generally, one often also displays an explicit factor of $1/M!$ when the interaction involves a sum of M factors of fields from the same symmetry multiplet, or when for this or any other reason the coupling coefficient is totally symmetric or antisymmetric under permutations of M boson or fermion fields.)

However, this cancellation of $M!$ factors is not always complete. For instance, consider a Feynman diagram in which the M identical fields in one interaction $\mathscr{H}_i(x)$ are paired with M corresponding field adjoints in a *single* other interaction $\mathscr{H}_j(y)$. (See Figure 6.2.) Then by following the above analysis, we find only $M!$ different pairings (since it makes no difference which of the field adjoints we call the first, second, ...), cancelling only one of the two factors of $1/M!$ in the two different interactions. In this case, we would have to insert an extra factor of $1/M!$ 'by hand' into the contribution of such a Feynman diagram.

Other combinatoric factors arise when some of the permutations of vertices have no effect on the Feynman diagram. We noted earlier that the factor $1/N!$ in the series (6.1.1) is usually cancelled by the sum over the $N!$ diagrams that differ only in the labelling of the N vertices. However, this cancellation is incomplete when relabelling the vertices does not yield a new diagram. This happens most commonly in the calculation of vacuum-to-vacuum S-matrix elements in a theory with a quadratic interaction $\mathscr{H} = \psi_\ell^\dagger M_{\ell\ell'}\psi_{\ell'}$, where M may depend on external fields. (The physical significance of such vacuum fluctuation diagrams is discussed in detail in Volume II.) The Feynman diagram of Nth order in $\mathscr{H}$ is a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

ring with N corners. (See Figure 6.3.) There are only $(N-1)!$ different diagrams here because a permutation of labels that moves each label to the next vertex around the ring yields the same diagram. Hence such a graph is accompanied with a factor

$$\frac{(N-1)!}{N!} = \frac{1}{N}. \tag{6.1.17}$$

(vi) In theories involving fermion fields, the use of Eqs. (6.1.4)–(6.1.6) to move annihilation and creation operators to the right and left introduces minus signs into the contribution of various pairings. To be specific, we get a minus sign wherever the permutation of the operators in Eq. (6.1.1) that is required to put all paired operators adjacent to one another (with annihilation operators just to the left of the paired creation operators) involves an odd number of interchanges of fermion operators. (This is because to compute the contribution of a certain pairing, we can first permute all operators in Eq. (6.1.1) so that each annihilation operator is just to the left of the creation operator with which it is paired, ignoring all commutators and anticommutators of unpaired operators, and then replace each product of paired operators with their commutators or anticommutators.) One immediate consequence is to produce the minus sign in the relative sign of the two terms in Eq. (6.1.14) for the fermion propagator. Whatever permutation puts the annihilation part $\psi^+(x)$ of a field in $\mathscr{H}(x)$ just to the left of the creation part $\psi^{+\dagger}(y)$ of a field adjoint in $\mathscr{H}(y)$, the permutation that puts the annihilation part $\psi^{-\dagger}(y)$ of the field adjoint just to the left of the creation part $\psi^-(x)$ of the field involves one extra interchange of fermion operators, yielding the minus sign in the second term of Eq. (6.1.14) for fermions.

In addition, minus signs can arise in the contribution of whole Feynman diagrams. As an example, let us take up a theory in which the sole interaction of fermions takes the form

$$\mathscr{H}(x) = \sum_{\ell mk} g_{\ell mk} \psi_\ell^\dagger(x) \psi_m(x) \phi_k(x)\,, \tag{6.1.18}$$

where $g_{\ell mk}$ are general constants, $\psi_\ell(x)$ are a set of complex fermion fields, and $\phi_m(x)$ are a set of real bosonic (but not necessarily scalar) fields. (Not only quantum electrodynamics, but the whole 'standard model' of weak, electromagnetic, and strong interactions, has fermionic interactions that can all be put in his form.) Let us first take up the process of fermion–fermion scattering, $12 \to 1'2'$, to second order in $\mathscr{H}$. The fermion operators in the second-order term in Eq. (6.1.1) appear in the order (with obvious abbreviations)

$$a(2')a(1')\psi^\dagger(x)\psi(x)\psi^\dagger(y)\psi(y)a^\dagger(1)a^\dagger(2)\quad . \tag{6.1.19}$$

There are two connected diagrams to this order, corresponding to the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

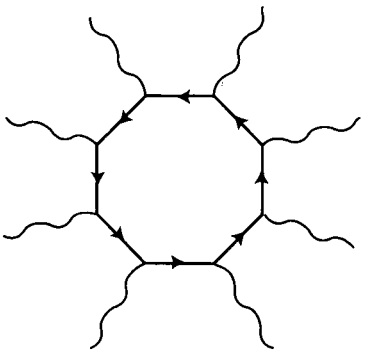

Figure 6.3. An eighth-order graph for the vacuum-to-vacuum amplitude with particles interacting only with an external field. In this diagram the external field is represented by wiggly lines. There are 7! such diagrams, differing only by relabelling the vertices, and not counting as different those labellings that simply rotate the ring. The factor 1/8! from the Dyson formula (6.1.1) is therefore not entirely cancelled here, leaving us with an extra factor 1/8.

pairings

$$[a(2')\psi^\dagger(x)]\ [a(1')\psi^\dagger(y)]\ [\psi(y)a^\dagger(1)]\ [\psi(x)a^\dagger(2)] \tag{6.1.20}$$

and

$$[a(1')\psi^\dagger(x)]\ [a(2')\psi^\dagger(y)]\ [\psi(y)a^\dagger(1)]\ [\psi(x)a^\dagger(2)]\ . \tag{6.1.21}$$

(See Figure 6.4.) To go from (6.1.19) to (6.1.20) requires an *even* permutation of fermionic operators. (For instance, move $\psi(x)$ past three operators to the right, and then move $a(1')$ past one operator to the right.) Thus there is no extra minus sign in the contribution of the pairing (6.1.20). This in itself is not so important; the *overall* sign of the S-matrix does not matter in transition rates, and in any case depends on sign conventions for the initial and final states. What is important is that the contributions of pairings (6.1.20) and (6.1.21) have opposite sign, as can be seen most easily by noting that the only difference between these two pairings is the interchange of two fermionic operators, $a(1')$ and $a(2')$. In fact, this relative minus sign is just what is required by Fermi statistics: it makes the scattering amplitude antisymmetric under the interchange of particles $1'$ and $2'$ (or 1 and 2).

However, it must not be thought that all sign factors can be related in such a simple way to the antisymmetry of the final or initial states, even in the lowest order of perturbation theory. To illustrate this point, let's now consider fermion–antifermion scattering, $1\,2^c \to 1'2'^c$, to second order in the same interaction (6.1.18). The fermionic operators in the second-order

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

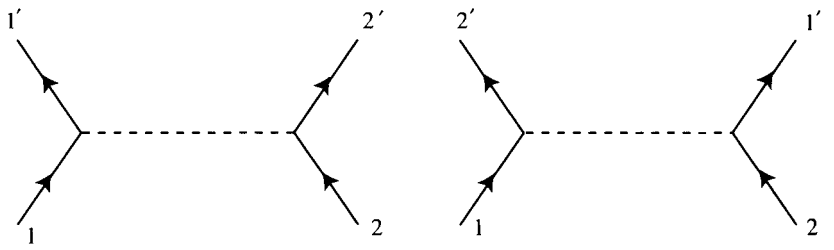

Figure 6.4. The connected second-order diagrams for fermion–fermion scattering in a theory with interaction (6.1.18). Here straight lines represent fermions; dotted lines are neutral bosons. There is a minus sign difference in the contributions of these two diagrams, arising from an extra interchange of fermion operators in the pairings represented by the second diagram.

term in Eq. (6.1.1) appear in the order:

$$a(2'^c)a(1')\psi^\dagger(x)\psi(x)\psi^\dagger(y)\psi(y)a^\dagger(1)a^\dagger(2^c) \quad . \tag{6.1.22}$$

Here again there are two Feynman diagrams to this order, corresponding to the pairings

$$[a(2'^c)\psi(x)]\ [a(1')\psi^\dagger(x)]\ [\psi(y)a^\dagger(1)]\ [\psi^\dagger(y)a^\dagger(2^c)] \tag{6.1.23}$$

and

$$[a(2'^c)\psi(x)]\ [a(1')\psi^\dagger(y)]\ [\psi(y)a^\dagger(1)]\ [\psi^\dagger(x)a^\dagger(2^c)]\ . \tag{6.1.24}$$

(See Figure 6.5.) To go from (6.1.22) to (6.1.23) requires an even permutation of fermionic operators (for instance, move $\psi(x)$ past two operators to the left and move $\psi^\dagger(y)$ past two operators to the right) so there is no extra minus sign in the contribution of the pairing (6.1.23). On the other hand, to go from (6.1.22) to (6.1.24) requires an odd permutation of fermionic operators (the same as for (6.1.23), *plus* the interchange of $\psi^\dagger(x)$ and $\psi^\dagger(y)$) so the contribution of this pairing does come with an extra minus sign.**

Additional signs are encountered when we consider contributions of higher order. In theories of the type considered here, in which the interactions of fermions all take the form (6.1.18), the fermion lines in

** Actually, this sign is not wholly unrelated to the requirements of Fermi statistics. The same field can destroy a particle and create an antiparticle, so there is a relation, known as 'crossing symmetry', between processes in which initial particles or antiparticles are exchanged with final antiparticles or particles. In particular the amplitudes for the process $1\,2^c \to 1'2'^c$ are related to those for the 'crossed' process $1\,2' \to 1'2$; the two pairings (6.1.23) and (6.1.24) just correspond to the two diagrams for this process, which differ by an interchange of 1 and 2′ (or 1′ and 2), so the antisymmetry of the scattering amplitude under interchange of initial (or final) particles naturally requires a minus sign in the relative contribution of these two pairings. However, crossing symmetry is not an ordinary symmetry (it involves an analytic continuation in kinematic variables) and it is difficult to use it with any precision for general processes.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

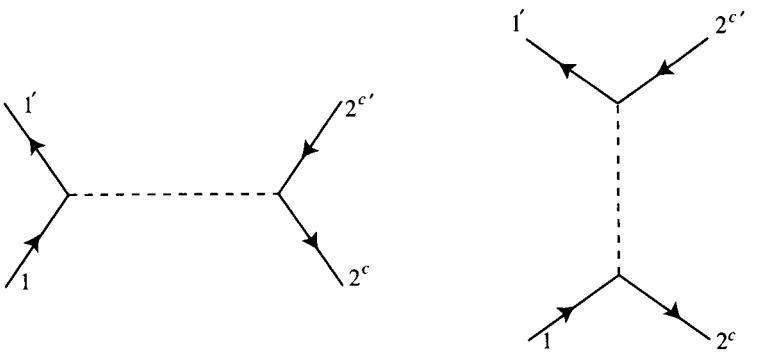

Figure 6.5. The connected second-order diagrams for fermion–antifermion scattering in a theory with interaction (6.1.18). Here straight lines represent fermions or antifermions, depending on the arrow direction; dotted lines are neutral bosons. There is again a minus sign difference in the contributions of these two diagrams, arising from an extra interchange of fermion operators in the pairings represented by the second diagram.

general Feynman diagrams form either chains of lines that pass through the diagram with arbitrary numbers of interactions with the boson fields, as in Figure 6.6, or else fermionic *loops*, like that shown in Figure 6.7. Consider the effect of adding a fermionic loop with M corners to the Feynman diagram for any process. This corresponds to the pairing of fermionic operators

$$[\psi(x_1)\bar{\psi}(x_2)]\ [\psi(x_2)\bar{\psi}(x_3)]\cdots[\psi(x_M)\bar{\psi}(x_1)]\ . \tag{6.1.25}$$

On the other hand, these operators appear in Eq. (6.1.1) in the order

$$\bar{\psi}(x_1)\psi(x_1)\bar{\psi}(x_2)\psi(x_2)\cdots\bar{\psi}(x_M)\psi(x_M)\ . \tag{6.1.26}$$

To go from (6.1.26) to (6.1.25) requires an odd permutation of fermionic operators (move $\bar{\psi}(x_1)$ to the right past $2M-1$ operators) so the contribution of each such fermionic loop is accompanied with a minus sign.

These rules yield the full S-matrix, including contributions from processes in which various clusters of particles interact in widely separated regions of spacetime. As discussed in Chapter 4, to calculate the part of the S-matrix that excludes such contributions, we should include only *connected* Feynman diagrams. In particular, this excludes lines passing clean through the diagram without interacting, which would yield the factors (6.1.13).

To make the Feynman rules perfectly clear, we will calculate the low-order contributions to the S-matrix for particle scattering in two different theories.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

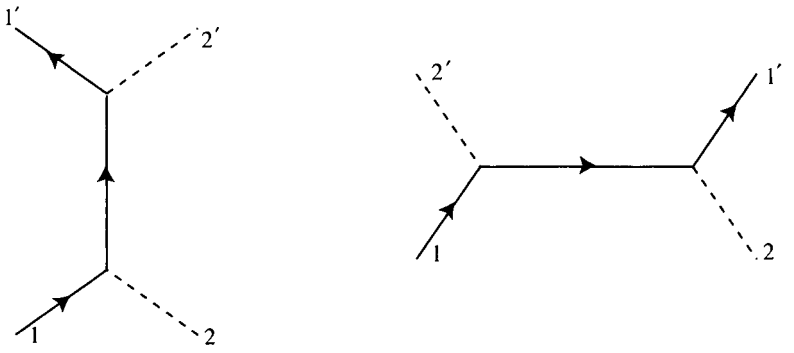

Figure 6.6. The connected second-order diagrams for boson–fermion scattering in a theory with interaction (6.1.18). Straight lines are fermions; dashed lines are neutral bosons.

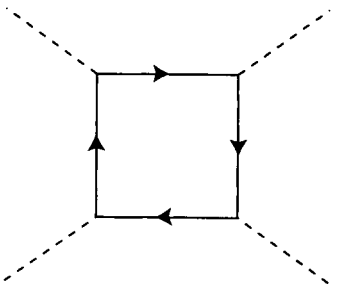

Figure 6.7. The lowest-order connected diagram for boson–boson scattering in a theory with interaction (6.1.18). Such fermion loop graphs yield an extra minus sign, arising from permutations of the paired fermion fields.

Theory I

Consider the theory of fermions and self-charge-conjugate bosons with interaction (6.1.18). The lowest-order connected diagrams for fermion–boson scattering are shown in Figure 6.6. Following the rules outlined in Figure 6.1, the corresponding S-matrix element is

$$
\begin{aligned}
&S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2'\,,\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} = \\
&(2\pi)^{-6}\sum_{k'l'm'klm}(-i)^2 g_{l'm'k'}\,g_{mlk}\,u_{l'}^{*}(\mathbf{p}_1{}'\sigma_1'n_1')\,u_l(\mathbf{p}_1\sigma_1 n_1) \\
&\times\int d^4x\int d^4y\,\Big(-i\Delta_{m'm}(y-x)\Big)e^{-ip_1'\cdot y}e^{ip_1\cdot x} \\
&\times\Big[e^{-ip_2'\cdot y}u_{k'}^{*}(\mathbf{p}_2{}'\sigma_2'n_2')e^{ip_2\cdot x}u_k(\mathbf{p}_2\sigma_2 n_2) \\
&+\;e^{-ip_2'\cdot x}u_k^{*}(\mathbf{p}_2{}'\sigma_2'n_2')e^{ip_2\cdot y}u_{k'}(\mathbf{p}_2\sigma_2 n_2)\Big]\,. \qquad (6.1.27)
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

(The labels 1 and 2 are used here for fermions and bosons, respectively.) For fermion–fermion scattering there are also two second-order diagrams, shown in Figure 6.4. They yield the S-matrix element

$$S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2'\,,\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} = (2\pi)^{-6}\sum_{k'l'm'klm}(-i)^2 g_{m'mk'}\,g_{l'lk}$$
$$\times u^*_{m'}(\mathbf{p}_2{}'\sigma_2' n_2')\,u^*_{l'}(\mathbf{p}_1{}'\sigma_1' n_1')\,u_m(\mathbf{p}_2\sigma_2 n_2)u_l(\mathbf{p}_1\sigma_1 n_1)$$
$$\times\int d^4x\int d^4y\,e^{-ip_2'\cdot x}e^{-ip_1'\cdot y}e^{ip_2\cdot x}e^{ip_1\cdot y}(-i)\Delta_{k'k}(x-y)$$
$$-[1'\rightleftharpoons 2'] \tag{6.1.28}$$

with the last term indicating subtraction of the preceding term with interchange of particles $1'$ and $2'$ (or equivalently 1 and 2). There are no second-order graphs for boson–boson scattering in this theory; the lowest-order graphs are of fourth order, such as that shown in Figure 6.7. More specific examples of formulas like Eqs. (6.1.27) and (6.1.28) will be given in Section 6.3, after we have had a chance to evaluate the propagators and go over to momentum space.

In the interaction (6.1.18) of the preceding example, the three fields are all different. It is instructive also to look at an example with a trilinear interaction in which all three fields are the same, or at least enter into the interaction in a symmetric way.

Theory II

Now take the interaction density to be a sum of terms that are trilinear in a set of *real* bosonic fields $\phi_\ell(x)$:

$$\mathscr{H}(x) = \frac{1}{3!}\sum_{\ell mn} g_{\ell mn}\phi_\ell(x)\phi_m(x)\phi_n(x) \tag{6.1.29}$$

with $g_{\ell mn}$ a real totally symmetric coupling coefficient. Suppose we want to consider a scattering process $1\ 2 \rightarrow 1'2'$ to second order in this interaction. Each of the two vertices must have two of the four external lines attached to it. (The only other possibility is that one of the external lines is attached to one vertex and three to the other vertex, but the vertex with three external lines attached to it would have no remaining lines to connect it to the other vertex so this would be a disconnected contribution.) The additional line required at each vertex must then just serve to connect the two vertices to each other. There are three graphs of this type, differing in whether the other external line that is attached to the same vertex as line 1 is line 2 or $1'$ or $2'$. (See Figure 6.8.) Following the rules given above, the contribution to the S-matrix from those three

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

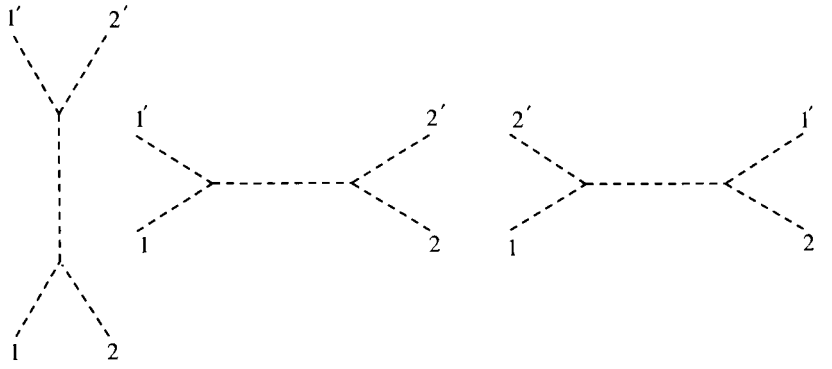

Figure 6.8. The connected-second order diagrams for boson–boson scattering in a theory with interaction (6.1.29).

diagrams is

$$
\begin{aligned}
&S_{\mathbf{p}'_1\sigma'_1 n'_1\ \mathbf{p}'_2\sigma'_2 n'_2\ ,\ \mathbf{p}_1\sigma_1 n_1\ \mathbf{p}_2\sigma_2 n_2} \\
&= (-i)^2(2\pi)^{-6} \sum_{\ell'\ell'' mm'm''} g_{\ell'\ell''} g_{mm'm''} \int d^4x \int d^4y \Big(-i\Delta_{\ell'' m''}(x,y)\Big) \\
&\quad \times \Big[u^*_{\ell'}(\mathbf{p}'_1\sigma'_1 n'_1)\, e^{-ip'_1\cdot x}\, u^*_{\ell''}(\mathbf{p}'_2\sigma'_2 n'_2)\, e^{-ip'_2\cdot x} \\
&\qquad\qquad \times u_m(\mathbf{p}_1\sigma_1 n_1) e^{ip_1\cdot y}\, u_{m'}(\mathbf{p}_2\sigma_2 n_2)\, e^{ip_2\cdot y} \\
&\quad + u^*_{\ell''}(\mathbf{p}'_1\sigma'_1 n'_1)\, e^{-ip'_1\cdot x}\, u_{\ell'}(\mathbf{p}_1\sigma_1 n_1) e^{ip_1\cdot x} \\
&\qquad\qquad \times u^*_{m'}(\mathbf{p}'_2\sigma'_2 n'_2)\, e^{-ip'_2\cdot y}\, u_m(\mathbf{p}_2\sigma_2 n_2)\, e^{ip_2\cdot y} \\
&\quad + u^*_{\ell''}(\mathbf{p}'_2\sigma'_2 n'_2)\, e^{-ip'_2\cdot x}\, u_{\ell'}(\mathbf{p}_1\sigma_1 n_1) e^{ip_1\cdot x} \\
&\qquad\qquad \times u^*_{m'}(\mathbf{p}'_1\sigma'_1 n'_1)\, e^{-ip'_1\cdot y}\, u_m(\mathbf{p}_2\sigma_2 n_2)\, e^{ip_2\cdot y} \Big]. \qquad (6.1.30)
\end{aligned}
$$

To be even more specific, if the bosons in this theory are spinless particles of a single species, then we write the interaction (6.1.29) in the form

$$\mathscr{H} = g\phi^3/3! \qquad (6.1.31)$$

and the S-matrix element (6.1.30) for scalar–scalar scattering is

$$
\begin{aligned}
&S_{\mathbf{p}'_1\ \mathbf{p}'_2\ ,\ \mathbf{p}_1\ \mathbf{p}_2} = \\
&\frac{ig^2}{(2\pi)^6\sqrt{16E'_1E'_2E_1E_2}} \int d^4x \int d^4y\ \Delta_F(x-y) \\
&\quad \times \Big[\exp(-i(p'_1+p'_2)\cdot x)\, \exp(i(p_1+p_2)\cdot y) \\
&\qquad + \exp(i(p_1-p'_1)\cdot x)\, \exp(i(p_2-p'_2)\cdot y) \\
&\qquad + \exp(i(p_1-p'_2)\cdot x)\, \exp(i(p_2-p'_1)\cdot y) \Big],
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

where $\Delta_F(x-y)$ is the scalar field propagator, calculated in the next section. There are no terms of third order in $\mathscr{H}(x)$, or of any odd order in $\mathscr{H}(x)$.

6.2 Calculation of the Propagator

We now turn to a calculation of the propagator (6.1.14), an essential ingredient in the Feynman rules that arises in the pairing of a field $\psi_\ell(x)$ with a field adjoint $\psi_m^\dagger(y)$. Inserting Eqs. (6.1.15) and (6.1.16) in Eq. (6.1.14), and using the commutation or anticommutation relations for annihilation and creation operators, we have immediately

$$-i\Delta_{\ell m}(x,y) = \theta(x-y)(2\pi)^{-3}\int d^3p \sum_\sigma u_\ell(\mathbf{p}\,\sigma\,n)u_m^*(\mathbf{p}\,\sigma\,n)e^{ip\cdot(x-y)}$$
$$\pm\,\theta(y-x)(2\pi)^{-3}\int d^3p \sum_\sigma v_m^*(\mathbf{p}\,\sigma\,n)v_\ell(\mathbf{p}\,\sigma\,n)e^{ip\cdot(y-x)}\,. \tag{6.2.1}$$

In the course of calculating commutators and anticommutators in Chapter 5, we showed that

$$\sum_\sigma u_\ell(\mathbf{p}\,\sigma\,n)u_m^*(\mathbf{p}\,\sigma\,n) = \left(2\sqrt{\mathbf{p}^2+m_n^2}\right)^{-1} P_{\ell m}\left(\mathbf{p},\sqrt{\mathbf{p}^2+m_n^2}\right)\,, \tag{6.2.2}$$

$$\sum_\sigma v_\ell(\mathbf{p}\,\sigma\,n)v_m^*(\mathbf{p}\,\sigma\,n) = \pm\left(2\sqrt{\mathbf{p}^2+m_n^2}\right)^{-1} P_{\ell m}\left(-\mathbf{p},-\sqrt{\mathbf{p}^2+m_n^2}\right)\,, \tag{6.2.3}$$

where $P_{\ell m}(\mathbf{p},\omega)$ is a polynomial in $\mathbf{p}$ and ω. (Here as in Eq. (6.2.1), the top and bottom signs refer to bosonic and fermionic fields, respectively.) For instance, if $\psi_\ell(x)$ and $\psi_m(y)$ are scalar fields $\phi(x)$ and $\phi(y)$ for a particle of spin zero, then we have simply

$$P(p) = 1\,. \tag{6.2.4}$$

If $\psi_\ell(x)$ and $\psi_m(y)$ are Dirac fields for a particle of spin $\frac{1}{2}$, then

$$P_{\ell m}(p) = \left[(-i\gamma_\mu p^\mu + m)\beta\right]_{\ell m}\,, \tag{6.2.5}$$

where ℓ and m are here four-valued Dirac indices. (The matrix β appears here because we are considering the pairing of $\psi_\ell(x)$ with $\psi_m^\dagger(y)$. It is absent in the pairing of $\psi(x)$ with $\bar\psi(y)\equiv\psi^\dagger(y)\beta$.) If $\psi_\ell(x)$ and $\psi_m(y)$ are vector fields $V_\mu(x)$ and $V_\nu(y)$ for a particle of spin one, then

$$P_{\mu\nu}(p) = \eta_{\mu\nu} + m^{-2}p_\mu p_\nu\,. \tag{6.2.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

More generally, if $\psi_\ell(x)$ and $\psi_m(y)$ are components of fields $\psi_{ab}(x)$ and $\psi_{\tilde{a}\tilde{b}}(y)$ for a particle of spin j, in the irreducible (A,B) and $(\tilde{A},\tilde{B})$ representations of the homogeneous Lorentz group, then

$$\begin{aligned} P_{ab,\tilde{a}\tilde{b}}(p) = \sum_{a'b'}\sum_{\tilde{a}'\tilde{b}'}\sum_{\sigma} & C_{AB}(j\sigma,a'b')C_{\tilde{A}\tilde{B}}(j\sigma,\tilde{a}'\tilde{b}') \\ & \times\left[\exp(-\theta\hat{p}\cdot\mathbf{J}^{(A)})\right]_{aa'}\left[\exp(+\theta\hat{p}\cdot\mathbf{J}^{(B)})\right]_{bb'} \\ & \times\left[\exp(-\theta\hat{p}\cdot\mathbf{J}^{(\tilde{A})})\right]_{\tilde{a}\tilde{a}'}\left[\exp(+\theta\hat{p}\cdot\mathbf{J}^{(\tilde{B})})\right]_{\tilde{b}\tilde{b}'}, \end{aligned} \tag{6.2.7}$$

where $\sinh\theta = |\mathbf{p}|/m$, while $a,b,\tilde{a},\tilde{b}$ run by unit steps from $-A$ to $+A$, $-B$ to $+B$, $-\tilde{A}$ to $+\tilde{A}$, and $-\tilde{B}$ to $+\tilde{B}$, respectively, and likewise for the running indices $a',b',\tilde{a}'$, and $\tilde{b}'$.

Inserting Eqs. (6.2.2) and (6.2.3) in Eq. (6.2.1) yields

$$\begin{aligned} -i\Delta_{\ell m}(x,y) = {} & \theta(x-y)P_{\ell m}\left(-i\frac{\partial}{\partial x}\right)\Delta_+(x-y) \\ & + \theta(y-x)P_{\ell m}\left(-i\frac{\partial}{\partial x}\right)\Delta_+(y-x), \end{aligned} \tag{6.2.8}$$

where $\Delta_+(x)$ is the function introduced in Chapter 5

$$\Delta_+(x) \equiv (2\pi)^{-3}\int d^3p\,(2p^0)^{-1}e^{ip\cdot x} \tag{6.2.9}$$

in which p^0 is taken as $+\sqrt{\mathbf{p}^2+m^2}$.

To go further, we must say a bit about how to extend the definition of the polynomial $P(p)$. Eqs. (6.2.2) and (6.2.3) only define $P(p)$ for four-momenta 'on the mass shell', i.e., with $p^0 = \pm\sqrt{\mathbf{p}^2+m^2}$. Any polynomial function of such four-momentum can always be taken as linear in p^0, because any power $(p^0)^{2\nu}$ or $(p^0)^{2\nu+1}$ can be written as $(\mathbf{p}^2+m^2)^\nu$ or $p^0(\mathbf{p}^2+m^2)^\nu$, respectively. Thus we can define a polynomial $P^{(L)}(q)$ by the conditions that

$$\begin{aligned} & P^{(L)}(p) = P(p) \qquad \left(\text{for } p^0 = \sqrt{\mathbf{p}^2+m^2}\right), \\ & P^{(L)}(q) = P^{(0)}(\mathbf{q}) + q^0P^{(1)}(\mathbf{q}) \quad (\text{for general } q^\mu), \end{aligned} \tag{6.2.10}$$

where $P^{(0,1)}$ are polynomials depending only on $\mathbf{q}$. We can now use the relations

$$\frac{\partial}{\partial x^0}\theta(x^0-y^0) = -\frac{\partial}{\partial x^0}\theta(y^0-x^0) = \delta(x^0-y^0) \tag{6.2.11}$$

(recall that $\theta(x)$ has a unit step at x^0, and is otherwise constant) to move

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

the derivative operators to the left of the θ functions in Eq. (6.2.8)

$$\Delta_{\ell m}(x,y) = P^{(L)}_{\ell m}\left(-i\frac{\partial}{\partial x}\right)\Delta_F(x-y) + \delta(x^0-y^0)P^{(1)}_{\ell m}(-i\nabla)\left[\Delta_+(x-y)-\Delta_+(y-x)\right], \quad (6.2.12)$$

where Δ_F is the 'Feynman propagator'

$$-i\Delta_F(x) \equiv \theta(x)\Delta_+(x) + \theta(-x)\Delta_+(-x)\,. \quad (6.2.13)$$

However, for $x^0 = 0$ the function $\Delta_+(x)$ is even in $\mathbf{x}$, since a change $\mathbf{x} \to -\mathbf{x}$ in Eq. (6.2.9) can be compensated by a change $\mathbf{p} \to -\mathbf{p}$ in the integration variable. We can therefore drop the second term in Eq. (6.2.12), and write simply

$$\Delta_{\ell m}(x,y) = P^{(L)}_{\ell m}\left(-i\frac{\partial}{\partial x}\right)\Delta_F(x-y)\,. \quad (6.2.14)$$

It will be most useful to use the expression of the Feynman propagator as a Fourier integral. The step functions in Eq. (6.2.13) have the Fourier representation*

$$\theta(t) = \frac{-1}{2\pi i}\int_{-\infty}^{\infty}\frac{\exp(-ist)}{s+i\epsilon}\,ds\,. \quad (6.2.15)$$

This can be combined with the Fourier integral (6.2.9) for $\Delta_+(x)$. We introduce new integration variables, $\mathbf{q} \equiv \mathbf{p}$, $q^0 = p^0 + s$ in the first term of Eq. (6.2.13), yielding

$$-i\Delta_F(x) = -\frac{1}{2\pi i}\int d^3q\int_{-\infty}^{\infty}dq^0\,\frac{\exp(i\mathbf{q}\cdot\mathbf{x} - iq^0x^0)}{(2\pi)^3 2\sqrt{\mathbf{q}^2+m^2}} \times\left[\left(q^0-\sqrt{\mathbf{q}^2+m^2}+i\epsilon\right)^{-1} + \left(-q^0-\sqrt{\mathbf{q}^2+m^2}+i\epsilon\right)^{-1}\right].$$

Combining denominators and adopting a four-dimensional notation, we have simply

$$\Delta_F(x) = (2\pi)^{-4}\int d^4q\,\frac{\exp(iq\cdot x)}{q^2+m^2-i\epsilon}\,, \quad (6.2.16)$$

where $q^2 \equiv \mathbf{q}^2 - (q^0)^2$. (In the denominator we have replaced $2\epsilon\sqrt{\mathbf{q}^2+m^2}$ with ϵ, because the only important thing about this quantity is that it

* To prove this, note that if $t > 0$ then the contour of integration can be closed with a large clockwise semi-circle in the lower half-plane, so the integral picks up a contribution of $-2\pi i$ from the pole at $s = -i\epsilon$. If $t < 0$ then the contour can be closed with a large counter-clockwise semi-circle in the upper half-plane, where the integrand is analytic, giving an integral equal to zero.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

is a positive infinitesimal.) This shows incidentally that Δ_F is a Green's function for the Klein–Gordon differential operator, in the sense that

$$(\Box - m^2)\Delta_F(x) = -\delta^4(x) \tag{6.2.17}$$

with boundary conditions specified by the $-i\epsilon$ in the denominator: as shown by Eq. (6.2.13), $\Delta_F(x)$ for $x^0 \to +\infty$ or $x^0 \to -\infty$ involves only positive or negative frequency terms, $\exp(-ix^0\sqrt{\mathbf{p}^2+m^2})$ or $\exp(+ix^0\sqrt{\mathbf{p}^2+m^2})$, respectively.

Inserting Eq. (6.2.16) in Eq. (6.2.14) now gives the propagator as

$$\Delta_{\ell m}(x,y) = (2\pi)^{-4}\int d^4q\,\frac{P^{(L)}_{\ell m}(q)e^{iq\cdot(x-y)}}{q^2+m^2-i\epsilon}\,. \tag{6.2.18}$$

There is one obvious problem with this expression. The polynomial $P(p)$ is Lorentz-covariant when p is on the mass shell, $p^2 = -m^2$, but in Eq. (6.2.18) we integrate over all q^μ, not restricted to the mass shell. The polynomial $P^{(L)}(q)$ is defined for general q^μ to be linear in q^0, a condition that clearly does not respect Lorentz covariance unless the polynomial is also linear in each spatial component q^i as well. We can instead always define our extension of the polynomial $P(p)$ to general four-momenta q^μ, which we shall call simply $P(q)$, in such a way that $P(q)$ *is* Lorentz-covariant for general q^μ, in the sense that

$$P_{\ell m}(\Lambda q) = D_{\ell\ell'}(\Lambda)D^*_{mm'}(\Lambda)P_{\ell' m'}(q)\,,$$

where $\Lambda^\mu{}_\nu$ is a general Lorentz transformation, and $D(\Lambda)$ is the appropriate representation of the Lorentz group. For instance, for scalar, Dirac, and four-vector fields, these covariant extensions are obviously provided by just replacing p^μ with a general four-momentum q^μ in Eqs. (6.2.4), (6.2.5), and (6.2.6). For the scalar and Dirac fields, these are already linear in q^0, so here there is no difference between $P^{(L)}(q)$ and $P(q)$:

$$P^{(L)}_{\ell m}(q) = P_{\ell m}(q) \qquad \text{(scalar, Dirac fields)}\,. \tag{6.2.19}$$

On the other hand, for the vector field of a spin one particle, the 00 components of the covariant polynomial $P_{\mu\nu}(q) \equiv \eta_{\mu\nu} + m^{-2}q_\mu q_\nu$ are quadratic in q^0, so here there *is* a difference:

$$\begin{aligned} P^{(L)}_{\mu\nu}(q) &= \eta_{\mu\nu} + m^{-2}\left[q_\mu q_\nu - \delta^0_\mu\delta^0_\nu(q_0^2 - \mathbf{q}^2 - m^2)\right] \\ &= P_{\mu\nu}(q) + m^{-2}(q^2+m^2)\delta^0_\mu\delta^0_\nu\,. \end{aligned} \tag{6.2.20}$$

(The extra term here is fixed by the two conditions that it must cancel the $(q_0)^2$ term in $P_{00}(q)$, and must vanish when q^μ is on the mass shell.) Inserting this in Eq. (6.2.18) gives the propagator of a vector field as

$$\Delta_{\mu\nu}(x,y) = (2\pi)^{-4}\int d^4q\,\frac{P_{\mu\nu}(q)e^{iq\cdot(x-y)}}{q^2+m^2-i\epsilon} + m^{-2}\delta^4(x-y)\delta^0_\mu\delta^0_\nu\,. \tag{6.2.21}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

The first term is manifestly covariant, and the second term, though not covariant, is local, so it can be cancelled by adding a local non-covariant term to the Hamiltonian density. Specifically, if $V_\mu(x)$ interacts with other fields through a term $V_\mu(x)J^\mu(x)$ in $\mathscr{H}(x)$, then the effect of the second term in Eq. (6.2.21) is to produce an effective interaction

$$-i\mathscr{H}_{eff}(x) = \tfrac{1}{2}\Big[-iJ^\mu(x)\Big]\Big[-iJ^\nu(x)\Big]\Big[-im^{-2}\delta^0_\mu\delta^0_\nu\Big] .$$

(The factors $-i$ are the usual ones which always accompany vertices and propagators. The factor $\frac{1}{2}$ is needed because there are two ways to pair other fields with $\mathscr{H}_{eff}(x)$, differing in the interchange of J^μ and J^ν.) Thus the effect of the non-covariant second term in Eq. (6.2.21) can be cancelled by adding to $\mathscr{H}(x)$ the non-covariant term

$$\mathscr{H}_{NC}(x) = -\mathscr{H}_{eff}(x) = \frac{1}{2m^2}\Big[J^0(x)\Big]^2 . \tag{6.2.22}$$

It is the singularity of the equal-time commutators of vector fields at zero separation that requires us to employ a wider class of interactions than those with a scalar density. A detailed non-perturbative proof of the Lorentz invariance of the S-matrix in this theory will be given in the next chapter.

It should not be thought that this is solely a phenomenon associated with spins $j \geq 1$. For instance, consider the vector field associated with a particle of spin $j = 0$, equal (as discussed in Chapter 5) to the derivative $\partial_\lambda\phi(x)$ of a scalar field. For the pairing of this field with a scalar $\phi^\dagger(y)$, the polynomial $P(p)$ on the mass shell is

$$P_\lambda(p) = ip_\lambda , \tag{6.2.23}$$

while the pairing of $\partial_\lambda\phi(x)$ with $\partial_\eta\phi^\dagger(y)$ yields a polynomial

$$P_{\lambda,\eta}(p) = p_\lambda p_\eta . \tag{6.2.24}$$

The covariant polynomials for general off-shell four-momenta q^μ are again obtained by just substituting q^μ for p^μ in Eqs. (6.2.23) and (6.2.24). Eq. (6.2.23) shows that $P_\lambda(q)$ is already linear in q_0, so here there is no difference between $P_\lambda(q)$ and $P^{(L)}_\lambda(q)$. However, for Eq. (6.2.24) there is a difference:

$$\begin{aligned} P^{(L)}_{\lambda,\eta}(q) &= q_\lambda q_\eta - (q_0^2 - \mathbf{q}^2 - m^2)\delta^0_\lambda\delta^0_\eta \\ &= P_{\lambda,\eta}(q) + (q^2 + m^2)\delta^0_\lambda\delta^0_\eta , \end{aligned} \tag{6.2.25}$$

so here the propagator is

$$\Delta_{\lambda,\eta}(x,y) = (2\pi)^{-4}\int d^4q\ \frac{q_\lambda q_\eta e^{iq\cdot x}}{q^2 + m^2 - i\epsilon} + \delta^0_\lambda\delta^0_\eta\delta^4(x-y) . \tag{6.2.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

Just as before, the non-covariant effects of the second term may be removed by adding to the interaction a non-covariant term

$$\mathscr{H}_{NC}(x) = \tfrac{1}{2}\left[J^0(x)\right]^2, \tag{6.2.27}$$

where $J^\mu(x)$ is here the current which multiplies $\partial_\mu\phi(x)$ in the covariant part of $\mathscr{H}(x)$.

It should be clear that (at least for massive particles) the effects of non-covariant parts of the propagator can always be cancelled in this way by adding non-covariant local terms to the Hamiltonian density. This is because the numerator $P^{(L)}_{\ell m}(q)$ in the propagator must equal the covariant polynomial $P_{\ell m}(q)$ when q^μ is on the mass shell, so the difference between $P^{(L)}_{\ell m}(q)$ and $P_{\ell m}(q)$ must contain a factor $q^2 + m^2$. This factor cancels the denominator $(q^2 + m^2 - i\epsilon)$ in the contribution of this difference to Eq. (6.2.18), so Eq. (6.2.18) always equals a covariant term plus a term proportional to the delta function $\delta^4(x-y)$ or its derivatives. The effect of the latter term may be cancelled by adding to the interaction a term quadratic in the currents to which the paired fields couple, or in their derivatives. In what follows, it will be assumed tacitly that such a term has been included in the interaction, and in consequence we shall use the *covariant* polynomial $P_{\ell m}(q)$ in the propagator (6.2.18), and will thus henceforth drop the label 'L'.

It may seem that this is a rather *ad hoc* procedure. Fortunately, in the canonical formalism discussed in the following chapter, the non-covariant term in the Hamiltonian density needed to cancel non-covariant terms in the propagator arises automatically. This, in fact, forms part of the motivation for introducing the canonical formalism.

* * *

Before closing this section, it may be useful to mention some other definitions of the propagator, equivalent to Eq. (6.2.1), that appear commonly in the literature. First, taking the vacuum expectation value of Eq. (6.1.14) gives

$$\begin{aligned} -i\Delta_{\ell m}(x,y) = {} & \theta(x-y)\left\langle\left[\psi^+_\ell(x), \psi^{+\dagger}_m(y)\right]_\mp\right\rangle_0 \\ & \pm \theta(y-x)\left\langle\left[\psi^{-\dagger}_m(y), \psi^-_\ell(x)\right]_\mp\right\rangle_0 . \end{aligned} \tag{6.2.28}$$

(Here $\langle AB\cdots\rangle_0$ denotes the vacuum expectation value $(\Phi_0, AB\cdots\Phi_0)$.) Both $\psi^+_\ell(x)$ and $\psi^{-\dagger}_m(y)$ annihilate the vacuum, so only one term in each commutator or anticommutator in Eq. (6.2.28) actually contributes to the propagator:

$$-i\Delta_{\ell m}(x,y) = \theta(x-y)\langle\psi^+_\ell(x)\psi^{+\dagger}_m(y)\rangle_0 \pm \theta(y-x)\langle\psi^{-\dagger}_m(y)\psi^-_\ell(x)\rangle_0 . \tag{6.2.29}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

Also, $\psi^{-\dagger}$ and ψ^{+} would annihilate a vacuum state on the right, and ψ^{-} and $\psi^{+\dagger}$ would annihilate the vacuum state on the left, so ψ^{+} and ψ^{-} may be replaced everywhere in Eq. (6.2.29) with the complete field $\psi = \psi^{+} + \psi^{-}$:

$$-i\Delta_{\ell m}(x,y) = \theta(x-y)\langle\psi_\ell(x)\psi_m^\dagger(y)\rangle_0 \pm \theta(y-x)\langle\psi_m^\dagger(y)\psi_\ell(x)\rangle_0 \,. \quad (6.2.30)$$

This is often written

$$-i\Delta_{\ell m}(x,y) = \langle T\{\psi_\ell(x)\psi_m^\dagger(y)\}\rangle_0 \,, \quad (6.2.31)$$

where T is a time-ordered product, whose definition is now extended** to all fields, with a minus sign for any odd permutation of fermionic operators.

6.3 Momentum Space Rules

The Feynman rules outlined in Section 6.1 specify how to calculate the contribution to the S-matrix of a given Nth order diagram, as the integral over N spacetime coordinates of a product of spacetime-dependent factors. For a final particle (or antiparticle) line with momentum p'^{μ} leaving a vertex with spacetime coordinate x^{μ}, we get a factor proportional to $\exp(-ip'\cdot x)$, and for an initial particle line with momentum p^{μ} entering a vertex with spacetime coordinate x^{μ}, we get a factor proportional to $\exp(+ip\cdot x)$. In Section 6.2 we saw that the factor associated with an internal line running from y to x can be expressed as a Fourier integral, over off-shell four-momenta q^{μ}, of an integrand proportional to $\exp(iq\cdot(x-y))$. We can think of q^{μ} as the four-momentum flowing along the internal line in the direction of the arrow from y to x. Hence the integral over each vertex's spacetime position merely yields a factor

$$(2\pi)^4\,\delta^4\left(\sum p + \sum q - \sum p' - \sum q'\right) \,, \quad (6.3.1)$$

where $\sum p'$ and $\sum p$ denote the total four-momentum of all the final or initial particles leaving or entering the vertex, and $\sum q'$ and $\sum q$ denote the total four-momentum of all the internal lines with arrows leaving or entering the vertex, respectively. Of course, in place of these integrals over x^{μ}s, we now have to do integrals over the Fourier variables q^{μ}, one for each internal line.

These considerations can be encapsulated in a new set of Feynman rules

** This is not inconsistent with our previous definition of the time-ordered product of Hamiltonian densities in Chapter 3, because the Hamiltonian density can only contain even numbers of fermionic field factors.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

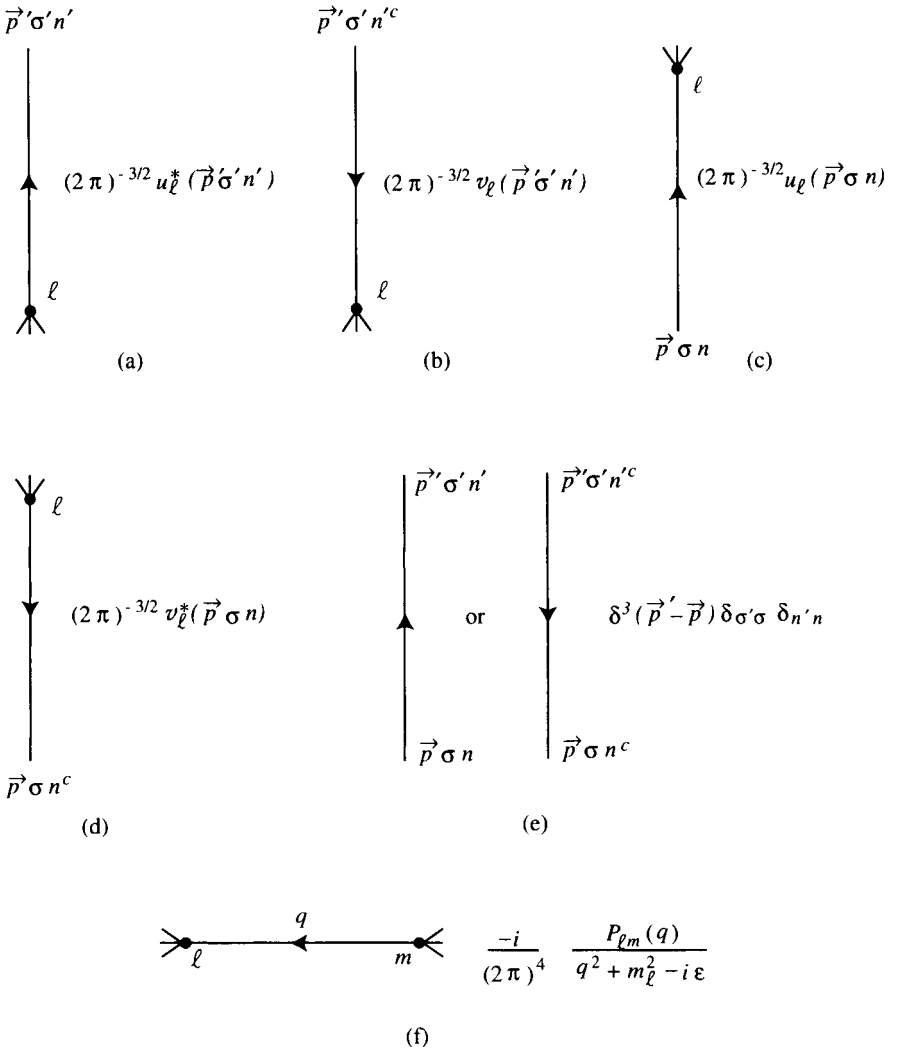

Figure 6.9. Graphical representation of pairings of operators arising in the momentum-space evaluation of the S-matrix. The expressions on the right are the factors that must be included in the momentum space integrand of the S-matrix for each line of the Feynman diagram.

(see Figure 6.9) for calculating contribution to the S-matrix as integrals over momentum variables:

(i) Draw all Feynman diagrams of the desired order, just as described in Section 6.1. However, instead of labelling each vertex with a spacetime coordinate, each internal line is now labelled with an off-mass-shell four-momentum, considered conventionally to flow in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

direction of the arrow (or in either direction for neutral particle lines without arrows.)

(ii) For each vertex of type i, include a factor

$$-i(2\pi)^4 g_i\,\delta^4\left(\sum p+\sum q-\sum p'-\sum q'\right) \tag{6.3.2}$$

with the momentum sums having the same meaning as in (6.3.1). This delta function ensures that the four-momentum is conserved at every point in the diagram. For each external line running upwards out of the diagram, include a factor $(2\pi)^{-3/2}u^*_\ell(\mathbf{p}'\sigma'n')$ or $(2\pi)^{-3/2}v_\ell(\mathbf{p}'\sigma'n')$, for arrows pointing upwards or downwards, respectively. For each external line running from below into the diagram, include a factor $(2\pi)^{-3/2}u_\ell(\mathbf{p}\sigma n)$ or $(2\pi)^{-3/2}v^*_\ell(\mathbf{p}\sigma n)$, for arrows pointing upwards or downwards, respectively. For each internal line with ends labelled ℓ and m, the arrow pointing from m to ℓ, and carrying a momentum label q^μ, include as a factor the coefficient of $e^{iq\cdot x}$ in the integral for $-i\Delta_{\ell m}(x)$:

$$-i(2\pi)^{-4}P_{\ell m}(q)\Big/(q^2+m_\ell^2-i\epsilon)\,. \tag{6.3.3}$$

A reminder: for scalars or antiscalars of four-momentum q, the us and vs are simply $(2q^0)^{-1/2}$, while the polynomial $P(q)$ is unity. For Dirac spinors of four-momentum p and mass M, the us and vs are the normalized Dirac spinors described in Section 5.5, and the polynomial $P(p)$ is the matrix $(-i\gamma_\mu p^\mu+M)\beta$.

(iii) Integrate the product of all these factors over the four-momenta carried by internal lines, and sum over all field indices ℓ, m, etc.

(iv) Add up the results obtained in this way from each Feynman diagram.

Additional combinatoric factors and fermionic signs may need to be included, as described in parts (v) and (vi) of Section 6.1. Examples will be given at the end of this section.

We have a four-momentum integration variable for every internal line, but many of these are eliminated by the delta functions associated with vertices. Since energy and momentum are separately conserved for each connected part of a Feynman diagram, there will be C delta functions left over in a graph with C connected parts. Hence in a diagram with I internal lines and V vertices, the number of independent four-momenta that are *not* fixed by the delta functions is $I-[V-C]$. This is clearly also the number L of independent loops:

$$L=I-V+C\,, \tag{6.3.4}$$

which is defined as the maximum number of internal lines that can be cut without disconnecting the diagram, because any such and only

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

such internal lines can be assigned an independent four-momentum. We can think of the independent momentum variables as characterizing the momenta that circulate in each loop. In particular a *tree* graph is one without loops; after taking the delta functions into account there are no momentum-space integrals left for such graphs.

For instance, in a theory with interaction (6.1.18), the S-matrix (6.1.27) for fermion–boson scattering is given by the momentum-space Feynman rules as

$$\begin{aligned}
&S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2',\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} = \\
&\sum_{k'l'm'klm}(-i)^2(2\pi)^8 g_{l'm'k'}\,g_{mlk}\,u^*_{l'}(\mathbf{p}_1{}'\sigma_1'n_1')\,u_l(\mathbf{p}_1\sigma_1 n_1) \\
&\times\int d^4q\,\left(-i(2\pi)^{-4}\frac{P_{m'm}(q)}{q^2+m_m^2-i\epsilon}\right) \\
&\times(2\pi)^{-6}\Big[u^*_{k'}(\mathbf{p}_2{}'\sigma_2'n_2')u_k(\mathbf{p}_2\sigma_2 n_2)\delta^4(p_1+p_2-q)\delta^4(q-p_{1'}-p_{2'}) \\
&\quad+\;u^*_k(\mathbf{p}_2{}'\sigma_2'n_2')u_{k'}(\mathbf{p}_2\sigma_2 n_2)\delta^4(p_2-p_{1'}+q)\delta^4(p_1-p_{2'}-q)\Big]\,,
\end{aligned}$$

with labels 1 and 2 here denoting fermions and bosons, respectively. The momentum-space integral here is trivial, and gives

$$\begin{aligned}
&S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2',\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} = i(2\pi)^{-2}\delta^4(p_1+p_2-p_1'-p_2') \\
&\times\sum_{k'l'm'klm} g_{l'm'k'}\,g_{mlk}\,u^*_{l'}(\mathbf{p}_1{}'\sigma_1'n_1')\,u_l(\mathbf{p}_1\sigma_1 n_1) \\
&\times\left[\frac{P_{m'm}(p_1+p_2)}{(p_1+p_2)^2+m_m^2-i\epsilon}\,u^*_{k'}(\mathbf{p}_2{}'\sigma_2'n_2')u_k(\mathbf{p}_2\sigma_2 n_2)\right. \\
&\quad\left.+\;\frac{P_{m'm}(p_{2'}-p_1)}{(p_{2'}-p_1)^2+m_m^2-i\epsilon}\,u^*_k(\mathbf{p}_2{}'\sigma_2'n_2')u_{k'}(\mathbf{p}_2\sigma_2 n_2)\right]\,. \qquad (6.3.5)
\end{aligned}$$

In the same way, the S-matrix element (6.1.28) for fermion–fermion scattering in the same theory is

$$\begin{aligned}
&S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2',\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} = i(2\pi)^{-2}\delta^4(p_1+p_2-p_1'-p_2') \\
&\times\sum_{k'l'm'klm} g_{m'mk'}\,g_{l'lk}\frac{P_{k'k}(p_{1'}-p_1)}{(p_{1'}-p_1)^2+m_k^2-i\epsilon} \\
&\times u^*_{m'}(\mathbf{p}_2{}'\sigma_2'n_2')\,u^*_{l'}(\mathbf{p}_1{}'\sigma_1'n_1')\,u_m(\mathbf{p}_2\sigma_2 n_2)u_l(\mathbf{p}_1\sigma_1 n_1) \\
&-[1'\rightleftharpoons 2']\,. \qquad (6.3.6)
\end{aligned}$$

These results illustrate the need for a more compact notation. We may define a fermion–boson coupling matrix

$$[\Gamma_k]_{lm}\equiv g_{lmk}\,. \qquad (6.3.7)$$

The matrix elements (6.3.5) and (6.3.6) for fermion–boson and fermion–

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

fermion scattering can then be rewritten in matrix notation as

$$
\begin{aligned}
S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2'\,,\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} &= i(2\pi)^{-2}\delta^4(p_1+p_2-p_1'-p_2')\sum_{k'k}\\
\times\Bigg[&\left(u^\dagger(\mathbf{p}_1{}'\sigma_1'n_1')\,\Gamma_{k'}\frac{P(p_1+p_2)}{(p_1+p_2)^2+M^2-i\epsilon}\Gamma_k u(\mathbf{p}_1\sigma_1 n_1)\right)\\
&\times u^*_{k'}(\mathbf{p}_2{}'\sigma_2'n_2')u_k(\mathbf{p}_2\sigma_2 n_2)\\
+&\left(u^\dagger(\mathbf{p}_1{}'\sigma_1'n_1')\,\Gamma_{k'}\frac{P(p_1-p_2')}{(p_1-p_2')^2+M^2-i\epsilon}\Gamma_k u(\mathbf{p}_1\sigma_1 n_1)\right)\\
&\times u^*_k(\mathbf{p}_2{}'\sigma_2'n_2')u_{k'}(\mathbf{p}_2\sigma_2 n_2)\Bigg]
\end{aligned}
\tag{6.3.8}
$$

and

$$
\begin{aligned}
S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2'\,,\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} &= i(2\pi)^{-2}\delta^4(p_1+p_2-p_1'-p_2')\\
&\times\sum_{k'k}\frac{P_{k'k}(p_{1'}-p_1)}{(p_{1'}-p_1)^2+m_k^2-i\epsilon}\\
&\times\left(u^\dagger(\mathbf{p}_2{}'\sigma_2'n_2')\,\Gamma_{k'}u(\mathbf{p}_2\sigma_2 n_2)\right)\left(u^\dagger(\mathbf{p}_1'\sigma_1'n_1')\Gamma_k\,u(\mathbf{p}_1\sigma_1 n_1)\right)\\
&-[1'\rightleftharpoons 2']\,,
\end{aligned}
\tag{6.3.9}
$$

where M^2 and m^2 are the diagonal mass matrices of the fermions and bosons in Eqs. (6.3.8) and (6.3.9), respectively. The general rule is that in using matrix notation, one writes coefficient functions, coupling matrices, and propagators in an order dictated by following lines *backwards* from the order indicated by the arrows. In the same notation, the S-matrix for boson–boson scattering in the same theory would be given by a sum of one-loop diagrams, shown in Figure 6.7:

$$
\begin{aligned}
S_{\mathbf{p}_1{}'\sigma_1'n_1'\,\mathbf{p}_2{}'\sigma_2'n_2'\,,\,\mathbf{p}_1\sigma_1 n_1\,\mathbf{p}_2\sigma_2 n_2} &= -(2\pi)^{-6}\delta^4(p_1+p_2-p_1'-p_2')\\
\times\sum_{k_1k_2k_1'k_2'}&u^*_{k_1'}(\mathbf{p}_1',\sigma_1',n_1')u^*_{k_2'}(\mathbf{p}_2',\sigma_2',n_2')u_{k_1}(\mathbf{p}_1,\sigma_1,n_1)u_{k_2}(\mathbf{p}_2,\sigma_2,n_2)\\
\times\int d^4q\ \mathrm{Tr}\Bigg\{&\Gamma_{k_2'}\frac{P(q)}{q^2+M^2-i\epsilon}\Gamma_{k_1'}\frac{P(q+p_1')}{(q+p_1')^2+M^2-i\epsilon}\\
&\times\Gamma_{k_1}\frac{P(q+p_1'-p_1)}{(q+p_1'-p_1)^2+M^2-i\epsilon}\Gamma_{k_2}\frac{P(q-p_2')}{(q-p_2')^2+M^2-i\epsilon}\Bigg\}\\
+\dots\,,&
\end{aligned}
\tag{6.3.10}
$$

where the ellipsis in the last line indicates terms obtained by permuting bosons 1′, 2′, and 2. The minus sign at the beginning of the right-hand side is the extra minus sign associated with fermionic loops. Note that after elimination of delta functions there is just one momentum-space

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

integral here, as appropriate for a diagram with one loop. We shall see how to do this sort of momentum-space integral in Chapter 11.

To make this more specific, consider a theory with a Dirac spinor field $\psi(x)$ of mass M and a pseudoscalar field $\phi(x)$ of mass m, interacting through the interaction $-ig\phi\bar{\psi}\gamma_5\psi$. (The factor $-i$ is inserted to make this interaction Hermitian for real coupling constants g.) Recall that the polynomial $P(q)$ for the scalar is just unity, while for the spinor it is $[-i\gamma_\mu q^\mu + M]\beta$. Also, the u for a scalar of energy E is $(2E)^{-1/2}$, while for the spinor u is the conventionally normalized Dirac spinor discussed in Section 5.5. Eqs. (6.3.8), (6.3.9), and (6.3.10) give the lowest-order connected S-matrix elements for fermion–boson scattering, fermion–fermion scattering, and boson–boson scattering:

$$\begin{aligned}
S_{\mathbf{p}_1'\sigma_1'\,\mathbf{p}_2'\,,\,\mathbf{p}_1\sigma_1\,\mathbf{p}_2} &= -i(2\pi)^{-2}g^2(4E_2'E_2)^{-1/2}\delta^4(p_1+p_2-p_1'-p_2') \\
&\times\Bigg[\left(\bar{u}(\mathbf{p}_1{}'\sigma_1')\,\gamma_5\frac{-i\gamma_\mu(p_1+p_2)^\mu+M}{(p_1+p_2)^2+M^2-i\epsilon}\gamma_5 u(\mathbf{p}_1\sigma_1)\right) \\
&+\left(\bar{u}(\mathbf{p}_1{}'\sigma_1')\,\gamma_5\frac{-i\gamma_\mu(p_1-p_2')^\mu+M}{(p_1-p_2')^2+M^2-i\epsilon}\gamma_5\,u(\mathbf{p}_1\,\sigma_1)\right)\Bigg], \\
S_{\mathbf{p}_1{}'\sigma_1'\,\mathbf{p}_2{}'\sigma_2'\,,\,\mathbf{p}_1\sigma_1\,\mathbf{p}_2\sigma_2} &= -i(2\pi)^{-2}g^2\delta^4(p_1+p_2-p_1'-p_2') \\
&\times\left(\bar{u}(\mathbf{p}_2'\,\sigma_2')\,\gamma_5\,u(\mathbf{p}_2\,\sigma_2)\right)\left(\bar{u}(\mathbf{p}_1'\,\sigma_1')\gamma_5\,u(\mathbf{p}_1\,\sigma_1)\right) \\
&\times\frac{1}{(p_{1'}-p_1)^2+m^2-i\epsilon} \\
&-[1'\rightleftharpoons 2'], \\
S_{\mathbf{p}_1'\,\mathbf{p}_2'\,,\,\mathbf{p}_1\,\mathbf{p}_2} &= -(2\pi)^{-6}\,g^4\,(16E_1E_2E_1'E_2')^{-1/2}\,\delta^4(p_1+p_2-p_1'-p_2') \\
&\times\int d^4q\ \mathrm{Tr}\Bigg\{\gamma_5\frac{-i\gamma_\mu q^\mu+M}{q^2+M^2-i\epsilon}\gamma_5\frac{-i\gamma_\mu(q+p_1')^\mu+M}{(q+p_1')^2+M^2-i\epsilon} \\
&\times\gamma_5\frac{-i\gamma_\mu(q+p_1'-p_1)^\mu+M}{(q+p_1'-p_1)^2+M^2-i\epsilon}\gamma_5\frac{-i\gamma_\mu(q-p_2')^\mu+M}{(q-p_2')^2+M^2-i\epsilon}\Bigg\}+\cdots\ ,
\end{aligned}$$

where in the last formula the ellipsis indicates a sum over permutations of particles 2, 1′, 2′. The factors β in the fermion propagator numerators have been used to replace $u^\dagger$ with $\bar{u}$.

* * *

Another useful topological result expresses a sort of conservation law of lines. For the moment we can think of all internal and external lines as being created at vertices and destroyed in pairs at the centers of internal lines or when external lines leave the diagram. (This has nothing to do with the directions of the arrows carried by these lines.) Equating the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

numbers of lines that are created and destroyed then gives

$$2I + E = \sum_i n_i V_i \,, \tag{6.3.11}$$

where I and E are the numbers of internal and external lines, V_i are the numbers of vertices of various types labelled i, and n_i are the number of lines attached to each vertex. (This also holds separately for fields of each type.) In particular, if all interactions involve the same number $n_i = n$ of fields, then this reads

$$2I + E = nV \,, \tag{6.3.12}$$

where V is the total number of all vertices. In this case, we can eliminate I from Eqs. (6.3.4) and (6.3.11), and find that for a connected (i.e., $C = 1$) graph the number of vertices is given by

$$V = \frac{2L + E - 2}{n - 2} \,. \tag{6.3.13}$$

For instance for a trilinear interaction the diagrams for a scattering process ($E = 4$) with $L = 0, 1, 2 \cdots$ has $V = 2, 4, 6 \cdots$ vertices. In general, the expansion in powers of the coupling constants is an expansion in increasing numbers of loops.

6.4 Off the Mass Shell

In the Feynman diagrams for any S-matrix element all external lines are 'on the mass shell'; that is, the four-momentum associated with an external line for a particle of mass m is constrained to satisfy $p_\mu p^\mu = -m^2$. It is often important also to consider Feynman diagrams 'off the mass shell', for which the external line energies like the energies associated with internal lines are free variables, unrelated to any three-momenta. For one thing, these arise as parts of larger Feynman diagrams; for instance, a loop appearing as an insertion in some internal line of a diagram could be regarded as a Feynman diagram with two external lines, both off the mass shell.

Of course, once we calculate the contribution of a given Feynman diagram off the mass shell, it is easy to calculate the associated S-matrix elements by going to the mass shell, taking the four-momentum p^μ flowing along the line *into* the diagram to have $p^0 = \sqrt{\mathbf{p}^2 + m^2}$ for particles in the initial state and $p^0 = -\sqrt{\mathbf{p}^2 + m^2}$ for particles in the final state, and including the appropriate external line factors $(2\pi)^{-3/2} u_\ell$ or $(2\pi)^{-3/2} v_\ell^*$ for initial particles or antiparticles and $(2\pi)^{-3/2} u_\ell^*$ or $(2\pi)^{-3/2} v_\ell$ for final particles or antiparticles. Indeed, when we come to the path integral approach in Chapter 9 we shall find it easiest first to derive the Feynman

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

rules for diagrams with *all* external lines off the mass shell, and then obtain S-matrix elements by letting the momenta associated with external lines approach their appropriate mass shells.

Feynman graphs with lines off the mass shell are just a special case of a wider generalization of the Feynman rules that takes into account the effects of various possible external fields. Suppose we add a sum of terms involving external fields $\epsilon_a(x)$ to the Hamiltonian, so that the interaction $V(t)$ that is used in the Dyson series (3.5.10) for the S-matrix is replaced with

$$V_\epsilon(t) = V(t) + \sum_a \int d^3x\, \epsilon_a(\mathbf{x}, t)\, o_a(\mathbf{x}, t)\,. \tag{6.4.1}$$

The 'currents' $o_a(t)$ have the usual time-dependence of the interaction picture

$$o_a(t) = \exp(iH_0 t) o_a(0) \exp(-iH_0 t)\,, \tag{6.4.2}$$

but are otherwise quite arbitrary operators. The S-matrix for any given transition $\alpha \to \beta$ then becomes a functional $S_{\beta\alpha}[\epsilon]$ of the c-number functions $\epsilon_a(t)$. The Feynman rules for computing this functional are given by an obvious extension of the usual Feynman rules. In addition to the usual vertices obtained from $V(t)$, we must include additional vertices: if $o_a(x)$ is a product of n_a field factors, then any o_a vertex with position label x must have n_a lines of corresponding types attached, and makes a contribution to the position-space Feynman rules equal to $-i\epsilon_a(x)$ times whatever numerical factors appear in $o_a(x)$. It follows then that the rth variational derivative of $S_{\beta\alpha}[\epsilon]$ with respect to $\epsilon_a(x), \epsilon_b(y)\cdots$ at $\epsilon = 0$ is given by position space diagrams with r additional vertices, to which are attached respectively $n_a, n_b \cdots$ internal lines, and no external lines. These vertices carry position labels $x, y \cdots$ over which we do *not* integrate; each such vertex makes a contribution equal to $-i$ times whatever numerical factors appear in the associated current o_a.

In particular, in the case where these currents are all single field factors, i.e.,

$$V_\epsilon(t) = V(t) + \sum_\ell \int d^3x\, \epsilon_\ell(\mathbf{x}, t)\, \psi_\ell(\mathbf{x}, t)\,,$$

the rth variational derivative of $S_{\beta\alpha}[\epsilon]$ with respect to $\epsilon_\ell(x), \epsilon_m(y)\cdots$ at $\epsilon = 0$ is given by position space diagrams with r additional vertices carrying spacetime labels $x, y \cdots$, to each of which is attached a single internal particle line of type $\ell, m \cdots$. These can be thought of as off-shell external lines, with the difference that their contribution to the matrix element is not a coefficient function like $(2\pi)^{-3/2} u_\ell(\mathbf{p}, \sigma) e^{ip\cdot x}$ or $(2\pi)^{-3/2} u^*_\ell(\mathbf{p}, \sigma) e^{-ip\cdot x}$ but a *propagator*, as well as a factor $-i$ from the vertex at the end of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

the line. We obtain a momentum space Feynman diagram with particles in states α and β on the mass shell plus r external lines of type $\ell, m \cdots$ carrying momenta $p, p' \cdots$ from the variational derivative

$$\left[\frac{\delta^r S_{\beta\alpha}[\epsilon]}{\delta\epsilon_\ell(x)\,\delta\epsilon_m(y)\cdots}\right]_{\epsilon=0}$$

by stripping away the propagators on each of the off-shell lines and then taking the appropriate Fourier transforms and multiplying with appropriate coefficient functions u_ℓ, u_ℓ^*, etc. and a factor $(-i)^r$.

It is very useful for a number of purposes to recognize that there is a simple relation between the sum of contributions from all perturbation theory diagrams for any off-shell amplitude and a matrix element, between eigenstates of the full Hamiltonian, of a time-ordered product of corresponding operators in the Heisenberg picture. This relation is provided by a theorem,[3] which states that to all orders of perturbation theory*

$$\left[\frac{\delta^r S_{\beta\alpha}[\epsilon]}{\delta\epsilon_a(x)\,\delta\epsilon_b(y)\cdots}\right]_{\epsilon=0} = \left(\Psi_\beta{}^-, T\left\{-iO_a(x), -iO_b(y)\cdots\right\}\Psi_\alpha{}^+\right), \tag{6.4.3}$$

where $O_a(x)$, etc. are the counterparts of $o_a(x)$ in the Heisenberg picture

$$O_a(\mathbf{x}, t) = \exp(iHt)\, o_a(\mathbf{x}, 0) \exp(-iHt) = \Omega(t)\, o_a(\mathbf{x}, t)\, \Omega^{-1}(t), \tag{6.4.4}$$

$$\Omega(t) \equiv e^{iHt}\, e^{-iH_0 t}, \tag{6.4.5}$$

and $\Psi_\beta{}^+$ and $\Psi_\beta{}^-$ are 'in' and 'out' eigenstates of the full Hamiltonian H, respectively.

Here is the proof. From Eq. (3.5.10), we see immediately that the left-hand side of Eq. (6.4.3) is

$$\left[\frac{\delta^r S[\epsilon]}{\delta\epsilon_{a_1}(x_1)\ldots\delta\epsilon_{a_r}(x_r)}\right]_{\epsilon=0} = \sum_{N=0}^{\infty} \frac{(-i)^{N+r}}{N!} \int_{-\infty}^{\infty} d\tau_1 \cdots d\tau_N$$
$$\times \left(\Phi_\beta,\ T\left\{V(\tau_1)\cdots V(\tau_N) o_{a_1}(x_1)\cdots o_{a_r}(x_r)\right\}\Phi_\alpha\right). \tag{6.4.6}$$

For definiteness, suppose that $x_1^0 \geq x_2^0 \geq \cdots \geq x_r^0$. Then we can denote by $\tau_{01}\cdots\tau_{0N_0}$ all τs greater than x_1^0; by $\tau_{11}\cdots\tau_{1N_1}$ all τs between x_1^0 and x_2^0, and so on; finally denoting by $\tau_{r1}\cdots\tau_{rN_r}$ all τs that are less than x_r^0.

* For a single O operator, this is a version of the Schwinger action principle.[4]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

Eq. (6.4.6) then becomes:

$$\left[\frac{\delta^r S[\epsilon]}{\delta\epsilon_{a_1}(x_1)\cdots\delta\epsilon_{a_r}(x_r)}\right]_{\epsilon=0} = \sum_{N=0}^{\infty}\frac{(-i)^{N+r}}{N!}\sum_{N_0 N_1\cdots N_r}\frac{N!\,\delta_{N,N_0+N_1+\cdots+N_r}}{N_0!N_1!\cdots N_r!}$$
$$\times\int_{x_1^0}^{\infty} d\tau_{01}\cdots d\tau_{0N_0}\int_{x_2^0}^{x_1^0} d\tau_{11}\cdots d\tau_{1N_1}\cdots\int_{-\infty}^{x_r^0} d\tau_{r1}\cdots d\tau_{rN_r}$$
$$\times\Big(\Phi_\beta,\ T\Big\{V(\tau_{01})\cdots V(\tau_{0N_0})\Big\}\,o_{a_1}(x_1)T\Big\{V(\tau_{11})\cdots V(\tau_{1N_1})\Big\}\,o_{a_2}(x_2)\cdots$$
$$\times\cdots o_{a_r}(x_r)\ T\Big\{V(\tau_{r1})\cdots V(\tau_{rN_r})\Big\}\Phi_\alpha\Big)\ .$$

The factor $N!/N_0!N_1!\cdots N_r!$ is the number of ways of sorting N τs into $r+1$ subsets, each containing $N_0,\ N_1,\cdots N_r$ of these τs. Instead of summing over $N_0, N_1,\cdots, N_r$, subject to the condition $N_0+N_1+\cdots+N_r = N$, and then summing over N, we can just sum independently over $N_0,\ N_1,\cdots, N_r$, setting N where it appears in $(-i)^N$ equal to $N_0+N_1+\cdots+N_r$. This gives

$$\left[\frac{\delta^r S[\epsilon]}{\delta\epsilon_{a_1}(x_1)\cdots\delta\epsilon_{a_r}(x_r)}\right]_{\epsilon=0} = (-i)^r\Big(\Phi_\beta, U(\infty,x_1^0)\,o_{a_1}(x_1)$$
$$\times\ U(x_1^0,x_2^0)\,o_{a_2}(x_2)\,U(x_2^0,x_3^0)\cdots o_{a_r}(x_r)\,U(x_r^0,-\infty)\Phi_\alpha\Big)\ , \qquad (6.4.7)$$

where

$$U(t',t) = \sum_{N=0}^{\infty}\frac{(-i)^N}{N!}\int_t^{t'} d\tau_1\cdots d\tau_N\ T\Big\{V(\tau_1)\cdots V(\tau_N)\Big\}\ . \qquad (6.4.8)$$

The operator $U(t',t)$ satisfies the differential equation

$$\frac{d}{dt'}U(t',t) = -iV(t')U(t',t) \qquad (6.4.9)$$

with the obvious initial condition

$$U(t,t) = 1\ . \qquad (6.4.10)$$

This has the solution

$$U(t',t) = \exp(iH_0t')\exp(-iH(t'-t))\exp(-iH_0t) = \Omega^{-1}(t')\Omega(t) \qquad (6.4.11)$$

with Ω given by Eq. (6.4.5). Inserting Eq. (6.4.11) in Eq. (6.4.7) and using

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

Eq. (6.4.4), we have

$$\left[\frac{\delta S[\epsilon]}{\delta\epsilon_{a_1}(x_1)\cdots\delta\epsilon_{a_r}(x_r)}\right]_{\epsilon=0}$$
$$= (-i)^r\Big(\Omega(\infty)\Phi_\beta,\ O_{a_1}(x_1)\cdots O_{a_r}(x_r)\Omega(-\infty)\Phi_\alpha\Big)\ . \qquad (6.4.12)$$

In deriving this result we supposed that $x_1^0 \geq x_2^0 \geq \cdots \geq x_r^0$, so we could just as well replace the product of operators on the right-hand-side with the time-ordered product:

$$\left[\frac{\delta S[\epsilon]}{\delta\epsilon_{a_1}(x_1)\cdots\delta\epsilon_{a_r}(x_r)}\right]_{\epsilon=0}$$
$$= (-i)^r\Big(\Omega(\infty)\Phi_\beta,\ T\Big\{O_{a_1}(x_1)\cdots O_{a_r}(x_r)\Big\}\Omega(-\infty)\Phi_\alpha\Big)\ . \qquad (6.4.13)$$

But now both sides are entirely symmetric (or for fermions antisymmetric) in the as and xs, so this relation holds whatever the order of the times $x_1^0 \ldots x_r^0$. Also, we saw in Section 3.1 that (in the sense of Eq. (3.1.12))

$$\Psi_\beta{}^{\pm} = \Omega(\mp\infty)\Phi_\beta\ . \qquad (6.4.14)$$

Hence Eq. (6.4.13) is the desired result (6.4.3).

Problems

1. Consider the theory of a real scalar field ϕ, with interaction (in the interaction picture) $V = g\int d^3x\,\phi(x)^3/3!$. Calculate the connected S-matrix element for scalar–scalar scattering to second order in g, doing all integrals. Use the results to calculate the differential cross-section for scalar–scalar scattering in the center-of-mass system.

2. Consider a theory involving a neutral scalar field $\phi(x)$ for a boson B and a complex Dirac field $\psi(x)$ for a fermion F, with interaction (in the interaction picture) $V = ig\int d^3x\,\bar\psi(x)\gamma_5\psi(x)\phi(x)$. Draw all the connected order-g^2 Feynman diagrams and calculate the corresponding S-matrix elements for the processes $F^c + B \to F^c + B$, $F + F^c \to F + F^c$, and $F^c + F \to B + B$ (where F^c is the antiparticle of F). Do all integrals.

3. Consider the theory of a real scalar field $\phi(x)$, with interaction $V = g\int d^3x\,\phi(x)^4/4!$. Calculate the S-matrix for scalar–scalar scattering to order g, and use the result to calculate the differential scattering cross-section. Calculate the correction terms in the S-matrix for scalar–scalar scattering to order g^2, expressing the result as an integral over a single four-momentum, but do all x-integrals.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

4. What is the contribution in Feynman diagrams from the contraction of the derivative $\partial_\mu \psi_\ell(x)$ of a Dirac field with the adjoint $\psi_m^\dagger(y)$ of the field?

5. Use the theorem of Section 6.4 to give expressions for the vacuum expectation values of Heisenberg picture operators $(\Psi_0, \Phi(x)\Psi_0)$ and $(\Psi_0, T\{\Phi(x), \Phi(y)\}\Psi_0)$ in the theory of Problem 1, to orders g and g^2, respectively.

References

1. F. J. Dyson, *Phys. Rev.* **75**, 486, 1736 (1949).

2. The formal statement of this result is known as *Wick's theorem*; see G. C. Wick, *Phys. Rev.* **80**, 268 (1950).

3. I do not know who first proved this theorem. It was known in the early 1950s to several theorists, including M. Gell-Mann and F. E. Low.

4. J. Schwinger, *Phys. Rev.* **82**, 914 (1951).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.008

7

The Canonical Formalism

Ever since the birth of quantum field theory in the papers of Born, Dirac, Fermi, Heisenberg, Jordan, and Pauli in the late 1920s, its development has been historically linked to the canonical formalism, so much so that it seems natural to begin any treatment of the subject today by postulating a Lagrangian and applying to it the rules of canonical quantization. This is the approach used in most books on quantum field theory. Yet historical precedent is not a very convincing reason for using this formalism. If we discovered a quantum field theory that led to a physically satisfactory S-matrix, would it bother us if it could not be derived by the canonical quantization of some Lagrangian?

To some extent this question is moot because, as we shall see in Section 7.1, all of the most familiar quantum field theories furnish canonical systems, and these can easily be put in a Lagrangian form. However, there is no proof that every conceivable quantum field theory can be formulated in this way. And even if it can, this does not in itself explain why we should *prefer* to use the Lagrangian formalism as a starting point in constructing various quantum field theories.

The point of the Lagrangian formalism is that it makes it easy to satisfy Lorentz invariance and other symmetries: a classical theory with a Lorentz-invariant Lagrangian density will when canonically quantized lead to a Lorentz-invariant quantum theory. That is, we shall see here that such a theory allows the construction of suitable quantum mechanical operators that satisfy the commutation relations of the Poincaré algebra, and therefore leads to a Lorentz-invariant S-matrix.

This is not so trivial. We saw in the previous chapter that in theories with derivative couplings or spins $j \geq 1$, it is not enough to take the interaction Hamiltonian as the integral over space of a scalar interaction density; we also need to add non-scalar terms to the interaction density to compensate for non-covariant terms in the propagators. The canonical formalism with a scalar Lagrangian density will automatically provide these extra terms. Later, when we come to non-Abelian gauge theories in Volume II, this extra convenience will become a necessity; it would be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

just about hopeless to try to guess at the form of the Hamiltonian in such theories without starting with a Lorentz-invariant and gauge-invariant Lagrangian density.

7.1 Canonical Variables

In this section we shall show that various quantum field theories that we have constructed so far satisfy the commutation rules and equations of motion of the Hamiltonian version of the canonical formalism. It is the Hamiltonian formalism that is needed to calculate the S-matrix (whether by operator or path-integral methods) but it is not always easy to choose Hamiltonians that yield a Lorentz-invariant S-matrix. In the balance of this chapter we shall take the Lagrangian version of the canonical formalism as our starting point, and use it to derive physically satisfactory Hamiltonians. The purpose of the present section is to identify the canonical fields and their conjugates in various field theories, to tell us how to separate the free-field terms in the Lagrangian, and incidentally to reassure us that the canonical formalism is indeed applicable to physically realistic theories.

We first show that the free fields constructed in Chapter 5 automatically provide a system of quantum operators $q^n(\mathbf{x},t)$ and canonical conjugates $p_n(\mathbf{x},t)$ that satisfy the familiar canonical commutation or anticommutation relations:

$$[q^n(\mathbf{x},t), p_{\bar{n}}(\mathbf{y},t)]_{\mp} = i\delta^3(\mathbf{x}-\mathbf{y})\delta^n_{\bar{n}}\,, \tag{7.1.1}$$

$$[q^n(\mathbf{x},t), q^{\bar{n}}(\mathbf{y},t)]_{\mp} = 0\,, \tag{7.1.2}$$

$$[p_n(\mathbf{x},t), p_{\bar{n}}(\mathbf{y},t)]_{\mp} = 0\,, \tag{7.1.3}$$

where the subscripts $\mp$ indicate that these are commutators if either of the particles created and destroyed by the two operators are bosons, and anticommutators if both particles are fermions. For instance, the real scalar field $\phi(\mathbf{x})$ for a self-charge-conjugate particle of zero spin was found in Section 5.2 to obey the commutation relation

$$[\phi(x), \phi(y)]_- = \Delta(x-y)\,,$$

where Δ is the function

$$\Delta(x) \equiv \int \frac{d^3k}{2k^0(2\pi)^3}[e^{ik\cdot x} - e^{-ik\cdot x}]$$

with $k^0 \equiv \sqrt{\mathbf{k}^2+m^2}$. We note that

$$\Delta(\mathbf{x},0) = 0\,, \qquad \dot{\Delta}(\mathbf{x},0) = -i\delta^3(\mathbf{x})\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

(A dot denotes the derivative with respect to the time x^0.) It is easy then to see that the field and its time-derivative $\dot{\phi}$ obey the equal-time commutation relations:

$$[\phi(\mathbf{x},t),\dot{\phi}(\mathbf{y},t)]_- = i\delta^3(\mathbf{x}-\mathbf{y})\,, \tag{7.1.4}$$

$$[\phi(\mathbf{x},t),\phi(\mathbf{y},t)]_- = 0\,, \tag{7.1.5}$$

$$[\dot{\phi}(\mathbf{x},t),\dot{\phi}(\mathbf{y},t)]_- = 0\,. \tag{7.1.6}$$

Therefore we may define canonical variables

$$q(\mathbf{x},t) \equiv \phi(\mathbf{x},t)\,, \qquad p(\mathbf{x},t) \equiv \dot{\phi}(\mathbf{x},t) \tag{7.1.7}$$

which satisfy the canonical commutation relations (7.1.1)–(7.1.3).

For the complex scalar field of a particle of spin zero with a distinct antiparticle, the commutation relations are

$$[\phi(x),\phi^\dagger(y)]_- = \Delta(x-y)\,, \qquad [\phi(x),\phi(y)]_- = 0\,.$$

We may therefore define the free-particle canonical variables as the complex operators

$$q(\mathbf{x},t) \equiv \phi(\mathbf{x},t)\,, \tag{7.1.8}$$

$$p(\mathbf{x},t) \equiv \dot{\phi}^\dagger(\mathbf{x},t)\,. \tag{7.1.9}$$

Equivalently, writing $\phi \equiv (\phi_1 + i\phi_2)/\sqrt{2}$ with ϕ_k Hermitian for $k = 1,2$, we have canonical variables

$$q^k(\mathbf{x},t) = \phi_k(\mathbf{x},t)\,, \tag{7.1.10}$$

$$p_k(\mathbf{x},t) = \dot{\phi}_k(\mathbf{x},t)\,, \tag{7.1.11}$$

and these satisfy the commutation relations (7.1.1)–(7.1.3).

For the real vector field of a particle of spin one, the commutation relations are given by Section 5.3 as

$$[v^\mu(x),v^\nu(y)]_- = \left[\eta^{\mu\nu} - \frac{\partial^\mu\partial^\nu}{m^2}\right]\Delta(x-y)\,.$$

(We are using v^μ rather than V^μ for the vector field because we want to reserve upper case letters here for the fields in the Heisenberg picture.) Here the free-particle canonical variables may be taken as

$$q^i(\mathbf{x},t) = v^i(\mathbf{x},t)\,, \tag{7.1.12}$$

$$p_i(\mathbf{x},t) = \dot{v}^i(\mathbf{x},t) + \frac{\partial v^0(\mathbf{x},t)}{\partial x^i}\,, \tag{7.1.13}$$

with $i = 1,2,3$. The reader may check that (7.1.12) and (7.1.13) satisfy the commutation relations (7.1.1)–(7.1.3). The field equations (5.3.36) and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

(5.3.38) together with Eq. (7.1.13) allow us to express v^0 in terms of the other variables as

$$v^0 = m^{-2}\nabla \cdot \mathbf{p}\,, \tag{7.1.14}$$

so v^0 is not regarded as one of the qs. The extension of these results to complex vector fields may be handled just as for complex scalar fields.

For the Dirac field of a non-Majorana spin $\frac{1}{2}$ particle, Section 5.5 shows that the anticommutator is

$$[\psi_n(x), \psi^\dagger_{\bar{n}}(y)]_+ = \left[(-\gamma^\mu \partial_\mu + m)\beta\right]_{n,\bar{n}} \Delta(x-y)$$

and

$$[\psi_n(x), \psi_{\bar{n}}(y)]_+ = 0\,.$$

Here it would be inconsistent to take ψ_n and $\psi^\dagger_n$ to be independent canonical variables, because their anticommutator does not vanish at equal times. It is conventional instead to define

$$q^n(x) \equiv \psi_n(x)\,, \tag{7.1.15}$$

$$p_n(x) \equiv i\psi^\dagger_n(x)\,. \tag{7.1.16}$$

It is easy then to see that (7.1.15) and (7.1.16) satisfy the canonical anticommutation relations (7.1.1)–(7.1.3).

For any system of operators that satisfy commutation or anticommutation relations like (7.1.1)–(7.1.3), we may define a quantum mechanical functional derivative: for an arbitrary bosonic functional $F[q(t), p(t)]$ of $q^n(\mathbf{x}, t)$ and $p_n(\mathbf{x}, t)$ at a fixed time t, we define*

$$\frac{\delta F[q(t), p(t)]}{\delta q^n(\mathbf{x}, t)} \equiv i\Big[p_n(\mathbf{x}, t)\,,\; F[q(t), p(t)]\Big]\,, \tag{7.1.17}$$

$$\frac{\delta F[q(t), p(t)]}{\delta p_n(\mathbf{x}, t)} \equiv i\Big[F[q(t), p(t)]\,,\; q^n(\mathbf{x}, t)\Big]\,. \tag{7.1.18}$$

This definition is motivated by the fact that if $F[q(t), p(t)]$ is written with all qs to the left of all ps, then (7.1.17) and (7.1.18) are respectively just the left- and right-derivatives with respect to q^n and p_n. That is, for an

* We are here using a notation that will be adopted henceforth; if $f(x, y)$ is a function of two classes of variables collectively called x and y, then $F[f(y)]$ indicates a functional that depends on the values of $f(x, y)$ for all x at fixed y. By a bosonic functional we mean one in which each term contains only even numbers of fermionic fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

arbitrary c-number** variation δq and δp of the qs and ps, we have

$$\delta F[q(t),p(t)] = \int d^3x \sum_n \left(\delta q^n(\mathbf{x},t) \frac{\delta F[q(t),p(t)]}{\delta q^n(\mathbf{x},t)} + \frac{\delta F[q(t),p(t)]}{\delta p_n(\mathbf{x},t)} \delta p_n(\mathbf{x},t) \right).$$

For more general functionals we need the definitions (7.1.17) and (7.1.18) to pin down various signs and equal-time commutators that may appear.

In particular, H_0 is the generator of time-translations on free-particle states in the sense that:

$$q^n(\mathbf{x},t) = \exp(iH_0t) q^n(\mathbf{x},0) \exp(-iH_0t) , \tag{7.1.19}$$

$$p_n(\mathbf{x},t) = \exp(iH_0t) p_n(\mathbf{x},0) \exp(-iH_0t) , \tag{7.1.20}$$

so the free-particle operators have the time-dependence

$$\dot{q}^n(\mathbf{x},t) = i[H_0, q^n(\mathbf{x},t)] = \frac{\delta H_0}{\delta p_n(\mathbf{x},t)} , \tag{7.1.21}$$

$$\dot{p}_n(\mathbf{x},t) = -i[p_n(\mathbf{x},t), H_0] = -\frac{\delta H_0}{\delta q^n(\mathbf{x},t)} . \tag{7.1.22}$$

We recognize these as the familiar dynamical equations in the Hamiltonian formalism.

The free-particle Hamiltonian is given as always by

$$H_0 = \sum_{n,\sigma} \int d^3k \; a^\dagger(\mathbf{k},\sigma,n) \, a(\mathbf{k},\sigma,n) \sqrt{\mathbf{k}^2 + m_n^2} . \tag{7.1.23}$$

This H_0 may be rewritten in terms of the qs and ps at time t. For instance, it is easy to see that for a real scalar field, Eq. (7.1.23) is equal up to a constant term to the functional

$$H_0 = \int d^3x \left[\tfrac{1}{2}p^2 + \tfrac{1}{2}(\nabla q)^2 + \tfrac{1}{2}m^2 q^2 \right] . \tag{7.1.24}$$

To be more precise, using (7.1.7) and the Fourier representation of the scalar field ϕ, we find that Eq. (7.1.24) becomes:

$$\begin{aligned} H_0 &= \tfrac{1}{2} \int d^3k \, k^0 \left[a(\mathbf{k}), a^\dagger(\mathbf{k}) \right]_+ \\ &= \int d^3k \, k^0 \left(a^\dagger(\mathbf{k}) a(\mathbf{k}) + \tfrac{1}{2}\delta^3(\mathbf{k}-\mathbf{k}) \right) . \end{aligned} \tag{7.1.25}$$

** Where q^n and p_n are bosonic or fermionic, δq^n and δp_n are understood to commute or anticommute with all fermionic operators, respectively, and to commute with all bosonic operators.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

This is the same as Eq. (7.1.23), except for the infinite constant term. Such terms only affect the zero of energy, and have no physical significance in the absence of gravity.† Explicit forms for H_0 as a functional of the q and p variables for other fields will be given in Section 7.5.

It is usual in textbooks on quantum field theory to derive Eq. (7.1.25) as a consequence of Eq. (7.1.24), which in turn is derived from a Lagrangian density. This seems to me backward, for Eq. (7.1.25) *must* hold; if some assumed free-particle Lagrangian did not give Eq. (7.1.25) up to a constant term, we would conclude that it was the wrong Lagrangian. Rather, we should ask what free-field Lagrangian gives Eq. (7.1.25) for spinless particles, or more generally, gives the free-particle Hamiltonian (7.1.23). This question may be answered by the well-known Legendre transformation from the Hamiltonian to the Lagrangian; the free-field Lagrangian is given by

$$L_0[q(t),\dot{q}(t)] = \sum_n \int d^3x\; p_n(\mathbf{x},t)\,\dot{q}^n(\mathbf{x},t) \;-\; H_0\,, \tag{7.1.26}$$

it being understood that p_n is replaced everywhere by its expression in terms of q^n and $\dot{q}^n$ (and, as we shall see, perhaps some auxiliary fields as well). For instance, from the Hamiltonian (7.1.24) and (7.1.7) we can derive the free-field Lagrangian for a scalar field:

$$\begin{aligned} L_0 &= \int d^3x \left[p\dot{q} - \tfrac{1}{2}p^2 - \tfrac{1}{2}(\nabla q)^2 - \tfrac{1}{2}m^2q^2\right] \\ &= \int d^3x \left[-\tfrac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2\right]. \end{aligned} \tag{7.1.27}$$

Whatever we suppose the complete Lagrangian of the scalar field may be, this is the term that must be separated out and treated as a term of zeroth order in perturbation theory. A similar exercise may be carried out for the other canonical systems described in this section, but from now on we shall content ourselves with guessing the form of the free-field Lagrangian and then confirming that it gives the correct free-particle Hamiltonian.

We have seen that various free-field theories can be formulated in canonical terms. It is then a short step to show that the same is true of the interacting fields. We can introduce canonical variables in what is called the 'Heisenberg picture', defined by

$$Q^n(\mathbf{x},t) \equiv \exp(iHt)q^n(\mathbf{x},0)\exp(-iHt), \tag{7.1.28}$$

$$P_n(\mathbf{x},t) \equiv \exp(iHt)p_n(\mathbf{x},0)\exp(-iHt), \tag{7.1.29}$$

† However, *changes* in such terms due to changes in the boundary conditions for the fields, as for instance quantizing in the space between parallel plates rather than in infinite space, are physically significant, and have even been measured.[1]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

where H is the full Hamiltonian. Because this is a similarity transformation that commutes with H, the total Hamiltonian is the same functional of the Heisenberg picture operators as it was of the qs and ps:

$$H[Q,P] = e^{iHt}H[q,p]e^{-iHt} = H[q,p]\,.$$

Also, because Eqs. (7.1.28)–(7.1.29) define a similarity transformation, the Heisenberg picture operators again satisfy the canonical commutation or anticommutation relations:

$$[Q^n(\mathbf{x},t), P_{\bar{n}}(\mathbf{y},t)]_{\mp} = i\delta^3(\mathbf{x}-\mathbf{y})\delta^n_{\bar{n}}\,, \tag{7.1.30}$$

$$[Q^n(\mathbf{x},t), Q^{\bar{n}}(\mathbf{y},t)]_{\mp} = 0\,, \tag{7.1.31}$$

$$[P_n(\mathbf{x},t), P_{\bar{n}}(\mathbf{y},t)]_{\mp} = 0\,. \tag{7.1.32}$$

However, they now have the time-dependence

$$\dot{Q}^n(\mathbf{x},t) = i[H, Q^n(\mathbf{x},t)] = \frac{\delta H}{\delta P_n(\mathbf{x},t)}\,, \tag{7.1.33}$$

$$\dot{P}_n(\mathbf{x},t) = -i[P_n(\mathbf{x},t), H] = -\frac{\delta H}{\delta Q^n(\mathbf{x},t)}\,. \tag{7.1.34}$$

For instance, we might take the Hamiltonian for a real scalar field as the free-particle term (7.1.24) plus the integral of a scalar interaction density $\mathscr{H}$, so that in terms of Heisenberg-picture variables

$$H = \int d^3x \left[\tfrac{1}{2}P^2 + \tfrac{1}{2}(\nabla Q)^2 + \tfrac{1}{2}m^2 Q^2 + \mathscr{H}(Q)\right]\,. \tag{7.1.35}$$

In this case the canonical conjugate to Q is given by the same formula as for free fields:

$$P = \dot{Q}\,. \tag{7.1.36}$$

However, as we shall see, the relation between the canonical conjugates $P_n(x)$ and the field variables and their time-derivatives is in general not the same as for the free-particle operators, but must be inferred from Eqs. (7.1.33) and (7.1.34).

7.2 The Lagrangian Formalism

Having seen that various realistic theories may be cast in the canonical formalism, we must now face the question of how to choose the Hamiltonian. As we will see in the next section, the easiest way to enforce Lorentz invariance and other symmetries is to choose a suitable Lagrangian and use it to derive the Hamiltonian. There is not much loss of generality in this;

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

given a realistic Hamiltonian, we can generally reconstruct a Lagrangian from which it could be derived, by reversing the process that we are going to describe here of deriving Hamiltonians from Lagrangians. (The derivation of Eq. (7.1.26) gives one example of this reconstruction.) But although we can go from Hamiltonians to Lagrangians or Lagrangians to Hamiltonians, it is easier to explore physically satisfactory theories by listing possible Lagrangians, rather than Hamiltonians.

The Lagrangian is, in general, a functional* $L[\Psi(t), \dot{\Psi}(t)]$ of a set of generic fields $\Psi^\ell(\mathbf{x}, t)$ and their time-derivatives $\dot{\Psi}^\ell(\mathbf{x}, t)$. The conjugate fields $\Pi_\ell(\mathbf{x}, t)$ are defined as the variational derivatives**

$$\Pi_\ell(\mathbf{x}, t) \equiv \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \dot{\Psi}^\ell(\mathbf{x}, t)} , \tag{7.2.1}$$

The equations of motion are

$$\dot{\Pi}_\ell(\mathbf{x}, t) = \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \Psi^\ell(\mathbf{x}, t)} . \tag{7.2.2}$$

These field equations can be usefully reformulated as a variational principle. We define a functional of $\Psi^\ell(x)$ over all spacetime, known as the *action*

$$I[\Psi] \equiv \int_{-\infty}^{\infty} dt\, L[\Psi(t), \dot{\Psi}(t)] . \tag{7.2.3}$$

Under an arbitrary variation of $\Psi(x)$, the change in $I[\Psi]$ is

$$\delta I[\Psi] = \int_{-\infty}^{\infty} dt \int d^3x \left[\frac{\delta L}{\delta \Psi^\ell(x)} \delta \Psi^\ell(x) + \frac{\delta L}{\delta \dot{\Psi}^\ell(x)} \delta \dot{\Psi}^\ell(x) \right] .$$

Assuming that $\delta \Psi^\ell(x)$ vanishes for $t \to \pm\infty$, we may integrate by parts, and write

$$\delta I[\Psi] = \int d^4x \left[\frac{\delta L}{\delta \Psi^\ell(x)} - \frac{d}{dt} \frac{\delta L}{\delta \dot{\Psi}^\ell(x)} \right] \delta \Psi^\ell(x) . \tag{7.2.4}$$

We see that the action is stationary with respect to all variations $\delta \Psi^\ell$ that vanish at $t \to \pm\infty$ if and only if the fields satisfy the field equations (7.2.2).

---

* Recall that in the notation we use for functionals, a functional like L in which we display the variable t is understood to depend on the fields $\Psi^\ell(\mathbf{x}, t)$ and $\dot{\Psi}^\ell(\mathbf{x}, t)$, with the undisplayed variables ℓ and $\mathbf{x}$ running over all their values at a fixed value of the displayed variable t. We use upper case Ψs and Πs to indicate that these are interacting rather than free fields.

** Because the Ψs and $\dot{\Psi}$s do not in general satisfy simple commutation or anticommutation relations, we cannot give a simple definition of the functional derivatives occuring here as we did for functional derivatives with respect to the Qs and Ps in the previous section. Instead, we will simply specify that the variational derivatives are what they would be for c-number variables, with minus signs and equal-time commutators or anticommutators supplied as needed to make the formulas correct quantum-mechanically. As far as I know, no important issues hinge on the details here.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Because the field equations are determined by the functional $I[\Psi]$, it is natural in trying to construct a Lorentz-invariant theory to make $I[\Psi]$ a scalar functional. In particular, since $I[\Psi]$ is a time-integral of $L[\Psi(t), \dot{\Psi}(t)]$, we guess that L should itself be a space-integral of an ordinary scalar function of $\Psi(x)$ and $\partial\Psi(x)/\partial x^\mu$, known as the *Lagrangian density* $\mathscr{L}$:

$$L[\Psi(t), \dot{\Psi}(t)] = \int d^3x \, \mathscr{L}\left(\Psi(\mathbf{x}, t), \nabla\Psi(\mathbf{x}, t), \dot{\Psi}(\mathbf{x}, t)\right) , \tag{7.2.5}$$

so that the action is

$$I[\Psi] = \int d^4x \, \mathscr{L}\Big(\Psi(x), \partial\Psi(x)/\partial x^\mu\Big) . \tag{7.2.6}$$

All field theories used in current theories of elementary particles have Lagrangians of this form.

Varying $\Psi^\ell(x)$ by an amount $\delta\Psi^\ell(x)$, and integrating by parts, we find a variation in L:

$$\begin{aligned}\delta L &= \int d^3x \left[\frac{\partial\mathscr{L}}{\partial\Psi^\ell}\delta\Psi^\ell + \frac{\partial\mathscr{L}}{\partial(\nabla\Psi^\ell)}\nabla\delta\Psi^\ell + \frac{\partial\mathscr{L}}{\partial\dot{\Psi}^\ell}\delta\dot{\Psi}^\ell\right] \\ &= \int d^3x \left[\left(\frac{\partial\mathscr{L}}{\partial\Psi^\ell} - \nabla\cdot\frac{\partial\mathscr{L}}{\partial(\nabla\Psi^\ell)}\right)\delta\Psi^\ell + \frac{\partial\mathscr{L}}{\partial\dot{\Psi}^\ell}\delta\dot{\Psi}^\ell\right] ,\end{aligned}$$

so (with obvious arguments suppressed)

$$\frac{\delta L}{\delta\Psi^\ell} = \frac{\partial\mathscr{L}}{\partial\Psi^\ell} - \nabla\cdot\frac{\partial\mathscr{L}}{\partial(\nabla\Psi^\ell)} , \tag{7.2.7}$$

$$\frac{\delta L}{\delta\dot{\Psi}^\ell} = \frac{\partial\mathscr{L}}{\partial\dot{\Psi}^\ell} . \tag{7.2.8}$$

The field equations (7.2.2) then read

$$\frac{\partial}{\partial x^\mu}\frac{\partial\mathscr{L}}{\partial(\partial\Psi^\ell/\partial x^\mu)} = \frac{\partial\mathscr{L}}{\partial\Psi^\ell} . \tag{7.2.9}$$

These are known as the *Euler–Lagrange equations*. As expected, if $\mathscr{L}$ is a scalar then these equations are Lorentz-invariant.

In addition to being Lorentz-invariant, the action I is required to be *real*. This is because we want just as many field equations as there are fields. By breaking up any complex fields into their real and imaginary parts, we can always think of I as being a functional only of a number of *real* fields, say N of them. If I were complex, with independent real and imaginary parts, then the real and imaginary parts of the conditions that I be stationary (the Euler–Lagrange equations) would yield $2N$ field equations for N fields, too many to be satisfied except in special cases. We will see in the next section that the reality of the action also ensures

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

that the generators of various symmetry transformations are Hermitian operators.

Although the Lagrangian formalism makes it easy to construct theories that will satisfy Lorentz invariance and other symmetries, to calculate the S-matrix we need a formula for the interaction Hamiltonian. In general, the Hamiltonian is given by the *Legendre transformation*

$$H = \sum_\ell \int d^3x\, \Pi_\ell(\mathbf{x},t)\dot{\Psi}^\ell(\mathbf{x},t) - L[\Psi(t),\dot{\Psi}(t)]\,. \tag{7.2.10}$$

Although Eq. (7.2.1) does not in general allow $\dot{\Psi}^\ell$ to be expressed uniquely in terms of Ψ^ℓ and Π_ℓ, it is easy to see that Eq. (7.2.10) has vanishing variational derivative with respect to $\dot{\Psi}^\ell$ for any $\dot{\Psi}^\ell$ satisfying Eq. (7.2.1), so in general it is a functional only of Ψ^ℓ and Π_ℓ. Its variational derivatives with respect to these variables are

$$\left.\frac{\delta H}{\delta\Psi^\ell(\mathbf{x},t)}\right|_\Pi = \int d^3y \sum_{\ell'} \Pi_{\ell'}(\mathbf{y},t) \left.\frac{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}{\delta\Psi^\ell(\mathbf{x},t)}\right|_\Pi - \left.\frac{\delta L}{\delta\Psi^\ell(\mathbf{x},t)}\right|_{\dot{\Psi}}$$
$$- \int d^3y \sum_{\ell'} \left.\frac{\delta L}{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}\right|_\Psi \left.\frac{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}{\delta\Psi^\ell(\mathbf{x},t)}\right|_\Pi ,$$

$$\left.\frac{\delta H}{\delta\Pi_\ell(\mathbf{x},t)}\right|_\Psi = \dot{\Psi}^\ell(\mathbf{x},t) + \int d^3y \sum_{\ell'} \Pi_{\ell'}(\mathbf{y},t) \left.\frac{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}{\delta\Pi^\ell(\mathbf{x},t)}\right|_\Psi$$
$$- \int d^3y \sum_{\ell'} \left.\frac{\delta L}{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}\right|_\Psi \left.\frac{\delta\dot{\Psi}^{\ell'}(\mathbf{y},t)}{\delta\Pi^\ell(\mathbf{x},t)}\right|_\Psi ,$$

where subscripts denote the quantities held fixed in these variational derivatives. Using the defining equation (7.2.1) for Π_ℓ, this simplifies to

$$\left.\frac{\delta H}{\delta\Psi^\ell(\mathbf{x},t)}\right|_\Pi = -\left.\frac{\delta L}{\delta\Psi^\ell(\mathbf{x},t)}\right|_{\dot{\Psi}} , \tag{7.2.11}$$

and

$$\left.\frac{\delta H}{\delta\Pi_\ell(\mathbf{x},t)}\right|_\Psi = \dot{\Psi}^\ell(\mathbf{x},t)\,. \tag{7.2.12}$$

The equations of motion (7.2.2) are then equivalent to

$$\left.\frac{\delta H}{\delta\Psi^\ell(\mathbf{x},t)}\right|_\Pi = -\dot{\Pi}_\ell(\mathbf{x},t)\,. \tag{7.2.13}$$

It is tempting now to identify the generic field variables Ψ^ℓ and their conjugates Π_ℓ with the canonical variables Q^n and P_n of the previous section, and impose on them the same canonical commutation relations (7.1.30)–(7.1.32), so that Eqs. (7.2.12) and (7.2.13) are the same as the Hamiltonian equations of motion (7.1.33) and (7.1.34). This is indeed the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

case for the simple example of the real scalar field Φ with non-derivative coupling. Consider the Lagrangian density[†]

$$\mathscr{L} = -\frac{1}{2}\,\partial_\mu\Phi\partial^\mu\Phi - \frac{m^2}{2}\Phi^2 - \mathscr{H}(\Phi)\,, \tag{7.2.14}$$

which can be obtained by adding a real function $-\mathscr{H}(\Phi)$ of Φ to the free-field Lagrangian density found in the previous section. The Euler–Lagrange equations here are

$$(\Box - m^2)\Phi = \mathscr{H}'(\Phi)\,. \tag{7.2.15}$$

From this Lagrangian density, we calculate a canonical conjugate to Φ:

$$\Pi = \frac{\partial\mathscr{L}}{\partial\dot{\Phi}} = \dot{\Phi}\,, \tag{7.2.16}$$

which is the same as Eq. (7.1.36) if we identify Φ and Π with the canonical variables Q and P. The Hamiltonian is now given by Eq. (7.2.10) as[‡]

$$\begin{aligned} H &= \int d^3x\,(\Pi\dot{\Phi} - \mathscr{L}) \\ &= \int d^3x\left[\tfrac{1}{2}\Pi^2 + \tfrac{1}{2}(\nabla\Phi)^2 + \tfrac{1}{2}m^2\Phi^2 + \mathscr{H}(\Phi)\right], \end{aligned} \tag{7.2.17}$$

which we recognize as the Hamiltonian (7.1.35). This little exercise should not be regarded as another derivation of this Hamiltonian, but rather as a validation of the Lagrangian (7.2.14) as a possible theory of scalar fields.

Matters are not always so simple. We have already seen in the previous section that there are field variables, such as the time component of a vector field or the Hermitian conjugate of a Dirac field, that are not canonical field variables Q^n and do not have canonical conjugates; yet Lorentz invariance dictates that these must appear in the Lagrangians for the vector and Dirac fields.

From the point of view of the Lagrangian formalism, the special character of field variables like the time component of a vector field or the Hermitian conjugate of a Dirac field arises from the fact that although they appear in the Lagrangian, their time-derivatives do not. We shall denote the field variables $\Psi^{\prime}$ whose time-derivatives do not appear in the Lagrangian as C^r; the remaining independent field variables are the

[†] We do not include a free constant factor in the term $-\frac{1}{2}\,\partial_\mu\Phi\partial^\mu\Phi$, because any such constant if positive can be absorbed into the normalization of Φ. As we shall see, a negative constant here would lead to a Hamiltonian that is not bounded below. The constant m is known as the bare mass. The most general Lagrangian that satisfies the principle of renormalizability (discussed in Chapter 12) is of this form, with $\mathscr{H}(\Phi)$ a quartic polynomial in Φ.

[‡] In order for H to be interpreted as an energy, it should be bounded below. The positivity of the first two terms shows that we guessed correctly as to the sign in the first term in Eq. (7.2.14). The remaining condition is that $\frac{1}{2}m^2\Phi^2 + \mathscr{H}(\Phi)$ must be bounded below as a function of Φ.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

canonical variables Q^n. The Q^n have canonical conjugates

$$P_n(\mathbf{x},t) = \frac{\delta L[Q(t),\dot{Q}(t),C(t)]}{\delta \dot{Q}^n(\mathbf{x},t)}\,, \tag{7.2.18}$$

and satisfy the commutation relations (7.1.30)–(7.1.32), but there are no canonical conjugates for the C^r. Because $\delta L/\delta \dot{C}^r = 0$, the Hamiltonian (7.2.10) is in general

$$H = \sum_n \int d^3x\, P_n \dot{Q}^n - L[Q(t),\dot{Q}(t),C(t)]\,, \tag{7.2.19}$$

but this is not yet useful until we express the C^r and $\dot{Q}^\ell$ in terms of the Qs and Ps. The equations of motion of the C^r involve only fields and their first time-derivatives

$$0 = \frac{\delta L[Q(t),\dot{Q}(t),C(t)]}{\delta C^r(\mathbf{x},t)}\,. \tag{7.2.20}$$

In the simple cases to be discussed in this chapter, these equations together with Eq. (7.2.18) can be solved to give the C^r and $\dot{Q}^\ell$ in terms of the Qs and Ps. Section 7.6 shows how in such cases one can avoid the task of actually solving for the C^r and $\dot{Q}^\ell$. In gauge theories like electrodynamics other methods must be used: either choosing a particular gauge, as in Chapter 8, or the more modern covariant methods to be discussed in Volume II.

Once we have derived a Hamiltonian as a functional of the Heisenberg picture Qs and Ps, to use perturbation theory we must make a transition to the interaction picture. The Hamiltonian is time-independent, so it can be written in terms of the P_n and Q^n at $t = 0$, which are equal to the corresponding operators p_n and q^n in the interaction picture at $t = 0$. The Hamiltonian derived in this way may then be expressed in terms of the qs and ps of the interaction picture, and split into two parts, a suitable free-particle term H_0 and an interaction V. Finally, the time-dependence equations (7.1.21) and (7.1.22) and the commutation or anticommutation relations (7.1.1)–(7.1.3) are used to express the qs and ps in $V(t)$ as linear combinations of annihilation and creation operators.

We shall present a number of examples of this procedure in Section 7.5; for the moment we will give only one example of the simplest type, the scalar field with Hamiltonian (7.2.17). We split H into a free-particle term and an interaction

$$H = H_0 + V \tag{7.2.21}$$

$$H_0 = \int d^3x \left[\tfrac{1}{2}\Pi^2 + \tfrac{1}{2}(\nabla\Phi)^2 + \tfrac{1}{2}m^2\Phi^2 \right] \tag{7.2.22}$$

$$V = \int d^3x\, \mathscr{H}(\Phi)\,. \tag{7.2.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Here Φ and Π are taken at the same time t, and H is independent of t, though H_0 and V usually are not.

We now pass to the interaction representation. Taking $t = 0$ in Eqs. (7.2.22) and (7.2.23), we can simply replace Φ, Π with the interaction picture variables ϕ, π, since they are defined by Eqs. (7.1.28) and (7.1.29) to be equal at that time. To calculate the interaction $V(t)$ in the interaction picture, we apply the similarity transformation (3.5.5)

$$\begin{aligned} V(t) &= \exp(iH_0 t)\, V\, \exp(-iH_0 t) \\ &= \int d^3x\, \mathscr{H}\Big(\phi(\mathbf{x}, t)\Big) . \end{aligned} \tag{7.2.24}$$

The same transformation applied to H_0 leaves it constant:

$$\begin{aligned} H_0 &= \exp(iH_0 t)\, H_0\, \exp(-iH_0(t)) \\ &= \int d^3x \left[\tfrac{1}{2}\pi^2(\mathbf{x}, t) + \tfrac{1}{2}\Big(\nabla\phi(\mathbf{x}, t)\Big)^2 + \tfrac{1}{2}m^2\phi^2 \right] . \end{aligned} \tag{7.2.25}$$

The relation between π and $\dot{\phi}$ is dictated by Eq. (7.1.21)

$$\dot{\phi}(\mathbf{x}, t) = \frac{\delta H_0}{\delta\pi(\mathbf{x}, t)} = \pi(\mathbf{x}, t) . \tag{7.2.26}$$

(This happens to be the same relation as in Eq. (7.2.16), but as we shall see this is not to be expected in general.) Also, the equation of motion for ϕ is dictated by Eq. (7.1.22):

$$\dot{\pi}(\mathbf{x}, t) = -\frac{\delta H_0}{\delta\phi(\mathbf{x}, t)} = +\nabla^2\phi(\mathbf{x}, t) - m^2\phi(\mathbf{x}, t) , \tag{7.2.27}$$

which together with Eq. (7.2.26) yields the field equations

$$(\Box - m^2)\phi(x) = 0 . \tag{7.2.28}$$

The general real solution may be expressed as

$$\phi(x) = (2\pi)^{-3/2} \int d^3p (2p^0)^{-1/2} \Big[e^{ip\cdot x} a(\mathbf{p}) + e^{-ip\cdot x} a^\dagger(\mathbf{p}) \Big] \tag{7.2.29}$$

with $p^0 = \sqrt{\mathbf{p}^2 + m^2}$ understood, and $a(\mathbf{p})$ some as-yet-unknown operator function of $\mathbf{p}$. Eq. (7.2.26) then gives the canonical conjugate as

$$\pi(x) = -i(2\pi)^{-3/2} \int d^3p\, (p^0/2)^{1/2} \Big[e^{ip\cdot x} a(\mathbf{p}) - e^{-ip\cdot x} a^\dagger(\mathbf{p}) \Big] . \tag{7.2.30}$$

In order to get the desired commutation relations,

$$\Big[\phi(\mathbf{x}, t),\, \pi(\mathbf{y}, t)\Big]_- = i\delta^3(\mathbf{x} - \mathbf{y}) , \tag{7.2.31}$$

$$\Big[\phi(\mathbf{x}, t),\, \phi(\mathbf{y}, t)\Big]_- = 0 , \tag{7.2.32}$$

$$\Big[\pi(\mathbf{x}, t),\, \pi(\mathbf{y}, t)\Big]_- = 0 , \tag{7.2.33}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

we must take the as to satisfy the familiar commutation relations

$$\left[a(\mathbf{p}),\ a^\dagger(\mathbf{p}')\right] = \delta^3(\mathbf{p}-\mathbf{p}'), \tag{7.2.34}$$

$$\left[a(\mathbf{p}),\ a(\mathbf{p}')\right] = 0\,. \tag{7.2.35}$$

Also, we have already shown in the previous section that using these expansions in Eq. (7.2.25) gives the usual formula (4.2.11) for the free-particle Hamiltonian, up to an inconsequential additive constant. As remarked before, these results should not be regarded so much as an alternative derivation of Eqs. (7.2.29), (7.2.34), and (7.2.35) (which were obtained in Chapter 5 on quite other grounds) but rather as a validation of the first two terms of Eq. (7.2.14) as the correct free-particle Lagrangian for a real scalar field. We can now proceed to use perturbation theory to calculate the S-matrix, taking (7.2.24) as $V(t)$, with the field $\phi(x)$ given by Eq. (7.2.29).

The procedures illustrated here will be carried out for examples that are more complicated and more interesting in Section 7.5.

* * *

In considering the various possible Lagrangian densities for physical theories it is common to apply integration by parts, treating as equivalent any Lagrangian densities that differ only by total derivatives $\partial_\mu \mathscr{F}^\mu$. It is obvious that such total derivative terms do not contribute to the action and hence do not affect the field equations. It is also obvious that a space-derivative term $\nabla\cdot\mathscr{F}$ in the Lagrangian density does not contribute to the Lagrangian and hence does not affect the quantum theory defined by the Lagrangian.¶ What is less obvious and worth noting here is that a *time*-derivative $\partial_0\mathscr{F}^0$ in the Lagrangian density also does not affect the quantum structure of the theory. To see this, let's first consider the effect of adding a term to the Lagrangian of the more general form

$$\Delta L(t) = \int d^3x\, D_{n,\mathbf{x}}[Q(t)]\,\dot{Q}^n(\mathbf{x},t)\,, \tag{7.2.36}$$

where D is an arbitrary n- and $\mathbf{x}$-dependent functional of the values of Q at a given time. This changes the formula for the conjugate variables $P(t)$ as functionals of $Q(t)$ and $\dot{Q}(t)$ by the amount

$$\Delta P_n(\mathbf{x},t) = \frac{\delta\,\Delta L(t)}{\delta\,\dot{Q}^n(\mathbf{x},t)} = D_{n,\mathbf{x}}[Q(t)]\,. \tag{7.2.37}$$

It follows that there is no change in the Hamiltonian as expressed as a

¶ This is under the usual assumption, that the fields vanish at infinity. These results do not necessarily apply when we allow fields of different topology, as discussed in Volume II.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

functional of the $Q(t)$ and $\dot{Q}(t)$:

$$\int d^3x\, \Delta P_n(\mathbf{x},t)\dot{Q}^n(\mathbf{x},t) - \Delta L(t) = 0. \tag{7.2.38}$$

Hence also there is no change in the Hamiltonian as expressed as a functional of the old canonical variables Q^n and P_n. However, the Hamiltonian is *not* the same functional of the *new* canonical variables Q^n and $P_n + \Delta P_n$ as it was of the Q^n and P_n, and in a theory described by the new Lagrangian $\mathscr{L} + \Delta\mathscr{L}$ it is the new canonical variables Q^n and $P_n + \Delta P_n$ rather than the Q^n and P_n that would satisfy the canonical commutation relations. The commutators of the Q^n with each other and of the Q^n with the P_m are given by the usual canonical relations, but the commutators of the P_n with each other are now

$$\begin{aligned}[P_n(\mathbf{x},t), P_m(\mathbf{y},t)] &= [P_n(\mathbf{x},t) + \Delta P_n(\mathbf{x},t), P_m(\mathbf{y},t) + \Delta P_m(\mathbf{y},t)] \\ &\quad -[\Delta P_n(\mathbf{x},t), P_m(\mathbf{y},t) + \Delta P_m(\mathbf{y},t)] \\ &\quad -[P_n(\mathbf{x},t) + \Delta P_n(\mathbf{x},t), \Delta P_m(\mathbf{y},t)] \\ &\quad +[\Delta P_n(\mathbf{x},t), \Delta P_m(\mathbf{y},t)] \\ &= -i\frac{\delta D_{n,\mathbf{x}}[Q(t)]}{\delta Q^m(\mathbf{y},t)} + i\frac{\delta D_{m,\mathbf{y}}[Q(t)]}{\delta Q^n(\mathbf{x},t)}\,. \end{aligned} \tag{7.2.39}$$

In general this doesn't vanish, but if the added term in the Lagrangian is a total time-derivative

$$\Delta L = \frac{dG}{dt} = \int d^3x\, \frac{\delta G[Q(t)]}{\delta Q^n(\mathbf{x},t)}\dot{Q}^n(\mathbf{x},t)\,, \tag{7.2.40}$$

then D in Eq. (7.2.36) is of the special form

$$D_{n,\mathbf{x}}[Q] = \frac{\delta G[Q(t)]}{\delta Q^n(\mathbf{x},t)}\,. \tag{7.2.41}$$

In this case the commutator (7.2.39) vanishes, so the variables Q^n and P_n satisfy the usual commutation relations. We have seen that a change of the form (7.2.36) in the Lagrangian does not change the form of the Hamiltonian as a functional of the Q^n and P_n, and since, as we have now shown, the commutation relations of these variables are also unchanged, the addition to the Lagrangian of the term (7.2.36) has no effect on the quantum structure of the theory. Different Lagrangian densities obtained from each other by partial integration may therefore be regarded as equivalent in quantum as well as classical field theory.

7.3 Global Symmetries

We now come to the real point of the Lagrangian formalism, that it provides a natural framework for the quantum mechanical implementation

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

of symmetry principles. This is because the dynamical equations in the Lagrangian formalism take the form of a variational principle, the principle of stationary action. Consider any infinitesimal transformation of the fields

$$\Psi^\ell(x) \to \Psi^\ell(x) + i\epsilon\, \mathscr{F}^\ell(x) \tag{7.3.1}$$

that leaves the action (7.2.3) invariant:

$$0 = \delta I = i\epsilon \int d^4x\, \frac{\delta I[\Psi]}{\delta \Psi^\ell(x)} \mathscr{F}^\ell(x)\,. \tag{7.3.2}$$

(With ϵ a constant, such symmetries are known as *global* symmetries. In general, $\mathscr{F}^\ell$ depends on the fields and their derivatives at x.) Of course Eq. (7.3.2) is automatically satisfied for *all* infinitesimal variations of the fields if the fields satisfy the dynamical equations; by an infinitesimal symmetry transformation we mean one that leaves the action invariant even when the dynamical equations are *not* satisfied. If we now consider the same transformation with ϵ an arbitrary function of position in spacetime:

$$\Psi^\ell(x) \to \Psi^\ell(x) + i\epsilon(x)\, \mathscr{F}^\ell(x), \tag{7.3.3}$$

then, in general, the variation of the action will not vanish, but it will have to be of the form

$$\delta I = -\int d^4x\, J^\mu(x) \frac{\partial \epsilon(x)}{\partial x^\mu} \tag{7.3.4}$$

in order that it should vanish when $\epsilon(x)$ *is* constant. If we *now* take the fields in $I[\Psi]$ to satisfy the field equations then I is stationary with respect to arbitrary field variations that vanish at large spacetime distances, including variations of the form (7.3.3), so in this case (7.3.4) should vanish. Integrating by parts, we see that $J^\mu(x)$ must satisfy a conservation law:

$$0 = \frac{\partial J^\mu(x)}{\partial x^\mu}\,. \tag{7.3.5}$$

It follows immediately that

$$0 = \frac{dF}{dt}\,, \tag{7.3.6}$$

where

$$F \equiv \int d^3x\, J^0\,. \tag{7.3.7}$$

There is one such conserved current J^μ and one constant of the motion F for each independent infinitesimal symmetry transformation. This represents a general feature of the canonical formalism, often referred to as Noether's theorem: *symmetries imply conservation laws.*

Many symmetry transformations leave the Lagrangian and not just the action invariant. This is the case, for instance, for translations and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

rotations in space and also isospin transformations and other internal symmetry transformations, though not for general Lorentz transformations. When the Lagrangian is invariant we can go further, and write an explicit formula for the conserved quantities F. Consider a field variation (7.3.3) in which $\epsilon(x)$ depends on t but not $\mathbf{x}$. In this case the variation in the action is

$$\delta I = i \int dt \int d^3x \left[\frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \Psi^\ell(\mathbf{x}, t)} \epsilon(t) \mathscr{F}^\ell(\mathbf{x}, t) + \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \dot{\Psi}^\ell(\mathbf{x}, t)} \frac{d}{dt} \left(\epsilon(t) \mathscr{F}^\ell(\mathbf{x}, t) \right) \right] . \tag{7.3.8}$$

The requirement that the Lagrangian be invariant under this transformation when ϵ is a constant yields

$$0 = \int d^3x \left[\frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \Psi^\ell(\mathbf{x}, t)} \mathscr{F}^\ell(\mathbf{x}, t) + \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \dot{\Psi}^\ell(\mathbf{x}, t)} \frac{d}{dt} \mathscr{F}^\ell(\mathbf{x}, t) \right] , \tag{7.3.9}$$

so for general fields (whether or not the field equations are satisfied) the variation in the action is

$$\delta I = i \int dt \int d^3x \, \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \dot{\Psi}^\ell(\mathbf{x}, t)} \dot{\epsilon}(t) \mathscr{F}^\ell(\mathbf{x}, t). \tag{7.3.10}$$

Comparing this with Eq. (7.3.4) gives

$$F = -i \int d^3x \frac{\delta L[\Psi(t), \dot{\Psi}(t)]}{\delta \dot{\Psi}^\ell(\mathbf{x}, t)} \mathscr{F}^\ell(\mathbf{x}, t). \tag{7.3.11}$$

Using the symmetry condition (7.3.9), the reader can easily check that this F is indeed time-independent for any fields that satisfy the dynamical equations (7.2.2).

Other symmetry transformations such as isospin rotations leave not only the action and the Lagrangian invariant but also the Lagrangian density. In such cases we can go even further, and write an explicit formula for the current $J^\mu(x)$. Writing the action as in Eq. (7.2.6) as the integral of the Lagrangian density, its variation under the transformation (7.3.3) with a general infinitesimal parameter $\epsilon(x)$ is

$$\delta I[\Psi] = i \int d^4x \left[\frac{\partial \mathscr{L}(\Psi(x), \partial_\mu \Psi(x))}{\partial \Psi^\ell(x)} \mathscr{F}^\ell(x) \epsilon(x) + \frac{\partial \mathscr{L}(\Psi(x), \partial_\mu \Psi(x))}{\partial(\partial_\mu \Psi^\ell(x))} \partial_\mu \left(\mathscr{F}^\ell(x) \epsilon(x) \right) \right] . \tag{7.3.12}$$

The invariance of the Lagrangian density when ϵ is a constant requires

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

that

$$0 = \frac{\partial \mathscr{L}(\Psi(x), \partial_\mu \Psi(x))}{\partial \Psi^\prime(x)} \mathscr{F}^\prime(x) + \frac{\partial \mathscr{L}(\Psi(x), \partial_\mu \Psi(x))}{\partial(\partial_\mu \Psi^\prime(x))} \partial_\mu \mathscr{F}^\prime(x) , \quad (7.3.13)$$

so for arbitrary fields the variation of the action is

$$\delta I[\Psi] = i \int d^4x \, \frac{\partial \mathscr{L}(\Psi(x), \partial_\mu \Psi(x))}{\partial(\partial_\mu \Psi^\prime(x))} \mathscr{F}^\prime(x) \partial_\mu \epsilon(x). \quad (7.3.14)$$

Comparison with Eq. (7.3.4) shows that

$$J^\mu = -i \frac{\partial \mathscr{L}}{\partial(\partial \Psi^\prime / \partial x^\mu)} \mathscr{F}^\prime . \quad (7.3.15)$$

Using the symmetry condition (7.3.13), it is easy to see directly that $\partial_\mu J^\mu$ vanishes when the fields satisfy the Euler–Lagrange equations (7.2.9). Note also that the integral of the time component of the current (7.3.15) has the previously derived value (7.3.11).

So far everything we have said would apply to classical as well as quantum mechanical field theories. The quantum properties of the conserved quantities F are most easily seen for symmetries of the Lagrangian (not necessarily the Lagrangian density) that transform the canonical fields $Q^n(\mathbf{x}, t)$ (that is, those of the $\Psi^\prime$ whose time derivatives appear in the Lagrangian) into $\mathbf{x}$-dependent functionals of themselves at the same time. For such transformations, we have

$$\mathscr{F}^n(\mathbf{x}, t) = \mathscr{F}^n[Q(t); \mathbf{x}] . \quad (7.3.16)$$

As we shall see, infinitesimal spatial translations and rotations as well as all infinitesimal internal symmetry transformations are of the form (7.3.1), (7.3.16), with $\mathscr{F}^n$ a linear functional of the Q^m, but we will not need to assume here that the symmetry is linear. For all such symmetries the operator F is not only conserved; it also acts in quantum mechanics as a *generator* of this symmetry.

To see this, note first that when $\Psi^\prime$ is a canonical field Q^n, the functional derivative $\delta L / \delta \dot{\Psi}^\prime$ is equal to the canonical conjugate P_n, while when $\Psi^\prime$ is an auxiliary field C^r, this functional derivative vanishes; hence we may rewrite Eq. (7.3.11) in the form

$$F = -i \int d^3x \, P_n(\mathbf{x}, t) \mathscr{F}^n(\mathbf{x}, t) = -i \int d^3x \, P_n(\mathbf{x}, t) \, \mathscr{F}^n[Q(t), \mathbf{x}] . \quad (7.3.17)$$

To calculate the commutator (not anticommutator) of F with a canonical field $Q^m(\mathbf{x}, t)$ at an arbitrary time t, we can invoke Eq. (7.3.6) to evaluate F as a functional of the Qs and Ps at the time t, and then use the equal-time

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

canonical commutation relations (7.1.30)–(7.1.32) to obtain*

$$[F, Q^n(\mathbf{x},t)]_- = -\mathscr{F}^n(\mathbf{x},t)\,. \tag{7.3.18}$$

It is in this sense that F is the generator of the transformation with Eq. (7.3.16). Eq. (7.3.17) and the canonical commutation rules give also

$$[F, P_n(\mathbf{x},t)]_- = \int d^3y\, P_m(\mathbf{y},t)\,\frac{\delta\mathscr{F}^m(Q(t);\mathbf{y})}{\delta Q^n(\mathbf{x},t)}\,. \tag{7.3.19}$$

Where F^m is linear, Eq. (7.3.19) tells us that P_n transforms contragrediently to Q^n.

As a first example, consider the symmetry transformation of spacetime translation:

$$\Psi^\ell(x) \to \Psi^\ell(x+\epsilon) = \Psi^\ell(x) + \epsilon^\mu\partial_\mu\Psi^\ell(x)\,. \tag{7.3.20}$$

This is of the form (7.3.1), with four independent parameters ϵ^μ and four corresponding transformation functions

$$\mathscr{F}^\ell_\mu = -i\partial_\mu\Psi^\ell\,. \tag{7.3.21}$$

In consequence we have four independent conserved currents, conventionally grouped together in the *energy–momentum tensor* $T^\mu{}_\nu$:

$$\partial_\mu T^\mu{}_\nu = 0 \tag{7.3.22}$$

from which we can derive time-independent quantities as the spatial integrals of the time components of the translation 'currents' (not to be confused with the canonical conjugate field variables $P_n(\mathbf{x},t)$):

$$P_\nu = \int d^3x\, T^0{}_\nu\,, \tag{7.3.23}$$

$$\frac{d}{dt}\,P_\nu = 0\,. \tag{7.3.24}$$

The Lagrangian is invariant under spatial translations, so in accordance with the above general results we can conclude that the spatial components of P_ν take the form

$$\mathbf{P} \equiv -\int d^3x\, P_n(\mathbf{x},t)\nabla Q^n(\mathbf{x},t)\,. \tag{7.3.25}$$

Using the equal-time commutation relations (7.1.30)–(7.1.32), we also find the commutator of this operator with the canonical fields and

* We are here assuming that for Q^n bosonic or fermionic the variation $\mathscr{F}^n$ is also respectively bosonic or fermionic, so that F is bosonic. The only exceptions are certain symmetries known as supersymmetries, for which F is fermionic and (7.3.18) is an anticommutator if Q^n is also fermionic.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

conjugates:

$$[\mathbf{P}\,,\, Q^n(\mathbf{x},t)]_- = i\nabla Q^n(\mathbf{x},t)\,, \tag{7.3.26}$$

$$[\mathbf{P}\,,\, P_n(\mathbf{x},t)]_- = i\nabla P_n(\mathbf{x},t)\,. \tag{7.3.27}$$

It follows that for any function $\mathcal{G}$ of Qs and Ps that does not also depend explicitly on $\mathbf{x}$, we have

$$\left[\mathbf{P},\, \mathcal{G}(x)\right] = i\nabla\mathcal{G}(x)\,. \tag{7.3.28}$$

These results show that the operator $\mathbf{P}$ is indeed the generator of space translations.

In contrast, time-translations do not leave the Lagrangian $L(t)$ invariant. However, we already know the generator of time-translations; it is the Hamiltonian $P^0 \equiv H$, which as we know satisfies the commutation relation

$$[H\,,\, \mathcal{G}(\mathbf{x},t)] = -i\dot{\mathcal{G}}(\mathbf{x},t) \tag{7.3.29}$$

for any function $\mathcal{G}$ of Heisenberg picture operators.

If we further assume that the Lagrangian is the integral of a Lagrangian density, then we may also obtain an explicit formula for the energy–momentum tensor $T^\mu{}_\nu$. However, the Lagrangian density $\mathscr{L}(x)$ is not invariant under spacetime translations, so we cannot use Eq. (7.3.15) here. Instead, note that the change in the action under a spacetime-dependent translation

$$\Psi^\ell(x) \to \Psi^\ell(x+\epsilon(x)) = \Psi^\ell(x) + \epsilon^\mu(x)\partial_\mu\Psi^\ell(x) \tag{7.3.30}$$

is

$$\delta I[\Psi] = \int d^4x \left(\frac{\partial\mathscr{L}}{\partial\Psi^\ell}\epsilon^\mu\partial_\mu\Psi^\ell + \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Psi^\ell)}\partial_\nu[\epsilon^\mu\partial_\mu\Psi^\ell]\right)\,. \tag{7.3.31}$$

The Euler–Lagrange equations (7.2.9) show that the terms proportional to ϵ add up to $\epsilon^\mu\partial_\mu\mathscr{L}$, so

$$\delta I[\Psi] = \int d^4x \left(\frac{\partial\mathscr{L}}{\partial x^\mu}\epsilon^\mu + \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Psi^\ell)}\partial_\mu\Psi^\ell\partial_\nu\epsilon^\mu\right)\,. \tag{7.3.32}$$

Integrating by parts, we see that this takes the form of Eq. (7.3.4)

$$\delta I = -\int d^4x\, T^\nu{}_\mu\partial_\nu\epsilon^\mu \tag{7.3.33}$$

with 'currents'

$$T^\nu{}_\mu = \delta^\nu_\mu\mathscr{L} - \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Psi^\ell)}\partial_\mu\Psi^\ell\,. \tag{7.3.34}$$

As a check, we may note that the spatial components of Eq. (7.3.23) are the same as our previous formula (7.3.25) for $\mathbf{P}$, while for $\mu = 0$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Eq. (7.3.23) gives the usual formula for the Hamiltonian:

$$H \equiv -P_0 = \int d^3x \left[\sum_n P_n \dot{Q}^n - \mathscr{L} \right] . \tag{7.3.35}$$

(A warning: the tensor $T^{\mu\nu}$ obtained by raising the second index in Eq. (7.3.34) is not in general symmetric, and therefore cannot be used as the right-hand side of the field equations of general relativity. The correct energy–momentum tensor to use as the source of the gravitational field is the symmetric tensor $\Theta^{\mu\nu}$ introduced in the next section.)

In many theories there are also one or more symmetry principles that state the invariance of the action, under a set of linear coordinate-independent transformations of the canonical fields

$$Q^n(x) \to Q^n(x) + i\epsilon^a (t_a)^n{}_m Q^m(x) \tag{7.3.36}$$

together with a set of suitable transformations on any auxiliary fields C^r:

$$C^r(x) \to C^r(x) + i\epsilon^a (\tau_a)^r{}_s C^s(x) . \tag{7.3.37}$$

Here t_a and τ_a are sets of Hermitian matrices furnishing some representations of the Lie algebra of the symmetry group, and we sum over repeated group indices a, b, etc. (For instance, in electrodynamics there is such a symmetry, for which the one matrix $t^n{}_m$ is diagonal, with the charges carried by each field on the main diagonal.) From any such symmetry, we can infer the existence of another set of conserved currents J_a^μ:

$$\partial_\mu J_a^\mu = 0 , \tag{7.3.38}$$

whose time components are the densities of a set of time-independent operators

$$T_a = \int d^3x \, J_a^0 . \tag{7.3.39}$$

When the Lagrangian as well as the action is invariant under the transformation (7.3.36), Eq. (7.3.11) provides an explicit formula for the T_a:

$$T_a = -i \int d^3x \, P_n(\mathbf{x}, t)(t_a)^n{}_m Q^m(\mathbf{x}, t) . \tag{7.3.40}$$

The equal-time commutation relations here give

$$\left[T_a , Q^n(x) \right] = -(t_a)^n{}_m Q^m(x), \tag{7.3.41}$$

$$\left[T_a , P_n(x) \right] = +(t_a)^m{}_n P_m(x) . \tag{7.3.42}$$

(Where t_a is diagonal, this tells us that Q^n and P_n respectively lower and raise the value of T_a by an amount equal to the nth diagonal element of t_a.) Using these results, we can calculate the commutator of T_a with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

another generator T_b:

$$[T_a, T_b]_- = i \int d^3x \left[- P_m (t_a)^m{}_n (t_b)^n{}_k Q^k + P_n (t_b)^n{}_k (t_a)^k{}_m Q^m \right] . \tag{7.3.43}$$

Thus, if the matrices t_a form a Lie algebra with structure constants $f_{ab}{}^c$,

$$[t_a, t_b]_- = i f_{ab}{}^c t_c , \tag{7.3.44}$$

then so do the quantum operators T_a:

$$[T_a, T_b]_- = i f_{ab}{}^c T_c . \tag{7.3.45}$$

This confirms that the quantities (7.3.40) are correctly normalized to qualify as generators of the symmetry group.

Where the Lagrangian is the integral of a Lagrangian density which is invariant under (7.3.36) and (7.3.37) we can go further, and use Eq. (7.3.15) to provide an explicit formula for the currents associated with these global symmetries:

$$J_a^\mu \equiv -i \frac{\partial \mathscr{L}}{\partial(\partial Q^n / \partial x^\mu)} (t_a)^n{}_m Q^m - i \frac{\partial \mathscr{L}}{\partial(\partial C^r / \partial x^\mu)} (\tau_a)^r{}_s C^s . \tag{7.3.46}$$

As an illustration, suppose we have *two* real scalar fields of equal mass, with Lagrangian density

$$\mathscr{L} = - \tfrac{1}{2} \partial_\mu \Phi_1 \partial^\mu \Phi_1 - \tfrac{1}{2} m^2 \Phi_1^2 - \tfrac{1}{2} \partial_\mu \Phi_2 \partial^\mu \Phi_2 - \tfrac{1}{2} m^2 \Phi_2^2 - \mathscr{H}(\Phi_1^2 + \Phi_2^2) . \tag{7.3.47}$$

This is invariant under a linear transformation like (7.3.36):

$$\delta \Phi_1 = -\epsilon \Phi_2 , \qquad \delta \Phi_2 = +\epsilon \Phi_1 ,$$

so there is a conserved current (7.3.46):

$$J^\mu = \Phi_2 \partial^\mu \Phi_1 - \Phi_1 \partial^\mu \Phi_2 .$$

The explicit formula (7.3.46) for the current can be used to derive other useful commutation relations. In particular, since the Lagrangian density does not involve time-derivatives of the auxiliary fields, we have

$$J_a^0 = -i P_n (t_a)^n{}_m Q^m . \tag{7.3.48}$$

We can then derive the equal-time commutators of general fields not only with the symmetry generators T_a, but also with the densities J_a^0:

$$[J_a^0(\mathbf{x}, t), Q^n(\mathbf{y}, t)] = -\delta^3(\mathbf{x} - \mathbf{y}) (t_a)^n{}_m Q^m(\mathbf{x}, t) , \tag{7.3.49}$$

$$[J_a^0(\mathbf{x}, t), P_m(\mathbf{y}, t)] = \delta^3(\mathbf{x} - \mathbf{y}) (t_a)^n{}_m P_n(\mathbf{x}, t) . \tag{7.3.50}$$

If the auxiliary fields are constructed as local functions of the Ps and Qs in such a way that they transform according to a representation of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

symmetry algebra with generators τ_a, then also

$$[J_a^0(\mathbf{x},t), C^r(\mathbf{y},t)] = -\delta^3(\mathbf{x}-\mathbf{y})\,(\tau_a)^r{}_s C^s(\mathbf{x},t)\,. \tag{7.3.51}$$

We often summarize Eqs. (7.3.49) and (7.3.51) in the single commutation relation

$$[J_a^0(\mathbf{x},t)\,,\ \Psi^\ell(\mathbf{y},t)] = -\delta^3(\mathbf{x}-\mathbf{y})\,(t_a)^\ell{}_{\ell'}\Psi^{\ell'}(\mathbf{x},t)\,. \tag{7.3.52}$$

Commutation relations like (7.3.49)–(7.3.51) will be used in Chapter 10 to derive relations called Ward identities for matrix elements involving the current J^μ.

7.4 Lorentz Invariance

We are now going to show that the Lorentz invariance of the Lagrangian density implies the Lorentz invariance of the S-matrix. Consider an infinitesimal Lorentz transformation

$$\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \omega^\mu{}_\nu\,, \tag{7.4.1}$$

$$\omega_{\mu\nu} = -\omega_{\nu\mu}\,. \tag{7.4.2}$$

According to the analysis of the previous section, the invariance of the action under such transformations tells us immediately that there is a set of conserved 'currents' $\mathscr{M}^{\rho\mu\nu}$:

$$\partial_\rho \mathscr{M}^{\rho\mu\nu} = 0\,, \tag{7.4.3}$$

$$\mathscr{M}^{\rho\mu\nu} = -\mathscr{M}^{\rho\nu\mu}\,, \tag{7.4.4}$$

one current for each independent component of $\omega_{\mu\nu}$. The integrals of the time-components of these 'currents' then provide us with a set of time-independent tensors:

$$J^{\mu\nu} \equiv \int d^3x\ \mathscr{M}^{0\mu\nu}\,, \tag{7.4.5}$$

$$\frac{d}{dt}J^{\mu\nu} = 0\,. \tag{7.4.6}$$

The $J^{\mu\nu}$ will turn out to be the generators of the homogeneous Lorentz group.

We would like to have an explicit formula for the tensor $\mathscr{M}^{\rho\mu\nu}$, but Lorentz transformations act on the coordinates and hence cannot leave the Lagrangian density invariant, so we cannot immediately use the results of the previous section. However, translation invariance allows us to formulate Lorentz invariance as a symmetry of the Lagrangian density

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

under a set of transformations on the fields and field-derivatives alone. The fields undergo the matrix transformation

$$\delta\Psi^\ell = \frac{i}{2}\omega^{\mu\nu}(\mathscr{J}_{\mu\nu})^\ell{}_m\Psi^m\,, \tag{7.4.7}$$

where $\mathscr{J}_{\mu\nu}$ are a set of matrices satisfying the algebra of the homogeneous Lorentz group

$$[\mathscr{J}_{\mu\nu},\ \mathscr{J}_{\rho\sigma}] = i\mathscr{J}_{\rho\nu}\eta_{\mu\sigma} - i\mathscr{J}_{\sigma\nu}\eta_{\mu\rho} - i\mathscr{J}_{\mu\sigma}\eta_{\nu\rho} + i\mathscr{J}_{\mu\rho}\eta_{\nu\sigma}\,. \tag{7.4.8}$$

For example, for a scalar field ϕ we have $\delta\phi = 0$, so $\mathscr{J}_{\mu\nu} = 0$, while for an irreducible field of type (A, B) we have

$$\mathscr{J}_{ij} = \epsilon_{ijk}(\mathscr{A}_k + \mathscr{B}_k)\,, \qquad \mathscr{J}_{k0} = -i(\mathscr{A}_k - \mathscr{B}_k)\,,$$

where $\mathscr{A}$ and $\mathscr{B}$ are spin matrices for spin A and B, respectively. We specially note that for a covariant vector field, $\delta V_\kappa = \omega_\kappa{}^\lambda V_\lambda$, so here

$$(\mathscr{J}_{\rho\sigma})_\kappa{}^\lambda = -i\eta_{\rho\kappa}\delta_\sigma{}^\lambda + i\eta_{\sigma\kappa}\delta_\rho{}^\lambda\,.$$

The derivative of a field that transforms as in Eq. (7.4.7) transforms like another such field, but with an extra vector index

$$\delta(\partial_\kappa\Psi_\ell) = \tfrac{1}{2}i\,\omega^{\mu\nu}(\mathscr{J}_{\mu\nu})_\ell{}^m\partial_\kappa\Psi_m + \omega_\kappa{}^\lambda\partial_\lambda\Psi_\ell\,. \tag{7.4.9}$$

The Lagrangian density is assumed to be invariant under the combined transformations Eqs. (7.4.7) and (7.4.9), so

$$0 = \frac{\partial\mathscr{L}}{\partial\Psi^\ell}\,\frac{i}{2}\,\omega^{\mu\nu}(\mathscr{J}_{\mu\nu})^\ell{}_m\Psi^m + \frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}\,\frac{i}{2}\,\omega^{\mu\nu}(\mathscr{J}_{\mu\nu})^\ell{}_m\partial_\kappa\Psi^m + \frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}\,\omega_\kappa{}^\lambda\partial_\lambda\Psi^\ell\,.$$

Setting the coefficient of $\omega^{\mu\nu}$ equal to zero gives

$$0 = \frac{i}{2}\,\frac{\partial\mathscr{L}}{\partial\Psi^\ell}\,(\mathscr{J}_{\mu\nu})^\ell{}_m\Psi^m + \frac{i}{2}\frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}\,(\mathscr{J}_{\mu\nu})^\ell{}_m\partial_\kappa\Psi^m + \frac{1}{2}\,\frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}\,(\eta_{\kappa\mu}\partial_\nu - \eta_{\kappa\nu}\partial_\mu)\Psi^\ell\,.$$

Using the Euler–Lagrange equations (7.2.9), and our formula (7.3.34) for the energy-momentum tensor $T_{\mu\nu}$, we may write this as

$$0 = \partial_\kappa\left[\frac{i}{2}\,\frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}\,(\mathscr{J}_{\mu\nu})^\ell{}_m\Psi^m\right] - \frac{1}{2}\,(T_{\mu\nu} - T_{\nu\mu})\,. \tag{7.4.10}$$

This immediately suggests the definition of a new energy–momentum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

tensor, known as the *Belinfante tensor*:[2]

$$\Theta^{\mu\nu} = T^{\mu\nu} - \frac{i}{2}\partial_\kappa\left[\frac{\partial\mathscr{L}}{\partial(\partial_\kappa\Psi^\ell)}(\mathscr{J}^{\mu\nu})^\ell{}_m\Psi^m - \frac{\partial\mathscr{L}}{\partial(\partial_\mu\Psi^\ell)}(\mathscr{J}^{\kappa\nu})^\ell{}_m\Psi^m - \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Psi^\ell)}(\mathscr{J}^{\kappa\mu})^\ell{}_m\Psi^m\right]. \quad (7.4.11)$$

The quantity in square brackets is manifestly antisymmetric in μ and κ, so $\Theta^{\mu\nu}$ satisfies the same conservation law as $T^{\mu\nu}$,

$$\partial_\mu\Theta^{\mu\nu} = 0\,. \quad (7.4.12)$$

For the same reason, when we set $\mu = 0$ in Eq. (7.4.11) the index κ runs over space components only, so the derivative term here drops out when we integrate over all space

$$\int \Theta^{0\nu}d^3x = \int T^{0\nu}d^3x = P^\nu\,, \quad (7.4.13)$$

where $P^0 \equiv H$. Thus $\Theta^{\mu\nu}$ can be regarded as *the* energy–momentum tensor, just as well as $T^{\mu\nu}$. However, Eq. (7.4.10) tells us that, unlike $T^{\mu\nu}$, the Belinfante tensor $\Theta^{\mu\nu}$ is not only conserved but also *symmetric*:

$$\Theta^{\mu\nu} = \Theta^{\nu\mu}\,. \quad (7.4.14)$$

It is $\Theta^{\mu\nu}$ rather than $T^{\mu\nu}$ that acts as a source of the gravitational field.[3] In consequence of the symmetry of $\Theta^{\mu\nu}$, we may construct one more conserved tensor density:

$$\mathscr{M}^{\lambda\mu\nu} \equiv x^\mu\Theta^{\lambda\nu} - x^\nu\Theta^{\lambda\mu}\,. \quad (7.4.15)$$

This is conserved, in the sense that

$$\partial_\lambda\mathscr{M}^{\lambda\mu\nu} = \Theta^{\mu\nu} - \Theta^{\nu\mu} = 0\,. \quad (7.4.16)$$

Thus Lorentz invariance allows us to define one more time-independent tensor

$$J^{\mu\nu} = \int \mathscr{M}^{0\mu\nu}d^3x = \int d^3x(x^\mu\Theta^{0\nu} - x^\nu\Theta^{0\mu})\,. \quad (7.4.17)$$

The rotation generator $J_k = \epsilon_{ijk}J^{ij}/2$ is not only time-independent, but also has no *explicit* time-dependence, so it commutes with the Hamiltonian

$$[H, \mathbf{J}] = 0\,. \quad (7.4.18)$$

Also, applying Eq. (7.3.28) to the function $\Theta^{0\nu}$, we have

$$[P_j, J_i] = \frac{1}{2}\epsilon_{i\ell k}[P_j, J^{\ell k}] = \frac{i}{2}\epsilon_{i\ell k}\int d^3x\left(x^\ell\frac{\partial}{\partial x^j}\Theta^{0k} - x^k\frac{\partial}{\partial x^j}\Theta^{0\ell}\right)$$
$$= -i\epsilon_{ijk}\int d^3x\,\Theta^{0k}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

and therefore

$$[P_j, J_i] = -i\,\epsilon_{ijk}\,P_k\ . \tag{7.4.19}$$

On the other hand, the 'boost' generator $K_k \equiv J^{k0}$, though time-independent, does explicitly involve the time coordinate

$$K_k = \int d^3x (x^k \Theta^{00} - x^0 \Theta^{0k})\ ,$$

or more explicitly

$$\mathbf{K} = -t\mathbf{P} + \int d^3x\ \mathbf{x}\,\Theta^{00}(\mathbf{x},t)\ . \tag{7.4.20}$$

Since this is a constant, we have $0 = \dot{\mathbf{K}} = -\mathbf{P} + i[H, \mathbf{K}]$, and therefore

$$[H, \mathbf{K}] = -i\,\mathbf{P}\ . \tag{7.4.21}$$

Also, applying Eq. (7.3.28) again gives

$$[P_j, K_k] = i \int d^3x\, x^k \frac{\partial}{\partial x^j}\,\Theta^{00} = -i\delta_{jk} \int d^3x\,\Theta^{00}$$

and therefore

$$[P_j, K_k] = -i\,\delta_{jk}\,H\ . \tag{7.4.22}$$

For any reasonable Lagrangian density, the operator (7.4.20) will be 'smooth' in the sense used in Section 3.3, i.e., the interaction terms in $e^{iH_0t} \int d^3x\ \mathbf{x}\,\Theta^{00}(\mathbf{x},0)e^{-iH_0t}$ vanish* for $t \to \pm\infty$. (Note that the interaction terms in $e^{iH_0t} \int d^3x\ \Theta^{00}(\mathbf{x},0)e^{-iH_0t}$ must vanish for $t \to \pm\infty$ in order to allow the introduction of 'in' and 'out' states and the S-matrix.) With this smoothness assumption and the commutation relation (7.4.21) in hand, we can repeat the arguments of Section 3.3, and conclude that the S-matrix is Lorentz-invariant.

* * *

The same arguments were also used in Section 3.3 to verify that the remaining commutation relations of the Lorentz group, those of the $J^{\mu\nu}$ with each other, take the proper form. This can also be shown directly for the commutators of the rotation generators, which here take the form

$$J^{ij} = \int d^3x\ \frac{\partial \mathscr{L}}{\partial \dot{\Psi}^\ell} \left(-x^i \partial_j \Psi^\ell + x^j \partial_i \Psi^\ell - i \left(\mathscr{J}^{ij} \right)^\ell{}_m \Psi^m \right)\ . \tag{7.4.23}$$

Since the Lagrangian density does not depend on the time-derivatives of the auxiliary fields, and the rotation generators do not mix canonical and

* When we say that some interaction-picture operator vanishes for $t \to \pm\infty$, we mean that its matrix elements between states that are smooth superpositions of energy eigenstates vanish in this limit.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

auxiliary fields, this can also be written as a sum over canonical fields alone:

$$J^{ij} = \int d^3x\, P_n \left(-x^i \partial_j Q^n + x^j \partial_i Q^n - i \left(\mathscr{J}^{ij} \right)^n{}_{n'} Q^{n'} \right) . \tag{7.4.24}$$

It follows immediately then from the canonical commutation relations that

$$[J^{ij}, Q^n(x)]_- = -i(-x_i \partial_j + x_j \partial_i) Q^n(x) - (\mathscr{J}^{ij})^n{}_{n'} Q^{n'}(x) , \tag{7.4.25}$$

$$[J^{ij}, P_n(x)]_- = i(-x_i \partial_j + x_j \partial_i) P^n(x) + (\mathscr{J}^{ij})^{n'}{}_n P_{n'}(x) . \tag{7.4.26}$$

These results can be used to derive the usual commutation relations of the J^{ij} with each other and other generators.** If there are no auxiliary fields then the same arguments may be applied to the 'boost' generators to complete the demonstration that the P^μ and $J^{\mu\nu}$ satisfy the commutation relations of the inhomogeneous Lorentz group. However the 'boost' matrices $\mathscr{J}^{i0}$ will in general mix canonical and auxiliary fields (such as the components V^i and V^0 of a vector field), so the direct proof of the commutation relations of the J^{i0} with each other has to be given on a case by case basis. Fortunately, this is not needed for the proof of the Lorentz invariance of the S-matrix given in Section 3.3.

7.5 Transition to Interaction Picture: Examples

At the end of Section 7.2 we showed how to use the Lagrangian of a simple scalar field theory to derive the structure of the interaction and the free fields it contains in the interaction picture. We will now turn to somewhat more complicated and revealing examples.

Scalar Field, Derivative Coupling

First let's consider a neutral scalar field, but now with derivative coupling. We take the Lagrangian as

$$\mathscr{L} = -\tfrac{1}{2} \partial_\mu \Phi \partial^\mu \Phi - \tfrac{1}{2} m^2 \Phi^2 - J^\mu \partial_\mu \Phi - \mathscr{H}(\Phi) , \tag{7.5.1}$$

where J^μ is either a c-number external current (unrelated to currents J^μ introduced earlier), or a functional of various fields other than Φ (in which case terms involving these other fields need to be added to (7.5.1)). The

** Also, since J^{ij} commutes with H and $P_n \dot{Q}^n$, it commutes with L. The commutator of J^{ij} with the auxiliary fields must thus be consistent with the rotational invariance of L.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

canonical conjugate to Φ is now

$$\Pi = \frac{\partial \mathscr{L}}{\partial \dot{\Phi}} = \dot{\Phi} - J^0 \tag{7.5.2}$$

and the Hamiltonian is

$$\begin{aligned} H &= \int d^3x[\Pi\dot{\Phi} - \mathscr{L}] \\ &= \int d^3x\Big[\Pi(\Pi + J^0) + \tfrac{1}{2}(\nabla\Phi)^2 - \tfrac{1}{2}(\Pi + J^0)^2 \\ &\quad + \tfrac{1}{2}m^2\Phi^2 + \mathbf{J}\cdot\nabla\Phi + J^0(\Pi + J^0) + \mathscr{H}(\Phi)\Big] . \end{aligned}$$

Collecting terms, we can write this as

$$H = H_0 + V, \tag{7.5.3}$$

$$H_0 = \int d^3x\Big[\tfrac{1}{2}\Pi^2 + \tfrac{1}{2}(\nabla\Phi)^2 + \tfrac{1}{2}m^2\Phi^2\Big] , \tag{7.5.4}$$

$$V = \int d^3x\Big[\Pi J^0 + \mathbf{J}\cdot\nabla\Phi + \tfrac{1}{2}(J^0)^2 + \mathscr{H}(\Phi)\Big] . \tag{7.5.5}$$

As explained in Section 7.2, we can pass to the interaction picture by simply replacing Π and Φ with π and ϕ (and likewise for any fields in the current J^μ, though we will not bother to indicate this explicitly):

$$H_0 = \int d^3x\left[\tfrac{1}{2}\pi^2(\mathbf{x},t) + \tfrac{1}{2}\Big(\nabla\phi(\mathbf{x},t)\Big)^2 + \tfrac{1}{2}m^2\phi^2(\mathbf{x},t)\right] , \tag{7.5.6}$$

$$\begin{aligned} V(t) = \int d^3x\Big[&\pi(\mathbf{x},t)J^0(\mathbf{x},t) + \mathbf{J}(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t) \\ &+ \tfrac{1}{2}\Big[J^0(\mathbf{x},t)\Big]^2 + \mathscr{H}(\phi(\mathbf{x},t))\Big] . \end{aligned} \tag{7.5.7}$$

The free-particle Hamiltonian is just the same as Eq. (7.2.25), and leads as in Section 7.2 to Eqs. (7.2.26)–(7.2.35). Indeed, whatever the total Hamiltonian may be, we *must* take (7.5.6) as the part we split off and call the free-particle part, with the remainder called the interaction, because as we have seen it is this form of the free-particle Hamiltonian that leads to the correct expansion (7.2.29) of the scalar field in terms of creation and annihilation operators that satisfy the commutation relations (7.2.34), (7.2.35). The last step is to replace π in the interaction Hamiltonian with its value $\dot{\phi}$ in the interaction picture (*not* its value $\dot{\phi} - J^0$ in the Heisenberg picture):

$$V(t) = \int d^3x\left[J^\mu(\mathbf{x},t)\partial_\mu\phi(\mathbf{x},t) + \tfrac{1}{2}\Big[J^0(\mathbf{x},t)\Big]^2 + \mathscr{H}(\phi(\mathbf{x},t)\right] . \tag{7.5.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

The extra non-invariant term in Eq. (7.5.8) is just what we saw in Section 6.2 is needed to cancel a non-invariant term in the propagator of $\partial\phi$.

Vector Field, Spin One

Similar results are obtained in the canonical quantization of the vector field V_μ for a particle of spin one. Let's here keep an open mind, and write the Lagrangian density in a fairly general form

$$\mathscr{L} = -\tfrac{1}{2}\alpha\,\partial_\mu V_\nu\,\partial^\mu V^\nu - \tfrac{1}{2}\beta\,\partial_\mu V_\nu\,\partial^\nu V^\mu - \tfrac{1}{2}m^2\,V_\mu V^\mu + J_\mu V^\mu\,, \qquad (7.5.9)$$

where α, β, and m^2 are so far arbitrary constants, and J_μ is either a c-number external current, or an operator depending on fields other than V^μ, in which case additional terms involving these fields must be added to $\mathscr{L}$. The Euler–Lagrange field equations for V_μ read

$$-\alpha\Box V_\nu - \beta\,\partial_\nu(\partial_\mu V^\mu) + m^2 V_\nu = J_\nu\,.$$

Taking the divergence gives

$$-(\alpha+\beta)\Box\partial_\lambda V^\lambda + m^2\partial_\lambda V^\lambda = \partial_\lambda J^\lambda\,. \qquad (7.5.10)$$

This is the equation for an ordinary scalar field with mass $m^2/(\alpha+\beta)$ and source $\partial_\lambda J^\lambda/(\alpha+\beta)$. We want to describe a theory containing only particles of spin one, not spin zero, so to avoid the appearance of $\partial_\lambda V^\lambda$ as an independently propagating scalar field, we take $\alpha = -\beta$, in which case $\partial_\lambda V^\lambda$ can be expressed in terms of an external current or other fields, as $\partial_\lambda J^\lambda/m^2$. The constant α can be absorbed in the definition of V_μ, so we can take $\alpha = -\beta = 1$, and therefore

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{2}m^2\,V_\mu V^\mu + J_\mu V^\mu\,, \qquad (7.5.11)$$

where

$$F_{\mu\nu} \equiv \partial_\mu V_\nu - \partial_\nu V_\mu\,. \qquad (7.5.12)$$

The derivative of the Lagrangian with respect to the time-derivative of the vector field is

$$\frac{\partial\mathscr{L}}{\partial\dot{V}_\mu} = -F^{0\mu}\,. \qquad (7.5.13)$$

This is non-vanishing for μ a spatial index i, so the V^i are canonical fields, with conjugates

$$\Pi^i = F^{i0} = \dot{V}^i + \partial_i V^0\,. \qquad (7.5.14)$$

On the other hand $F^{00} = 0$, so $\dot{V}^0$ does not appear in the Lagrangian, and V^0 is therefore an auxiliary field. This causes no serious difficulty: the fact that $\partial\mathscr{L}/\partial\dot{V}^0$ vanishes means that the field equation for V^0 involves no second time-derivatives, and can therefore be used as a constraint that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

eliminates a field variable. Specifically, the Euler–Lagrange equation for $v = 0$ is

$$\partial_i F^{i0} = m^2 V^0 - J^0 \tag{7.5.15}$$

or using Eq. (7.5.14)

$$V^0 = \frac{1}{m^2}(\nabla \cdot \boldsymbol{\Pi} + J^0)\,. \tag{7.5.16}$$

Now let us calculate the Hamiltonian $H = \int d^3x\,(\boldsymbol{\Pi}\cdot\dot{\mathbf{V}} - \mathscr{L})$ for this theory. Eq. (7.5.14) allows us to write $\dot{\mathbf{V}}$ in terms of $\boldsymbol{\Pi}$ and J^0:

$$\dot{\mathbf{V}} = -\nabla V^0 + \boldsymbol{\Pi} \;=\; \boldsymbol{\Pi} - \frac{1}{m^2}\nabla(\nabla\cdot\boldsymbol{\Pi} + J^0),$$

so

$$\begin{aligned} H = \int d^3x \Big[& \boldsymbol{\Pi}^2 + m^{-2}(\nabla\cdot\boldsymbol{\Pi})(\nabla\cdot\boldsymbol{\Pi} + J^0) \\ & - \tfrac{1}{2}\boldsymbol{\Pi}^2 + \tfrac{1}{2}(\nabla\times\mathbf{V})^2 + \tfrac{1}{2}m^2\mathbf{V}^2 \\ & - \tfrac{1}{2}m^{-2}(\nabla\cdot\boldsymbol{\Pi} + J^0)^2 - \mathbf{J}\cdot\mathbf{V} + m^{-2}J^0(\nabla\cdot\boldsymbol{\Pi} + J^0)\Big]\,. \end{aligned}$$

Again, we split this up into a free-particle term H_0 and interaction V:

$$H = H_0 + V\,, \tag{7.5.17}$$

and pass to the interaction picture by replacing the Heisenberg-picture quantities $\mathbf{V}$ and $\boldsymbol{\Pi}$ with their interaction-picture counterparts $\mathbf{v}$ and π (and, though not shown explicitly, likewise for whatever fields and conjugates are present in J^μ):

$$H_0 = \int d^3x \left[\frac{1}{2}\boldsymbol{\pi}^2 + \frac{1}{2m^2}(\nabla\cdot\boldsymbol{\pi})^2 + \frac{1}{2}(\nabla\times\mathbf{v})^2 + \frac{m^2}{2}\mathbf{v}^2\right], \tag{7.5.18}$$

$$V = \int d^3x \left[-\mathbf{J}\cdot\mathbf{v} + m^{-2}J^0\nabla\cdot\boldsymbol{\pi} + \frac{1}{2m^2}(J^0)^2\right]. \tag{7.5.19}$$

The relation between $\boldsymbol{\pi}$ and $\mathbf{v}$ is then

$$\dot{\mathbf{v}} = \frac{\delta H_0(\mathbf{v},\boldsymbol{\pi})}{\delta\boldsymbol{\pi}} = \boldsymbol{\pi} - m^{-2}\nabla(\nabla\cdot\boldsymbol{\pi}) \tag{7.5.20}$$

and the 'field equation' is

$$\dot{\boldsymbol{\pi}} = -\frac{\delta H_0(\mathbf{v},\boldsymbol{\pi})}{\delta\mathbf{v}} = +\nabla^2\mathbf{v} - \nabla(\nabla\cdot\mathbf{v}) - m^2\mathbf{v}\,. \tag{7.5.21}$$

Since V^0 is not an independent field variable, it is not related by a similarity transformation to any interaction-picture object v^0. Instead, we can *invent* a quantity

$$v^0 \equiv m^{-2}\nabla\cdot\boldsymbol{\pi}\,. \tag{7.5.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Eq. (7.5.20) then allows us to write $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi} = \dot{\mathbf{v}} + \nabla v^0 \,. \tag{7.5.23}$$

Inserting this in Eqs. (7.5.22) and (7.5.21) gives our field equations in the form

$$\nabla^2 v^0 + \nabla \cdot \dot{\mathbf{v}} - m^2 v^0 = 0 \,,$$
$$\nabla^2 \mathbf{v} - \nabla(\nabla \cdot \mathbf{v}) - \ddot{\mathbf{v}} - \nabla \dot{v}^0 - m^2 \mathbf{v} = 0 \,.$$

These can be combined in the covariant form

$$\Box v^\mu - \partial^\mu \partial_\nu v^\nu - m^2 v^\mu = 0 \,. \tag{7.5.24}$$

Taking the divergence gives

$$\partial_\mu v^\mu = 0 \tag{7.5.25}$$

and hence

$$(\Box - m^2) v^\mu = 0 \,. \tag{7.5.26}$$

A real vector field satisfying Eqs. (7.5.25) and (7.5.26) can be expressed as a Fourier transform

$$v^\mu(x) = (2\pi)^{-3/2} \sum_\sigma \int d^3p \, (2p^0)^{-1/2} \Big\{ e^\mu(\mathbf{p}, \sigma) a(\mathbf{p}, \sigma) e^{ip \cdot x} + e^{\mu *}(\mathbf{p}, \sigma) a^\dagger(\mathbf{p}, \sigma) \, e^{-ip \cdot x} \Big\} \,, \tag{7.5.27}$$

where $p^0 = \sqrt{\mathbf{p}^2 + m^2}$; the $e^\mu(\mathbf{p}, \sigma)$ for $\sigma = +1, 0, -1$ are three independent vectors satisfying

$$p_\mu e^\mu(\mathbf{p}, \sigma) = 0 \tag{7.5.28}$$

and normalized so that

$$\sum_\sigma e^\mu(\mathbf{p}, \sigma) e^{\nu *}(\mathbf{p}, \sigma) = \eta^{\mu\nu} + p^\mu p^\nu / m^2 \quad ; \tag{7.5.29}$$

and the $a(\mathbf{p}, \sigma)$ are operator coefficients. It is straightforward using Eqs. (7.5.23), (7.5.27), and (7.5.29) to calculate that $\mathbf{v}$ and $\boldsymbol{\pi}$ satisfy the correct commutation relations

$$\Big[v^i(\mathbf{x}, t), \, \pi^j(\mathbf{y}, t) \Big] = i \, \delta_{ij} \, \delta^3(\mathbf{x} - \mathbf{y}) \,,$$
$$\Big[v^i(\mathbf{x}, t), \, v^j(\mathbf{y}, t) \Big] = \Big[\pi^i(\mathbf{x}, t), \, \pi^j(\mathbf{y}, t) \Big] = 0 \,, \tag{7.5.30}$$

provided that $a(\mathbf{p}, \sigma)$ and $a^\dagger(\mathbf{p}, \sigma)$ satisfy the commutation relations

$$\Big[a(\mathbf{p}, \sigma), \, a^\dagger(\mathbf{p}', \sigma') \Big] = \delta^3(\mathbf{p}' - \mathbf{p}) \delta_{\sigma' \sigma} \,, \tag{7.5.31}$$

$$\Big[a(\mathbf{p}, \sigma), \, a(\mathbf{p}', \sigma') \Big] = 0 \,. \tag{7.5.32}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

We already know that the vector field for a spin one particle must take the form (7.5.27), so our derivation of these results serves to verify that Eq. (7.5.18) gives the correct free-particle Hamiltonian for a massive particle of spin one. It is easy to check also that Eq. (7.5.18) may be written (up to a constant term) in the standard form of a free-particle energy, as $\sum_\sigma \int d^3p\, p^0\, a^\dagger(\mathbf{p},\sigma)a(\mathbf{p},\sigma)$. Finally, using Eq. (7.5.22) in Eq. (7.5.19) yields the interaction in the interaction picture

$$V(t) = \int d^3x \left[-J_\mu v^\mu + \frac{1}{2m^2}(J^0)^2\right] . \tag{7.5.33}$$

The extra non-invariant term in Eq. (7.5.33) is just what we found in Chapter 6 is needed to cancel a non-invariant term in the propagator of the vector field.

Dirac field, Spin One Half

For the Dirac field of a particle of spin 1/2, we tentatively take the Lagrangian as

$$\mathscr{L} = -\bar{\Psi}(\gamma^\mu\partial_\mu + m)\Psi - \mathscr{H}(\bar{\Psi},\Psi) \tag{7.5.34}$$

with $\mathscr{H}$ a real function of $\bar{\Psi}$ and Ψ. This is not real, but the action is, because

$$\bar{\Psi}\gamma^\mu\partial_\mu\Psi - (\bar{\Psi}\gamma^\mu\partial_\mu\Psi)^\dagger = \bar{\Psi}\gamma^\mu\partial_\mu\Psi + (\partial_\mu\bar{\Psi})\gamma^\mu\Psi = \partial_\mu(\bar{\Psi}\gamma^\mu\Psi) .$$

Hence the field equations obtained by requiring the action to be stationary with respect to $\bar{\Psi}$ are the adjoints of those obtained by requiring the action to be stationary with respect to Ψ, as necessary if we are to avoid having too many field equations. The canonical conjugate to Ψ is

$$\Pi = \frac{\partial\mathscr{L}}{\partial\dot{\Psi}} = -\bar{\Psi}\gamma^0 , \tag{7.5.35}$$

so we should not regard $\bar{\Psi}$ as a field like Ψ, but rather as proportional to the canonical conjugate of Ψ. The Hamiltonian is

$$H = \int d^3x[\Pi\dot{\Psi} - \mathscr{L}] = \int d^3x\left[\Pi\gamma^0[\boldsymbol{\gamma}\cdot\boldsymbol{\nabla} + m]\Psi + \mathscr{H}\right] .$$

We write this as

$$H = H_0 + V , \tag{7.5.36}$$

where

$$H_0 = \int d^3x\, \Pi\gamma^0[\boldsymbol{\gamma}\cdot\boldsymbol{\nabla} + m]\Psi , \tag{7.5.37}$$

$$V = \int d^3x\, \mathscr{H}(\bar{\Psi},\Psi) . \tag{7.5.38}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

We now pass to the interaction picture. Since Eq. (7.5.35) does not involve the time, the similarity transformation (7.1.28), (7.1.29) yields immediately

$$\pi = -\bar{\psi}\,\gamma^0 \,. \tag{7.5.39}$$

Likewise, H_0 and $V(t)$ can be calculated by replacing Ψ and Π with ψ and π in Eqs. (7.5.37) and (7.5.38). This gives the equation of motion

$$\dot{\psi} = \frac{\delta H_0}{\delta \pi} = \gamma^0(\boldsymbol{\gamma}\cdot\boldsymbol{\nabla} + m)\psi \tag{7.5.40}$$

or more neatly

$$(\gamma^\mu \partial_\mu + m)\psi = 0 \,. \tag{7.5.41}$$

(The other equation of motion, $\dot{\pi} = -\delta H_0 \big/ \delta\psi$, yields just the adjoint of this one.) Any field satisfying Eq. (7.5.41) can be written as a Fourier transform

$$\psi(x) = (2\pi)^{-3/2} \int d^3p \sum_\sigma \left\{ u(\mathbf{p},\sigma)\, e^{ip\cdot x} a(\mathbf{p},\sigma) + \; v(\mathbf{p},\sigma) e^{-ip\cdot x} b^\dagger(\mathbf{p},\sigma) \right\} , \tag{7.5.42}$$

where $p^0 \equiv \sqrt{\mathbf{p}^2 + m^2}$; $a(\mathbf{p},\sigma)$ and $b^\dagger(\mathbf{p},\sigma)$ are operator coefficients; and $u(\mathbf{p}, \pm\frac{1}{2})$ are the two independent solutions of

$$(i\gamma^\mu p_\mu + m)\, u(\mathbf{p},\sigma) = 0 \tag{7.5.43}$$

and likewise

$$(-i\gamma^\mu p_\mu + m)\, v(\mathbf{p},\sigma) = 0 \tag{7.5.44}$$

normalized so that*

$$\sum_\sigma u(\mathbf{p},\sigma)\bar{u}(\mathbf{p},\sigma) = \frac{(-i\gamma^\mu p_\mu + m)}{2p^0} \,, \tag{7.5.45}$$

$$\sum_\sigma v(\mathbf{p},\sigma)\bar{v}(\mathbf{p},\sigma) = -\frac{(i\gamma^\mu p_\mu + m)}{2p^0} \,. \tag{7.5.46}$$

In order to obtain the desired anticommutators

$$\begin{aligned} \left[\psi_\alpha(\mathbf{x},t),\, \bar{\psi}_\beta(\mathbf{y},t)\right]_+ &= \left[\psi_\alpha(\mathbf{x},t),\, \pi_\gamma(\mathbf{y},t)\right]_+ (\gamma^0)_{\gamma\beta} \\ &= i\,(\gamma^0)_{\alpha\beta}\delta^3(\mathbf{x}-\mathbf{y}) \,, \end{aligned} \tag{7.5.47}$$

$$\left[\psi_\alpha(\mathbf{x},t),\, \psi_\beta(\mathbf{y},t)\right]_+ = 0 \,, \tag{7.5.48}$$

* The matrix $i\gamma^\mu p_\mu$ has eigenvalues $\pm m$, so $\Sigma u\bar{u}$ and $\Sigma v\bar{v}$ must be proportional to the projection matrices $(-i\gamma^\mu p_\mu + m)/2m$ and $(i\gamma^\mu p_\mu + m)/2m$, respectively. The proportionality factor may be adjusted up to a sign by absorbing it in the definition of u and v. The overall sign is determined by positivity: $\mathrm{Tr}\,\Sigma u\bar{u}\beta = \Sigma u^\dagger u$ and $\mathrm{Tr}\,\Sigma v\bar{v}\beta = \Sigma v^\dagger v$ must be positive.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

we must adopt the anticommutation relations

$$\left[a(\mathbf{p},\sigma),\, a^\dagger(\mathbf{p}',\sigma')\right]_+ = \left[b(\mathbf{p},\sigma),\, b^\dagger(\mathbf{p}',\sigma')\right]_+ = \delta^3(\mathbf{p}'-\mathbf{p})\delta_{\sigma'\sigma}\,, \tag{7.5.49}$$

$$\left[a(\mathbf{p},\sigma),\, a(\mathbf{p}',\sigma')\right]_+ = \left[b(\mathbf{p},\sigma),\, b(\mathbf{p}',\sigma')\right]_+ =$$

$$\left[a(\mathbf{p},\sigma),\, b(\mathbf{p}',\sigma')\right]_+ = \left[a(\mathbf{p},\sigma),\, b^\dagger(\mathbf{p}',\sigma')\right]_+ = 0\,, \tag{7.5.50}$$

and their adjoints. These agree with the results obtained in Chapter 5, thus verifying that (7.5.37) is the correct free-particle Hamiltonian for spin $\frac{1}{2}$. In terms of the as and bs, this Hamiltonian is

$$H_0 = \sum_\sigma \int d^3p\; p^0 \Big(a^\dagger(\mathbf{p},\sigma)a(\mathbf{p},\sigma) - b(\mathbf{p},\sigma)b^\dagger(\mathbf{p},\sigma)\Big)\,. \tag{7.5.51}$$

We can rewrite this as a more conventional free-particle Hamiltonian, plus another infinite c-number**

$$H_0 = \sum_\sigma \int d^3\mathbf{p}\, p^0 \Big[a^\dagger(\mathbf{p},\sigma)a(\mathbf{p},\sigma) + b^\dagger(\mathbf{p},\sigma)b(\mathbf{p},\sigma) - \delta^3(\mathbf{p}-\mathbf{p})\Big]\,. \tag{7.5.52}$$

The c-number term in Eq. (7.5.52) is only important if we worry about gravitational phenomena; otherwise here, as for the scalar field, we can throw it away, since it only affects the zero of energy with respect to which all energies are measured. With this understanding, H_0 is a positive operator, just as for bosons.

7.6 Constraints and Dirac Brackets

The chief obstacle to deriving the Hamiltonian from the Lagrangian is the occurrence of constraints. The standard analysis of this problem is that of Dirac,[5] whose terminology we will follow here. Dirac's analysis is not really needed for the simple theories discussed in this chapter, where it is easy to identify the unconstrained canonical variables. We shall use the theory of a real massive vector field for illustration here, returning to Dirac's approach in the next chapter, where it will be actually useful.

Primary constraints are either imposed on the system (as when in the next chapter we choose a gauge for the electromagnetic field) or arise from the structure of the Lagrangian itself. For an example of the latter type, consider the Lagrangian (7.5.11) of a massive vector field V^μ interacting

** Note the negative sign of the c-number term. The conjectured symmetry known as *supersymmetry*[4] connects the numbers of boson and fermion fields, in such a way that the c-numbers in H_0 all cancel.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

with a current J_μ:

$$\mathscr{L} = -\tfrac{1}{4} F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{2}m^2 V_\mu V^\mu + J_\mu V^\mu \tag{7.6.1}$$

where

$$F_{\mu\nu} \equiv \partial_\mu V_\nu - \partial_\nu V_\mu \,. \tag{7.6.2}$$

Suppose we try to treat all four components of V^μ on the same basis. We should then define the conjugates

$$\Pi^\mu \equiv \frac{\partial \mathscr{L}}{\partial(\partial_0 V_\mu)} = -F^{0\mu} \,. \tag{7.6.3}$$

We immediately find the primary constraint:

$$\Pi_0 = 0 \,. \tag{7.6.4}$$

More generally, we encounter primary constraints whenever the equations $\Pi_\ell = \delta L/\delta\partial_0\Psi^\ell$ cannot be solved to give all the $\partial_0\Psi^\ell$ (at least locally) in terms of Π_ℓ and Ψ^ℓ. This will be the case if and only if the matrix $\delta^2 L/\delta(\partial_0\Psi^\ell)\,\delta(\partial_0\Psi^m)$ has vanishing determinant. Such Lagrangians are called *irregular*.

Then there are *secondary constraints*, which arise from the requirement that the primary constraints be consistent with the equations of motion. For the massive vector field, this is just the Euler–Lagrange equation (7.5.16) for V^0:

$$\partial_i \Pi_i = m^2 V^0 - J^0 \,. \tag{7.6.5}$$

Here we are finished, but in other theories we might encounter further constraints by requiring consistency of the secondary constraints with the field equations, and so on. The distinction between primary, secondary, etc. constraints is not important; we will treat them all together here.

There is another distinction between certain types of constraint that is more important. The constraints we have found for the massive vector field are of a type known as *second class*, for which there is a universal prescription for the commutation relations. To explain the distinction between first and second class constraints, and the prescription used to deal with second class constraints, it is useful first to recall the definition of the Poisson brackets of classical mechanics.

Consider any Lagrangian $L(\Psi, \dot{\Psi})$ that depends on a set of variables $\Psi^a(t)$ and their time-derivatives $\dot{\Psi}^a(t)$. (The Lagrangians of quantum field theory are a special case, with the index a running over all pairs of ℓ and $\mathbf{x}$.) We can define canonical conjugates for *all* of these variables by

$$\Pi_a \equiv \frac{\partial L}{\partial \dot{\Psi}^a} \,. \tag{7.6.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

The Πs and Ψs will in general not be independent variables, but may instead be related by various constraint equations, both primary and secondary. The Poisson bracket is then defined by

$$[A,B]_{\rm P} \equiv \frac{\partial A}{\partial \Psi^a}\frac{\partial B}{\partial \Pi_a} - \frac{\partial B}{\partial \Psi^a}\frac{\partial A}{\partial \Pi_a} \tag{7.6.7}$$

with the constraints ignored in calculating the derivatives with respect to Ψ^a and Π_a. In particular, we always have $[\Psi^a,\Pi_b]_{\rm P} = \delta^a_b$. (Here and below all fields are taken at the same time, and time arguments are everywhere dropped.) These brackets have the same algebraic properties as commutators:

$$[A,B]_{\rm P} = -[B,A]_{\rm P}, \tag{7.6.8}$$

$$[A,BC]_{\rm P} = [A,B]_{\rm P}C + B[A,C]_{\rm P}, \tag{7.6.9}$$

including the Jacobi identity

$$[A,[B,C]_{\rm P}]_{\rm P} + [B,[C,A]_{\rm P}]_{\rm P} + [C,[A,B]_{\rm P}]_{\rm P} = 0\,. \tag{7.6.10}$$

If we could adopt the usual commutation relations $[\Psi^a,\Pi_b] = i\delta^a_b$, $[\Psi^a,\Psi^b] = [\Pi_a,\Pi_b] = 0$, then the commutator of any two functions of the Ψs and Πs would be just $[A,B] = i[A,B]_{\rm P}$. But the constraints do not always allow this.

The constraints may in general be expressed in the form $\chi_N = 0$, where the χ_N are a set of functions of the Ψs and Πs. Because we are including secondary constraints along with the primary constraints, the set of all the constraints is necessarily consistent with the equations of motion $\dot{A} = [A,H]_{\rm P}$, and therefore

$$[\chi_N,H]_{\rm P} = 0 \tag{7.6.11}$$

when the constraint equations $\chi_N = 0$ are satisfied.

We call a constraint *first class* if its Poisson bracket with all the other constraints vanishes when (*after* calculating the Poisson brackets) we impose the constraints. We shall see a simple example of such a constraint in the quantization of the electromagnetic field in the next chapter, where the first class constraint arises from a symmetry of the action, electromagnetic gauge invariance. In fact, the set of first class constraints $\chi_N = 0$ is always associated with a group of symmetries, under which an arbitrary quantity A undergoes the infinitesimal transformation

$$\delta_N A \equiv \sum_N \epsilon_N\,[\chi_N,A]_{\rm P}\,. \tag{7.6.12}$$

(In field theory these are local transformations, because the index N con-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

tains a spacetime coordinate.) Eq. (7.6.11) shows that this transformation leaves the Hamiltonian invariant, and for first class constraints it also respects all other constraints. Such first class constraints can be eliminated by a choice of gauge.

After all of the first class constraints have been eliminated by a choice of gauge, the remaining constraint equations $\chi_N = 0$ are such that no linear combination $\sum_N u_N[\chi_N, \chi_M]_P$ of the Poisson brackets of these constraints with each other vanishes. It follows that the matrix of the Poisson brackets of the remaining constraints is non-singular:

$$\text{Det}\, C \neq 0\,, \tag{7.6.13}$$

where

$$C_{NM} \equiv [\chi_N, \chi_M]_P\,. \tag{7.6.14}$$

Constraints of this sort are called *second class.* Note that there must always be an even number of second class constraints, because an antisymmetric matrix of odd dimensionality necessarily has vanishing determinant.

As we have seen, in the case of the massive real vector field the constraints are

$$\chi_{1\mathbf{x}} = \chi_{2\mathbf{x}} = 0\,, \tag{7.6.15}$$

where

$$\chi_{1\mathbf{x}} = \Pi_0(\mathbf{x})\,, \qquad \chi_{2\mathbf{x}} = \partial_i \Pi_i(\mathbf{x}) - m^2 V^0(\mathbf{x}) - J^0(\mathbf{x})\,. \tag{7.6.16}$$

The Poisson bracket of these constraints is

$$C_{1\mathbf{x},2\mathbf{y}} = -C_{2\mathbf{y},1\mathbf{x}} = [\chi_{1\mathbf{x}}, \chi_{2\mathbf{y}}]_P = m^2 \delta^3(\mathbf{x} - \mathbf{y}) \tag{7.6.17}$$

and, of course,

$$C_{1\mathbf{x},1\mathbf{y}} = C_{2\mathbf{x},2\mathbf{y}} = 0\,. \tag{7.6.18}$$

This 'matrix' is obviously non-singular, so the constraints (7.6.15) are second class.

Dirac suggested that when all constraints are second class, the commutation relations will be given by

$$[A, B] = i[A, B]_D\,, \tag{7.6.19}$$

where $[A, B]_D$ is a generalization of the Poisson bracket known as the *Dirac bracket*:

$$[A, B]_D \equiv [A, B]_P - [A, \chi_N]_P\, (C^{-1})^{NM}\, [\chi_M, B]_P\,. \tag{7.6.20}$$

(Here N and M are compound indices including the position in space,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

taking values like $1, \mathbf{x}$ and $2, \mathbf{x}$ in the vector field example.) He noted that the Dirac bracket like the Poisson bracket satisfies the same algebraic relations as the commutators

$$[A, B]_D = -[B, A]_D , \tag{7.6.21}$$

$$[A, BC]_D = [A, B]_D C + B[A, C]_D , \tag{7.6.22}$$

$$[A, [B, C]_D]_D + [B, [C, A]_D]_D + [C, [A, B]_D]_D = 0 , \tag{7.6.23}$$

and also the relations

$$[\chi_N, B]_D = 0 \tag{7.6.24}$$

which make the commutation relations (7.6.19) consistent with the constraints $\chi_N = 0$. Also, the Dirac brackets are unchanged if we replace the χ_N with any functions χ'_N for which the equations $\chi'_N = 0$ and $\chi_N = 0$ define the same submanifold of phase space. But all these agreeable properties do not prove that the commutators are actually given by Eq. (7.6.19) in terms of the Dirac brackets.

This issue is illuminated if not settled by a powerful theorem proved by Maskawa and Nakajima.[6] They showed that for any set of canonical variables Ψ^a, Π_a governed by second class constraints, it is always possible by a canonical transformation* to construct two sets of variables Q^n, $\mathscr{Q}^r$ and their respective conjugates P_n, $\mathscr{P}_r$, such that the constraints read $\mathscr{Q}^r = \mathscr{P}_r = 0$. Using these coordinates to calculate Poisson brackets, and redefining the constraint functions as $\chi_{1r} = \mathscr{Q}^r$, $\chi_{2r} = \mathscr{P}_r$, we have

$$C_{1r,2s} = [\mathscr{Q}^r, \mathscr{P}_s]_P = \delta^r_s ,$$

$$C_{1r,1s} = [\mathscr{Q}^r, \mathscr{Q}^s]_P = 0 , \qquad C_{2r,2s} = [\mathscr{P}_r, \mathscr{P}_s]_P = 0 ,$$

and for any functions A, B

$$[A, \chi_{1r}]_P = -\frac{\partial A}{\partial \mathscr{P}_r} , \qquad [A, \chi_{2r}]_P = \frac{\partial A}{\partial \mathscr{Q}^r} ,$$

This C-matrix has inverse $C^{-1} = -C$, so the Dirac brackets (7.6.20) are

* Recall that by a canonical transformation, we mean a transformation from a set of phase space coordinates Ψ^a, Π_a to some other phase space coordinates $\tilde{\Psi}^a$, $\tilde{\Pi}_a$, such that $[\tilde{\Psi}^a, \tilde{\Pi}_b]_P = \delta^a_b$ and $[\tilde{\Psi}^a, \tilde{\Psi}^b]_P = [\tilde{\Pi}_a, \tilde{\Pi}_b]_P = 0$, the Poisson brackets being calculated in terms of the Ψ^a and Π_a. It follows that the Poisson brackets for any functions A, B are the same whether calculated in terms of Ψ^a and Π_a or in terms of $\tilde{\Psi}^a$ and $\tilde{\Pi}_a$. It also follows that if Ψ^a and Π_a satisfy the Hamiltonian equations of motion, then so do $\tilde{\Psi}^a$ and $\tilde{\Pi}_a$, with the same Hamiltonian. The Lagrangian is changed by a canonical transformation, but only by a time-derivative, which does not affect the action.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

here

$$\begin{aligned}[A,B]_D &= [A,B]_P + [A,\chi_{1r}]_P[\chi_{2r},B]_P - [A,\chi_{2r}]_P[\chi_{1r},B]_P \\ &= [A,B]_P - \frac{\partial A}{\partial \mathscr{Q}^r}\frac{\partial B}{\partial \mathscr{P}_r} + \frac{\partial B}{\partial \mathscr{Q}^r}\frac{\partial A}{\partial \mathscr{P}_r} \\ &= \frac{\partial A}{\partial Q^n}\frac{\partial B}{\partial P_n} - \frac{\partial B}{\partial Q^n}\frac{\partial A}{\partial P_n}\,. \end{aligned} \tag{7.6.25}$$

In other words, *the Dirac bracket is equal to the Poisson bracket calculated in terms of the reduced set of unconstrained canonical variables* Q^n, P_n. If we assume that these unconstrained variables satisfy the canonical commutation relations, then the commutators of general operators A, B are given by Eq. (7.6.19) in terms of the Dirac brackets.**

We now return to the massive vector field, to see how it can be quantized using Dirac brackets. This is a case where it is easy to express the constrained variables V^0 and Π_0 in terms of the unconstrained ones† V_i and Π_i; we have simply $\Pi_0 = 0$, and V^0 is given by Eq. (7.6.5). From Eqs. (7.6.17) and (7.6.18), we see that C_{NM} here has the inverse

$$(C^{-1})^{1\mathbf{x},2\mathbf{y}} = -(C^{-1})^{2\mathbf{y},1\mathbf{x}} = -m^{-2}\delta^3(\mathbf{x}-\mathbf{y}), \tag{7.6.26}$$

$$(C^{-1})^{1\mathbf{x},1\mathbf{y}} = (C^{-1})^{2\mathbf{x},2\mathbf{y}} = 0\,. \tag{7.6.27}$$

Therefore the Dirac prescription (7.6.19), (7.6.20) yields the equal-time commutators

$$\begin{aligned}[A,B] &= i[A,B]_{\mathrm{P}} \\ &\quad + im^{-2}\int d^3z\left([A,\Pi_0(\mathbf{z})]_{\mathrm{P}}\,[\partial_i\Pi_i(\mathbf{z}) - m^2V^0(\mathbf{z}) - J^0(\mathbf{z}),B]_{\mathrm{P}} - A\leftrightarrow B\right). \end{aligned} \tag{7.6.28}$$

By definition, we have

$$[V^\mu(\mathbf{x}),\Pi_\nu(\mathbf{y})]_{\mathrm{P}} = \delta^3(\mathbf{x}-\mathbf{y})\delta^\mu_\nu\,, \quad [V^\mu(\mathbf{x}),V^\nu(\mathbf{y})]_{\mathrm{P}} = [\Pi_\mu(\mathbf{x}),\Pi_\nu(\mathbf{y})]_{\mathrm{P}} = 0\,. \tag{7.6.29}$$

Hence

$$[V^i(\mathbf{x}),V^j(\mathbf{y})] = [V^0(\mathbf{x}),V^0(\mathbf{y})] = 0\,,$$

** It is still an open question whether we should adopt canonical commutation relations for the unconstrained variables Q^n, P_n constructed by the Maskawa-Nakajima canonical transformation. Ultimately, the test of such canonical commutation relations is their consistency with the free-field commutation relations derived in Chapter 5, but to apply this test we need to know what the Q^n and P_n are. In the Appendix to this chapter we display two large classes of theories in which we can identify a set of unconstrained Qs and Ps, such that the Dirac commutation relations (7.6.19) follow from the ordinary canonical commutation relations of the Qs and Ps. We shall also show that in these cases, the Hamiltonian defined in terms of the unconstrained Ψs and Πs may be written just as well in terms of the constrained variables.

† This is a special case of the theories discussed in Part A of the Appendix.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

$$
\begin{aligned}
[V^i(\mathbf{x}), V^0(\mathbf{y})] &= -im^{-2}\partial_i\delta^3(\mathbf{x}-\mathbf{y})\,, \\
[V^i(\mathbf{x}), \Pi_j(\mathbf{y})] &= i\delta^i_j\delta^3(\mathbf{x}-\mathbf{y})\,, \\
[V^0(\mathbf{x}), \Pi_j(\mathbf{y})] &= [V^\mu(\mathbf{x}), \Pi_0(\mathbf{y})] = 0\,, \\
[\Pi^\mu(\mathbf{x}), \Pi^\nu(\mathbf{y}] &= 0\,.
\end{aligned}
\tag{7.6.30}
$$

These are indeed just the commutation relations that we would find by assuming that the unconstrained variables satisfy the usual canonical commutation relations $[V^i(\mathbf{x}), \Pi_j(\mathbf{y})] = i\delta^i_j\delta^3(\mathbf{x}-\mathbf{y})$, $[V^i(\mathbf{x}), V^j(\mathbf{y})] = [\Pi_i(\mathbf{x}), \Pi_j(\mathbf{y})] = 0$, and using the constraints to evaluate the commutators involving Π_0 and V^0.

7.7 Field Redefinitions and Redundant Couplings*

Observables like masses and S-matrix elements are independent of some of the coupling parameters in any action, known as the *redundant* parameters. This is because changes in these parameters can be undone by simply redefining the field variables. A continuous redefinition of the fields, such as an infinitesimal local transformation $\Psi^\ell(x) \to \Psi^\ell(x) + \epsilon F^\ell(\Psi(x), \partial_\mu\Psi(x), \cdots)$, clearly cannot affect any *observable* of the theory,** though, of course, it would change the values of matrix elements of the fields themselves.

How can we tell whether some variation in the parameters of a theory can be cancelled by a field redefinition? A continuous local field redefinition will produce a change in the action of the form

$$
\delta I[\Psi] = \epsilon \sum_\ell \int d^4x \, \frac{\delta I[\Psi]}{\delta\Psi^\ell(x)} \, F^\ell(\Psi(x), \partial\Psi(x), \cdots) \quad . \tag{7.7.1}
$$

So any change δg_i in the coupling parameters g_i, for which the change in the action is of the form

$$
\sum_i \frac{\partial I}{\partial g_i}\,\delta g_i = -\epsilon \sum_\ell \int d^4x \, \frac{\delta I[\Psi]}{\delta\Psi^\ell(x)} \, F^\ell(\Psi(x), \partial\Psi(x), \cdots)\,, \tag{7.7.2}
$$

may be compensated by a field redefinition

$$
\Psi^\ell(x) \to \Psi^\ell(x) + \epsilon F^\ell(\Psi(x), \partial_\mu\Psi(x), \cdots)\,,
$$

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** For instance, the theorem of Section 10.2 shows that as long as we multiply by the correct field renormalization constants, S-matrix elements can be obtained from the vacuum expectation value of a time-ordered product of *any* operators that have non-vanishing matrix elements between the vacuum and the one-particle states of the particles participating in the reaction.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

and therefore can have no effect on any observables. In other words, *a coupling parameter is redundant if the change in the action when we vary this parameter vanishes when we use the field equations* $\delta I/\delta\Psi^i = 0$.

For example, suppose we write the Lagrangian density of a scalar field theory in the form

$$\mathscr{L} = -\tfrac{1}{2}Z\,(\partial^\mu\Phi\,\partial_\mu\Phi + m^2\Phi^2) - \tfrac{1}{24}gZ^2\Phi^4\,.$$

The constant Z is a redundant coupling, because

$$\frac{\partial I}{\partial Z} = \tfrac{1}{2}\int d^4x\,\Phi(\Box\Phi - m^2\Phi - \tfrac{1}{6}gZ\Phi^3)\,,$$

and this vanishes when we use the field equation

$$\Box\Phi - m^2\Phi = \tfrac{1}{6}gZ\Phi^3\,.$$

On the other hand, neither the bare mass m nor the coupling g are redundant, and no function of m and g is redundant.

In this example, the field redefinition needed to compensate for a change of Z is a simple rescaling, in which F is proportional to Φ. (For this reason Z is called a field renormalization constant.) This is the most general field transformation that leaves the general form of this action invariant. But for the more general actions considered in Sections 12.3 and 12.4, with arbitrary numbers of fields and derivatives, we would have to consider non-linear as well as linear field redefinitions, and an infinite subset of the parameters of the theory would be redundant.

Appendix Dirac Brackets from Canonical Commutators

In this Appendix we shall show, in theories of two types, that the formula giving commutators as Dirac brackets times i follows from the usual canonical commutation relations for a reduced set of variables.

A

Suppose (as in the case of a massive vector field V^μ) that the quantum variables Ψ^a and Π_a appearing in the Lagrangian L may be divided into two classes:* one set Q^n of independent canonical variables (like $V^i(\mathbf{x})$) with independent canonical conjugates $P_n = \partial L/\partial\dot{Q}^n$; and another set

* We are again using a compact notation, in which labels like a, n, and r include a space coordinate $\mathbf{x}$ as well as discrete indices. Repeated labels are summed and integrated. All quantum variables are understood to be evaluated at the same time, with the common time argument dropped everywhere. The quantities $\mathscr{Q}^r$ are the same as the C^r introduced in Section 7.2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

$\mathscr{Q}^r(\mathbf{x})$ (like V^0) whose time-derivatives do not appear in the Lagrangian. The primary constraints are the conditions $\chi_{1r} = 0$, where

$$\chi_{1r} = \mathscr{P}_r \tag{7.A.1}$$

are the variables conjugate to the $\mathscr{Q}^r$. The secondary constraints arise from the equations of motion $0 = \partial L/\partial \mathscr{Q}^r$ for the $\mathscr{Q}^r$; we suppose that these constraints can be 'solved' — i.e., they may be written in the form $\chi_{2r} = 0$, with χ_{2r} in the form

$$\chi_{2r} = \mathscr{Q}^r - f^r(Q, P)\,. \tag{7.A.2}$$

(An example is provided by Eq. (7.6.5), which gives V^0 in terms of the independent Ps (here, the Π_i) and Qs.) We assume that the independent Qs and Ps satisfy the usual canonical commutation rules:

$$[Q^n, P_m] = i\delta^n_m\,, \qquad [Q^n, Q^m] = [P_n, P_m] = 0\,. \tag{7.A.3}$$

The constraint $\chi_{2r} = 0$ yields the commutators involving $\mathscr{Q}$:

$$[\mathscr{Q}^r, Q^n] = -i\frac{\partial f^r}{\partial P_n}\,, \qquad [\mathscr{Q}^r, P_n] = i\frac{\partial f^r}{\partial Q^n}\,, \tag{7.A.4}$$

$$[\mathscr{Q}^r, \mathscr{Q}^s] = i\Gamma^{rs}\,, \tag{7.A.5}$$

where Γ^{rs} is the Poisson bracket

$$\Gamma^{rs} \equiv [f^r, f^s]_{\mathrm{P}}\,, \tag{7.A.6}$$

and, of course, all commutators involving $\mathscr{P}_r$ vanish:

$$[\mathscr{P}_r, Q^n] = [\mathscr{P}_r, P_n] = [\mathscr{P}_r, \mathscr{Q}^s] = [\mathscr{P}_r, \mathscr{P}_s] = 0\,. \tag{7.A.7}$$

Now let us compare these commutators with the Dirac brackets. The Poisson brackets of the constraint functions are

$$C_{1r,1s} \equiv [\chi_{1r}, \chi_{1s}]_{\mathrm{P}} = 0\,, \tag{7.A.8}$$

$$C_{1r,2s} \equiv -C_{2s,1r} \equiv [\chi_{1r}, \chi_{2s}]_{\mathrm{P}} = -\delta^s_r\,, \tag{7.A.9}$$

$$C_{2r,2s} \equiv [\chi_{2r}, \chi_{2s}]_{\mathrm{P}} = [f^r(Q,P), f^s(Q,P)]_{\mathrm{P}} \equiv \Gamma^{rs}\,. \tag{7.A.10}$$

(In the example of the massive vector field Γ^{rs} vanishes, but the discussion here will apply also for non-vanishing Γ^{rs}.) It is easy to see that the C-matrix has the inverse

$$\begin{aligned} (C^{-1})^{1r,1s} &= \Gamma^{rs}\,, \qquad (C^{-1})^{2r,2s} = 0\,, \\ (C^{-1})^{1r,2s} &= -(C^{-1})^{2s,1r} = \delta^r_s\,. \end{aligned} \tag{7.A.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Also, the Poisson brackets of any function A with the constraint functions are

$$[A, \chi_{1r}]_{\mathrm{P}} = \frac{\partial A}{\partial \mathcal{Q}^r}, \qquad [A, \chi_{2r}]_{\mathrm{P}} = -\frac{\partial A}{\partial \mathcal{P}_r} - [A, f^r(Q,P)]_{\mathrm{P}}.$$

Hence the Dirac bracket is

$$[A,B]_{\mathrm{D}} = [A,B]_{\mathrm{P}} - \frac{\partial A}{\partial \mathcal{Q}^r}\frac{\partial B}{\partial \mathcal{P}_r} + \frac{\partial B}{\partial \mathcal{Q}^r}\frac{\partial A}{\partial \mathcal{P}_r} + \frac{\partial A}{\partial \mathcal{Q}^r}\Gamma^{rs}\frac{\partial B}{\partial \mathcal{Q}^s} - \frac{\partial A}{\partial \mathcal{Q}^r}[B, f^r]_{\mathrm{P}} + [A, f^r]_{\mathrm{P}}\frac{\partial B}{\partial \mathcal{Q}^r}. \quad (7.A.12)$$

Now, if A and B are both functions only of the independent canonical variables Q^n and P_n, then $\partial A/\partial \mathcal{Q}^r = \partial B/\partial \mathcal{Q}^r = 0$, so the Dirac bracket is equal to the Poisson bracket. In particular,

$$[Q^n, P_m]_{\mathrm{D}} = \delta^n_m, \qquad [Q^n, Q^m]_{\mathrm{D}} = [P_n, P_m]_{\mathrm{D}} = 0. \quad (7.A.13)$$

Where A is $\mathcal{Q}^r$ and B is a function of Qs and Ps, it is only the fifth term on the right-hand side of Eq. (7.A.12) that contributes. In particular

$$[\mathcal{Q}^r, Q^n]_{\mathrm{D}} = -\frac{\partial f^r}{\partial P_n}, \qquad [\mathcal{Q}^r, P_n]_{\mathrm{D}} = +\frac{\partial f^r}{\partial Q^n}. \quad (7.A.14)$$

Where both A and B are $\mathcal{Q}$s, we have only the fourth term

$$[\mathcal{Q}^r, \mathcal{Q}^s]_{\mathrm{D}} = \Gamma^{rs}. \quad (7.A.15)$$

Finally, where A is $\mathcal{P}_r$ and B is anything, we have only the first and third terms, which cancel:

$$[\mathcal{P}_r, B]_{\mathrm{D}} = [\mathcal{P}_r, B]_{\mathrm{P}} + \frac{\partial B}{\partial \mathcal{Q}^r} = 0. \quad (7.A.16)$$

Comparison of Eqs. (7.A.13)–(7.A.16) with Eqs. (7.A.3)–(7.A.7) shows that in all cases, the commutators are equal to the Dirac brackets times i. This is only to be expected, because as remarked in Section 7.6 all Dirac brackets involving the constraint functions vanish, so the Dirac brackets involving $\mathcal{Q}^r$ and/or $\mathcal{P}_s$ are given by using the constraint equations to express $\mathcal{Q}^r$ and/or $\mathcal{P}_s$ in terms of the independent Qs and Ps.

B

Next consider the case where the constraints take the form of conditions $\chi_{1r}(\Psi) = 0$ on the Ψ^a, which can be solved by expressing them in terms of a smaller set of unconstrained variables Q^n, and an equal number of separate conditions $\chi_{2r}(\Pi) = 0$ on the Π_a, which can be solved by expressing the Π_a in terms of a smaller set of unconstrained P_n. (We will see an example in the next chapter, where the constraints on the Ψ^a are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

gauge fixing conditions that are used to eliminate first class constraints, and the constraints on the Π_a are secondary constraints arising from the consistency of the first class constraints with the field equations.) We assume that the unconstrained variables satisfy the usual canonical commutation relations $[Q^n, P_m] = i\delta^n_m$, $[Q^n, Q^m] = [P_n, P_m] = 0$. The constrained and unconstrained momenta are related by

$$P_n = \frac{\partial L}{\partial \dot{Q}^n} = \frac{\partial L}{\partial \dot{\Psi}^b}\frac{\partial \Psi^b}{\partial Q^n} = \Pi_b \frac{\partial \Psi^b}{\partial Q^n} \,. \tag{7.A.17}$$

It follows that

$$[\Psi^a, \Pi_b]\frac{\partial \Psi^b}{\partial Q^n} = [\Psi^a, P_n] = i\frac{\partial \Psi^a}{\partial Q^n}$$

or, in other words,

$$\{[\Psi^a, \Pi_b] - i\delta^a_b\}\frac{\partial \Psi^b}{\partial Q^n} = 0 \,. \tag{7.A.18}$$

Now, the constraint $\chi_{1r}(\Psi) = 0$ is satisfied for $\Psi^a = \Psi^a(Q)$ for all Q, so

$$\frac{\partial \chi_{1r}}{\partial \Psi^b}\frac{\partial \Psi^b}{\partial Q^n} = 0 \,. \tag{7.A.19}$$

Furthermore, the vectors $(V_r)_b \equiv \partial\chi_{1r}/\partial\Psi^b$ form a complete set perpendicular to all the vectors $(U_n)^b \equiv \partial\Psi^b/\partial Q^n$, because if there were some other vector V_b with $V_b(U_n)^b = 0$ for all n, then there would be additional constraints on the Ψ^a. Hence Eq. (7.A.18) implies that

$$[\Psi^a, \Pi_b] = i\delta^a_b + ic^a_r\frac{\partial \chi_{1r}}{\partial \Psi^b} \tag{7.A.20}$$

with some unknown coefficients c^a_r. To determine these coefficients, we make use of the other constraint, that $\chi_{2r}(\Pi) = 0$. It follows that

$$0 = [\Psi^a, \chi_{2r}(\Pi)] = i[\Psi^a, \Pi_b]\frac{\partial \chi_{2r}(\Pi)}{\partial \Pi_b} \,.$$

Using Eq. (7.A.20), we have then

$$\frac{\partial \chi_{2r}(\Pi)}{\partial \Pi_a} = -c^a_s\frac{\partial \chi_{1s}(\Psi)}{\partial \Psi^b}\frac{\partial \chi_{2r}(\Pi)}{\partial \Pi_b} \,. \tag{7.A.21}$$

We recognize the factor multiplying c^a_s as the Poisson bracket

$$\frac{\partial \chi_{1s}(\Psi)}{\partial \Psi^b}\frac{\partial \chi_{2r}(\Pi)}{\partial \Pi_b} = [\chi_{1s}, \chi_{2r}]_{\mathrm{P}} \equiv C_{1s,2r} \,.$$

Also, since χ_{1s} depends only on the Ψ and χ_{2r} depends only on the Π, these are the only non-vanishing Poisson brackets of constraints, so

$$C_{1r,1s} = C_{2r,2s} = 0 \,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

Thus Eq. (7.A.21) may be written

$$\frac{\partial \chi_N}{\partial \Pi_a} = -c_s^a \, C_{1s,N} \tag{7.A.22}$$

with N running over all constraint functions. For second class constraints, this has the unique solution

$$c_s^a = -\frac{\partial \chi_N}{\partial \Pi_a}(C^{-1})^{N,1s} = -\frac{\partial \chi_{2r}}{\partial \Pi_a}(C^{-1})^{2r,1s} \, . \tag{7.A.23}$$

Using this in Eq. (7.A.20) shows that

$$[\Psi^a, \Pi_b] = i\left[\delta_b^a - \frac{\partial \chi_{2r}}{\partial \Pi_a}(C^{-1})^{2r,1s}\frac{\partial \chi_{1s}}{\partial \Psi^b}\right] \, . \tag{7.A.24}$$

The Poisson brackets of Ψ^a and Π_b with the constraint functions are

$$\begin{aligned} [\Psi^a, \chi_{1r}]_{\mathrm{P}} &= 0 \, , & [\Psi^a, \chi_{2r}]_{\mathrm{P}} &= \frac{\partial \chi_{2r}}{\partial \Pi_a} \, , \\ [\chi_{1r}, \Pi_b]_{\mathrm{P}} &= \frac{\partial \chi_{1r}}{\partial \Psi^b} \, , & [\chi_{2r}, \Pi_b]_{\mathrm{P}} &= 0 \, , \end{aligned} \tag{7.A.25}$$

so the quantity in brackets on the right-hand side of Eq. (7.A.24) is the Dirac bracket

$$[\Psi^a, \Pi_b] = i[\Psi^a, \Pi_b]_{\mathrm{D}} \tag{7.A.26}$$

as was to be shown. Also, we can easily see that because C^{-1} has no 11 or 22 components, the other Dirac brackets are

$$[\Psi^a, \Psi^b]_{\mathrm{D}} = [\Pi_a, \Pi_b]_{\mathrm{D}} = 0 \, , \tag{7.A.27}$$

so trivially

$$[\Psi^a, \Psi^b] = i[\Psi^a, \Psi^b]_{\mathrm{D}} \, , \qquad [\Pi_a, \Pi_b] = i[\Pi_a, \Pi_b]_{\mathrm{D}} \, . \tag{7.A.28}$$

* * *

In addition to commutation rules, we also need an explicit formula for the Hamiltonian. The usual canonical formalism tells us to take

$$H = P_n \dot{Q}^n - L \, , \tag{7.A.29}$$

the sum running over independent canonical variables. In theories of both types considered in this Appendix, this Hamiltonian may be written in terms of the constrained variables as

$$H = \Pi_a \dot{\Psi}^a - L \, . \tag{7.A.30}$$

For theories of type A, this is trivial; the sum over a runs over values n, for which $\Psi^n = Q^n$ and $\Pi_n = P_n$ are the independent canonical variables,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

together with values r, for which $\Pi_r = \mathscr{P}_r = 0$. For theories of type B, we note that Eq. (7.A.17) gives

$$P_n \dot{Q}^n = \Pi_b \frac{\partial \Psi^b}{\partial Q^n} \dot{Q}^n = \Pi_b \dot{\Psi}^b$$

which again yields Eq. (7.A.30).

Problems

1. Consider the theory of a set of real scalar fields Φ^n, with Lagrangian density $\mathscr{L} = -\frac{1}{2} \sum_{mn} \partial_\mu \Phi^n \partial^\mu \Phi^m f_{nm}(\Phi)$, where $f_{nm}(\Phi)$ is an arbitrary real non-singular matrix function of the field. (This is called the *non-linear σ-model.*) Carry out the canonical quantization of this theory. Derive the interaction $V[\phi(t), \dot{\phi}(t)]$ in the interaction picture.

2. Consider a theory of real scalar fields Φ^n and Dirac fields Ψ^i, with Lagrangian density $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1$, where $\mathscr{L}_0$ is the usual free-field Lagrangian density, and $\mathscr{L}_1$ is an interaction term involving Φ^n and Ψ^i, but not their derivatives. Derive an explicit expression for the symmetric energy–momentum tensor $\Theta^{\mu\nu}$.

3. In the theory described in Problem 2, suppose that the Lagrangian density is invariant under a global infinitesimal symmetry $\delta\Phi^n = i\epsilon \sum_m t^n{}_m \Phi^m$, $\delta\Psi^i = i\epsilon \sum_j \tau^i{}_j \Psi^j$. Derive an explicit expression for the conserved current associated with this symmetry.

4. Consider the theory of a complex scalar field Φ and a real vector field V^μ, with Lagrangian density

$$\mathscr{L} = -(D_\mu \Phi)^\dagger D^\mu \Phi - \tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} - \tfrac{1}{2} m^2 V_\mu V^\mu - \mathscr{H}(\Phi^\dagger \Phi) ,$$

where $D_\mu \equiv \partial_\mu - igV_\mu$ and $F_{\mu\nu} \equiv \partial_\mu V_\nu - \partial_\nu V_\mu$, and $\mathscr{H}$ is an arbitrary function. Carry out the canonical quantization of this theory. Derive the interaction in the interaction picture.

5. In the theory of Problem 4, derive expressions for the symmetric energy–momentum tensor $\Theta^{\mu\nu}$ and for the conserved current associated with the symmetry under $\delta\Phi = i\epsilon\Phi$, $\delta V^\mu = 0$.

6. Prove that the Dirac bracket satisfies the Jacobi identity (7.6.23).

7. Prove that the Dirac bracket is independent of the choice of constraint functions χ_N used to describe a given submanifold of phase space.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

References

1. H. B. G. Casimir, *Proc. K. Ned. Akad. Wet.* **51**, 635 (1948); M. J. Spaarnay, *Nature* **180**, 334 (1957).

2. F. Belinfante, Physica **6**, 887 (1939); also see L. Rosenfeld, *Mémoires de l'Academie Roy. Belgique* **6**, 30 (1930).

3. See, e.g., S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Chapter 12.

4. See, e.g., J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, Princeton, 1983), and original references quoted therein.

5. P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, New York, 1964). Also see P. A. M. Dirac, *Can. J. Math.* **2**, 129 (1950); *Proc. Roy. Soc. London*, ser. A, **246**, 326 (1958); P. G. Bergmann, *Helv. Phys. Acta Suppl.* IV, 79 (1956).

6. T. Maskawa and H. Nakajima, *Prog. Theor. Phys.* **56**, 1295 (1976). I am grateful to J. Feinberg for bringing this reference to my attention.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.009

8
Electrodynamics

The original approach to quantum electrodynamics was to take for granted Maxwell's classical theory of electromagnetism, and quantize it. It will probably not surprise the reader that this book will follow a different path. We shall first infer the need for a principle of gauge invariance from the peculiar difficulties that arise in formulating a quantum theory of massless particles with spin, and then deduce the main features of electrodynamics from the gauge invariance principle. After that we shall follow a more conventional modern approach, in which one takes gauge invariance as the starting point and uses it to deduce the existence of a vector potential describing massless particles of unit spin.

It is too soon to tell which of these two alternatives corresponds to the logical order of nature. Most theorists have tended to take gauge invariance as a starting point, but in modern string theories[1] the argument runs the other way; one first notices a state of mass zero and unit spin among the normal modes of a string, and then from that deduces the gauge invariance of the effective field theory that describes such particles. At any rate, as we shall see, using either approach one is led to the quantized version of Maxwell's theory, still the paradigmatic example of a successful quantum field theory.

8.1 Gauge Invariance

Let's start by recalling the problems encountered in constructing covariant free fields for a massless particle of helicity ± 1. We saw in Section 5.9 that there is no difficulty in constructing an antisymmetric tensor free field $f_{\mu\nu}(x)$ for such particles. This field can be expressed in terms of the four-potential $a_\mu(x)$, given by Eq. (5.9.23), through the familiar relation

$$f_{\mu\nu}(x) = \partial_\mu a_\nu(x) - \partial_\nu a_\mu(x) \,. \tag{8.1.1}$$

However, Eq. (5.9.31) shows that the $a_\mu(x)$ transforms as a four-vector

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

only up to a gauge transformation

$$U_0(\Lambda)a_\mu(x)\,U_0^{-1}(\Lambda) = \Lambda_\mu{}^\nu a_\nu(\Lambda x) + \partial_\mu\Omega(x,\Lambda)\,. \tag{8.1.2}$$

There is, in fact, no way to construct a true four-vector as a linear combination of the creation and annihilation operators for helicity ± 1. This is one way of understanding the presence of singularities at $m = 0$ in the propagator of a massive vector field

$$\Delta_{\mu\nu}(x,y) = (2\pi)^{-4}\int d^4q\; e^{iq\cdot(x-y)}\,\frac{\eta_{\mu\nu} + q_\mu q_\nu/m^2}{q^2 + m^2 - i\epsilon}\,,$$

which prevent us from dealing with massless particles of helicity ± 1 by simply passing to the limit $m \to 0$ of the theory of a massive particle of spin one.

We could avoid these problems by demanding that all interactions involve only* $F_{\mu\nu}(x) \equiv \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)$ and its derivatives, not $A_\mu(x)$, but this is not the most general possibility, and not the one realized in nature. Instead of banishing $A_\mu(x)$ from the action, we shall require instead that the part of the action I_M for matter and its interaction with radiation be invariant under the general gauge transformation

$$A_\mu(x) \to A_\mu(x) + \partial_\mu\epsilon(x) \tag{8.1.3}$$

(at least when the matter fields satisfy the field equations) so that the extra term in Eq. (8.1.2) should have no effect. The change in the matter action under the transformations (8.1.3) may be written

$$\delta I_M = \int d^4x\,\frac{\delta I_M}{\delta A_\mu(x)}\partial_\mu\epsilon(x)\,. \tag{8.1.4}$$

Hence the Lorentz invariance of I_M requires that

$$\partial_\mu\frac{\delta I_M}{\delta A_\mu(x)} = 0\,. \tag{8.1.5}$$

This is trivially true if I_M involves only $F_{\mu\nu}(x)$ and its derivatives, along with matter fields. In this case

$$\frac{\delta I_M}{\delta A_\mu(x)} = 2\partial_\nu\,\frac{\delta I_M}{\delta F_{\mu\nu}(x)}\,.$$

But if I_M involves $A_\mu(x)$ itself then Eq. (8.1.5) is a non-trivial constraint on the theory.

Now, what sort of theory will provide conserved currents to which we can couple the field $A^\mu(x)$? We saw in Section 7.3 that infinitesimal internal

* We now use A_μ and $F_{\mu\nu}$ for the electromagnetic potential vector and the field strength tensor because these are interacting fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

symmetries of the action imply the existence of conserved currents. In particular, if the transformation**

$$\delta\Psi^\ell(x) = i\epsilon(x)\, q_\ell\, \Psi^\ell(x) \tag{8.1.6}$$

leaves the matter action invariant for a constant ϵ, then for general infinitesimal functions $\epsilon(x)$ the change in the matter action must take the form

$$\delta I_M = -\int d^4x\, J^\mu(x)\partial_\mu\epsilon(x)\,. \tag{8.1.7}$$

When the matter fields satisfy their field equations, the matter action is stationary with respect to *any* variation of the Ψ_ℓ, so in this case (8.1.7) must vanish, and hence

$$\partial_\mu J^\mu = 0\,. \tag{8.1.8}$$

In particular, we saw in Section 7.3 that if I_M is the integral of a function $\mathscr{L}_M$ of Ψ^ℓ and $\partial_\mu\Psi^\ell$, then the conserved current is given by†

$$J^\mu = -i\sum_\ell \frac{\partial\mathscr{L}_M}{\partial(\partial_\mu\Psi^\ell)}\, q_\ell\, \Psi^\ell\,,$$

and this generates the transformations (8.1.6) in the sense that

$$[Q\,,\, \Psi^\ell(x)] = -q_\ell\, \Psi^\ell(x)\,, \tag{8.1.9}$$

where Q is the time-independent charge operator

$$Q = \int d^3x\, J^0\,. \tag{8.1.10}$$

We can therefore construct a Lorentz-invariant theory by coupling the vector field A_μ to the conserved current J^μ, in the sense that $\delta I_M/\delta A_\mu(x)$ is taken to be proportional to $J^\mu(x)$. Any constant of proportionality may be absorbed into the definition of the overall scale of the charges q_ℓ, so we may simply set these quantities equal:

$$\frac{\delta I_M}{\delta A_\mu(x)} = J^\mu(x)\,. \tag{8.1.11}$$

The conservation of electric charge only allows us to fix the values of all charges in terms of the value of any one of them, conventionally taken to

** Because the field transformation matrix is taken now to be diagonal it is not convenient here to use the summation convention for sums over field indices, so there is no sum over ℓ in Eq. (8.1.6).

† Here Ψ^ℓ is understood to run over all independent fields other than A^μ. We use a capital psi to indicate that these are Heisenberg-picture fields, whose time-dependence includes the effects of interactions. Of course, this Ψ^ℓ is not to be confused with a state-vector or wave function.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

be the electron charge, denoted $-e$. It is Eq. (8.1.11) that gives a definite meaning‡ to the value of e.

The requirement (8.1.11) may be restated as a principle of invariance:[1a] the matter action is invariant under the joint transformations

$$\delta A_\mu(x) = \partial_\mu \epsilon(x) , \tag{8.1.12}$$

$$\delta \Psi_\ell(x) = i\epsilon(x) q_\ell \Psi_\ell(x) . \tag{8.1.13}$$

A symmetry of this type with an arbitrary function $\epsilon(x)$ is called a *local symmetry*, or a gauge invariance of the second kind. A symmetry under a transformation with ϵ constant is called a *global* symmetry, or a gauge invariance of the first kind. Several exact local symmetries are now known, but the only purely global symmetries appear to be accidents enforced by other principles. (See Section 12.5.)

We have not yet said anything about the action for photons themselves. As a guess, we can take this to be the same as for massive vector fields, but with $m = 0$:

$$I_\gamma = -\tfrac{1}{4} \int d^4x \, F_{\mu\nu} F^{\mu\nu} . \tag{8.1.14}$$

This is the same as the action used in classical electrodynamics, but its real justification is that it is (up to a constant) the unique gauge-invariant functional that is quadratic in $F_{\mu\nu}$, without higher derivatives. Also, as we will see in the next section, it leads to a consistent quantum theory. If there are any terms in the action with higher derivatives and/or of higher order in $F_{\mu\nu}$ they can be lumped into what we have called the matter action. Using Eqs. (8.1.11) and (8.1.14), the field equation for electromagnetism now reads

$$0 = \frac{\delta}{\delta A_\nu} [I_\gamma + I_M] = \partial_\mu F^{\mu\nu} + J^\nu . \tag{8.1.15}$$

We recognize these as the usual inhomogeneous Maxwell equations, with current J^ν. There are also other, homogeneous, Maxwell equations

$$0 = \partial_\mu F_{\nu\epsilon} + \partial_\epsilon F_{\mu\nu} + \partial_\nu F_{\epsilon\mu} , \tag{8.1.16}$$

which follow directly from the definition $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$.

In the above discussion, we have started with the existence of massless spin one particles, and have been led to infer the invariance of the matter action under a local gauge transformation (8.1.12), (8.1.13). As usually presented, the derivation runs in the opposite direction. That is, one starts

‡ Of course, Eq. (8.1.11) fixes the definition of e only after we have defined how we are normalizing $A_\mu(x)$. The question of electromagnetic field normalization is taken up in Section 10.4.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

with a global internal symmetry

$$\delta\Psi^\ell(x) = i\epsilon q^\ell \Psi^\ell(x) \tag{8.1.17}$$

and asks what must be done to promote this to a local symmetry

$$\delta\Psi^\ell(x) = i\epsilon(x) q_\ell \Psi^\ell(x)\,. \tag{8.1.18}$$

If the Lagrange density $\mathscr{L}$ depended only on fields $\Psi^\ell(x)$ and not on their derivatives then it would make no difference whether ϵ is constant or not; invariance with ϵ constant would imply invariance with ϵ a function of spacetime position. But all realistic Lagrangians do involve field derivatives, and here we have the problem that derivatives of fields transform differently from fields themselves:

$$\delta\,\partial_\mu\Psi^\ell(x) = i\epsilon(x) q_\ell \partial_\mu\Psi^\ell(x) + i\,q_\ell\Psi^\ell(x)\partial_\mu\epsilon(x)\,. \tag{8.1.19}$$

In order to cancel the second term here, we 'invent' a vector field $A_\mu(x)$ with transformation rule

$$\delta A_\mu(x) = \partial_\mu\epsilon(x) \tag{8.1.20}$$

and require that the Lagrangian density depend on $\partial_\mu\Psi^\ell$ and A_μ only in the combination

$$D_\mu\Psi^\ell \equiv \partial_\mu\Psi^\ell - iq_\ell A_\mu\Psi^\ell\,, \tag{8.1.21}$$

which transforms just like Ψ^ℓ

$$\delta\,D_\mu\Psi^\ell(x) = i\epsilon(x) q_\ell D_\mu\Psi^\ell(x)\,. \tag{8.1.22}$$

A matter Lagrangian density $\mathscr{L}_M(\Psi, D\Psi)$ that is formed only out of Ψ^ℓ and $D_\mu\Psi^\ell$ will be invariant under the transformations (8.1.18), (8.1.20), with $\epsilon(x)$ an arbitrary function, if it is invariant with ϵ a constant. With the Lagrangian of this form, we have

$$\frac{\delta I_M}{\delta A_\mu} = \sum_\ell \frac{\partial\mathscr{L}_M}{\partial D_\mu\Psi^\ell}(-iq_\ell\Psi^\ell) = -i\sum_\ell \frac{\partial\mathscr{L}_M}{\partial\,\partial_\mu\Psi^\ell} q_\ell\Psi^\ell\,,$$

which is the same as Eq. (8.1.11). (We could also include $F_{\mu\nu}$ and its derivatives in $\mathscr{L}_M$.) From this point of view, the masslessness of the particles described by A_μ is a consequence of gauge invariance rather than an assumption: a term $-\frac{1}{2}m^2A_\mu A^\mu$ in the Lagrangian density would violate gauge invariance.

8.2 Constraints and Gauge Conditions

There are aspects of electrodynamics that stand in the way of quantizing the theory as we did for various theories of massive particles in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

previous chapter. As usual, we may define the canonical conjugates to the electromagnetic vector potential by

$$\Pi^{\mu} \equiv \frac{\partial \mathscr{L}}{\partial(\partial_0 A_\mu)} . \tag{8.2.1}$$

Quantization by the usual rules would give

$$[A_\mu(\mathbf{x}, t), \Pi^{\nu}(\mathbf{y}, t)] = i\delta_\mu^{\nu}\delta^3(\mathbf{x} - \mathbf{y}) .$$

But this is not possible here, because A_μ and Π^{ν} are subject to several constraints.

The first constraint arises from the fact that the Lagrangian density is independent* of the time-derivative of A_0, and therefore

$$\Pi^0(x) = 0 . \tag{8.2.2}$$

This is called a *primary constraint*, because it follows directly from the structure of the Lagrangian. There is also a *secondary constraint* here, which follows from the field equation for the quantity fixed by the primary constraint:**

$$\partial_i \Pi^i = -\partial_i \frac{\partial \mathscr{L}}{\partial F_{i0}} = -\frac{\partial \mathscr{L}}{\partial A_0} = -J^0 , \tag{8.2.3}$$

the time-derivative term dropping out because $F_{00} = 0$. Even though the matter Lagrangian may generally depend on A^0, the charge density depends only on the canonical matter fields† Q^n and their canonical conjugates P_n:

$$J^0 = -i\sum_{\ell} \frac{\partial \mathscr{L}}{\partial(\partial_0 \Psi^\ell)} q_\ell \Psi^\ell = -i\sum_n P_n q_n Q^n . \tag{8.2.4}$$

Hence Eq. (8.2.3) is a functional relation among canonical variables. Both Eq. (8.2.2) and Eq. (8.2.3) are inconsistent with the usual assumptions that $[A_\mu(\mathbf{x}, t), \Pi^{\nu}(\mathbf{y}, t)] = i\delta_\mu^{\nu}\delta^3(\mathbf{x} - \mathbf{y})$ and $[Q^n(\mathbf{x}, t), \Pi^{\nu}(\mathbf{y}, t)] = [P_n(\mathbf{x}, t), \Pi^{\nu}(\mathbf{y}, t)] = 0$.

We encountered a similar problem in the theory of the massive vector field. In that case we found two equivalent ways of dealing with it: either by the method of Dirac brackets or, more directly, by treating only

* For $\mathscr{L}_\gamma = -F_{\mu\nu}F^{\mu\nu}/4$, we have $\partial\mathscr{L}_\gamma/\partial(\partial_0 A_\mu) = -F^{0\mu}$, which vanishes for $\mu = 0$ because $F^{\mu\nu}$ is antisymmetric. For matter Lagrangians $\mathscr{L}_M$ that involve only Ψ^ℓ and $D_\mu\Psi^\ell$, the prescription (8.1.21) tells us that $\mathscr{L}_M$ does not depend on any derivatives of any A^{ν}. Even if the matter Lagrangian depends also on $F_{\mu\nu}$, $\partial\mathscr{L}_M/\partial(\partial_\nu A_\mu)$ will be again antisymmetric in μ and ν, and therefore will vanish for $\mu = \nu = 0$.

** As usual, i, j, etc. run over the values 1, 2, 3.

† Due to exhaustion of alphabetic resources, I have had to adopt a notation here that is different from that of the previous chapter. The symbols Q^n and P_n are now reserved for the canonical matter fields and their canonical conjugates, respectively, while the canonical electromagnetic fields and canonical conjugates are A_i and Π_i.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

A_i and Π^i as canonical variables, solving the analog of Eq. (8.2.3) to calculate A^0 in terms of these variables. It is clear that here we cannot use Dirac brackets; the constraint functions χ here are Π^0 and $\partial_i\Pi_i + J^0$ (as compared with $\partial_i\Pi_i - m^2A^0 + J^0$) and these obviously have vanishing Poisson brackets. In Dirac's terminology, the constraints (8.2.2) and (8.2.3) are *first class*. Nor can we eliminate A^0 as a dynamical variable by solving for it in terms of the other variables. Instead of giving A^0 for all time, Eq. (8.2.3) is a mere initial condition; if Eq. (8.2.3) is satisfied at one time, then it is satisfied for all times, because (using the field equations for the other fields A^i), we have

$$\begin{aligned}\partial_0\left[\partial_i\frac{\partial\mathscr{L}}{\partial F_{i0}} - J^0\right] &= -\partial_i\partial_0\frac{\partial\mathscr{L}}{\partial F_{0i}} - \partial_0 J^0\\ &= +\partial_i\partial_j\frac{\partial\mathscr{L}}{\partial F_{ji}} - \partial_i J_i - \partial_0 J^0\end{aligned}$$

and the current conservation condition then gives

$$\partial_0\left[\partial_i\frac{\partial\mathscr{L}}{\partial F_{i0}} - J^0\right] = 0\,. \tag{8.2.5}$$

It should not be surprising that we still have four components of A^μ with only three field equations, because this theory has a local gauge symmetry that makes it, in principle, impossible to infer the values of the fields at arbitrary times from their values and rates of change at any one time. Given any solution $A_\mu(\mathbf{x},t)$ of the field equations, we can always find another solution $A_\mu(\mathbf{x},t) + \partial_\mu\epsilon(\mathbf{x},t)$ with the same value and time-derivative at $t = 0$ (by choosing ϵ so that its first and second derivatives vanish there) but which differs from $A_\mu(\mathbf{x},t)$ at later times.

Because of this partial arbitrariness of $A_\mu(\mathbf{x},t)$, it is not possible to apply the canonical quantization procedure directly to A_μ (or, as for finite mass, to **A**). Of the various approaches to this difficulty, two are particularly useful. One is the modern Lorentz-invariant method of BRST-quantization, to be discussed in Volume II. The other, which will be followed here, is to exploit the gauge invariance of the theory, to 'choose a gauge'. That is, we make a finite gauge transformation

$$A_\mu(x) \to A_\mu(x) + \partial_\mu\lambda(x)\,, \qquad \Psi_\ell(x) \to \exp\left(iq_\ell\lambda(x)\right)\Psi_\ell(x)$$

to impose a condition on $A_\mu(x)$ that will allow us to apply the methods of canonical quantization. There are various gauges that have been found useful in various applications:‡

‡ Here Φ is any complex scalar field with $q \neq 0$; this gauge condition is used when the gauge symmetry is spontaneously broken by a non-vanishing vacuum expectation value of Φ.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Lorentz (or Landau) gauge:	$\partial_\mu A^\mu = 0$
Coulomb gauge:	$\nabla \cdot \mathbf{A} = 0$
Temporal gauge:	$A^0 = 0$
Axial gauge:	$A^3 = 0$
Unitarity gauge:	Φ real

The canonical quantization procedure works most easily in the axial or Coulomb gauge, but of course Coulomb gauge keeps manifest rotation invariance in a way that axial gauge does not, so we will adopt Coulomb gauge here.[2]

To check that this is possible, note that if A^μ does not satisfy the Coulomb gauge condition, then the gauge-transformed field $A^\mu + \partial^\mu \lambda$ will, provided we choose λ so that $\nabla^2 \lambda = -\nabla \cdot \mathbf{A}$. From now on, we assume that this transformation has been made, so that

$$\nabla \cdot \mathbf{A} = 0 \,. \tag{8.2.6}$$

It will be convenient henceforth to limit ourselves to theories in which the matter Lagrangian $\mathscr{L}_M$ may depend on matter fields and their time-derivatives and also on A^μ but not on derivatives of A^μ. (The standard theories of the electrodynamics of scalar and Dirac fields have Lagrangians of this type.) Then the only term in the Lagrangian that depends on $F_{\mu\nu}$ is the kinematic term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, and the constraint equation (8.2.3) reads

$$-\partial_i F^{i0} = J^0 \,. \tag{8.2.7}$$

Together with the Coulomb gauge condition (8.2.6), this yields

$$-\nabla^2 A^0 = J^0 \,, \tag{8.2.8}$$

which can be solved to give

$$A^0(\mathbf{x}, t) = \int d^3y \, \frac{J^0(\mathbf{y}, t)}{4\pi|\mathbf{x} - \mathbf{y}|} \,. \tag{8.2.9}$$

The remaining degrees of freedom are A^i, with $i = 1, 2, 3$, subject to the gauge condition $\nabla \cdot \mathbf{A} = 0$.

As mentioned earlier, the charge density depends only on the canonical matter fields Q^n and their canonical conjugates P_n, so Eq. (8.2.9) represents an explicit solution for the auxiliary field A^0.

8.3 Quantization in Coulomb Gauge

There is still an impediment to the canonical quantization of electrodynamics in the Coulomb gauge. Even after we use Eq. (8.2.9) to eliminate A^0 (and Π_0) from the list of canonical variables, we cannot apply the usual

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

canonical commutation relations to A^i and Π_i, because there are two remaining constraints on these variables.* One of them is the Coulomb gauge condition

$$\chi_{1\mathbf{x}} \equiv \partial_i A^i(\mathbf{x}) = 0 \,. \tag{8.3.1}$$

The other is the secondary constraint Eq. (8.2.3), which requires that

$$\chi_{2\mathbf{x}} \equiv \partial_i \Pi^i(\mathbf{x}) + J^0(\mathbf{x}) = 0 \,. \tag{8.3.2}$$

Neither constraint is consistent with the usual commutation relations $[A_i(\mathbf{x}), \Pi_j(\mathbf{y})] = i\delta_{ij}\delta^3(\mathbf{x}-\mathbf{y})$, because operating on the right-hand side with either $\partial/\partial x^i$ or $\partial/\partial y^j$ does not give zero.

These constraints are of a type known as *second class*, for which there is a universal prescription for the commutation relations, discussed in Section 7.6. Note that the constraint functions have the Poisson brackets

$$\begin{aligned} C_{1\mathbf{x},2\mathbf{y}} &= -C_{2\mathbf{y},1\mathbf{x}} \equiv [\chi_{1\mathbf{x}}, \chi_{2\mathbf{y}}]_{\mathrm{P}} = -\nabla^2\delta^3(\mathbf{x}-\mathbf{y}) \,, \\ C_{1\mathbf{x},1\mathbf{y}} &\equiv [\chi_{1\mathbf{x}}, \chi_{1\mathbf{y}}]_{\mathrm{P}} = 0 \,, \\ C_{2\mathbf{x},2\mathbf{y}} &\equiv [\chi_{2\mathbf{x}}, \chi_{2\mathbf{y}}]_{\mathrm{P}} = 0 \,, \end{aligned} \tag{8.3.3}$$

where here, for any functionals U and V,

$$[U,V]_{\mathrm{P}} \equiv \int d^3x \left[\frac{\delta U}{\delta A^i(\mathbf{x})} \frac{\delta V}{\delta \Pi_i(\mathbf{x})} - \frac{\delta V}{\delta A^i(\mathbf{x})} \frac{\delta U}{\delta \Pi_i(\mathbf{x})} \right] .$$

The 'matrix' C_{NM} is non-singular, which identifies these as second class constraints. Also, the field variables A^i may be expressed in terms of independent canonical variables, which may, for instance, be taken as $Q_{1\mathbf{x}} = A^1(\mathbf{x})$, $Q_{2\mathbf{x}} = A^2(\mathbf{x})$, with A^3 given by the solution of Eq. (8.3.1):

$$A^3(\mathbf{x}) = -\int^{x^3} ds \, [\partial_1 A^1(x^1, x^2, s) + \partial_2 A^2(x^1, x^2, s)] \,.$$

Using Eq. (8.3.2), the canonical conjugates Π_i to A^i may likewise be expressed in terms of the canonical conjugates $P_{1\mathbf{x}}$ and $P_{2\mathbf{x}}$ to $Q_{1\mathbf{x}}$ and $Q_{2\mathbf{x}}$. In such cases, Part B of the Appendix to the previous chapter tells us that if the independent variables $Q_{1\mathbf{x}}$, $Q_{2\mathbf{x}}$, $P_{1\mathbf{x}}$, and $P_{2\mathbf{x}}$ satisfy the usual canonical commutation relations, then the commutators of the constrained variables and their canonical conjugates are given (aside from a factor i) by the corresponding Dirac brackets (7.6.20). This prescription has the great advantage that we do not have to use explicit expressions for the dependent variables in terms of the independent ones.

* In this section i, j, etc. run over the values 1,2,3. We continue the practice of taking all operators at the same time, and omitting the time argument.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

To calculate the Dirac brackets, we note that the matrix C has the inverse

$$(C^{-1})_{1\mathbf{x},2\mathbf{y}} = -(C^{-1})_{2\mathbf{y},1\mathbf{x}} = -\int \frac{d^3k}{(2\pi)^3}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{\mathbf{k}^2} = -\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\,,$$
$$(C^{-1})_{1\mathbf{x},1\mathbf{y}} = (C^{-1})_{2\mathbf{x},2\mathbf{y}} = 0\,. \tag{8.3.4}$$

Also, the non-vanishing Poisson brackets of the A^i and Π_i with the constraint functions are

$$[A^i(\mathbf{x}), \chi_{2\mathbf{y}}]_{\mathrm{P}} = -\frac{\partial}{\partial x^i}\delta^3(\mathbf{x}-\mathbf{y})$$

and

$$[\Pi_i(\mathbf{x}), \chi_{1\mathbf{y}}]_{\mathrm{P}} = +\frac{\partial}{\partial x^i}\delta^3(\mathbf{x}-\mathbf{y})\,.$$

Hence according to Eqs. (7.6.19) and (7.6.20), the equal-time commutators are

$$\begin{aligned}\left[A^i(\mathbf{x}), \Pi_j(\mathbf{y})\right] &= i\delta^i_j\delta^3(\mathbf{x}-\mathbf{y}) + i\frac{\partial^2}{\partial x^j \partial x^i}\left(\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\right)\,,\\ \left[A^i(\mathbf{x}),\, A^j(\mathbf{y})\right] &= \left[\Pi_i(\mathbf{x}),\, \Pi_j(\mathbf{y})\right] = 0\,.\end{aligned} \tag{8.3.5}$$

Note that these are consistent with the Coulomb gauge conditions (8.3.1) and (8.3.2), as is guaranteed by the general properties of the Dirac bracket.

Now, what is $\boldsymbol{\Pi}$ in electrodynamics? For the class of theories discussed in the previous section where only the kinematic term $-\frac{1}{4}\int d^3x F_{\mu\nu}F^{\mu\nu}$ in the Lagrangian depends on $\dot{\mathbf{A}}$, varying the Lagrangian with respect to $\dot{\mathbf{A}}$ without worrying about the constraint $\nabla\cdot\mathbf{A}=0$ gives

$$\Pi_j = \frac{\delta L}{\delta \dot{A}^j(\mathbf{x})} = \dot{A}^j(\mathbf{x}) + \frac{\partial}{\partial x^j}A^0(\mathbf{x})\,. \tag{8.3.6}$$

But with $\mathbf{A}$ constrained by the condition $\nabla\cdot\mathbf{A}=0$, variational derivatives with respect to $\dot{\mathbf{A}}$ are not really well defined. If the variation of L under a change $\delta\dot{\mathbf{A}}$ in $\dot{\mathbf{A}}$ is $\delta L = \int d^3x\, \mathscr{P}\cdot\delta\dot{\mathbf{A}}$, then since $\nabla\cdot\delta\dot{\mathbf{A}}=0$, we also have $\delta L = \int d^3x\,[\mathscr{P}+\nabla\mathscr{F}]\cdot\delta\dot{\mathbf{A}}$ for any scalar function $\mathscr{F}(\mathbf{x})$. Thus all we can conclude from inspection of the Lagrangian is that $\boldsymbol{\Pi}$ equals $\dot{\mathbf{A}}(\mathbf{x})+\nabla A^0(\mathbf{x})$ plus the gradient of some scalar. This ambiguity is removed by condition (8.3.2), which requires that $\nabla\cdot\boldsymbol{\Pi} = -J^0 = \nabla^2 A^0$. Because $\nabla\cdot\dot{\mathbf{A}}=0$, we conclude that Eq. (8.3.6) does indeed give the correct formula for Π^i.

Although the commutation relations (8.3.5) are reasonably simple, we must face the complication that $\boldsymbol{\Pi}$ does not commute with matter fields and their canonical conjugates. If F is any functional of these matter degrees of freedom, then its Dirac bracket with $\mathbf{A}$ vanishes, but its Dirac

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

bracket with $\mathbf{\Pi}$ is

$$\begin{aligned}[F,\mathbf{\Pi}(\mathbf{z})]_{\mathrm{D}} &= -\int d^3x\, d^3y\, [F,\chi_{2\mathbf{x}}]_{\mathrm{P}} \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}[\chi_{1\mathbf{y}},\mathbf{\Pi}(\mathbf{z})]_{\mathrm{P}} \\ &= -\int d^3x\, d^3y\, [F,J^0(\mathbf{x})]_{\mathrm{P}} \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\nabla\delta^3(\mathbf{y}-\mathbf{z}) \\ &= -\int d^3y\, [F,A^0(\mathbf{y})]_{\mathrm{P}}\nabla\delta^3(\mathbf{y}-\mathbf{z}) \\ &= [F,\nabla A^0(\mathbf{z})]_{\mathrm{P}} = [F,\nabla A^0(\mathbf{z})]_{\mathrm{D}}\,.\end{aligned}$$

In order to facilitate the transition to the interaction picture, instead of expressing the Hamiltonian in terms of $\mathbf{A}$ and $\mathbf{\Pi}$, we shall write it in terms of $\mathbf{A}$ and $\mathbf{\Pi}_\perp$, where $\mathbf{\Pi}_\perp$ is the solenoidal part of $\mathbf{\Pi}$:

$$\mathbf{\Pi}_\perp \equiv \mathbf{\Pi} - \nabla A^0 = \dot{\mathbf{A}}\,, \tag{8.3.7}$$

for which $[F,\mathbf{\Pi}_\perp(\mathbf{z})]$ vanishes. By using the facts that $\mathbf{\Pi}_\perp(\mathbf{x})$ commutes with $\mathbf{\Pi}(\mathbf{y}) - \mathbf{\Pi}_\perp(\mathbf{y}) = \nabla A^0(\mathbf{y})$ and that $\partial_i A^0(\mathbf{x})$ commutes with $\partial_j A^0(\mathbf{y})$, it is easy to see that $\mathbf{\Pi}_\perp(\mathbf{x})$ satisfies the same commutation relations (8.3.5) as $\mathbf{\Pi}(\mathbf{x})$, and also the simple constraint

$$\nabla\cdot\mathbf{\Pi}_\perp = 0\,. \tag{8.3.8}$$

Now we need to construct a Hamiltonian. According to the general results of the Appendix to Chapter 7, we can apply the usual relation between the Hamiltonian and Lagrangian using the constrained variables $\mathbf{A}$ and $\mathbf{\Pi}_\perp$, without first having explicitly to write the Hamiltonian in terms of the unconstrained Qs and Ps. In electrodynamics, this gives

$$H = \int d^3x\, \left[\Pi_{\perp i}\dot{A}^i + P_n\dot{Q}^n - \mathscr{L}\right]\,, \tag{8.3.9}$$

where, as mentioned earlier, Q^n and P_n are to be understood as the matter canonical fields and their canonical conjugates. (We can use $\mathbf{\Pi}_\perp$ in place of $\mathbf{\Pi}$ in Eq. (8.3.9) because $\nabla\cdot\mathbf{A} = 0$.)

To be specific, consider a theory with a Lagrangian density of the form

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + J_\mu A^\mu + \mathscr{L}_{\text{matter}}\,, \tag{8.3.10}$$

where J_μ is a current that does not involve A^μ, and $\mathscr{L}_{\text{matter}}$ is the Lagrangian for whatever other fields do appear in J^μ, aside from their electromagnetic interactions, which are given explicitly by the term $J_\mu A^\mu$ in Eq. (8.3.10). (The electrodynamics of spin $\frac{1}{2}$ particles has a Lagrangian of this form, but the electrodynamics of spinless particles is more complicated.) Replacing $\dot{\mathbf{A}}$ everywhere with $\mathbf{\Pi}_\perp$, this gives a Hamiltonian (8.3.9) of the form

$$H = \int d^3x\, \left[\mathbf{\Pi}_\perp^2 + \tfrac{1}{2}(\nabla\times\mathbf{A})^2 - \tfrac{1}{2}(\mathbf{\Pi}_\perp + \nabla A^0)^2 - \mathbf{J}\cdot\mathbf{A} + J^0A^0\right] + H_{\mathrm{M}}\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

where H_M is the Hamiltonian for matter fields, excluding their electromagnetic interactions

$$H_M \equiv \int d^3x \, (P_n \dot{Q}^n - \mathscr{L}_{\text{matter}}) .$$

Using the solution (8.2.9) for A^0, this is

$$H = \int d^3x \left[\tfrac{1}{2}\boldsymbol{\Pi}_\perp^2 + \tfrac{1}{2}(\nabla \times \mathbf{A})^2 - \mathbf{J} \cdot \mathbf{A} + \tfrac{1}{2} J^0 A^0 \right] + H_M . \tag{8.3.11}$$

The term $\frac{1}{2}J^0A^0$ may look peculiar, but this is nothing but the familiar Coulomb energy

$$\begin{aligned} V_{\text{Coul}} &= \tfrac{1}{2} \int d^3x \, J^0 \, A^0 \\ &= \tfrac{1}{2} \int d^3x \int d^3y \frac{J^0(\mathbf{x}) J^0(\mathbf{y})}{4\pi |\mathbf{x} - \mathbf{y}|} . \end{aligned} \tag{8.3.12}$$

The reader can verify, using the commutation relations (8.3.5), that the rate of change of any operator function F of $\mathbf{A}$ and $\boldsymbol{\Pi}$ is given by $i\dot{F} = [F, H]$, as it should be.

8.4 Electrodynamics in the Interaction Picture

We now break up the Hamiltonian (8.3.11) into a free-particle term H_0 and an interaction V

$$H = H_0 + V , \tag{8.4.1}$$

$$H_0 = \int d^3x \left[\frac{1}{2} \boldsymbol{\Pi}_\perp^2 + \frac{1}{2}(\nabla \times \mathbf{A})^2 \right] + H_{\text{matter},0} , \tag{8.4.2}$$

$$V = - \int d^3x \, \mathbf{J} \cdot \mathbf{A} + V_{\text{Coul}} + V_{\text{matter}} , \tag{8.4.3}$$

where $H_{\text{matter},0}$ and V_{matter} are the free-particle and interaction terms in H_{matter}, and V_{Coul} is the Coulomb interaction (8.3.12). The total Hamiltonian (8.4.1) is time-independent, so Eqs. (8.4.2) and (8.4.3) can be evaluated at any time we like (as long as both are evaluated at the same time), in particular at $t = 0$. As in Chapter 7, the transition to the interaction picture is made by applying the similarity transformation

$$\begin{aligned} V(t) &= \exp(iH_0 t) \, V[\mathbf{A}, \boldsymbol{\Pi}_\perp, Q, P]_{t=0} \exp(-iH_0 t) \\ &= V[\mathbf{a}(t), \boldsymbol{\pi}(t), q(t), p(t)] , \end{aligned} \tag{8.4.4}$$

where P here denotes the canonical conjugates to the matter fields Q, and any operator $o(\mathbf{x}, t)$ in the interaction picture is related to its value $O(\mathbf{x}, 0)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

in the Heisenberg picture at $t = 0$ by

$$o(\mathbf{x}, t) = \exp(iH_0 t)\ O(\mathbf{x}, 0)\ \exp(-iH_0 t), \tag{8.4.5}$$

so that

$$i\ \dot{o}(\mathbf{x}, t) = [o(\mathbf{x}, t), H_0]\ . \tag{8.4.6}$$

(We are dropping the subscript $\perp$ on $\boldsymbol{\pi}(x)$.) Since Eq. (8.4.5) is a similarity transformation, the equal-time commutation relations are the same as in the Heisenberg picture:

$$\left[a^i(\mathbf{x}, t), \pi^j(\mathbf{y}, t)\right] = i\left[\delta_{ij}\delta^3(\mathbf{x} - \mathbf{y}) + \frac{\partial^2}{\partial x^i \partial x^j}\frac{1}{4\pi|\mathbf{x} - \mathbf{y}|}\right], \tag{8.4.7}$$

$$\left[a^i(\mathbf{x}, t), a^j(\mathbf{y}, t)\right] = 0, \tag{8.4.8}$$

$$\left[\pi^i(\mathbf{x}, t), \pi^j(\mathbf{y}, t)\right] = 0, \tag{8.4.9}$$

and likewise for the matter fields and their conjugates. For the same reason, the constraints (8.2.6) and (8.3.8) still apply

$$\nabla \cdot \mathbf{a} = 0\ , \tag{8.4.10}$$

$$\nabla \cdot \boldsymbol{\pi} = 0\ . \tag{8.4.11}$$

To establish the relation between $\boldsymbol{\pi}$ and $\dot{\mathbf{a}}$, we must use Eq. (8.4.6) to evaluate $\dot{\mathbf{a}}$:

$$\begin{aligned} i\dot{a}_i(\mathbf{x}, t) &= [a_i(\mathbf{x}, t), H_0] \\ &= i\int d^3y\ \left[\delta_{ij}\delta^3(\mathbf{x} - \mathbf{y}) + \frac{\partial^2}{\partial x^i \partial x^j}\frac{1}{4\pi|\mathbf{x} - \mathbf{y}|}\right]\ \pi_j(\mathbf{y}, t)\ . \end{aligned}$$

We can replace $\partial/\partial x^j$ in the second term with $-\partial/\partial y^j$, integrate by parts, and use Eq. (8.4.11), yielding

$$\dot{\mathbf{a}} = \boldsymbol{\pi} \tag{8.4.12}$$

just as in the Heisenberg picture. The field equation is likewise determined by

$$\begin{aligned} i\dot{\pi}_i(\mathbf{x}, t) &= [\pi_i(\mathbf{x}, t), H_0] \\ &= -i\int d^3y\ \left[\delta_{ij}\delta^3(\mathbf{x} - \mathbf{y}) + \frac{\partial^2}{\partial x^i \partial x^j}\frac{1}{4\pi|\mathbf{x} - \mathbf{y}|}\right] \\ &\quad \times (\nabla \times \nabla \times \mathbf{a}(\mathbf{y}, t))_j\ , \end{aligned}$$

which (using Eqs. (8.4.10) and (8.4.12)) just yields the usual wave equation

$$\Box\ \mathbf{a} = 0\ . \tag{8.4.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Since A^0 is not an independent Heisenberg-picture field variable, but rather a functional (8.2.9) of the matter fields and their canonical conjugates that vanishes in the limit of zero charges, we do not introduce any corresponding operator a^0 in the interaction picture, but rather take

$$a^0 = 0\,. \tag{8.4.14}$$

The most general real solution of Eqs. (8.4.10), (8.4.13), and (8.4.14) may be written

$$a^\mu(x) = (2\pi)^{-3/2} \int \frac{d^3p}{\sqrt{2p^0}} \sum_\sigma \left[e^{ip\cdot x} e^\mu(\mathbf{p},\sigma)\; a(\mathbf{p},\sigma) + e^{-ip\cdot x} e^{\mu *}(\mathbf{p},\sigma)\; a^\dagger(\mathbf{p},\sigma) \right], \tag{8.4.15}$$

where $p^0 \equiv |\mathbf{p}|$; $e^\mu(\mathbf{p},\sigma)$ are any two independent 'polarization vectors' satisfying

$$\mathbf{p}\cdot\;\mathbf{e}(\mathbf{p},\sigma) = 0\,, \tag{8.4.16}$$

$$e^0(\mathbf{p},\sigma) = 0\,, \tag{8.4.17}$$

and $a(\mathbf{p},\sigma)$ are a pair of operator coefficients, with σ a two-valued index. By adjusting the normalization of $a(\mathbf{p},\sigma)$, we can normalize the $e^\mu(\mathbf{p},\sigma)$ so that the completeness relation reads

$$\sum_\sigma e^i(\mathbf{p},\sigma) e^j(\mathbf{p},\sigma)^* = \delta_{ij} - p_i\; p_j/|\mathbf{p}|^2\,. \tag{8.4.18}$$

For instance, we could take the $\mathbf{e}(\mathbf{p},\sigma)$ to be the same polarization vectors that we encountered in Section 5.9:

$$e^\mu(\mathbf{p},\pm 1) = R(\hat{\mathbf{p}}) \begin{bmatrix} 1/\sqrt{2} \\ \pm i/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}, \tag{8.4.19}$$

where $R(\hat{\mathbf{p}})$ is a standard rotation that carries the three-axis into the direction of $\mathbf{p}$. Using Eqs. (8.4.18) and (8.4.12), we can easily see that the commutation relations (8.4.7)–(8.4.9) are satisfied if (and in fact only if) the operator coefficients in Eq. (8.4.15) satisfy

$$\left[a(\mathbf{p},\sigma),\; a^\dagger(\mathbf{p}',\sigma') \right] = \delta^3(\mathbf{p}-\mathbf{p}')\;\delta_{\sigma\sigma'}, \tag{8.4.20}$$

$$\left[a(\mathbf{p},\sigma),\; a(\mathbf{p}',\sigma') \right] = 0\,. \tag{8.4.21}$$

As remarked before for massive particles, this result should be regarded not so much as an alternative derivation of Eqs. (8.4.20) and (8.4.21), but rather as a verification that Eq. (8.4.2) gives the correct Hamiltonian for free massless particles of helicity ± 1. In the same spirit one can also use Eqs. (8.4.12) and (8.4.15) in Eq. (8.4.2) to calculate the free-photon

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Hamiltonian

$$H_0 = \int d^3p \sum_\sigma \tfrac{1}{2}p^0 \left[a(\mathbf{p},\sigma),\, a^\dagger(\mathbf{p},\sigma)\right]_+$$
$$= \int d^3p \sum_\sigma p^0 \left(a^\dagger(\mathbf{p},\sigma)a(\mathbf{p},\sigma) + \tfrac{1}{2}\delta^3(\mathbf{p}-\mathbf{p})\right) \qquad (8.4.22)$$

which (aside from an inconsequential infinite c-number term) is just what we should expect.

Finally, we record that the interaction (8.4.4) in the interaction picture is

$$V(t) = -\int d^3x\; j_\mu(\mathbf{x},t)\; a^\mu(\mathbf{x},t) + V_{\text{Coul}}(t) + V_{\text{matter}}(t)\,, \qquad (8.4.23)$$

where in terms of the current J in the Heisenberg picture

$$j_\mu(\mathbf{x},t) \equiv \exp(iH_0t)\; J_\mu(\mathbf{x},0) \exp(-iH_0t)\,, \qquad (8.4.24)$$

while $V_{\text{Coul}}(t)$ is the Coulomb term

$$V_{\text{Coul}}(t) = \exp(iH_0t)V_{\text{Coul}} \exp(-iH_0t)$$
$$= \frac{1}{2}\int d^3x\; d^3y\; \frac{j^0(\mathbf{x},t)j^0(\mathbf{y},t)}{4\pi|\mathbf{x}-\mathbf{y}|} \qquad (8.4.25)$$

and $V_{\text{matter}}(t)$ is the non-electromagnetic part of the matter field interaction in the interaction picture:

$$V_{\text{matter}}(t) = \exp(iH_0t)\; V_{\text{matter}} \exp(-iH_0t)\,. \qquad (8.4.26)$$

We have written $j_\mu a^\mu$ instead of $\mathbf{j}\cdot\mathbf{a}$ in Eq. (8.4.23), but these are equal because a^μ has been defined to have $a^0 = 0$.

8.5 The Photon Propagator

The general Feynman rules described in Chapter 6 dictate that an internal photon line in a Feynman diagram contributes a factor to the corresponding term in the S-matrix, given by the propagator:

$$-i\Delta_{\mu\nu}(x-y) \equiv (\Phi_{\text{VAC}},\; T\,\{a_\mu(x),\, a_\nu(y)\}\;\Phi_{\text{VAC}})\;, \qquad (8.5.1)$$

where T as usual denotes a time-ordered product. Inserting our formula (8.4.15) for the electromagnetic potential then yields

$$-i\Delta_{\mu\nu}(x-y) = \int \frac{d^3p}{(2\pi)^3 2|\mathbf{p}|} P_{\mu\nu}(\mathbf{p}) \left[e^{ip\cdot(x-y)}\theta(x-y) + e^{ip\cdot(y-x)}\theta(y-x)\right]\,, \qquad (8.5.2)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

where

$$P_{\mu\nu}(\mathbf{p}) \equiv \sum_{\sigma=\pm 1} e_\mu(\mathbf{p},\sigma)\, e_\nu(\mathbf{p},\sigma)^* \tag{8.5.3}$$

and p^μ in the exponentials is taken with $p^0 = |\mathbf{p}|$. We recall from Eqs. (8.4.18) and (8.4.17) that

$$\begin{aligned} P_{ij}(\mathbf{p}) &= \delta_{ij} - \frac{p^i p^j}{|\mathbf{p}|^2}\,, \\ P_{0i}(\mathbf{p}) &= P_{i0}(\mathbf{p}) = P_{00}(\mathbf{p}) = 0\,. \end{aligned} \tag{8.5.4}$$

As we saw in Chapter 6, the theta functions in Eq. (8.5.2) may be expressed as integrals over an independent time-component q^0 of an off-shell four-momentum q^μ, so that Eq. (8.5.2) may be rewritten

$$\Delta_{\mu\nu}(x-y) = (2\pi)^{-4} \int d^4q\, \frac{P_{\mu\nu}(\mathbf{q})}{q^2 - i\epsilon}\, e^{iq\cdot(x-y)}\,. \tag{8.5.5}$$

Thus in using the Feynman rules in momentum space, the contribution of an internal photon line carrying four-momentum q that runs between vertices where the photon is created and destroyed by fields a^μ and a^ν is

$$\frac{-i}{(2\pi)^4}\, \frac{P_{\mu\nu}(\mathbf{q})}{q^2 - i\epsilon}\,. \tag{8.5.6}$$

It will be very useful (though apparently perverse) to rewrite Eq. (8.5.4) as

$$P_{\mu\nu}(\mathbf{q}) = \eta_{\mu\nu} + \frac{q^0 q_\mu n_\nu + q^0 q_\nu n_\mu - q_\mu q_\nu + q^2 n_\mu n_\nu}{|\mathbf{q}|^2}\,, \tag{8.5.7}$$

where $n^\mu \equiv (0,0,0,1)$ is a fixed time-like vector, q^2 as usual is $\mathbf{q}^2 - (q^0)^2$, but q^0 is here entirely arbitrary. We shall choose q^0 in Eq. (8.5.7) to be given by four-momentum conservation: it is the difference of the matter p^0s flowing in and out of the vertex where the photon line is created. The terms proportional to q_μ and/or q_ν then do not contribute to the S-matrix, because the factors q_μ or q_ν act like derivatives ∂_μ and ∂_ν, and the photon fields a_μ and a_ν are coupled to currents j^μ and j^ν that satisfy the conservation condition $\partial_\mu j^\mu = 0$.* The term proportional to $n_\mu n_\nu$ contains a factor q^2 that cancels the q^2 in the denominator of the propagator, yielding a term that is the same as would be produced by a term in the action:

$$-i\frac{1}{2} \int d^4x \int d^4y\, [-ij^0(x)][-ij^0(y)] \frac{-i}{(2\pi)^4} \int \frac{d^4q}{|\mathbf{q}|^2}\, e^{iq\cdot(x-y)}\,.$$

* This argument as given here is little better than hand-waving. The result has been justified by a detailed analysis of Feynman diagrams,[3] but the easiest way to treat this problem is by path-integral methods, as discussed in Section 9.6.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

The integral over q^0 here yields a delta function in time, so this is equivalent to a correction to the interaction Hamiltonian $V(t)$, of the form

$$-\frac{1}{2}\int d^3x \int d^3y \, \frac{j^0(\mathbf{x},t)\, j^0(\mathbf{y},t)}{4\pi|\mathbf{x}-\mathbf{y}|} \, .$$

This is just right to cancel the Coulomb interaction (8.4.25). Our result is that the photon propagator can be taken effectively as the covariant quantity

$$\Delta^{\text{eff}}_{\mu\nu}(x-y) = (2\pi)^{-4} \int d^4q \frac{\eta_{\mu\nu}}{q^2 - i\epsilon} e^{iq\cdot(x-y)} \tag{8.5.8}$$

with the Coulomb interaction dropped from now on. We see that the apparent violation of Lorentz invariance in the instantaneous Coulomb interaction is cancelled by another apparent violation of Lorentz invariance, that as noted in Section 5.9 the fields $a^\mu(x)$ are not four-vectors, and therefore have a non-covariant propagator. From a practical point of view, the important point is that in the momentum space Feynman rules, the contribution of an internal photon line is simply given by

$$\frac{-i}{(2\pi)^4} \frac{\eta_{\mu\nu}}{q^2 - i\epsilon} \tag{8.5.9}$$

and the Coulomb interaction is dropped.

8.6 Feynman Rules for Spinor Electrodynamics

We are now in a position to state the Feynman rules for calculating the S-matrix in quantum electrodynamics. For definiteness, we will consider the electrodynamics of a single species of spin $\frac{1}{2}$ particles of charge $q = -e$ and mass m. We will call these fermions electrons, but the same formalism applies to muons and other such particles. The simplest gauge- and Lorentz-invariant Lagrangian for this theory is*

$$\mathscr{L} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \bar{\Psi}\left(\gamma^\mu[\partial_\mu + ie\,A_\mu] + m\right)\Psi \, . \tag{8.6.1}$$

The electric current four-vector is then simply

$$J^\mu = \frac{\partial \mathscr{L}}{\partial A_\mu} = -ie\,\bar{\Psi}\gamma^\mu\Psi \, . \tag{8.6.2}$$

* In Chapter 12 we will discuss reasons why more complicated terms are excluded from the Lagrangian density.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

The interaction (8.4.23) in the interaction picture is here

$$V(t) = +ie \int d^3x \ (\bar{\psi}(\mathbf{x}, t)\gamma^\mu \psi(\mathbf{x}, t)) \ a_\mu(\mathbf{x}, t) + V_{\text{Coul}}(t) \,. \tag{8.6.3}$$

(There is no V_{matter} here.) As we have seen, the Coulomb term $V_{\text{Coul}}(t)$ just serves to cancel a part of the photon propagator that is non-covariant and local in time.

Following the general results of Section 6.3, we can state the momentum space Feynman rules for the connected part of the S-matrix in this theory as follows:

(i) Draw all Feynman diagrams with up to some given number of vertices. The diagrams consist of electron lines carrying arrows and photon lines without arrows, with the lines joined at vertices, at each of which there is one incoming and one outgoing electron line and one photon line. There is one external line coming into the diagram from below or going upwards out of the diagram for each particle in the initial or final states, respectively; electrons are represented by external lines carrying arrows pointing upwards into or out of the diagram, while positrons are represented by lines carrying arrows pointing downwards into or out of the diagram. There are also as many internal lines as are needed to give each vertex the required number of attached lines. Each internal line is labelled with an off-mass-shell four-momentum flowing in a definite direction along the line (taken conventionally to flow along the direction of the arrow for electron lines.) Each external line is labelled with the momentum and spin z-component or helicity of the electron or photon in the initial and final states.

(ii) Associate factors with the components of the diagram as follows:

Vertices

Label each vertex with a four-component Dirac index α at the electron line with its arrow coming into the vertex, a Dirac index β at the electron line with its arrow going out of the vertex, and a spacetime index μ at the photon line. For each such vertex, include a factor

$$(2\pi)^4 \, e \, (\gamma^\mu)_{\beta\alpha} \delta^4(k - k' + q) \,, \tag{8.6.4}$$

where k and k' are the electron four-momenta entering and leaving the vertex, and q is the photon four-momentum entering the vertex (or minus the photon momentum leaving the vertex).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

External lines:

Label each external line with the three-momentum **p** and spin z-component or helicity σ of the particle in the initial or final state. For each line for an electron in the final state running out of a vertex carrying a Dirac label β on this line, include a factor*

$$\frac{\bar{u}_\beta(\mathbf{p},\sigma)}{(2\pi)^{3/2}} \,. \tag{8.6.5}$$

For each line for a positron in the final state running into a vertex carrying a Dirac label α on this line, include a factor

$$\frac{v_\alpha(\mathbf{p},\sigma)}{(2\pi)^{3/2}} \,. \tag{8.6.6}$$

For each line for an electron in the initial state running into a vertex carrying a Dirac label α on this line, include a factor

$$\frac{u_\alpha(\mathbf{p},\sigma)}{(2\pi)^{3/2}} \,. \tag{8.6.7}$$

For each line for a positron in the initial state running out of a vertex carrying a Dirac label β on this line, include a factor

$$\frac{\bar{v}_\beta(\mathbf{p},\sigma)}{(2\pi)^{3/2}} \,. \tag{8.6.8}$$

The us and vs are the four-component spinors discussed in Section 5.5. For each line for a photon in the final state connected to a vertex carrying a spacetime label μ on this line, include a factor

$$\frac{e^*_\mu(\mathbf{p},\sigma)}{(2\pi)^{3/2}\sqrt{2p^0}} \,. \tag{8.6.9}$$

For each line for a photon in the initial state connected to a vertex carrying a spacetime label μ on this line, include a factor

$$\frac{e_\mu(\mathbf{p},\sigma)}{(2\pi)^{3/2}\sqrt{2p^0}} \,. \tag{8.6.10}$$

The e_μ are the photon polarization four-vectors described in the previous section.

Internal lines:

For each internal electron line carrying a four-momentum k and running from a vertex carrying a Dirac label β to another vertex carrying a Dirac

* A matrix β has been extracted from the interaction in (8.6.4), so that $\bar{u}$ and $\bar{v}$ appear instead of $u^\dagger$ and $v^\dagger$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

label α, include a factor

$$\frac{-i}{(2\pi)^4}\,\frac{[-i\not{k}+m]_{\alpha\beta}}{k^2+m^2-i\epsilon}\,. \tag{8.6.11}$$

(We are here using the very convenient 'Dirac slash' notation; for any four-vector v^μ, $\not{v}$ denotes $\gamma_\mu v^\mu$.) For each internal photon line carrying a four-momentum q that runs between two vertices carrying spacetime labels μ and ν include a factor

$$\frac{-i}{(2\pi)^4}\,\frac{\eta_{\mu\nu}}{q^2-i\epsilon}\,. \tag{8.6.12}$$

(iii) Integrate the product of all these factors over the four-momenta carried by the internal lines, and sum over all Dirac and spacetime indices.

(iv) Add up the results obtained in this way from each Feynman diagram.

Additional combinatoric factors and fermionic signs may need to be included, as described in parts (v) and (vi) of Section 6.1.

The difficulty of evaluating Feynman diagrams increases rapidly with the number of internal lines and vertices, so it is important to have some idea of what numerical factors tend to suppress the contributions of the more complicated diagrams. We shall estimate these numerical factors including not only the factors of the electronic charge e associated with vertices, but also the factors of 2 and π from vertices, propagators, and momentum space integrals.

Consider a connected Feynman diagram with V vertices, I internal lines, E external lines, and L loops. These quantities are not independent, but are subject to relations already used in Section 6.3:

$$L = I - V + 1\,, \qquad 2I + E = 3V\,.$$

There is a factor $e(2\pi)^4$ from each vertex, a factor $(2\pi)^{-4}$ from each internal line, and a four-dimensional momentum space integral for each loop. The volume element in four-dimensional Euclidean space in terms of a radius parameter κ is $\pi^2\kappa^2 d\kappa^2$, so each loop contributes a factor π^2. Thus the diagram will contain a factor

$$(2\pi)^{4V} e^V (2\pi)^{-4I} \pi^{2L} = (2\pi)^4 e^{E-2} \left(\frac{e^2}{16\pi^2}\right)^L\,.$$

The number E of external lines is fixed for a given process, so we see that the expansion parameter that governs the suppression of Feynman graphs for each additional loop is

$$\frac{e^2}{16\pi^2} = \frac{\alpha}{4\pi} = 5.81 \times 10^{-4}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Fortunately this is small enough that good accuracy can usually be obtained from Feynman diagrams with at most a few loops.

* * *

We must say a little more about the spin states of photons and electrons in realistic experiments, where not every particle in the initial and final states has a definite known helicity or spin z-component. This consideration is especially important for photons, which in practice are often characterized by a state of transverse or elliptical polarization rather than helicity. As we saw in the previous section, for photons of helicity ± 1, the polarization vectors are

$$e(\mathbf{p}, \pm 1) = R(\hat{\mathbf{p}}) \begin{bmatrix} 1/\sqrt{2} \\ \pm i/\sqrt{2} \\ 0 \\ 0 \end{bmatrix} ,$$

where $R(\hat{\mathbf{p}})$ is the standard rotation that takes the z-axis to the $\mathbf{p}$ direction. These are not the only possible photon states; in general, a photon state can be a linear combination of helicity states $\Psi_{\mathbf{p},\pm 1}$

$$\alpha_+ \Psi_{\mathbf{p},+1} + \alpha_- \Psi_{\mathbf{p},-1} \tag{8.6.13}$$

which is properly normalized if

$$|\alpha_+|^2 + |\alpha_-|^2 = 1 . \tag{8.6.14}$$

To calculate the S-matrix element for absorbing or emitting such a photon, we simply replace $e_\mu(\mathbf{p}, \pm 1)$ in the Feynman rules with

$$e_\mu(\mathbf{p}) = \alpha_+ \, e_\mu(\mathbf{p}, +1) + \alpha_- \, e_\mu(\mathbf{p}, -1) . \tag{8.6.15}$$

The polarization vectors for definite helicity satisfy the normalization condition

$$e^*_\mu(\mathbf{p}, \lambda') \, e^\mu(\mathbf{p}, \lambda) = \delta_{\lambda' \lambda} \tag{8.6.16}$$

and therefore in general

$$e^*_\mu(\mathbf{p}) \, e^\mu(\mathbf{p}) = 1 . \tag{8.6.17}$$

The two extreme cases are *circular polarization*, for which $\alpha_- = 0$ or $\alpha_+ = 0$, and *linear polarization*, for which $|\alpha_+| = |\alpha_-| = 1/\sqrt{2}$. For linear polarization, by an adjustment of the overall phase of the state (8.6.13), we can make α_+ and α_- complex conjugates, so that they can be expressed as

$$\alpha_\pm = \exp(\mp i\phi) \Big/ \sqrt{2} . \tag{8.6.18}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Then in the Feynman rules we should use a polarization vector

$$e_\mu(\mathbf{p}) = R(\hat{\mathbf{p}}) \begin{bmatrix} \cos\phi \\ \sin\phi \\ 0 \\ 0 \end{bmatrix} . \tag{8.6.19}$$

That is, ϕ is the azimuthal angle of the photon polarization in the plane perpendicular to $\mathbf{p}$. Note that the photon polarization vector here is *real*, which is only possible for linear polarization. In between the extremes of circular and linear polarization are the states of *elliptic* polarization, for which $|\alpha_+|$ and $|\alpha_-|$ are non-zero and unequal.

More generally, an initial photon may be prepared in a statistical mixture of spin states. In the most general case, an initial photon may have any number of possible polarization vectors $e_\mu^{(r)}(\mathbf{p})$, each with probability P_r. The rate for absorbing such a photon in a given process will then be of the form

$$\Gamma = \sum_r P_r |e_\mu^{(r)}(\mathbf{p})\, M^\mu|^2 = M^{\mu *} M^\nu \rho_{\nu\mu} , \tag{8.6.20}$$

where ρ is the *density matrix*

$$\rho_{\nu\mu} \equiv \sum_r P_r\, e_\nu^{(r)}(\mathbf{p})\, e_\mu^{(r)*}(\mathbf{p}) . \tag{8.6.21}$$

Since ρ is obviously a Hermitian positive matrix of unit trace (because $\sum_r P_r = 1$) with $\rho_{\nu 0} = \rho_{0\mu} = 0$ and $\rho_{\nu\mu} p^\nu = \rho_{\nu\mu} p^\mu = 0$, it may be written as

$$\rho_{\nu\mu} = \sum_{s=1,2} \lambda_s\, e_\nu(\mathbf{p}; s)\, e_\mu^*(\mathbf{p}; s) , \tag{8.6.22}$$

where $e_\mu(\mathbf{p}; s)$ are the two orthonormal eigenvectors of ρ with

$$e_0(\mathbf{p}; s) = e_\mu(\mathbf{p}; s) p^\mu = 0 \tag{8.6.23}$$

and λ_s are the corresponding eigenvalues, with

$$\lambda_s \geq 0 , \qquad \sum_{s=1,2} \lambda_s = 1 .$$

We may then write the rate for the photon absorption process as

$$\Gamma = \sum_{s=1,2} \lambda_s |e_\nu(\mathbf{p}; s) M^\nu|^2 . \tag{8.6.24}$$

Thus any statistical mixture of initial photon states is always equivalent to having just two orthonormal polarizations $e_\nu(\mathbf{p}; s)$ with probabilities λ_s.

In particular, if we know nothing whatever about the initial photon polarization, then the two probabilities λ_s for the polarization vectors

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

$e_v(\mathbf{p};s)$ are equal, so that $\lambda_1 = \lambda_2 = \frac{1}{2}$, and the density matrix (and hence the absorption rate) is an average over initial polarizations

$$\rho_{ij} = \tfrac{1}{2} \sum_{s=1,2} e_i(\mathbf{p};s)\, e_j^*(\mathbf{p};s) = \tfrac{1}{2} \left(\delta_{ij} - \hat{p}_i \hat{p}_j\right) . \tag{8.6.25}$$

Fortunately, this result does not depend on the particular pair of polarization vectors $e_i(\mathbf{p};s)$ over which we average; for unpolarized photons we can average the absorption rate over any pair of orthonormal polarization vectors. Similarly, if we make no attempt to measure the polarization of a photon in the final state, then the rate may be calculated by summing over any pair of orthonormal final photon polarization vectors.

The same remarks apply to electrons and positrons; if (as is usually the case) we make no attempt to prepare an electron or positron so that some spin states are more likely than others, then the rate is to be calculated by *averaging* over any two orthonormal initial spin states, such as those with spin z-component $\sigma = \pm\frac{1}{2}$; if we make no attempt to measure a final electron's or positron's spin state, then we must *sum* the rate over any two orthonormal final spin states, such as those with spin z-component $\sigma = \pm\frac{1}{2}$. Such sums may be performed using the relations (5.5.37) and (5.5.38):

$$\sum_\sigma u_\alpha(\mathbf{p},\sigma)\bar{u}_\beta(\mathbf{p},\sigma) = \left(\frac{-i\not{p} + m}{2p^0}\right)_{\alpha\beta} , \tag{8.6.26}$$

$$\sum_\sigma v_\alpha(\mathbf{p},\sigma)\bar{v}_\beta(\mathbf{p},\sigma) = \left(\frac{-i\not{p} - m}{2p^0}\right)_{\alpha\beta} , \tag{8.6.27}$$

where $p^0 = \sqrt{\mathbf{p}^2 + m^2}$. For instance, if the initial state contains an electron with momentum $\mathbf{p}$ and spin z-component σ, and a positron with momentum $\mathbf{p}'$ and spin z-component σ', then the S-matrix element for the process will be of the form $(\bar{v}_\alpha(\mathbf{p}',\sigma')\,\mathcal{M}_{\alpha\beta}\,u_\beta(\mathbf{p},\sigma))$. Hence if neither electron nor positron spins are observed, the rate will be proportional to

$$\begin{aligned} &\tfrac{1}{4} \sum_{\sigma',\sigma} \left|(\bar{v}_\alpha(\mathbf{p}',\sigma')\,\mathcal{M}_{\alpha\beta}\,u_\beta(\mathbf{p},\sigma))\right|^2 \\ &\quad = \tfrac{1}{4}\mathrm{Tr}\left\{\beta\,\mathcal{M}^\dagger\,\beta\left(\frac{-i\not{p}' - m}{2p^{0\prime}}\right)\mathcal{M}\left(\frac{-i\not{p} + m}{2p^0}\right)\right\} . \end{aligned}$$

Techniques for the calculation of such traces are described in the Appendix to this chapter.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

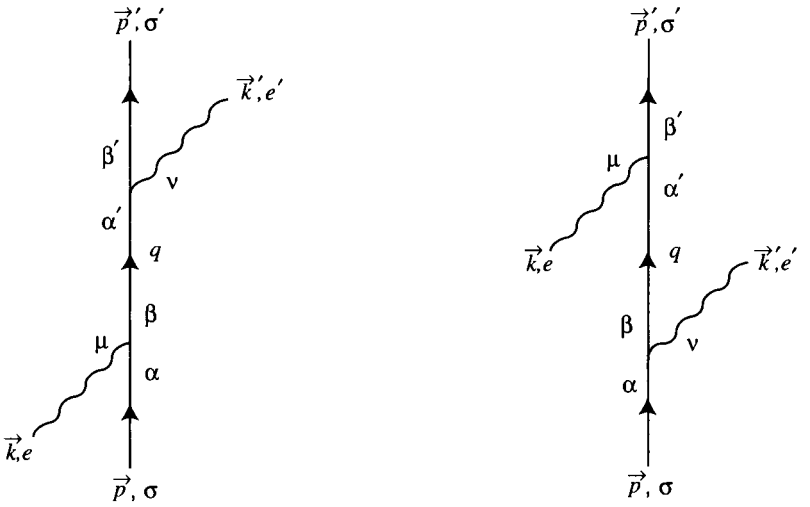

Figure 8.1. The two lowest-order Feynman diagrams for Compton scattering. Straight lines are electrons; wavy lines are photons.

8.7 Compton Scattering

As an example of the methods described in this chapter, we shall consider here the scattering of a photon by an electron (or other particle of spin $\frac{1}{2}$ and charge $-e$), to lowest order in e. We label the initial and final photon momenta and polarization vectors by k^μ, e^μ and k'^μ, e'^μ, where $k^0 = |\mathbf{k}|$ and $k^{0'} = |\mathbf{k}'|$. Also, the initial and final electron momenta and spin z-components are labelled p^μ, σ and p'^μ, σ', where $p^0 = \sqrt{\mathbf{p}^2 + m^2}$ and $p'^0 = \sqrt{\mathbf{p}'^2 + m^2}$, with m the electron mass. The lowest order Feynman diagrams for this process are shown in Figure 8.1. Using the rules outlined in the previous section, the corresponding S-matrix element is

$$
\begin{aligned}
&S(\mathbf{p},\sigma \,+\, \mathbf{k},e \,\rightarrow\, \mathbf{p}',\sigma' \,+\, \mathbf{k}',e') = \\
&\frac{\bar{u}(\mathbf{p}',\sigma')_{\beta'}}{(2\pi)^{3/2}} \frac{e_\nu^{*'}}{(2\pi)^{3/2}\sqrt{2k^{0'}}} \frac{u(\mathbf{p},\sigma)_\alpha}{(2\pi)^{3/2}} \frac{e_\mu}{(2\pi)^{3/2}\sqrt{2k^0}} \\
&\times \int d^4q \left[\frac{-i}{(2\pi)^4}\right] \left[\frac{-i\,\not{q} + m}{q^2 + m^2 - i\epsilon}\right]_{\alpha'\beta} \\
&\times \Big\{ \left[e(2\pi)^4 \gamma^\nu_{\beta'\alpha'} \delta^4(q - p' - k')\right] \left[e(2\pi)^4 \gamma^\mu_{\beta\alpha} \delta^4(q - p - k)\right] \\
&+ \left[e(2\pi)^4 \gamma^\mu_{\beta'\alpha'} \delta^4(q + k - p')\right] \left[e(2\pi)^4 \gamma^\nu_{\beta\alpha} \delta^4(q + k' - p)\right] \Big\} . \qquad (8.7.1)
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Performing the (trivial) q-integral, collecting factors of i and 2π, and rewriting the result in matrix notation, we have more simply

$$S = \frac{-ie^2\delta^4(p'+k'-p-k)}{(2\pi)^2\sqrt{2k^{0\prime}\cdot 2k^0}}\bar{u}(\mathbf{p}',\sigma')\left[\not{\epsilon}^{*\prime}\left(\frac{-i(\not{p}+\not{k})+m}{(p+k)^2+m^2}\right)\not{\epsilon}\right.$$
$$\left. + \not{\epsilon}\left(\frac{-i(\not{p}-\not{k}')+m}{(p-k')^2+m^2}\right)\not{\epsilon}^{*\prime}\right]u(\mathbf{p},\sigma)\,. \tag{8.7.2}$$

(Here $\not{\epsilon}^*$ means $e^*_\mu\gamma^\mu$, not $(\not{\epsilon})^*$. Also, we drop the $-i\epsilon$, because the denominators here do not vanish.) Because $p^2 = -m^2$ and $k^2 = k'^2 = 0$, the denominators can be simplified

$$(p+k)^2+m^2 = 2p\cdot k\,, \tag{8.7.3}$$
$$(p-k')^2+m^2 = -2p\cdot k'\,. \tag{8.7.4}$$

Also, the 'Feynman amplitude' M is defined in general by Eq. (3.3.2), which (because some scattering is assumed to take place) here reads

$$S = -2\pi i\delta^4(p'+k'-p-k)M\,, \tag{8.7.5}$$

so

$$M = \frac{e^2}{4(2\pi)^3\sqrt{k^0k^{0\prime}}}\bar{u}(\mathbf{p}'\sigma')\left\{\not{\epsilon}^{*\prime}\left[-i(\not{p}+\not{k})+m\right]\not{\epsilon}\Big/p\cdot k\right.$$
$$\left. - \not{\epsilon}\left[-i(\not{p}-\not{k}')+m\right]\not{\epsilon}^{*\prime}\Big/p\cdot k'\right\}u(\mathbf{p},\sigma)\,. \tag{8.7.6}$$

The differential cross-section is given in terms of M by Eq. (3.4.15), which here reads

$$d\sigma = (2\pi)^4u^{-1}|M|^2\delta^4(p'+k'-p-k)d^3p'd^3k'\,. \tag{8.7.7}$$

Since one of the particles here is massless, Eq. (3.4.17) for the initial velocity gives

$$u = |p\cdot k|\Big/p^0k^0\,. \tag{8.7.8}$$

To go further, it will be convenient to adopt a specific coordinate frame. Since electrons in atoms move non-relativistically, the laboratory frame for high-energy (X ray or gamma ray) photon–electron scattering experiments is usually (though not always) one in which the initial electron can be taken to be at rest. We will adopt this frame here, so that

$$\mathbf{p} = 0\,, \qquad p^0 = m\,. \tag{8.7.9}$$

The velocity (8.7.8) is then simply

$$u = 1\,. \tag{8.7.10}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

To save writing, we denote the photon energies by

$$\omega = k^0 = |\mathbf{k}| = -p \cdot k \big/ m \,, \tag{8.7.11}$$

$$\omega' = k^{0'} = |\mathbf{k}'| = -p \cdot k' \big/ m \,. \tag{8.7.12}$$

The three-momentum delta function in Eq. (8.7.7) just serves to eliminate the differential d^3p', setting $\mathbf{p}' = \mathbf{k} - \mathbf{k}'$. This leaves the remaining energy delta function

$$\delta(p'^0 + k'^0 - p^0 - k^0) = \delta\left(\sqrt{(\mathbf{k} - \mathbf{k}')^2 + m^2} + \omega' - m - \omega\right) \,. \tag{8.7.13}$$

This fixes ω' to satisfy

$$\sqrt{\omega^2 - 2\omega\omega' \cos\theta + \omega'^2 + m^2} = \omega + m - \omega' \,,$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$. Squaring both sides and cancelling ω'^2 terms gives*

$$\omega' = \omega \, \frac{m}{m + \omega(1 - \cos\theta)} \equiv \omega_c(\theta) \,. \tag{8.7.14}$$

The energy delta function (8.7.13) can be written

$$\begin{aligned} \delta(p'^0 + k'^0 - p^0 - k^0) &= \frac{\delta(\omega' - \omega_c(\theta))}{|\partial[\sqrt{\omega^2 - 2\omega\omega'\cos\theta + \omega'^2 + m^2} + \omega']/\partial\omega'|} \\ &= \frac{\delta(\omega' - \omega_c(\theta))}{|(\omega' - \omega\,\cos\theta)/p'^0 + 1|} \\ &= \frac{p'^0\omega'}{m\omega}\delta(\omega' - \omega_c(\theta)) \,. \end{aligned} \tag{8.7.15}$$

Also, the differential d^3k' can be written

$$d^3k' = \omega'^2 d\omega' d\Omega \,, \tag{8.7.16}$$

where $d\Omega$ is the solid angle into which the final photon is scattered. The final delta function in Eq. (8.7.15) just serves to eliminate the differential $d\omega'$ in Eq. (8.7.16), leaving us with a differential cross-section

$$d\sigma = (2\pi)^4 |M|^2 \, \frac{p'^0\omega'^3}{m\omega} \, d\Omega \tag{8.7.17}$$

with $p'^0 = m + \omega - \omega'$, and ω' given by Eq. (8.7.14).

---

* Equivalently, there is an increase in wavelength

$$\frac{1}{\omega'} - \frac{1}{\omega} = \frac{1 - \cos\theta}{m} \,.$$

The verification of this formula in the scattering of X rays by electrons by A.H. Compton in 1922–3 played a key role in confirming Einstein's 1905 proposal of a quantum of light, which soon after Compton's experiments came to be known as the photon.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Usually we do not measure the spin z-component of the initial or final electron. In such cases, we must sum over σ' and average over σ, or in other words take half the sum over σ and σ':

$$d\bar{\sigma}(\mathbf{p}+\mathbf{k},e\to\mathbf{p}'+\mathbf{k}',e') \equiv \tfrac{1}{2}\sum_{\sigma',\sigma} d\sigma(\mathbf{p},\sigma+\mathbf{k},e\to\mathbf{p}',\sigma'+\mathbf{k}',e')\,. \tag{8.7.18}$$

To calculate this, we use the standard formula

$$\sum_{\sigma} u_\alpha(\mathbf{p},\sigma)\bar{u}_\beta(\mathbf{p},\sigma) = \frac{(-i\not{p}+m)_{\alpha\beta}}{2p^0} \tag{8.7.19}$$

and likewise for the sum over σ'. It follows that for an arbitrary 4×4 matrix A

$$\begin{aligned}\sum_{\sigma,\sigma'} |\bar{u}(\mathbf{p}',\sigma')Au(\mathbf{p},\sigma)|^2 &= \sum_{\sigma,\sigma'} (\bar{u}(\mathbf{p}',\sigma')Au(\mathbf{p},\sigma))(\bar{u}(\mathbf{p},\sigma)\beta A^\dagger\beta u(\mathbf{p}',\sigma')) \\ &= \sum_{\sigma,\sigma'} A_{\beta\alpha}u_\alpha(\mathbf{p},\sigma)\bar{u}_\gamma(\mathbf{p},\sigma)(\beta A^\dagger\beta)_{\gamma\delta}u_\delta(\mathbf{p}',\sigma')\bar{u}_\beta(\mathbf{p}',\sigma') \\ &= \mathrm{Tr}\left\{A\left(\frac{-i\not{p}+m}{2p^0}\right)\beta A^\dagger\beta\left(\frac{-i\not{p}'+m}{2p'^0}\right)\right\}. \end{aligned} \tag{8.7.20}$$

Recalling that $\beta\gamma_\mu^\dagger\beta = -\gamma_\mu$, Eq. (8.7.6) gives now

$$\begin{aligned}\sum_{\sigma,\sigma'}|M|^2 = \frac{e^4}{64(2\pi)^6\omega\omega' p^0 p'^0} \\ \times \mathrm{Tr}\left[\left\{\not{\epsilon}'^*\frac{[-i(\not{p}+\not{k})+m]}{p\cdot k}\not{\epsilon} - \not{\epsilon}\frac{[-i(\not{p}-\not{k}')+m]}{p\cdot k'}\not{\epsilon}'^*\right\}(-i\not{p}+m)\right. \\ \left.\times\left\{\not{\epsilon}^*\frac{[-i(\not{p}+\not{k})+m]}{p\cdot k}\not{\epsilon}' - \not{\epsilon}'\frac{[-i(\not{p}-\not{k}')+m]}{p\cdot k'}\not{\epsilon}^*\right\}(-i\not{p}'+m)\right]. \end{aligned} \tag{8.7.21}$$

(Recall again that $\not{\epsilon}^*$ means $e_\mu^*\gamma^\mu$, not $(e_\mu\gamma^\mu)^*$, and likewise for $\not{\epsilon}'^*$.) We work in a 'gauge' in which

$$e\cdot p = e^*\cdot p = e'\cdot p = e'^*\cdot p = 0 \tag{8.7.22}$$

such as for instance Coulomb gauge in the laboratory frame, where $e^0 = e'^0 = 0$ and $\mathbf{p} = 0$. This implies that

$$\begin{aligned}[-i\not{p}+m]\not{\epsilon}[-i\not{p}+m] &= \not{\epsilon}[i\not{p}+m]\,[-i\not{p}+m] \\ &= \not{\epsilon}(\not{p}^2+m^2) = \not{\epsilon}(p_\mu p^\mu+m^2) = 0\end{aligned}$$

and likewise for $\not{\epsilon}'^*$, $\not{\epsilon}'$, and $\not{\epsilon}^*$. Eq. (8.7.21) can therefore be written in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

the greatly simplified form

$$\sum_{\sigma,\sigma'}|M|^2 = \frac{-e^4}{64(2\pi)^6\omega\omega' p^0 p'^0}\mathrm{Tr}\left[\left\{\frac{\not{\epsilon}'^* \not{k} \not{\epsilon}}{p\cdot k} + \frac{\not{\epsilon} \not{k}' \not{\epsilon}'^*}{p\cdot k'}\right\}(-i\not{p} + m)\right.$$
$$\left.\times\left\{\frac{\not{\epsilon}^* \not{k} \not{\epsilon}'}{p\cdot k} + \frac{\not{\epsilon}' \not{k}' \not{\epsilon}^*}{p\cdot k'}\right\}(-i\not{p}' + m)\right] . \tag{8.7.23}$$

The trace of any product of an odd number of gamma matrices vanishes, so this breaks up into terms of zeroth and second order in m:

$$\sum_{\sigma,\sigma'}|M|^2 = \frac{e^4}{64(2\pi)^6\omega\omega' p^0 p'^0}\left(\frac{T_1}{(p\cdot k)^2} + \frac{T_2}{(p\cdot k)(p\cdot k')} + \frac{T_3}{(p\cdot k)(p\cdot k')}\right.$$
$$\left. + \frac{T_4}{(p\cdot k')^2} - \frac{m^2 t_1}{(p\cdot k)^2} - \frac{m^2 t_2}{(p\cdot k)(p\cdot k')} - \frac{m^2 t_3}{(p\cdot k)(p\cdot k')} - \frac{m^2 t_4}{(p\cdot k')^2}\right) \tag{8.7.24}$$

where

$$T_1 = \mathrm{Tr}\left\{\not{\epsilon}'^* \not{k} \not{\epsilon} \not{p} \not{\epsilon}^* \not{k} \not{\epsilon}' \not{p}'\right\} , \tag{8.7.25}$$
$$T_2 = \mathrm{Tr}\left\{\not{\epsilon}'^* \not{k} \not{\epsilon} \not{p} \not{\epsilon}' \not{k}' \not{\epsilon}^* \not{p}'\right\} , \tag{8.7.26}$$
$$T_3 = \mathrm{Tr}\left\{\not{\epsilon} \not{k}' \not{\epsilon}'^* \not{p} \not{\epsilon}^* \not{k} \not{\epsilon}' \not{p}'\right\} , \tag{8.7.27}$$
$$T_4 = \mathrm{Tr}\left\{\not{\epsilon} \not{k}' \not{\epsilon}'^* \not{p} \not{\epsilon}' \not{k}' \not{\epsilon}^* \not{p}'\right\} , \tag{8.7.28}$$
$$t_1 = \mathrm{Tr}\left\{\not{\epsilon}'^* \not{k} \not{\epsilon} \not{\epsilon}^* \not{k} \not{\epsilon}'\right\} , \tag{8.7.29}$$
$$t_2 = \mathrm{Tr}\left\{\not{\epsilon}'^* \not{k} \not{\epsilon} \not{\epsilon}' \not{k}' \not{\epsilon}^*\right\} , \tag{8.7.30}$$
$$t_3 = \mathrm{Tr}\left\{\not{\epsilon} \not{k}' \not{\epsilon}'^* \not{\epsilon}^* \not{k} \not{\epsilon}'\right\} , \tag{8.7.31}$$
$$t_4 = \mathrm{Tr}\left\{\not{\epsilon} \not{k}' \not{\epsilon}'^* \not{\epsilon}' \not{k}' \not{\epsilon}^*\right\} . \tag{8.7.32}$$

The Appendix to this chapter shows how to calculate any trace $\mathrm{Tr}\{\not{a}\not{b}\not{c}\not{d}\ldots\}$ as a sum of products of scalar products of the four-vectors $a, b, c, d, \ldots$. In general, traces of products of 6 or 8 gamma matrices like the t_k or T_k would be given by a sum of 15 or 105 terms, respectively, but fortunately here most scalar products vanish; in addition to Eq. (8.7.22), we also have $k\cdot k = k'\cdot k' = 0$. (Furthermore, $e\cdot e^* = e'\cdot e'^* = 1$.) To simplify the calculation further, let us specialize to the case of *linear* polarization, where e^μ and e'^μ are real. Dropping the asterisks in Eqs. (8.7.25)–(8.7.32), we have then

$$T_1 = \mathrm{Tr}\left\{\not{\epsilon}' \not{k} \not{\epsilon} \not{p} \not{\epsilon} \not{k} \not{\epsilon}' \not{p}'\right\} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Since $e^\mu p_\mu = 0$ and $e^\mu e_\mu = 1$, we have

$$\not{e}\not{p}\not{e} = -\not{p}\not{e}\not{e} = -\not{p}$$

so

$$T_1 = -\mathrm{Tr}\left\{\not{e}'\not{k}\not{p}\not{k}\not{e}'\not{p}'\right\} .$$

Also, $k^\mu k_\mu = 0$, so

$$\not{k}\not{p}\not{k} = -\not{k}\not{k}\not{p} + 2\not{k}p\cdot k = 2\not{k}p\cdot k$$

and hence

$$T_1 = -2p\cdot k\ \mathrm{Tr}\left\{\not{e}'\not{k}\not{e}'\not{p}'\right\} .$$

Using Eq. (8.A.6), this is

$$T_1 = -8p\cdot k\ [2e'\cdot k\ e'\cdot p' - k\cdot p'] .$$

It is convenient to make the substitutions

$$e'\cdot p' = e'\cdot[p+k-k'] = e'\cdot k$$
$$k\cdot p' = -\tfrac{1}{2}(p'-k)^2 - \tfrac{1}{2}m^2 = -\tfrac{1}{2}(p-k')^2 - \tfrac{1}{2}m^2 = p\cdot k'$$

so

$$T_1 = -16\ p\cdot k(e'\cdot k)^2 + 8\ p\cdot k\ p\cdot k' . \tag{8.7.33}$$

A similar (though more lengthy) calculation gives

$$\begin{aligned} T_2 = T_3 = {} & -8(e\cdot k')^2(p\cdot k) + 16\ (e\cdot e')^2 p\cdot k' p\cdot k + 8(e\cdot e')^2 k\cdot k' m^2 \\ & -8(e\cdot e')m^2(k\cdot e')(k'\cdot e) + 8(e'\cdot k)^2 p\cdot k' \\ & -4(k\cdot p)^2 + 4(k\cdot k')(p\cdot p') - 4(k\cdot p')(p\cdot k') , \end{aligned} \tag{8.7.34}$$

$$T_4 = 16\ p\cdot k'(e\cdot k')^2 + 8(p\cdot k)(p\cdot k') , \tag{8.7.35}$$

$$t_1 = t_4 = 0 , \tag{8.7.36}$$

$$t_2 = t_3 = -8\ e\cdot e'\ k\cdot e'\ k'\cdot e + 8(k\cdot k')(e\cdot e')^2 - 4(k\cdot k') . \tag{8.7.37}$$

Combining all these terms in Eq. (8.7.24) gives

$$\sum_{\sigma,\sigma'} |M|^2 = \frac{e^4}{64(2\pi)^6\omega\omega' p^0 p'^0}\left[\frac{8(k\cdot k')^2}{(k\cdot p)(k'\cdot p)} + 32(e\cdot e')^2\right] . \tag{8.7.38}$$

All this applies in any Lorentz frame. In the *laboratory* frame, we have the special results

$$k\cdot k' = \omega\omega'(\cos\theta - 1) = m\omega\omega'\left(\frac{1}{\omega} - \frac{1}{\omega'}\right) ,$$
$$p\cdot k = -m\omega \qquad p\cdot k' = -m\omega' .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Combining Eq. (8.7.38) with Eq. (8.7.17), the laboratory frame cross-section is

$$\tfrac{1}{2}\sum_{\sigma,\sigma'} d\sigma(\mathbf{p},\sigma + \mathbf{k},e \to \mathbf{p}',\sigma' + \mathbf{k}',e') = \frac{e^4\omega'^2 d\Omega}{64\pi^2 m^2\omega^2}$$
$$\times \left[\frac{\omega}{\omega'} + \frac{\omega'}{\omega} - 2 + 4(e\cdot e')^2\right]. \tag{8.7.39}$$

This is the celebrated formula derived (using old-fashioned perturbation theory) by O. Klein and Y. Nishina[4] in 1929.

As discussed in Section 8.6, if the incoming photon is (as usual) not prepared in a state with any particular polarization, then we must average over two orthonormal values of **e**. This average gives

$$\tfrac{1}{2}\sum_{e} e_i e_j = \tfrac{1}{2}(\delta_{ij} - \hat{\mathbf{k}}_i\hat{\mathbf{k}}_j)$$

and the differential cross-section is then

$$\tfrac{1}{4}\sum_{e,\sigma,\sigma'} d\sigma(\mathbf{p},\sigma + \mathbf{k},e \to \mathbf{p}',\sigma' + \mathbf{k},e') = \frac{e^4\omega'^2 d\Omega}{64\pi^2 m^2\omega^2}\left[\frac{\omega}{\omega'} + \frac{\omega'}{\omega} - 2(\hat{\mathbf{k}}\cdot e')^2\right]. \tag{8.7.40}$$

We see that the scattered photon is preferentially polarized in a direction perpendicular to the incident as well as the final photon direction, i.e., perpendicular to the plane in which the scattering takes place. This is a well-known result, responsible among other things for the polarization of light from eclipsing binary stars.**

To calculate the cross-section for experiments in which the final photon polarization is not measured, we must sum Eq. (8.7.40) over e', using

$$\sum_{e'} e'_i e'_j = \delta_{ij} - \hat{\mathbf{k}}'_i\hat{\mathbf{k}}'_j .$$

This gives

$$\tfrac{1}{4}\sum_{e,e',\sigma,\sigma'} d\sigma(\mathbf{p},\sigma + \mathbf{k},e \to \mathbf{p}',\sigma' + \mathbf{k}',e')$$
$$= \frac{e^4\omega'^2 d\Omega}{32\pi^2 m^2\omega^2}\left[\frac{\omega}{\omega'} + \frac{\omega'}{\omega} - 1 + \cos^2\theta\right], \tag{8.7.41}$$

where θ is the angle between $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}'$. In the non-relativistic case, $\omega << m$,

** The light from one of the stars is polarized when it is scattered by free electrons in the outer atmosphere of the other, cooler, star when both are along the same line of sight. This polarization is normally undetectable because it cancels when the astronomer adds up light from all parts of the star's disk. The polarization has been observed in eclipsing binary stars at times when the cooler star blocks the light from just one side of the hotter star.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

Eq. (8.7.41) gives

$$\tfrac{1}{4}\sum_{e,e',\sigma,\sigma'} d\sigma = \frac{e^4\, d\Omega}{32\pi^2 m^2}(1+\cos^2\theta)\,. \tag{8.7.42}$$

The solid angle integral is

$$\int [1+\cos^2\theta]d\Omega = \int_0^{2\pi} d\phi \int_0^{\pi} [1+\cos^2\theta]\ \sin\theta\ d\theta = \frac{16\pi}{3}\,,$$

giving a total cross-section for $\omega \ll m$:

$$\sigma_{\mathrm{T}} = \frac{e^4}{6\pi m^2}\,. \tag{8.7.43}$$

This is often written $\sigma_{\mathrm{T}} = 8\pi r_0^2/3$, where $r_0 = e^2/4\pi m = 2.818 \times 10^{-13}$ cm is known as the *classical electron radius*. Expression (8.7.43) is called the *Thomson cross-section*, after J. J. Thomson, the discoverer of the electron. Eqs. (8.7.42) and (8.7.43) were originally derived using classical mechanics and electrodynamics, by calculating the reradiation of light by a non-relativistic point charge in a plane wave electromagnetic field.

8.8 Generalization : *p*-form Gauge Fields*

The antisymmetric field strength tensor $F_{\mu\nu}$ of electromagnetism is a special case of a general class of tensors of special importance in physics and mathematics. A *p-form* is an antisymmetric covariant tensor of rank p. From a p-form $t_{\mu_1,\mu_2,\cdots\mu_p}$ one may construct a $(p+1)$-form called the *exterior derivative*** dt by taking the derivative and then antisymmetrizing with respect to all indices:

$$\begin{aligned}(dt)_{\mu_1\mu_2\cdots\mu_{p+1}} &\equiv \partial_{[\mu_1} t_{\mu_2\mu_3\cdots\mu_{p+1}]} \\ &\equiv \partial_{\mu_1} t_{\mu_2\mu_3\cdots\mu_{p+1}} - \partial_{\mu_2} t_{\mu_1\mu_3\cdots\mu_{p+1}} + \cdots + (-1)^p \partial_{\mu_{p+1}} t_{\mu_1\mu_2\cdots\mu_p}\end{aligned} \tag{8.8.1}$$

with square brackets indicating antisymmetrization with respect to the indices within the brackets. Because derivatives commute, repeated exterior derivatives vanish

$$d(dt) = 0\,. \tag{8.8.2}$$

A p-form whose exterior derivative vanishes is called *closed*, while a p-form that is itself an exterior derivative is called *exact*. From Eq. (8.8.2)

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** Exterior derivatives and p-forms play a special role in general relativity, in part because the exterior derivative of a tensor transforms like a tensor even though it is calculated using ordinary rather than covariant derivatives.[5]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

it follows that any exact p-form is closed; a famous theorem[6] of Poincaré states that in a region that is smoothly contractible to a point, any closed p-form is exact.† For instance, the homogeneous Maxwell equations (8.1.16) tell us that the electromagnetic field strength two-form $F_{\mu\nu}$ is closed; Poincaré's theorem then shows that it is also exact, so that it can be written as an exterior derivative, i.e., as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Again using Eq. (8.8.2), we see that the two-form $F_{\mu\nu}$ is invariant if A_μ is changed by an exterior derivative, i.e., by a gauge transformation $\delta A_\mu = \partial_\mu \Omega$.

The formalism of p-forms and exterior derivatives makes it natural to consider the possibility of massless particles described by p-form gauge fields‡ $A_{\mu_1 \cdots \mu_p}$, with an invariance under gauge transformations

$$\delta A = d\Omega \tag{8.8.3}$$

or in more detail

$$\delta A_{\mu_1 \cdots \mu_p} = \partial_{[\mu_1} \Omega_{\mu_2 \cdots \mu_p]},$$

where $\Omega_{\mu_1 \cdots \mu_{p-1}}$ is an arbitrary $(p-1)$-form. From such a p-form gauge field, we can construct a gauge-invariant field strength tensor given by

$$F = dA \tag{8.8.4}$$

or in detail

$$F_{\mu_1 \cdots \mu_{p+1}} = \partial_{[\mu_1} A_{\mu_2 \cdots \mu_{p+1}]} \,. \tag{8.8.5}$$

(Alternatively, we can start with a $(p+1)$-form F, and from an assumed condition $dF = 0$ infer the existence of a p-form A with $F = dA$.) By analogy with electrodynamics, we might expect the Lagrangian density for A to take the form

$$\mathscr{L} = -\frac{1}{2(p+1)} F_{\mu_1 \cdots \mu_{p+1}} F^{\mu_1 \cdots \mu_{p+1}} + J^{\mu_1 \cdots \mu_p} A_{\mu_1 \cdots \mu_p} \,, \tag{8.8.6}$$

where J is an antisymmetric tensor current (either a c-number, or a function of fields other than A) that in order to make the action gauge-invariant must satisfy the conservation condition

$$\partial_{\mu_1} J^{\mu_1 \cdots \mu_p} = 0 \,. \tag{8.8.7}$$

† In multiply connected spaces closed forms are not necessarily exact; although it is possible to write a closed p-form as an exterior derivative locally, this cannot in general be done smoothly throughout the space. The set of closed p-forms, modulo exact p-forms, makes up what is called the pth de Rham cohomology group of the space. There is a deep relation between the de Rham cohomology groups of a space and its topology,[6] which will be discussed further in Volume II.

‡ We are speaking loosely in calling $A_{\mu_1 \cdots \mu_p}$ a p-form, because for $F = dA$ to be a tensor it is only necessary for A to be a tensor up to a gauge transformation. In fact, we have already seen that in four spacetime dimensions, it is not possible to construct a four-vector field from the creation and annihilation operators of physical massless particles of helicity ± 1, so we have to deal with an $A^\mu(x)$ that according to Eq. (8.1.2) transforms as a four-vector only up to a gauge transformation.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

The Euler–Lagrange equations are then

$$\partial_\mu F^{\mu\mu_1\cdots\mu_p} = -J^{\mu_1\cdots\mu_p} \,. \tag{8.8.8}$$

These p-form gauge fields play an important role in theories with more than four spacetime dimensions. For instance, in the simplest string theories in 26 spacetime dimensions, there is a normal mode of the string represented at low energies by a two-form gauge field $A_{\mu\nu}$. But in four spacetime dimensions, p-forms offer no new possibilities.

To see this, note first that in D spacetime dimensions there are no antisymmetric tensors with more than D indices, so in general we must take $p+1 \leq D$. Like any other $(p+1)$-form with $p+1 \leq D$, the field strength F may be expressed in terms of a dual $(D-p-1)$-form $\mathscr{F}$, as

$$F^{\mu_1\cdots\mu_{p+1}} = \epsilon^{\mu_1\cdots\mu_D} \mathscr{F}_{\mu_{p+2}\cdots\mu_D} \,. \tag{8.8.9}$$

Likewise, the p-form current J may be expressed in terms of a dual $(D-p)$-form current $\mathscr{J}$, as

$$J^{\mu_1\cdots\mu_p} = \epsilon^{\mu_1\cdots\mu_D} \mathscr{J}_{\mu_{p+1}\cdots\mu_D} \,. \tag{8.8.10}$$

The field equation (8.8.8) and conservation condition (8.8.7) then read simply

$$d\mathscr{F} = \mathscr{J} \,, \qquad d\mathscr{J} = 0 \,. \tag{8.8.11}$$

Because the dual current $\mathscr{J}$ is closed, it may be written in terms of a $(D-p-1)$-form $\mathscr{S}$ as

$$\mathscr{J} = d\mathscr{S} \,. \tag{8.8.12}$$

Eqs. (8.8.11) and (8.8.12) tell us that the difference of $\mathscr{F}$ and $\mathscr{S}$ is closed, and therefore according to Poincaré's theorem, may be written as

$$\mathscr{F} = \mathscr{S} + d\phi \,, \tag{8.8.13}$$

with ϕ a $(D-p-2)$-form. There is an exception for the case $p = D-1$, where $\mathscr{F}$ and $\mathscr{S}$ are zero-forms, i.e., scalars, and the condition $d\mathscr{F} = d\mathscr{S}$ simply tells us that $\mathscr{F}$ and $\mathscr{S}$ differ only by a constant. In this case the gauge field describes no degrees of freedom at all. We may therefore concern ourselves only with the cases $p \leq D-2$.

For $p \leq D-2$, the homogeneous 'Maxwell' equations $dF = 0$ read

$$\partial_{\mu_1} \mathscr{F}^{\mu_1\cdots\mu_{D-p-1}} = 0 \,, \tag{8.8.14}$$

which with Eq. (8.8.13) yields the field equation for ϕ:

$$\partial_{\mu_1} (d\phi)^{\mu_1\cdots\mu_{D-p-1}} = -\partial_{\mu_1} \mathscr{S}^{\mu_1\cdots\mu_{D-p-1}} \,. \tag{8.8.15}$$

This is invariant under a new set of gauge transformations $\phi \to \phi + d\omega$, except that where $D-p-2 = 0$ the gauge transformation that leaves F

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

invariant is $\phi \to \phi + c$, where c is an arbitrary constant. *We see that in D spacetime dimensions, the theory of a p-form gauge field A is equivalent to the theory of a $(D-p-2)$-form gauge field ϕ.*

We can now understand why p-form gauge fields offer no new possibilities in four spacetime dimensions. As we have seen, we need only consider the cases $p \leq D-2$, or $p = 0$, 1, or 2. A zero-form gauge field is a scalar S, for which Eq. (8.8.5) reads $F_\mu = \partial_\mu S$, and the field equations (8.8.8) read simply $\Box S = -J$. The gauge invariance here is invariance under a shift $S \to S + c$, with c a constant. This is just the theory of a massless scalar field with only derivative interactions. A one-form gauge field is a four-vector $A^\mu(x)$ coupled to a conserved four-vector current, just as in electrodynamics. Finally, according to the general result quoted above, a two-form gauge field in four spacetime dimensions is equivalent to a zero-form gauge field, which as we have seen is equivalent to a massless derivatively coupled scalar field.

Appendix Traces

In calculating S-matrix elements and transition rates for processes involving particles of spin $\frac{1}{2}$, we often encounter traces of products of Dirac gamma matrices. It will therefore be useful to give formulas for these traces that can be used in all such calculations.

For products of *even* numbers of gamma matrices, the trace is given by

$$\mathrm{Tr}\{\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_{2N}}\} = 4\sum_{\text{pairings}}\delta_P\prod_{\text{pairs}}\eta_{\text{paired }\mu\text{s}}\,. \tag{8.A.1}$$

Here the sum is over all different ways of pairing the indices $\mu_1,\cdots\mu_{2N}$. A pairing can be regarded as a permutation of the integers $1, 2,\cdots 2N$ into some order $P1, P2,\cdots P\cdot(2N)$, in which we pair μ_{P1} with μ_{P2}, μ_{P3} with μ_{P4}, and so on. Permuting pairs or permuting μs within a pair yields the same pairing, so the number of different pairings is

$$(2N)!/N!2^N = (2N-1)(2N-3)\cdots 1 \equiv (2N-1)!!\,. \tag{8.A.2}$$

We can avoid summing over equivalent pairings by requiring that

$$P1 < P2,\ P3 < P4,\cdots,\ P\cdot(2N-1) < P\cdot(2N) \tag{8.A.3}$$

and

$$P1 < P3 < P5 < \cdots\,. \tag{8.A.4}$$

With this convention, the factor δ_P is $+1$ or -1 according to whether the pairing involves an even or odd permutation of indices. The product in Eq.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

(8.A.1) is over all N pairs, the nth pair contributing a factor $\eta_{\mu_{P\cdot(2n-1)}\mu_{P\cdot(2n)}}$. For instance (writing $\mu, \nu, \rho, \sigma, \cdots$ in place of $\mu_1, \mu_2, \mu_3, \mu_4, \cdots$), for $N = 1, 2$, and 3 we have*

$$\mathrm{Tr}\{\gamma_\mu\gamma_\nu\} = 4\eta_{\mu\nu}\,, \tag{8.A.5}$$

$$\mathrm{Tr}\{\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\} = 4\left[\eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho}\right], \tag{8.A.6}$$

$$\begin{aligned}\mathrm{Tr}\{\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\gamma_\kappa\gamma_\eta\} = 4\Big[&\eta_{\mu\nu}\eta_{\rho\sigma}\eta_{\kappa\eta} - \eta_{\mu\nu}\eta_{\rho\kappa}\eta_{\sigma\eta} + \eta_{\mu\nu}\eta_{\rho\eta}\eta_{\sigma\kappa}\\ &-\eta_{\mu\rho}\eta_{\nu\sigma}\eta_{\kappa\eta} + \eta_{\mu\rho}\eta_{\nu\kappa}\eta_{\sigma\eta} - \eta_{\mu\rho}\eta_{\nu\eta}\eta_{\sigma\kappa} + \eta_{\mu\sigma}\eta_{\nu\rho}\eta_{\kappa\eta} - \eta_{\mu\sigma}\eta_{\nu\kappa}\eta_{\rho\eta}\\ &+\eta_{\mu\sigma}\eta_{\nu\eta}\eta_{\rho\kappa} - \eta_{\mu\kappa}\eta_{\nu\rho}\eta_{\sigma\eta} + \eta_{\mu\kappa}\eta_{\nu\sigma}\eta_{\rho\eta} - \eta_{\mu\kappa}\eta_{\nu\eta}\eta_{\rho\sigma} + \eta_{\mu\eta}\eta_{\nu\rho}\eta_{\sigma\kappa}\\ &-\eta_{\mu\eta}\eta_{\nu\sigma}\eta_{\rho\kappa} + \eta_{\mu\eta}\eta_{\nu\kappa}\eta_{\rho\sigma}\Big]\,.\end{aligned} \tag{8.A.7}$$

For an odd number of gamma matrices, the result is much simpler

$$\mathrm{Tr}\{\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_{2N+1}}\} = 0\,. \tag{8.A.8}$$

The proof of Eq. (8.A.1) is by mathematical induction. First note that

$$\mathrm{Tr}\{\gamma_\mu\gamma_\nu\} = -\mathrm{Tr}\{\gamma_\nu\gamma_\mu\} + 2\mathrm{Tr}\{\eta_{\mu\nu}\,1\} = -\mathrm{Tr}\{\gamma_\mu\gamma_\nu\} + 8\eta_{\mu\nu}\,,$$

so $\mathrm{Tr}\{\gamma_\mu\gamma_\nu\} = 4\eta_{\mu\nu}$, in agreement with Eq. (8.A.1). Next, suppose that Eq. (8.A.1) is true for $N \le M - 1$. We then have

$$\begin{aligned}\mathrm{Tr}\{\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_{2M}}\} &= 2\eta_{\mu_1\mu_2}\mathrm{Tr}\{\gamma_{\mu_3}\cdots\gamma_{\mu_{2M}}\} - \mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_1}\gamma_{\mu_3}\cdots\gamma_{\mu_{2M}}\}\\ &= 2\eta_{\mu_1\mu_2}\,\mathrm{Tr}\{\gamma_{\mu_3}\cdots\gamma_{\mu_{2M}}\} - 2\eta_{\mu_1\mu_3}\,\mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_4}\cdots\gamma_{\mu_{2N}}\}\\ &\quad + \mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\mu_1}\gamma_{\mu_4}\cdots\gamma_{\mu_{2M}}\}\\ &= 2\eta_{\mu_1\mu_2}\,\mathrm{Tr}\{\gamma_{\mu_3}\cdots\gamma_{\mu_{2M}}\} - 2\eta_{\mu_1\mu_3}\,\mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_4}\cdots\gamma_{\mu_{2M}}\}\\ &\quad + 2\eta_{\mu_1\mu_4}\,\mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\mu_5}\cdots\gamma_{\mu_{2M}}\} - \cdots\\ &\quad + 2\eta_{\mu_1\mu_{2M}}\,\mathrm{Tr}\{\gamma_{\mu_2}\cdots\gamma_{\mu_{2M-1}}\} - \mathrm{Tr}\{\gamma_{\mu_2}\cdots\gamma_{\mu_{2M}}\gamma_{\mu_1}\}\,.\end{aligned}$$

All commutators have zero trace, so the last term subtracted here is the same as the left-hand side, and so

$$\begin{aligned}\mathrm{Tr}\{\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_{2M}}\} &= \eta_{\mu_1\mu_2}\,\mathrm{Tr}\{\gamma_{\mu_3}\cdots\gamma_{\mu_{2M}}\}\\ &-\eta_{\mu_1\mu_3}\,\mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_4}\cdots\gamma_{\mu_{2M}}\} + \eta_{\mu_1\mu_4}\,\mathrm{Tr}\{\gamma_{\mu_2}\gamma_{\mu_3}\gamma_{\mu_5}\cdots\gamma_{\mu_{2M}}\}\\ &-\cdots+\eta_{\mu_1\mu_{2M}}\,\mathrm{Tr}\{\gamma_{\mu_2}\cdots\gamma_{\mu_{2M-1}}\}\,.\end{aligned} \tag{8.A.9}$$

If we assume that Eq. (8.A.1) correctly gives the trace of any product of $2N - 2$ Dirac matrices, then Eq. (8.A.9) shows that Eq. (8.A.1) also correctly gives the trace of any product of $2N$ Dirac matrices.

The easiest way to see that the trace of an odd number of Dirac matrices vanishes is to note that $-\gamma_\mu$ is related to γ_μ by a similarity transformation, $-\gamma_\mu = \gamma_5\gamma_\mu(\gamma_5)^{-1}$. Traces are unaffected by such similarity

* There are now computer programs[7] available for the calculation of traces of products of large numbers of Dirac matrices.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

transformations, so the trace of an odd number of Dirac matrices is equal to minus itself, and hence vanishes.

One occasionally encounters another class of traces, of the form

$$\mathrm{Tr}\{\gamma_5\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_n}\}.$$

This vanishes for odd n for the same reason as given above for traces without a γ_5. It also vanishes for $n = 0$ and $n = 2$:

$$\mathrm{Tr}\{\gamma_5\} = 0 \tag{8.A.10}$$

$$\mathrm{Tr}\{\gamma_5\gamma_\mu\gamma_\nu\} = 0\,. \tag{8.A.11}$$

(To see this just recall that $\gamma_5 \equiv i\gamma_0\gamma_1\gamma_2\gamma_3$, and note that there is no way of pairing the indices in $\mathrm{Tr}\{\gamma_0\gamma_1\gamma_2\gamma_3\}$ or in $\mathrm{Tr}\{\gamma_0\gamma_1\gamma_2\gamma_3\gamma_\mu\gamma_\nu\}$ so that the spacetime indices in each pair are equal.) For $n = 4$ it is possible to pair the indices in $\mathrm{Tr}\{\gamma_0\gamma_1\gamma_2\gamma_3\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\}$ so that the spacetime indices in each pair are equal, but only if μ, ν, ρ, σ are some permutation of $0, 1, 2, 3$. Furthermore this trace must be odd under permutations of μ, ν, ρ, σ since gamma matrices with different indices anticommute. Thus the trace $\mathrm{Tr}\{\gamma_5\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\}$ must be proportional to the totally antisymmetric tensor $\epsilon_{\mu\nu\rho\sigma}$. The constant of proportionality may be worked out by letting μ, ν, ρ, σ take the values $0, 1, 2, 3$, and recalling that $\epsilon_{0123} \equiv -1$. In this way we find

$$\mathrm{Tr}\{\gamma_5\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\} = 4i\epsilon_{\mu\nu\rho\sigma}\,. \tag{8.A.12}$$

The trace of products of γ_5 with six, eight, or more Dirac matrices may be calculated by the same methods used above to verify Eq. (8.A.1).

Problems

1. Calculate the differential and total cross-sections for the process $e^+e^- \to \mu^+\mu^-$ to lowest order in e. Assume that electron and muon spins are not observed. Use the simplest Lagrangian for the electrodynamics of electrons and muons.

2. Carry out the canonical quantization of the theory of a charged scalar field Φ and its interaction with electromagnetism, with Lagrangian density:

$$\mathscr{L} = -(D_\mu\Phi)^\dagger(D^\mu\Phi) - m^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}\,,$$

where

$$D_\mu\Phi \equiv \partial_\mu\Phi - iqA_\mu\Phi\,, \qquad F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu\,.$$

Use Coulomb gauge. Express the Hamiltonian in terms of the fields $\mathbf{A}$, Φ, and $\Phi^\dagger$ and their canonical conjugates. Evaluate the interaction

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

$V(t)$ in the interaction-picture in terms of the interaction picture fields and their derivatives.

3. Use the results of Problem 2 to calculate the differential and total cross-sections for photon scattering by a massive charged scalar particle to lowest order in e.

4. Write a gauge-invariant Lagrangian for a charged massive vector field interacting with the electromagnetic field.

5. Calculate the differential cross-section for electron–electron scattering to lowest order in e. Assume that final and initial spins are not measured.

References

1. See, e.g., M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge University Press, Cambridge, 1987): Section 2.2.

1*a*. V. Fock, *Z. f. Phys.* **39**, 226 (1927); H. Weyl, *Z. Phys.* **56**, 330 (1929). The term 'gauge invariance' derives from an analogy with earlier speculations about scale invariance by H. Weyl, in *Raum, Zeit, Materie*, 3rd ed. (Springer-Verlag, Berlin, 1920). Also see F. London, *Z. f. Phys.* **42**, 375 (1927). This history has been reviewed by C. N. Yang, talk at City College (unpublished).

2. The use of Coulomb gauge in electrodynamics was strongly advocated by Schwinger on pretty much the same grounds as here: that we ought not to introduce photons with helicities other than ± 1. See J. Schwinger, *Phys. Rev.* **78**, 1439 (1948); **127**, 324 (1962); *Nuovo Cimento* **30**, 278 (1963).

3. R. P. Feynman, *Phys. Rev.* **101**, 769 (1949): Section 8.

4. O. Klein and Y. Nishina, *Z. f. Phys.* **52**, 853 (1929); Y. Nishina, *ibid.*, 869 (1929); also see I. Tamm, *Z. f. Phys.* **62**, 545 (1930).

5. See, e.g., S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Section 4.11.

6. For a readable general introduction to the geometry and topology of p-forms, see H. Flanders, *Differential Forms* (Academic Press, New York, 1963).

7. T. West, *Comput. Phys. Commun.* **77**, 286 (1993).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.010

9

Path-Integral Methods

In Chapters 7 and 8 we applied the canonical quantization operator formalism to derive the Feynman rules for a variety of theories. In many cases, such as the scalar field with derivative coupling or the vector field with zero or non-zero mass, the procedure though straightforward was rather awkward. The interaction Hamiltonian turned out to contain a covariant term, equal to the negative of the interaction term in the Lagrangian, plus a non-covariant term, which served to cancel non-covariant terms in the propagator. In the case of electrodynamics this non-covariant term (the Coulomb energy) turned out to be not even spatially local, though it is local in time. Yet the final results are quite simple: the Feynman rules are just those we should obtain with covariant propagators, and using the negative of the interaction term in the Lagrangian to calculate vertex contributions. The awkwardness in obtaining these simple results, which was bad enough for the theories considered in Chapters 7 and 8, becomes unbearable for more complicated theories, like the non-Abelian gauge theories to be discussed in Volume II, and also general relativity. One would very much prefer a method of calculation that goes directly from the Lagrangian to the Feynman rules in their final, Lorentz-covariant form.

Fortunately, such a method does exist. It is provided by the path-integral approach to quantum mechanics. This was first presented in the context of non-relativistic quantum mechanics in Feynman's Princeton Ph. D. thesis,[1] as a means of working directly with a Lagrangian rather than a Hamiltonian. In this respect, it was inspired by earlier work of Dirac.[2] The path-integral approach played a part (along with inspired guesswork) in Feynman's later derivation of his diagrammatic rules.[3] However, although Feynman diagrams became widely used in the 1950s, most physicists (including myself) tended to derive them using the operator methods of Schwinger and Tomonaga, which were shown by Dyson in 1949 to lead to the same diagrammatic rules that had been obtained by Feynman by his own methods.

The path-integral approach was revived in the late 1960s, when Faddeev

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

and Popov[4] and De Witt[5] showed how to apply it to non-Abelian gauge theories and general relativity. For most theorists, the turning point came in 1971, when 't Hooft[6] used path-integral methods to derive the Feynman rules for spontaneously broken gauge theories (discussed in Volume II), including in particular the theory of weak and electromagnetic interactions, in a gauge that made the high energy behavior of these theories transparent. Soon after, as also discussed in Volume II, it was discovered that the path-integral method allows us to take account of contributions to the S-matrix that have an essential singularity at zero coupling constant and therefore cannot be discovered in any finite order of perturbation theory. Since then, the path-integral methods described here have become an indispensable part of the equipment of all physicists who make use of quantum field theory.

At this point the reader may be wondering why if the path-integral method is so convenient we bothered in Chapter 7 to introduce the canonical formalism. Indeed, Feynman seems at first to have thought of his path-integral approach as a substitute for the ordinary canonical formulation of quantum mechanics. There are two reasons for starting with the canonical formalism. The first is a point of principle: although the path-integral formalism provides us with manifestly Lorentz-invariant diagrammatic rules, it does not make clear why the S-matrix calculated in this way is unitary. As far as I know, the only way to show that the path-integral formalism yields a unitary S-matrix is to use it to reconstruct the canonical formalism, in which unitarity is obvious. There is a kind of conservation of trouble here; we can use the canonical approach, in which unitarity is obvious and Lorentz invariance obscure, or the path-integral approach, which is manifestly Lorentz-invariant but far from manifestly unitary. Since the path-integral approach is here derived from the canonical approach, we know that the two approaches yield the same S-matrix, so that the S-matrix must indeed be both Lorentz-invariant and unitary.

The second reason for introducing the canonical formalism first is more practical: there are important theories in which the simplest version of the Feynman path-integral method, in which propagators and interaction vertices are taken directly from the Lagrangian, is simply wrong. One example is the non-linear σ-model, with Lagrangian density $\mathscr{L} = -\frac{1}{2}g_{k\ell}(\phi)\partial_\mu\phi^k\partial^\mu\phi^\ell$. In such theories, using the naive Feynman rules derived directly from the Lagrangian density would yield an S-matrix that is not only wrong but even non-unitary, and that also depends on the way in which we define the scalar field.[7] In this chapter we shall derive the path-integral formalism from the canonical formalism, and in this way we will see what additional sorts of vertices are needed to supplement the simplest version of the Feynman path-integral method.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

9.1 The General Path-Integral Formula

We start with a general quantum mechanical system, with Hermitian operator 'coordinates' Q_a and conjugate 'momenta' P_b, satisfying the canonical commutation relations*:

$$[Q_a, P_b] = i\delta_{ab}\,, \tag{9.1.1}$$

$$[Q_a, Q_b] = [P_a, P_b] = 0\,. \tag{9.1.2}$$

(We shall restrict ourselves in this and the next three sections to bosonic operators, which satisfy commutation rather than anticommutation relations. Our results will be generalized to include fermionic operators in Section 9.5.) In a field theory, the index a consists of a position $\mathbf{x}$ and a discrete Lorentz and species index m, and we conventionally write

$$Q_{\mathbf{x},m} \equiv Q_m(\mathbf{x}), \tag{9.1.3}$$

$$P_{\mathbf{x},m} \equiv P_m(\mathbf{x})\,. \tag{9.1.4}$$

Also, the Kronecker delta in Eq. (9.1.1) is interpreted in a field theory as

$$\delta_{\mathbf{x},m\,;\,\mathbf{y},n} \equiv \delta^3(\mathbf{x}-\mathbf{y})\delta_{mn}\,. \tag{9.1.5}$$

However, for the present it will be convenient to use the more compact notation of Eqs. (9.1.1) and (9.1.2). These are 'Schrödinger-picture' operators, taken at a fixed time (say, $t = 0$). The time-dependent operators in the Heisenberg picture will be considered a little later.

Since the Q_a all commute, we can find a simultaneous eigenstate $|q\rangle$, with eigenvalues q_a:

$$Q_a|q\rangle = q_a\,|q\rangle\,. \tag{9.1.6}$$

(We are using lower case qs and ps here to denote eigenvalues rather than to denote operators in the interaction picture as in Chapter 7, but since we will not be using the interaction picture in this chapter no confusion should arise.) The eigenvectors can be taken to be orthonormal,

$$\langle q'|q\rangle = \prod_a \delta(q'_a - q_a) \equiv \delta(q'-q)\,, \tag{9.1.7}$$

so that the completeness relation reads

$$1 = \int \prod_a dq_a\,|q\rangle\,\langle q|\,. \tag{9.1.8}$$

* We are tacitly assuming here that any first class constraints are eliminated by choosing a gauge, and any remaining second class constraints are 'solved' by writing the constrained degrees of freedom in terms of the unconstrained Q_a and P_a, as in Section 7.6. The direct application of path-integral methods to constrained systems is described by Faddeev.[8]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

Similarly, we can find a complete orthonormal set of eigenstates of the P_a:

$$P_a|p\rangle = p_a|p\rangle \,, \tag{9.1.9}$$

$$\langle p'|p\rangle = \prod_a \delta(p'_a - p_a) \equiv \delta(p' - p) \,, \tag{9.1.10}$$

$$1 = \int \prod_a dp_a \, |p\rangle \, \langle p| \,. \tag{9.1.11}$$

As usual, it follows from Eq. (9.1.1) that these two complete sets of eigenstates have the scalar product**

$$\langle q \, | p\rangle = \prod_a \frac{1}{\sqrt{2\pi}} \exp(iq_a p_a) \,. \tag{9.1.12}$$

In the Heisenberg picture, the Q and P operators are given a time-dependence

$$Q_a(t) \equiv \exp(iHt)\, Q_a \exp(-iHt) \,, \tag{9.1.13}$$

$$P_a(t) \equiv \exp(iHt)\, P_a \exp(-iHt) \,, \tag{9.1.14}$$

where H is the total Hamiltonian. These have eigenstates $|q;t\rangle$ and $|p;t\rangle$

$$Q_a(t)|q;t\rangle = q_a|q;t\rangle \,, \tag{9.1.15}$$

$$P_a(t)|p;t\rangle = p_a|p;t\rangle \,, \tag{9.1.16}$$

given by

$$|q;t\rangle = \exp(iHt)|q\rangle \,, \tag{9.1.17}$$

$$|p;t\rangle = \exp(iHt)|p\rangle \,. \tag{9.1.18}$$

(Note that $|q;t\rangle$ is the eigenstate of $Q_a(t)$ with eigenvalue q_a, and *not* the result of letting the state $|q\rangle$ evolve for a time t. This is why its time-dependence is given by a factor $\exp(iHt)$ rather than $\exp(-iHt)$.) These states obviously satisfy the completeness and orthonormality conditions

$$\langle q';t|q;t\rangle = \delta(q' - q) \,, \tag{9.1.19}$$

$$\langle p';t|p;t\rangle = \delta(p' - p) \,, \tag{9.1.20}$$

$$\int \prod_a dq_a \, |q;t\rangle \, \langle q;t| = 1 \,, \tag{9.1.21}$$

$$\int \prod_a dp_a \, |p;t\rangle \, \langle p;t| = 1 \,, \tag{9.1.22}$$

** The proof follows the same lines as in the quantum mechanics of point particles. From Eq. (9.1.1), we see that P_b acts as $-i\partial/\partial q_b$ on wave functions in a q-basis. The right-hand side of Eq. (9.1.12) is then seen to be the wave function in this basis of an eigenstate of P. The factor $\prod 1/\sqrt{2\pi}$ is fixed by the normalization requirement, Eq. (9.1.10).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

and also

$$\langle q;t|p;t\rangle = \prod_a \frac{1}{\sqrt{2\pi}} \exp(iq_a p_a) \,. \tag{9.1.23}$$

If, by measurements at time t, we find that our system is in a definite state $|q;t\rangle$, then the probability amplitude for measurements at time t' to give a state $|q';t'\rangle$ is the scalar product $\langle q';t'|q;t\rangle$. Our central dynamical problem is to calculate this scalar product.

This is easy when t' and t are infinitesimally close, say $t' = \tau + d\tau$ and $t = \tau$. Using Eq. (9.1.17), we have

$$\langle q';\tau + d\tau|q;\tau\rangle = \langle q';\tau|\exp(-iHd\tau)|q;\tau\rangle \,. \tag{9.1.24}$$

The Hamiltonian H is given as a function $H(Q,P)$, but since (9.1.13) and (9.1.14) are similarity transformations, and H commutes with itself, it can equally well be written as the *same* function of $Q(t)$ and $P(t)$

$$H \equiv H(Q,P) = e^{iHt}H(Q,P)e^{-iHt} = H\Big(Q(t),\ P(t)\Big) \,. \tag{9.1.25}$$

This function can be written in various different forms, with different constant coefficients, by using the commutation relations (9.1.1) and (9.1.2) to move the Qs and Ps past each other. It will be convenient to adopt a standard form, in which all Qs appear to the *left* of all Ps. For instance, given a term in the Hamiltonian of form $P_a Q_b P_c$, we would rewrite it as $P_a Q_b P_c = Q_b P_a P_c - i\,\delta_{ab} P_c$. With this convention, the $Q_a(t)$s in the Hamiltonian in Eq. (9.1.24) may be replaced[†] with their eigenvalues q'_a. To deal with the $P(t)$, we use Eq. (9.1.23) to expand $|q;\tau\rangle$ in P-eigenstates $|p;\tau\rangle$, and find

$$\begin{aligned}\langle q';\tau + d\tau \mid q;\tau\rangle &= \int \prod_a dp_a \,\langle q';\tau|\exp\Big(-iH(Q(\tau),P(\tau))d\tau\Big)|p;\tau\rangle \\ &\qquad \times\langle p;\tau|q;\tau\rangle \\ &= \int \prod_a \frac{dp_a}{2\pi}\ \exp\left[-iH(q',p)d\tau + i\sum_a (q'_a - q_a)p_a\right] \,, \end{aligned} \tag{9.1.26}$$

with each p_a integrated from $-\infty$ to $+\infty$.

Now let's return to the more general case of a finite time-interval. To calculate $\langle q';t'|q;t\rangle$, with $t < t'$, we break up the time-interval from t to t' into steps $t, \tau_1, \tau_2, \cdots \tau_N, t'$, with

$$\tau_{k+1} - \tau_k = d\tau = (t' - t)/(N+1) \tag{9.1.27}$$

[†] This is only possible because with $d\tau$ infinitesimal, $\exp(-iH\,d\tau)$ is linear in H.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

and sum over a complete set of states $|q;\tau_k\rangle$ at each time τ_k:

$$\langle q';t'|q;t\rangle = \int dq_1 \cdots dq_N \langle q';t'|q_N;\tau_N\rangle\langle q_N;\tau_N|q_{N-1};\tau_{N-1}\rangle \cdots \langle q_1;\tau_1|q;t\rangle. \tag{9.1.28}$$

Inserting Eq. (9.1.26), this becomes

$$\langle q';t'|q;t\rangle = \int \left[\prod_{k=1}^{N}\prod_a dq_{k,a}\right]\left[\prod_{k=0}^{N}\prod_a dp_{k,a}/2\pi\right]$$
$$\times \exp\left[i\sum_{k=1}^{N+1}\left\{\sum_a (q_{k,a}-q_{k-1,a})p_{k-1,a} - H(q_k,p_{k-1})d\tau\right\}\right], \tag{9.1.29}$$

where

$$q_0 \equiv q\,, \qquad q_{N+1} \equiv q'\,. \tag{9.1.30}$$

Our result, Eq. (9.1.29), can be put in a much more elegant form. Define smooth interpolating functions, $q(\tau)$ and $p(\tau)$, such that

$$q_a(\tau_k) \equiv q_{k,a}\,, \qquad p_a(\tau_k) \equiv p_{k,a}\,. \tag{9.1.31}$$

In the limit $d\tau \to 0$ (i.e., $N \to \infty$), the argument of the exponential in Eq. (9.1.29) becomes just an integral over τ

$$\sum_{k=1}^{N+1}\left\{\sum_a (q_{k,a}-q_{k-1,a})p_{k-1,a} - H(q_k,p_{k-1})d\,\tau\right\}$$
$$= \sum_{k=1}^{N+1}\left\{\sum_a \dot{q}_a(\tau_k)p_a(\tau_k) - H\Big(q(\tau_k),p(\tau_k)\Big)\right\}d\,\tau + O(d\tau^2)$$
$$\to \int_t^{t'}\left\{\sum_a \dot{q}_a(\tau)p_a(\tau) - H\Big(q(\tau),p(\tau)\Big)\right\}d\,\tau\,. \tag{9.1.32}$$

Further, we may define integrals *over the functions* $q(\tau), p(\tau)$ by

$$\int \prod_{\tau,a} dq_a(\tau)\prod_{\tau,b}\frac{dp_b(\tau)}{2\pi}\cdots \equiv \lim_{d\tau\to 0}\int \prod_{k,a} dq_{k,a}\prod_{k,b}\frac{dp_{k,b}}{2\pi}\cdots\,. \tag{9.1.33}$$

Eq. (9.1.29) then becomes a constrained path integral

$$\langle q';t'|q;t\rangle = \int_{\substack{q_a(t)=q_a\\ q_a(t')=q'_a}} \prod_{\tau,a} dq_a(\tau)\prod_{\tau,b}\frac{dp_b(\tau)}{2\pi}$$
$$\times \exp\left[i\int_t^{t'} d\tau\left\{\sum_a \dot{q}_a(\tau)p_a(\tau) - H\Big(q(\tau),p(\tau)\Big)\right\}\right]. \tag{9.1.34}$$

This is called a *path integral*, because we integrate over all paths that take $q(\tau)$ from q at $\tau = t$ to q' at $\tau = t'$, as well as over all $p(\tau)$. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

great advantage of writing matrix elements in this way is that, as shown in Section 9.3, the path integrals are easy to calculate when expanded in powers of the coupling constants in H.

The path-integral formalism allows us to calculate not only transition probability amplitudes like $\langle q';t'|q;t\rangle$, but also the matrix elements between states $\langle q';t'|$ and $|q,t\rangle$ of time-ordered products of general operators $\mathcal{O}(P(t),Q(t))$. It will be convenient to define these operators with (unlike H) all Ps moved to the *left* and all Qs to the *right*. Then by inserting any such operator $\mathcal{O}\Big(P(\tau),Q(\tau)\Big)$ in Eq. (9.1.26), we have

$$\begin{aligned}\langle q';\tau+d\tau|\mathcal{O}\Big(P(\tau),Q(\tau)\Big)|q;\tau\rangle &= \int\prod_d dp_a \\ &\times\langle q';\tau|\exp\Big(-iH\Big(Q(\tau),P(\tau)\Big)d\tau\Big)\ |p;\tau\rangle\langle p;\tau|\mathcal{O}\Big(P(\tau),Q(\tau)\Big)|q;\tau\rangle \\ &= \int\prod_a\frac{dp_a}{2\pi}\exp\left[-iH(q',p)d\tau+i\sum_a(q'_a-q_a)p_a\right]\mathcal{O}(p,q)\,. \qquad (9.1.35)\end{aligned}$$

In order to calculate the matrix element of a product $\mathcal{O}_A\Big(P(t_A),Q(t_A)\Big)\,\mathcal{O}_B\Big(P(t_B),Q(t_B)\Big)\cdots$ of operators with $t_A>t_B>\cdots$, we can insert the $\mathcal{O}$-operators between the appropriate states in Eq. (9.1.28), and use Eq. (9.1.35). For instance, if the time t_A falls between τ_k and τ_{k+1}, then insert $\mathcal{O}_A\Big(P(t_A),Q(t_A)\Big)$ between $\langle q_{k+1};\tau_{k+1}|$ and $|q_k;\tau_k\rangle$. Note that in Eq. (9.1.28) each successive sum over states is at a later time, so this is only possible because of our assumption that $t_A>t_B>\cdots$. Following the same steps as before, we now find the general path-integral formula

$$\begin{aligned}&\langle q',t'|\mathcal{O}_A\Big(P(t_A),Q(t_A)\Big)\mathcal{O}_B\Big(P(t_B),Q(t_B)\Big)\cdots|q,t\rangle \\ &=\int\limits_{\substack{q_a(t)=q_a\\ q_a(t')=q'_a}}\prod_{\tau,a}dq_a(\tau)\prod_{\tau,b}\frac{dp_b(\tau)}{2\pi}\ \mathcal{O}_A\Big(p(t_A),q(t_A)\Big)\mathcal{O}_B\Big(p(t_B),q(t_B)\Big)\cdots \\ &\qquad\times\exp\left[i\int_t^{t'}d\tau\left\{\sum_a\dot{q}_a(\tau)p_a(\tau)-H\Big(q(\tau),p(\tau)\Big)\right\}\right]\,. \qquad (9.1.36)\end{aligned}$$

This result is only valid if the times are ordered, with

$$t'>t_A>t_B>\cdots>t\,. \qquad (9.1.37)$$

However, nothing on the right-hand side of Eq. (9.1.36) refers to the order of time-arguments. Hence if we are presented with a path integral like the right-hand side of Eq. (9.1.36), with $t_A,t_B,\cdots$ in arbitrary order (all between t and t', with $t<t'$), then this path integral will equal a matrix element like the left-hand side of Eq. (9.1.36), but with the operators arranged in order (from left to right) of decreasing time. That is, for

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$t_A, t_B, \cdots$ in arbitrary order, we have

$$\langle q', t' | T\left\{ \mathcal{O}_A\Big(P(t_A), Q(t_A)\Big), \mathcal{O}_B\Big(P(t_B), Q(t_B)\Big), \cdots \right\} | q, t\rangle$$
$$= \int_{\substack{q_a(t) = q_a \\ q_a(t') = q'_a}} \prod_{\tau,a} dq_a(\tau) \prod_{\tau,b} \frac{dp_b(\tau)}{2\pi}\ \mathcal{O}_A\Big(p(t_A), q(t_A)\Big)\mathcal{O}_B\Big(p(t_B), q(t_B)\Big)\cdots$$
$$\times \exp\left[i \int_t^{t'} d\tau \left\{ \sum_a \dot{q}_a(\tau) p_a(\tau) - H\Big(q(\tau), p(\tau)\Big) \right\} \right] , \quad (9.1.38)$$

where T denotes the usual time-ordered product.

It should perhaps be stressed that the c-number functions $q_a(\tau)$, $p_a(\tau)$ in Eq. (9.1.38) are mere variables of integration, and in particular are *not* constrained to obey the equations of motion of classical Hamiltonian dynamics

$$\dot{q}_a(\tau) - \frac{\partial H(q(\tau), p(\tau))}{\partial p_a(\tau)} = 0 , \quad (9.1.39)$$

$$\dot{p}_a(\tau) + \frac{\partial H(q(\tau), p(\tau))}{\partial q_a(\tau)} = 0 . \quad (9.1.40)$$

(For this reason, the Hamiltonian $H(q(\tau), p(\tau))$ in Eq. (9.1.38) is *not* constant in τ.) Nevertheless, there is a limited sense in which path integrals do respect these equations of motion. Suppose that one of the functions in Eq. (9.1.38), say $\mathcal{O}_A\Big(p(t_A), q(t_A)\Big)$, happens to be the left-hand side of either Eq. (9.1.39) or Eq. (9.1.40). We note that (for $t < t_A < t'$)

$$\left(\dot{q}_a(t_A) - \frac{\partial H(q(t_A), p(t_A))}{\partial p_a(t_A)} \right) \exp\Big(iI[q,p] \Big) = -i \frac{\delta}{\delta p_a(t_A)} \exp\Big(iI[q,p] \Big),$$
$$\left(\dot{p}_a(t_A) + \frac{\partial H(q(t_A), p(t_A))}{\partial q_a(t_A)} \right) \exp\Big(iI[q,p] \Big) = i \frac{\delta}{\delta q_a(t_A)} \exp\Big(iI[q,p] \Big),$$

where iI is the argument of the exponential in Eq. (9.1.38):

$$I[q,p] \equiv \int_t^{t'} d\tau \left\{ \sum_a \dot{q}_a(\tau) p_a(\tau) - H\Big(q(\tau), p(\tau)\Big) \right\} .$$

As long as t_A does not approach t or t', the integrations over $q_a(t_A)$ and $p_a(t_A)$ are unconstrained, and so with reasonable assumptions about the convergence of these integrals, the integral of such variational derivatives must vanish. Hence the path integral (9.1.38) vanishes if $\mathcal{O}_A(p,q)$ is taken to be the left-hand side of either of the equations of motion (9.1.39) or (9.1.40).

This simple rule applies only if the integration variables $q_a(t_A)$, $p_a(t_A)$ are independent of any of the variables $q_a(t_B)$, $p_a(t_B)$, etc. appearing in any of the other functions $\mathcal{O}_B$, $\mathcal{O}_C$, etc. in Eq. (9.1.38), and hence

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

only if we prohibit t_A from approaching t_B, t_C, etc. as well as t or t'. When t_A approaches, say, t_B, the path integral will be found to involve a non-zero term proportional to $\delta(t_A - t_B)$ or its derivatives. These delta functions are the same as would be found in the operator formalism from time-derivatives of the step functions implicit in the definition of the time-ordered product.

In evaluating the path integrals (9.1.34) and (9.1.38), we only need to know the classical Hamiltonian, the c-number function $H(q,p)$. If we were to define a theory by the path integrals, the question would naturally arise, which of many possible quantum mechanical Hamiltonians $H(Q,P)$ (differing in the order of Qs and Ps) governed the quantum theory that corresponds to these path integrals. Our derivation has provided an answer: the quantum Hamiltonian is to be taken with all Qs on the left, all Ps on the right. But it would be a mistake to give this prescription too much significance. There are a great many ways of interpreting the measure $\Pi\, dq_a(\tau)\ \Pi\, dp_b(\tau)$ appearing in path integrals like (9.1.34) or (9.1.38). Our prescription, of putting all Qs to the left of all Ps, is appropriate only if the measure is interpreted according to Eqs. (9.1.31)–(9.1.33). Other measures would lead to other prescriptions for operator ordering. The question is not an urgent one, because different prescriptions for ordering operators in the Hamiltonian just correspond to different choices of the constants that appear as coefficients of the various terms in the Hamiltonian, and we generally formulate theories with these constants left as arbitrary parameters anyway.

It is difficult to use the general path integral in Eq. (9.1.38) for numerical calculations or as a source of rigorous theorems. For these purposes it is better to use the path-integral method to calculate amplitudes in Euclidean space, where t is replaced with an imaginary quantity $-ix_4$, and the argument of the exponential in Eq. (9.1.38) is a negative real quantity. In this way, instead of jagged paths producing rapid oscillations of the integrand from one path to another, all jagged paths are exponentially suppressed. Though we shall not go into it here, quantum field theory may be formulated from the beginning in terms of Feynman amplitudes in Euclidean spacetime.[8a] Under certain plausible assumptions, it is possible to reconstruct the Feynman amplitudes in Minkowskian spacetime from their Euclidean counterparts.[8b] But we may as well stick to the Minkowski space formulation of the path integral if we are only going to use it to calculate Feynman amplitudes in perturbation theory.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

9.2 Transition to the S-Matrix

As already mentioned, we can easily convert the general quantum mechanical results of Section 9.1 to a notation appropriate to quantum field theory, by letting the index a run over points $\mathbf{x}$ in space and over a spin-and-species index m, and replacing $Q_a(t)$ and $P_a(t)$ with $Q_m(\mathbf{x},t)$ and $P_m(\mathbf{x},t)$, respectively. Eq. (9.1.38) then reads*

$$\begin{aligned}\langle q',t'|T\{\mathcal{O}_A[P(t_A),Q(t_A)],\ \mathcal{O}_B[P(t_B),Q(t_B)],\cdots\}|q,t\rangle \\ = \int_{\substack{q_m(\mathbf{x},t)=q_m(\mathbf{x}) \\ q_m(\mathbf{x},t')=q'_m(\mathbf{x})}} \prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau) \prod_{\tau,\mathbf{x},m} \frac{dp_m(\mathbf{x},\tau)}{2\pi} \\ \times \mathcal{O}_A\Big[p(t_A),q(t_A)\Big]\ \mathcal{O}_B\Big[p(t_B),q(t_B)\Big]\cdots \\ \times \exp\left[i\int_t^{t'} d\tau\left\{\int d^3x \sum_m \dot{q}_m(\mathbf{x},\tau)p_m(\mathbf{x},t) - H\Big[q(\tau),p(\tau)\Big]\right\}\right]. \quad (9.2.1)\end{aligned}$$

However, in field theory Eq. (9.2.1) is not exactly what we want. Experimentalists do not measure probability amplitudes for transitions between eigenstates $\langle q',t'|$ and $|q,t\rangle$ of the quantum field Q, but rather S-matrix elements, the probability amplitudes for transitions between states that at $t\to-\infty$ or $t\to+\infty$ contain definite numbers of particles of various types. These are called 'in' and 'out' states, $|\alpha,\text{ in}\rangle$ and $|\beta,\text{ out}\rangle$, where α and β denote sets of particles characterized by the various particles' momenta, spin z-component (or helicity), and species. To calculate a matrix element of a time-ordered product (perhaps empty) between such states, we need to multiply Eq. (9.2.1) by the 'wave functions' $\langle\beta,\text{ out}|q',t'\rangle$ and $\langle q,t|\alpha,\text{in}\rangle$ at any fixed times t and t', taken for convenience here to be $-\infty$ and $+\infty$, respectively, and then perform an integral over the 'arguments' $q_m(\mathbf{x})$ and $q'_m(\mathbf{x})$ of these wave functions. But instead of constraining the path integral over $q_m(\mathbf{x},\tau)$ by the conditions

$$q_m(\mathbf{x},+\infty)=q'_m(\mathbf{x}), \qquad q_m(\mathbf{x},-\infty)=q_m(\mathbf{x}), \quad (9.2.2)$$

and then integrating over $q'_m(\mathbf{x})$ and $q_m(\mathbf{x})$, we can just as well do an unconstrained integral over $q_m(\mathbf{x},\tau)$ (and also over $p_m(\mathbf{x},\tau)$), and set the arguments of the wave functions equal to the values given by Eq. (9.2.2):

* We are now writing H and the $\mathcal{O}$s with square brackets, to remind us that $H[q(t),p(t)]$ and $\mathcal{O}[p(t),q(t)]$ are *functionals* of $q_m(\mathbf{x},t)$ and $p_m(\mathbf{x},t)$ at a fixed time t.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$$\begin{aligned}
&\langle\beta, \text{out}|T\left\{\mathcal{O}_A\left[P(t_A), Q(t_A)\right], \mathcal{O}_B\left[P(t_B), Q(t_B)\right], \cdots\right\}|\alpha, \text{ in}\rangle \\
&= \int \prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau) \prod_{\tau,\mathbf{x},m} (dp_m(\mathbf{x},\tau)/2\pi) \\
&\quad \times \mathcal{O}_A\left[p(t_A), q(t_A)\right] \mathcal{O}_B\left[p(t_B), q(t_B)\right] \cdots \\
&\quad \times \exp\left[i\int_{-\infty}^{\infty} d\tau \left\{\int d^3x \sum_m \dot{q}_m(\mathbf{x},\tau)p_m(\mathbf{x},\tau) - H\left[q(\tau), p(\tau)\right]\right\}\right] \\
&\quad \times \langle\beta, \text{ out}|q(+\infty); +\infty\rangle \langle q(-\infty); -\infty|\alpha, \text{ in}\rangle . \qquad (9.2.3)
\end{aligned}$$

Incidentally, this result leads immediately** to Eq. (6.4.3), a theorem that we use repeatedly to relate sums of off-shell Feynman graphs to matrix elements of Heisenberg-picture operators between exact energy eigenstates.

It is necessary now to consider how to calculate the wave functions appearing as the final pair of factors in Eq. (9.2.3). Let's first consider the simplest and most important case, the vacuum. (We saw in Section 6.4 that S-matrix elements may be easily calculated from the vacuum expectation values of time-ordered products.) We assume as usual that for $t \to \pm\infty$, matrix elements may be calculated as if there were no interactions. The 'in' and 'out' vacua may thus be defined by the conditions

$$\begin{aligned}
a_{\text{in}}(\mathbf{p},\sigma,n)|\text{VAC, in}\rangle &= 0 , \\
a_{\text{out}}(\mathbf{p},\sigma,n)|\text{VAC, out}\rangle &= 0 ,
\end{aligned} \qquad (9.2.4)$$

where a_{in} and a_{out} are the operators appearing in the coefficients of $\exp(i\mathbf{p}\cdot\mathbf{x} - iEt)$ in the plane-wave expansion of the operator $Q_m(\mathbf{x},t)$ at $t \to -\infty$ and $t \to +\infty$, respectively. For instance, for the real scalar field of a neutral spinless particle, we have in effect

$$\Phi(\mathbf{x},t) \overset{t\to\mp\infty}{\longrightarrow} (2\pi)^{-3/2} \int d^3p\, (2E)^{-1/2}\left[a_{\substack{\text{in}\\\text{out}}}(\mathbf{p})e^{ip\cdot x} + \text{H.c.}\right] , \qquad (9.2.5)$$

** It is only necessary to note that, for a Hamiltonian $H[P(t),Q(t)] + \sum_A \int d^3x\, \epsilon_A(\mathbf{x},t)\mathcal{O}_A(\mathbf{x},t)$, the S-matrix is given by Eq. (9.2.3) as

$$\begin{aligned}
\langle\beta, \text{out}|\alpha, \text{ in}\rangle_\epsilon = &\int \prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau) \prod_{\tau,\mathbf{x},m} (dp_m(\mathbf{x},\tau)/2\pi) \\
&\times \exp\left[i\int_{-\infty}^{\infty} d\tau \left\{\int d^3x\, \dot{q}_m(\mathbf{x},\tau)p_m(\mathbf{x},\tau) - H\left[q(\tau), p(\tau)\right]\right.\right. \\
&\left.\left. - \sum_A \int d^3x\, \epsilon_A(\mathbf{x},\tau)O_A(\mathbf{x},\tau)\right\}\right] \\
&\times \langle\beta, \text{ out}|q(+\infty); +\infty\rangle \langle q(-\infty); -\infty|\alpha, \text{ in}\rangle .
\end{aligned}$$

The left-hand side of Eq. (6.4.3) is the derivative of this expression with respect to ϵ_a, ϵ_b, etc., at $\epsilon = 0$, which yields the right-hand side of Eq. (9.2.3), and using Eq. (9.2.3) again then immediately gives the right-hand side of Eq. (6.4.3).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$$\Pi(\mathbf{x},t) \overset{t\to\mp\infty}{\longrightarrow} \dot{\Phi}(\mathbf{x},t)$$
$$\overset{t\to\mp\infty}{\longrightarrow} -i(2\pi)^{-3/2}\int d^3p\,(E/2)^{1/2}\left[a_{\substack{\text{in}\\\text{out}}}(\mathbf{p})e^{ip\cdot x} - \text{H.c.}\right] \tag{9.2.6}$$

where $p^0 \equiv E \equiv \sqrt{\mathbf{p}^2+m^2}$, and we here use the conventional Φ and Π rather than Q and P for a scalar field, and drop the unnecessary labels m,σ,n. Inverting the Fourier transforms and taking a linear combination of the resulting expressions, we have

$$a_{\substack{\text{in}\\\text{out}}}(\mathbf{p}) = \lim_{t\to\mp\infty}\frac{e^{iEt}}{(2\pi)^{3/2}}\int d^3x\,e^{-i\mathbf{p}\cdot\mathbf{x}} \times\left[\sqrt{\frac{E}{2}}\Phi(\mathbf{x},t) + i\sqrt{\frac{1}{2E}}\Pi(\mathbf{x},t)\right]. \tag{9.2.7}$$

As mentioned in Section 9.1, the 'momentum' $\Pi(\mathbf{x},t)$ acts on wave-functions in a ϕ-basis as the variational derivative $-i\delta/\delta\phi(\mathbf{x},t)$, so in this basis the conditions (9.2.4) read

$$0 = \int d^3x\,e^{-i\mathbf{p}\cdot\mathbf{x}}\left[\frac{\delta}{\delta\phi(\mathbf{x})} + E(\mathbf{p})\phi(\mathbf{x})\right]\left\langle\phi(t\to\mp\infty);\mp\infty\middle|\text{VAC},{}^{\text{in}}_{\text{out}}\right\rangle . \tag{9.2.8}$$

The analogous ordinary differential equation has a well-known Gaussian solution, so let's try a Gaussian ansatz here:

$$\left\langle\phi(t\to\mp\infty);\mp\infty\middle|\text{VAC},{}^{\text{in}}_{\text{out}}\right\rangle = \mathcal{N}\exp\left(-\tfrac{1}{2}\int d^3x\,d^3y\,\mathscr{E}(\mathbf{x},\mathbf{y})\phi(\mathbf{x})\phi(\mathbf{y})\right), \tag{9.2.9}$$

with kernel $\mathscr{E}$ and constant $\mathcal{N}$ to be determined. Substituting this in Eq. (9.2.8), we see that the functional differential equation for the vacuum wave functional is satisfied if for all ϕ

$$0 = \int d^3x\,e^{-i\mathbf{p}\cdot\mathbf{x}}\left[\int d^3y\,\mathscr{E}(\mathbf{x},\mathbf{y})\phi(\mathbf{y}) - E(\mathbf{p})\phi(\mathbf{x})\right] \tag{9.2.10}$$

or, in other words, if

$$\int d^3x\,e^{-i\mathbf{p}\cdot\mathbf{x}}\mathscr{E}(\mathbf{x},\mathbf{y}) = E(\mathbf{p})\,e^{-i\mathbf{p}\cdot\mathbf{y}} . \tag{9.2.11}$$

The solution is easily found by inverting the Fourier transform

$$\mathscr{E}(\mathbf{x},\mathbf{y}) = (2\pi)^{-3}\int d^3p\,e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}E(\mathbf{p}) . \tag{9.2.12}$$

(Recall that $E(\mathbf{p}) \equiv \sqrt{\mathbf{p}^2+m^2}$). This is actually the most useful representation for the kernel $\mathscr{E}$, but we may note in passing that for $\mathbf{x}\neq\mathbf{y}$, $\mathscr{E}$ may also be written in terms of a Hankel function of negative order

$$\mathscr{E}(\mathbf{x},\mathbf{y}) = \frac{m}{2\pi^2 r}\frac{d}{dr}\left(\frac{1}{r}K_{-1}(mr)\right), \tag{9.2.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

where $r \equiv |\mathbf{x}-\mathbf{y}|$. The constant $\mathcal{N}$ in Eq. (9.2.9) may be formally obtained from the normalization condition for the vacuum state, but we will not need this result.

According to Eq. (9.2.9), in calculating vacuum expectation values in the theory of a scalar field, the product of the last two factors in Eq. (9.2.3) is

$$\begin{aligned}
&\langle \text{VAC, out}|\phi(\infty);+\infty\rangle\ \langle\phi(-\infty);-\infty|\text{VAC, in}\rangle \\
&\quad = |\mathcal{N}|^2\ \exp\left(-\tfrac{1}{2}\int d^3x\, d^3y\, \mathscr{E}(\mathbf{x},\mathbf{y})\Big[\phi(\mathbf{x},+\infty)\phi(\mathbf{y},+\infty)\right. \\
&\qquad\qquad \left.+\phi(\mathbf{x},-\infty)\phi(\mathbf{y},-\infty)\Big]\right) \qquad (9.2.14)\\
&\quad = |\mathcal{N}|^2\ \exp\left(-\tfrac{1}{2}\epsilon\int d^3x\, d^3y \int_{-\infty}^{\infty} d\tau\, \mathscr{E}(\mathbf{x},\mathbf{y})\phi(\mathbf{x},\tau)\phi(\mathbf{y},\tau)e^{-\epsilon|\tau|}\right),
\end{aligned}$$

where ϵ is a positive infinitesimal. To obtain the final expression, we have used the fact that for any reasonably smooth function $f(\tau)$,

$$f(+\infty)+f(-\infty) = \lim_{\epsilon\to 0+}\epsilon\int_{-\infty}^{\infty} d\tau f(\tau)\, e^{-\epsilon|\tau|}\,. \qquad (9.2.15)$$

Inserting Eq. (9.2.14) in Eq. (9.2.3) now gives

$$\begin{aligned}
&\Big\langle \text{VAC, out}\Big|T\left\{\mathcal{O}_A\Big[\Pi(t_A),\Phi(t_A)\Big],\ \mathcal{O}_B\Big[\Pi(t_B),\Phi(t_B)\Big],\cdots\right\}\Big|\text{VAC, in}\Big\rangle \\
&= |\mathcal{N}|^2\int\prod_{\tau,\mathbf{x}} d\phi(\mathbf{x},\tau)\prod_{\tau,\mathbf{x}}(d\pi(\mathbf{x},\tau)/2\pi)\ \ \mathcal{O}_A\Big[\pi(t_A),\phi(t_A)\Big] \\
&\quad\times\mathcal{O}_B\Big[\pi(t_B),\phi(t_B)\Big]\cdots\exp\left[i\int_{-\infty}^{\infty}d\tau\left\{\int d^3x\,\dot{\phi}(\mathbf{x},\tau)\,\pi(\mathbf{x},\tau)\right.\right. \\
&\quad\left.\left.-H\Big[\phi(\tau),\pi(\tau)\Big]+\ \tfrac{1}{2}i\epsilon\int d^3x\, d^3y\ \mathscr{E}(\mathbf{x},\mathbf{y})e^{-\epsilon|\tau|}\phi(\mathbf{x},\tau)\phi(\mathbf{y},\tau)\right\}\right]. \qquad (9.2.16)
\end{aligned}$$

We shall see in Section 9.4 that the whole effect of the last term in the argument of the exponential in Eq. (9.2.16) is to provide the $-i\epsilon$ in the denominator of the scalar field propagator in momentum space, $[p^2+m^2-i\epsilon]^{-1}$. We will not go into the corresponding details for fields of general spin, but will simply state that in general

$$\begin{aligned}
&\Big\langle \text{VAC, out}|T\left\{\mathcal{O}_A\Big[P_A(t_A),Q(t_A)\Big],\ \mathcal{O}_B\Big[P_B(t_B),Q(t_B)\Big],\cdots\right\}|\text{VAC, in}\Big\rangle \\
&= |\mathcal{N}|^2\int\left[\prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau)\right]\left[\prod_{\tau,\mathbf{x},m}\frac{dp_m(\mathbf{x},\tau)}{2\pi}\right]\ \mathcal{O}_A\Big[p(t_A),q(t_A)\Big] \\
&\quad\times\mathcal{O}_B\Big[p(t_B),q(t_B)\Big]\cdots\exp\left[i\int_{-\infty}^{\infty}d\tau\left\{\int d^3x\ \sum_m \dot{q}_m(\mathbf{x},\tau)p_m(\mathbf{x},\tau)\right.\right. \\
&\qquad\left.\left.-H\Big[q(\tau),p(\tau)\Big]+\ i\epsilon\ \text{terms}\right\}\right], \qquad (9.2.17)
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

where the '$i\epsilon$ terms' just have the effect of putting the correct $-i\epsilon$ in the denominators of all propagators.

This is a good place to mention that field-independent factors in Eq. (9.2.17), like the constant $|\mathcal{N}|^2$, are not important. This is because such factors contribute also to the matrix element $\langle \text{VAC, out}|\text{VAC, in}\rangle$. In calculating the *connected* part of vacuum expectation values of time-ordered products (or the S-matrix) we eliminate the contribution of disconnected vacuum fluctuation subgraphs by dividing by $\langle \text{VAC, out}|\text{VAC, in}\rangle$, and any constant factors in the vacuum expectation values cancel in this ratio.

We could go on and calculate matrix elements between multi-particle states, by inserting the appropriate 'wave functionals' in Eq. (9.2.3). These can be calculated by applying the adjoints of annihilation operators such as (9.2.7) to the vacuum state; just as for the harmonic oscillator, these wave functionals turn out to be Hermite polynomials in the field times the vacuum Gaussian. We do not need to work all this out here, because as shown in Section 6.4, the vacuum expectation values (9.2.17) are all we need in order to be able to calculate S-matrix elements.

9.3 Lagrangian Version of the Path-Integral Formula

The integrand in the exponential in Eqs. (9.1.38) or (9.2.17) looks like the Lagrangian L associated with the Hamiltonian H. This appearance is somewhat misleading because here the 'momenta' $p_a(t)$ or $p_n(\mathbf{x}, t)$ are independent variables, not yet related to $q_a(t)$ or $q_n(\mathbf{x}, t)$ or their derivatives. However, there is a large and important class of theories in which the integral over the 'momenta' can be done by just replacing them with the values dictated by the canonical formalism, in which case the integrand in the exponential in the path integrals really is the Lagrangian.

These are theories with a Hamiltonian that is quadratic in the 'momenta' — in the language of field theory

$$H[Q,P] = \tfrac{1}{2}\sum_{nm}\int d^3x\, d^3y\, A_{\mathbf{x}n,\mathbf{y}m}[Q]\, P_n(\mathbf{x})\, P_m(\mathbf{y}) + \sum_n \int d^3x\, B_{\mathbf{x}n}[Q]\, P_n(\mathbf{x}) + C[Q] \tag{9.3.1}$$

with a 'matrix' A that is real, symmetric, positive, and non-singular. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

argument of the exponential in Eq. (9.2.17) is then quadratic in the ps:

$$\begin{aligned}&\int d\tau \left\{ \int d^3x \sum_n p_n(\mathbf{x},\tau)\dot{q}_n(\mathbf{x},\tau) - H\Big[q(\tau),p(\tau)\Big] \right\} \\ &= -\tfrac{1}{2}\sum_{nm} \int d^3x\, d^3y\, d\tau\, d\tau'\, \mathscr{A}_{\tau\mathbf{x}n,\tau'\mathbf{y}m}[q]\; p_n(\mathbf{x},\tau)\; p_m(\mathbf{y},\tau') \\ &\quad -\sum_n \int d^3x \int d\tau\, \mathscr{B}_{\tau\mathbf{x}n}[q]\, p_n(\mathbf{x},\tau) \;-\; \mathscr{C}[q]\,, \end{aligned} \tag{9.3.2}$$

where

$$\mathscr{A}_{\tau\mathbf{x}n,\tau'\mathbf{y}m}[q] \equiv A_{\mathbf{x}n,\mathbf{y}m}[q(\tau)]\delta(\tau-\tau')\,, \tag{9.3.3}$$

$$\mathscr{B}_{\tau\mathbf{x}n}[q] \equiv B_{\mathbf{x}n}[q(\tau)] - \dot{q}_n(\mathbf{x},\tau)\,, \tag{9.3.4}$$

$$\mathscr{C}[q] \equiv \int d\tau\, C[q(\tau)]\,. \tag{9.3.5}$$

Now, in general the integral of the exponential of a quadratic expression like (9.3.2) will be proportional to the exponential evaluated at the stationary point of its argument. For a finite number of real variables ξ_s, this formula reads

$$\begin{aligned}&\int_{-\infty}^{\infty} \left(\prod_s d\xi_s\right) \exp\left\{ -\tfrac{1}{2}i\sum_{sr} \mathscr{A}_{sr}\xi_s\xi_r - i\sum_s \mathscr{B}_s\xi_s - i\mathscr{C} \right\} \\ &= (\mathrm{Det}\,[i\mathscr{A}/2\pi])^{-1/2} \exp\left\{ -i\,\tfrac{1}{2}\sum_{sr} \mathscr{A}_{sr}\bar{\xi}_s\bar{\xi}_r - i\sum_s \mathscr{B}_s\bar{\xi}_s - i\mathscr{C} \right\}, \end{aligned} \tag{9.3.6}$$

where $\bar{\xi}$ is the stationary point

$$\bar{\xi}_s \;=\; -\sum_r (\mathscr{A}^{-1})_{sr}\mathscr{B}_r\,. \tag{9.3.7}$$

(For a proof of this formula, see the Appendix to this chapter.) Hence, as long as the $\mathcal{O}_A, \mathcal{O}_B$, etc. in Eq. (9.2.17) are independent of the ps, for such a Hamiltonian we can evaluate the path integral over the ps in Eq. (9.2.17) by setting these variables at the stationary point of the quadratic expression in the argument of the exponential. But the variational derivative of this quadratic is

$$\begin{aligned}&\frac{\delta}{\delta p_n(\mathbf{x},\tau)} \int_{-\infty}^{\infty} d\tau \left\{ \int d^3x\; \dot{q}_n(\mathbf{x},\tau)p_n(\mathbf{x},\tau) - H\Big[q(\tau),p(\tau)\Big] + i\epsilon \text{ terms} \right\} \\ &\qquad = \dot{q}_n(\mathbf{x},\tau) - \frac{\delta}{\delta p_n(\mathbf{x},\tau)} H\Big[q(\tau),p(\tau)\Big]\,. \end{aligned}$$

(The $i\epsilon$ terms depend only on the qs.) Thus the stationary 'point' $\bar{p}_n(\mathbf{x},t)$ where this vanishes is just the value of $p_n(\mathbf{x},t)$ dictated by the canonical

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

formula

$$\dot{q}_n(\mathbf{x},\tau) = \left[\frac{\delta H\left[q(\tau),p(\tau)\right]}{\delta p_n(\mathbf{x},\tau)}\right]_{p=\bar{p}} . \tag{9.3.8}$$

With $p_n(\mathbf{x},t)$ set equal to this value, the argument of the exponential in Eq. (9.2.17) *is* the ordinary Lagrangian

$$L\left[q(\tau),\dot{q}(\tau)\right] \equiv \int d^3x \left(\sum_n \dot{q}_n(\mathbf{x},\tau)\bar{p}_n(\mathbf{x},t) - H\left[q(\tau),\bar{p}(\tau)\right]\right) \tag{9.3.9}$$

and we can write Eq. (9.2.17) as

$$\begin{aligned}
&\left\langle \text{VAC, out}\middle| T\left\{\mathcal{O}_A\left[Q(t_A)\right],\mathcal{O}_B\left[Q(t_B)\right],\cdots\right\}\middle|\text{VAC, in}\right\rangle \\
&= |\mathcal{N}|^2 \int \prod_{\tau,\mathbf{x},n} dq_n(\mathbf{x},\tau)\left(\text{Det}\left[2i\pi\mathcal{A}[q]\right]\right)^{-1/2} \\
&\quad \times \mathcal{O}_A\left[q(t_A)\right]\ \mathcal{O}_B\left[q(t_B)\right]\cdots \\
&\quad \times \exp\left[i\int_{-\infty}^{\infty} d\tau\left\{L\left[q(\tau),\dot{q}(\tau)\right] + i\epsilon \text{ terms}\right\}\right] .
\end{aligned} \tag{9.3.10}$$

(We have combined the $1/2\pi$ factors in the integrals over the p_n with the determinant coming from Eq. (9.3.6).) This is the desired Lagrangian form of the path-integral formula.

In deriving Eq. (9.3.10), it was necessary to assume that the operators $\mathcal{O}_A,\mathcal{O}_B,\cdots$ were independent of the canonical 'momenta'. This is not as restrictive as it may seem. For instance, in a scalar field theory for which the canonical conjugate to Φ is $\Pi = \dot{\Phi}$, it is possible to calculate the matrix element of a time-ordered product of operators, one of which is $\dot{\Phi}(t)$, by taking the difference of matrix elements in which this operator is replaced with $\Phi(t+d\tau)$ and $\Phi(t)$, and then dividing by $d\tau$, with $d\tau \to 0$. Equivalently, as long as t is not equal to any of the other time-arguments of the operator in Eq. (9.3.10), we can simply differentiate Eq. (9.3.10) with respect to t.

The one serious remaining complication in Eq. (9.3.10) is the determinant of $\mathcal{A}[q]$. If $\mathcal{A}[q]$ is field-independent, then this is no problem; we have already noted that overall constants make no contribution to the connected parts of vacuum expectation values, in which we divide by a vacuum-vacuum amplitude proportional to the same constant factor. This is the case for instance for the theory of a set of scalar fields Φ_n with non-derivative coupling to each other and/or derivative coupling to external currents J_n. The Lagrangian density here is

$$\mathcal{L} = -\sum_n \left[\tfrac{1}{2}\partial_\lambda\Phi_n\partial^\lambda\Phi_n + J_n{}^\lambda\partial_\lambda\Phi_n\right] - V(\Phi) .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

An obvious extension of the results of Section 7.5 from one to several derivatively coupled scalars shows that this Lagrangian implies the Hamiltonian

$$H = \int d^3x \sum_n \Big[\tfrac{1}{2}\Pi_n^2 + \tfrac{1}{2}(\nabla\Phi_n)^2 + \mathbf{J}_n \cdot \nabla\Phi_n + J_n{}^0\Pi_n + \tfrac{1}{2}(J_n{}^0)^2 \Big] + \int d^3x\, V(\Phi) .$$

(The Φ_n are taken to be real scalars, but complex scalars can be accommodated by separating them into real and imaginary parts.) In general there is a non-trivial term that is linear in the Π_n, but the coefficient of the quadratic term is a constant, just the unit 'matrix':

$$\mathscr{A}_{xn,x'n'} = \delta^4(x - x')\,\delta_{nn'} .$$

The factor $\Big(\mathrm{Det}\big[2i\pi\mathscr{A}[q]\big]\Big)^{-1/2}$ in Eq. (9.3.10) is here a field-independent constant, and therefore is without effect.

However, matters are not always so simple. As a second example, let us consider the so-called non-linear σ-model, with Lagrangian density

$$\mathscr{L} = -\tfrac{1}{2}\sum_{nm} \partial_\lambda\Phi_n\partial^\lambda\Phi_m\big[\delta_{nm} + U_{nm}(\Phi)\big] - V(\Phi) .$$

A straightforward calculation gives the Hamiltonian as

$$H = \int d^3x \left[\tfrac{1}{2}\Pi_n\Big(1 + U(\Phi)\Big)^{-1}_{nm}\Pi_m + \tfrac{1}{2}\nabla\Phi_n \cdot \nabla\Phi_m \left(1 + U(\Phi)\right)_{nm} + V(\Phi) \right].$$

Here $\mathscr{A}$ is the field-dependent quantity

$$\mathscr{A}_{nx,my} = \big[1 + U(\Phi(x))\big]^{-1}_{nm}\delta^4(x - y) .$$

In cases of this sort, the determinant may be reexpressed as a contribution to the effective Lagrangian, using the relation $\mathrm{Det}\,\mathscr{A} = \exp\,\mathrm{Tr}\ln\mathscr{A}$. By replacing the continuum of spacetime positions with a discrete lattice of points surrounded by separate regions of very small spacetime volume Ω, we may interpret the delta function in $\mathscr{A}_{nx,my}$ as $\delta^4(x - y) = \Omega^{-1}\delta_{x,y}$, so that

$$(\ln\mathscr{A})_{nx,my} = \delta_{x,y}\big[-\ln(1 + U(\Phi(x))) - 1\cdot\ln\Omega\big]_{nm}$$

with the logarithm of a matrix defined now by its power series expansion

$$\ln(1 + U) = U - \frac{U^2}{2} + \frac{U^3}{3} - \cdots .$$

To evaluate the trace, we note that $\sum_x \cdots = \Omega^{-1}\int d^4x \cdots$. The determi-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

nant factor here is then

$$\text{Det}\,\mathscr{A} \propto \exp\left[-\Omega^{-1}\int d^4x\ \text{tr}\ \ln\left[1+U(\Phi(x))\right]\right],$$

where 'tr' is to be understood as the trace in an ordinary matrix sense. The constant of proportionality (which arises from the $-\ln\Omega$ term) is field-independent and therefore of no present interest. We can regard this determinant as providing a correction to the effective Lagrangian density

$$\Delta\mathscr{L} = -\tfrac{1}{2}i\Omega^{-1}\text{tr}\ \ln\left[1+U(\Phi(x))\right].$$

The factor Ω^{-1} may be written as an ultraviolet divergent integral

$$\Omega^{-1} = \delta^4(x-x) = (2\pi)^{-4}\int d^4p\ \cdot\ 1\ .$$

We shall not show this here, but the extra terms in the Feynman diagrams for this theory contributed by $\Delta\mathscr{L}$ could also have been derived in the canonical formalism by taking account of the equal-time-commutator terms in the propagator of time-derivatives of the scalar field.[7] Ignoring this correction would lead to a spurious dependence of the S-matrix on the way that the scalar field is defined, and would also be inconsistent with any symmetries of the Lagrangian under transformations of the scalar fields.

Even where the factor $(\text{Det}\,\mathscr{A})^{-1/2}$ in the path-integral formula (9.3.10) is field-independent, the Lagrangian in this formula may not be the same as the one with which we started. As an example, let's consider the theory of a set of real vector fields, with Lagrangian density

$$\mathscr{L} = -\sum_n\Big[\tfrac{1}{4}\left(\partial_\mu A_{n\lambda}-\partial_\lambda A_{n\mu}\right)\left(\partial^\mu A_n{}^\lambda-\partial^\lambda A_n{}^\mu\right) + \tfrac{1}{2}m_n^2\,A_{n\lambda}A_n{}^\lambda + J_n{}^\lambda A_{n\lambda}\Big],$$

where the currents $J_n{}^\mu$ are either externally produced c-number quantities or depend on other fields (in which case terms describing these other fields are to be added to the Lagrangian). By a simple extension of the results of Section 7.5, we see that the Hamiltonian is

$$H = \int d^3x\sum_n\left[\frac{1}{2}\boldsymbol{\Pi}_n^2+\frac{1}{2}(\nabla\times\mathbf{A}_n)^2+\frac{1}{2}m_n^2\mathbf{A}_n{}^2 + \frac{1}{2m_n^2}(\nabla\cdot\boldsymbol{\Pi}_n)^2+\mathbf{J}_n\cdot\mathbf{A}_n-\frac{1}{m_n^2}J_n^0\nabla\cdot\boldsymbol{\Pi}_n+\frac{1}{2m_n^2}(J_n^0)^2\right]$$

again with the understanding that other terms must be added involving any fields that appear in $J_n{}^\mu$. Here the coefficient of the quadratic term is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

somewhat more complicated than in our first example:

$$\mathscr{A}_{nix,mjy} = \delta_{nm}\left[\delta_{ij}\delta^4(x-y) - \frac{1}{2m_n^2}\nabla_i\nabla_j\delta^4(x-y)\right] ,$$

but it is field-independent, so that the factor $(\mathrm{Det}\,\mathscr{A})^{-1/2}$ has no effect. On the other hand, the Lagrangian (9.3.9) is here not the one with which we started; it is expressed entirely in terms of $\mathbf{A}$ and its spacetime derivatives, with no dependence on any time-component A^0. For this reason, the Lorentz invariance of Eq. (9.3.10) is far from obvious.

To remedy this, we may reintroduce the auxiliary field. Suppose we add to the Hamiltonian a term

$$\Delta H = -\tfrac{1}{2}\sum_n m_n^2 \int d^3x \left[A_n^0 - m_n^{-2}\nabla\cdot\mathbf{\Pi}_n + m_n^{-2}J_n^0\right]^2$$

and integrate over the A_n^0 as well as the $\mathbf{A}_n$ and $\mathbf{\Pi}_n$. This can only introduce a field-independent overall factor, since ΔH is a quadratic in A^0 (with a field-independent coefficient in the term of second order in A^0) whose stationary value vanishes. However, suppose that we now integrate over the $\mathbf{\Pi}_n$ *before* integrating over the A_n^0. The Hamiltonian in the path integral (9.2.17) is here replaced with

$$H + \Delta H = \int d^3x \sum_n \Big[\tfrac{1}{2}\mathbf{\Pi}_n^2 + \tfrac{1}{2}(\nabla\times\mathbf{A}_n)^2 + \tfrac{1}{2}m_n^2\mathbf{A}_n{}^2 - \tfrac{1}{2}m_n^2(A_n^0)^2 + \mathbf{J}_n\cdot\mathbf{A}_n - J_n^0 A_n^0 + A_n^0\nabla\cdot\mathbf{\Pi}_n\Big] .$$

This is still quadratic in $\mathbf{\Pi}_n$ with a field-independent (and somewhat simpler) coefficient of the quadratic term, so the integral over the $\mathbf{\Pi}_n$s can be done by just replacing $\mathbf{\Pi}_n$ with its value at the stationary point of the functional $\sum_n \int d^3x\,\mathbf{\Pi}_n\cdot\dot{\mathbf{A}}_n - H - \Delta H$:

$$\mathbf{\Pi}_n = \dot{\mathbf{A}}_n + \nabla A_n^0 .$$

With $\mathbf{\Pi}_n$ eliminated in this way, $\sum_n \int d^3x\,\mathbf{\Pi}_n\cdot\dot{\mathbf{A}}_n - H - \Delta H$ is just the Lorentz-invariant Lagrangian with which we started.

In order to take account of the possible need to introduce auxiliary fields like A_n^0, from now on we shall write the path-integral formula after elimination of the canonical conjugates in terms of fields ψ_ℓ that include both canonical fields q_n and auxiliary fields c_r:

$$\begin{aligned}&\left\langle \mathrm{VAC,out}\left|T\left\{\mathscr{O}_A\left[\Psi_A(t_A)\right],\mathscr{O}_B\left[\Psi_B(t_B)\right],\cdots\right\}\right|\mathrm{VAC,in}\right\rangle\\ &\propto \int\prod_{\tau,\mathbf{x},n} d\psi_n(\mathbf{x},\tau)\;\mathscr{O}_A\left[\psi(t_A)\right]\;\mathscr{O}_B\left[\psi(t_B)\right]\cdots\\ &\times\exp\left[i\int_{-\infty}^{\infty}d\tau\left\{L\left[\psi(\tau),\dot\psi(\tau)\right]+i\epsilon\text{ terms}\right\}\right] ,\end{aligned} \tag{9.3.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

it now being understood that L includes any terms arising from a possible field-dependent factor $(\mathrm{Det}\,\mathscr{A})^{-1/2}$.

9.4 Path-Integral Derivation of Feynman Rules

We are now ready to use the path-integral formalism to derive the Feynman rules in a wide class of theories. We will concentrate here on the vacuum expectation values of time-ordered products of field operators (and their adjoints),

$$M_{\ell_A\ell_B\cdots}(x_A x_B\cdots)=\frac{\langle \mathrm{VAC,out}\,|\,T\{\Psi_{\ell_A}(x_A),\Psi_{\ell_B}(x_B)\cdots\}|\,\mathrm{VAC,in}\rangle}{\langle \mathrm{VAC,out}|\mathrm{VAC,in}\rangle} \tag{9.4.1}$$

from which S-matrix elements may be obtained (as shown in Section 6.4) by stripping off the final propagators associated with each field, replacing them with the coefficient functions that multiply creation or annihilation operators in the corresponding free fields, and summing over the indices on these coefficient functions.

For the simpler theories whose Hamiltonian is quadratic in the Πs, Eq. (9.3.11) gives

$$M_{\ell_A\ell_B\cdots}(x_A x_B\cdots)=\frac{\int\left[\prod_{x,\ell}d\psi_\ell(x)\right]\psi_{\ell_A}(x_A)\,\psi_{\ell_B}(x_B)\cdots e^{iI[\psi]}}{\int\prod_{x,\ell}d\psi_\ell(x)\,e^{iI[\psi]}}\,, \tag{9.4.2}$$

where $I[\psi]$ is the action

$$I[\psi]=\int_{-\infty}^{\infty}d\tau\left\{L\Big[\psi(\tau),\dot\psi(\tau)\Big]+i\epsilon\text{ terms}\right\} \tag{9.4.3}$$

with L now including any terms that may arise from a field-dependent determinant in Eq. (9.3.10).

Let us now suppose that the Lagrangian is the integral of a Lagrangian density, consisting of a quadratic term $\mathscr{L}_0$ which would be present in the absence of interactions, plus a Lagrangian interaction density $\mathscr{L}_1$:

$$\begin{aligned}L\Big[\psi(\tau),\dot\psi(\tau)\Big]=\int d^3x\,\Big[&\mathscr{L}_0\Big(\psi(\vec x,\tau),\,\partial_\mu\psi(\vec x,\tau)\Big)\\&+\mathscr{L}_1\Big(\psi(\vec x,\tau),\,\partial_\mu\psi(\vec x,\tau)\Big)\Big]\,.\end{aligned} \tag{9.4.4}$$

That is, the action (9.4.3) is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$$I[\psi] = I_0[\psi] + I_1[\psi], \tag{9.4.5}$$

$$I_0[\psi] = \int d^4x\ \mathscr{L}_0\Big(\psi(x), \partial_\mu\psi(x)\Big) + i\epsilon \text{ terms}, \tag{9.4.6}$$

$$I_1[\psi] = \int d^4x\ \mathscr{L}_1\Big(\psi(x), \partial_\mu\psi(x)\Big)\ . \tag{9.4.7}$$

Since $\mathscr{L}_0$ and the '$i\epsilon$ terms' are quadratic in the fields, we may always write I_0 in the generalized quadratic form

$$I_0[\psi] = -\tfrac{1}{2}\int d^4x\ d^4x' \sum_{\ell,\ell'} \mathscr{D}_{\ell x,\ell' x'}\ \psi_\ell(x)\ \psi_{\ell'}(x')\ . \tag{9.4.8}$$

For instance, for a real scalar field of mass m, the unperturbed Lagrangian is

$$\mathscr{L}_0 = -\tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2 \tag{9.4.9}$$

and the $i\epsilon$ terms in I_0 are given by Eq. (9.2.16), as

$$\tfrac{1}{2}i\epsilon\int dt\int d^3x\, d^3x'\ \mathscr{E}(\mathbf{x},\mathbf{x}')\ \phi(\mathbf{x},t)\phi(\mathbf{x}',t) \tag{9.4.10}$$

so here

$$\mathscr{D}_{x,x'} = \frac{\partial}{\partial x^\mu}\ \frac{\partial}{\partial x'_\mu}\ \delta^4(x-x') + m^2\ \delta^4(x-x') - i\epsilon\mathscr{E}(\mathbf{x},\mathbf{x}')\delta(t-t')\ . \tag{9.4.11}$$

(We are now dropping the factor $e^{-\epsilon|\tau|}$ in the $i\epsilon$ term, since it produces a correction of higher order in ϵ.) To deal with interactions, we will expand the exponential in powers of I_1,

$$\exp(iI[\psi]) = \exp(iI_0[\psi])\sum_{N=0}^{\infty}\ \frac{i^N}{N!}\ \Big(I_1[\psi]\Big)^N \tag{9.4.12}$$

and then expand I_1 in powers of the fields. The general integrals that we encounter in the numerator and the denominator of Eq. (9.4.2) are of the form

$$\mathscr{I}_{\ell_1\ell_2\cdots}(x_1x_2\cdots) \equiv \int\left(\prod_{\ell,x} d\psi_\ell(x)\right)\ e^{iI_0[\psi]}\ \psi_{\ell_1}(x_1)\ \psi_{\ell_2}(x_2)\cdots\ , \tag{9.4.13}$$

where the field factors $\psi_{\ell_1}(x_1), \psi_{\ell_2}(x_2)$, etc. arise from $I_1[\psi]$ and/or from the field factors $\psi_{\ell_A}(x_A)$ etc. originally present in the numerator of Eq. (9.4.2). With $I_0[\psi]$ of the form (9.4.8), the integral (9.4.13) is of the same form as the integral evaluated in the Appendix to this chapter, with the discrete index s replaced with the pair of labels ℓ, x. We can therefore

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

use Eqs. (9.A.12) and (9.A.15), which give here

$$\mathscr{I}_{\ell_1\ell_2\cdots}(x_1x_2\cdots) = \left[\mathrm{Det}\left(\frac{i\mathscr{D}}{2\pi}\right)\right]^{-1/2} \sum_{\substack{\text{pairings of}\\ \text{fields}}} \prod_{\text{pairs}} \left[-i\mathscr{D}^{-1}\right]_{\text{paired fields}} . \tag{9.4.14}$$

This just amounts to the coordinate-space Feynman rules for calculating the numerator of Eq. (9.4.2) in their covariant form: we expand in the interaction I_1, and then sum over the ways of pairing the fields in the I_1s with each other and with the fields $\psi_{\ell_A}(x_A)$, etc., with the contribution of each pairing being given by the spacetime integral of the product of the coefficients of the fields in $I_1[\psi]$ and the product of the 'propagators' $-i\Delta$, where

$$\Delta_{\ell_1\ell_2}(x_1, x_2) = (\mathscr{D}^{-1})_{\ell_1 x_1, \ell_2 x_2} . \tag{9.4.15}$$

(The factor $[\mathrm{Det}(i\mathscr{D}/2\pi)]^{-1/2}$ in Eq. (9.4.14) actually represents the contribution of graphs with unlimited numbers of single loops unattached to any other lines, but in any case this factor cancels in the ratio (9.4.2).)

It remains to calculate the propagators (9.4.15). We interpret Eq. (9.4.15) as an integral equation

$$\sum_{\ell_2} \int d^4x_2\, \mathscr{D}_{\ell_1 x_1, \ell_2 x_2} \Delta_{\ell_2 \ell_3}(x_2, x_3) = \delta^4(x_1 - x_3)\delta_{\ell_1\ell_3} . \tag{9.4.16}$$

In the absence of external fields, translation invariance will make $\mathscr{D}$ necessarily a function only of $x_1 - x_2$, which can be written as a Fourier integral

$$\mathscr{D}_{\ell_1 x_1, \ell_2 x_2} \equiv (2\pi)^{-4} \int d^4p\, e^{ip\cdot(x_1 - x_2)}\, \mathscr{D}_{\ell_1\ell_2}(p) . \tag{9.4.17}$$

The solution of Eq. (9.4.16) is then

$$\Delta_{\ell_1\ell_2}(x_1, x_2) = (2\pi)^{-4} \int d^4p\, e^{ip\cdot(x_1 - x_2)}\, \mathscr{D}^{-1}_{\ell_1\ell_2}(p) , \tag{9.4.18}$$

where $\mathscr{D}^{-1}$ is the ordinary inverse of the matrix $\mathscr{D}$. As we will see, the $i\epsilon$ terms have the effect of making the inverse well-defined for all real values of p. We have thus reduced the problem of calculating the propagator to that of taking the inverse of a finite matrix.

First consider a massive scalar field, for which the kernel $\mathscr{D}$ takes the form (9.4.11). We can write this as a Fourier integral

$$\mathscr{D}_{x,y} = (2\pi)^{-4} \int d^4p\, e^{ip\cdot(x-y)} \Big(p^2 + m^2 - i\epsilon\, E(\mathbf{p})\Big) ,$$

so the propagator is

$$\Delta(x, y) = (2\pi)^{-4} \int d^4p\, e^{ip\cdot(x-y)} \Big(p^2 + m^2 - i\epsilon E(\mathbf{p})\Big)^{-1} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

We recognize this as the same scalar propagator previously obtained by operator methods. (The difference between ϵ and $\epsilon E(\mathbf{p})$ is immaterial, since both are just positive infinitesimals.)

For a second example, consider a real massive vector field. The unperturbed Lagrangian is

$$\mathscr{L}_0 = -\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) - \tfrac{1}{2}m^2 A_\mu A^\mu .$$

We can again write $I_0[\psi]$ in the form (9.4.8), with kernel

$$\begin{aligned}\mathscr{D}_{\rho x,\sigma y} &= \left[\eta_{\rho\sigma}\frac{\partial^2}{\partial x^\mu \partial y_\mu} - \frac{\partial^2}{\partial x^\sigma \partial y^\rho} + m^2\eta_{\rho\sigma}\right]\delta^4(x-y) + i\epsilon \text{ terms} \\ &= (2\pi)^{-4}\int d^4p\, e^{ip\cdot(x-y)}\left[\eta_{\rho\sigma}p^2 - p_\rho p_\sigma + m^2\eta_{\rho\sigma} + i\epsilon \text{ terms}\right] .\end{aligned}$$

We will not bother to show it here, but the '$+\ i\epsilon$ terms' here take the simple form $-i\epsilon E(\mathbf{p})\eta_{\rho\sigma}$. The vector field propagator is then given by simply inverting the 4×4 matrix in the integrand

$$\Delta_{\rho\sigma}(x,y) = (2\pi)^{-4}\int \frac{d^4p\, e^{ip\cdot(x-y)}}{p^2 + m^2 - i\epsilon E(\mathbf{p})}\left[\eta_{\rho\sigma} + \frac{p_\rho p_\sigma}{m^2}\right] .$$

(Terms proportional to ϵ are dropped in the numerator. They are important in the denominator in defining how the integrand is to be treated near the mass shell, $p^2 = -m^2$.) This is the same as the propagator derived by operator methods, except that the non-covariant terms proportional to $\delta(x^0 - y^0)$ are now absent. These non-covariant terms were previously needed to cancel non-covariant terms in the interaction Hamiltonian, but the vertex contributions in the Feynman rules are now obtained directly by inspection of the covariant Lagrangian, and no such cancellation is needed.

Theories with derivative coupling are equally simple. The factor arising from the pairing of a field derivative $\partial_\mu\psi_\ell(x)$ with any other field $\psi_m(y)$ (perhaps itself a field derivative) is

$$\begin{aligned}\langle\partial_\mu\psi_\ell(x)\,\psi_m(y)\rangle &= \frac{\int\left[\prod_{x,\ell} d\psi_\ell(x)\right]\partial_\mu\psi_\ell(x)\,\psi_m(y)\,e^{iI[\psi]}}{\int\left[\prod_{x,\ell} d\psi_\ell(x)\right]e^{iI[\psi]}} \\ &= \frac{\partial}{\partial x^\mu}\langle\psi_\ell(x)\,\psi_m(y)\rangle . \qquad (9.4.19)\end{aligned}$$

Such propagators have no non-covariant pieces. For instance, for a real scalar field, the pairing of $\partial_\mu\phi$ with $\partial_\nu\phi$ gives a momentum space propagator $k_\mu k_\nu/(k^2 + m^2 - i\epsilon)$. Also, as we saw in the previous section, vertices in the theory of a scalar field with derivative couplings to other fields may be read off from the Lagrangian, and are separately covariant.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

9.5 Path Integrals for Fermions

We now turn to the problem of extending the path-integral formalism to cover theories containing fermions as well as bosons. It would be easy to proceed in a purely formal way, by analogy with the bosonic case, with the justification that this gives the 'right' Feynman rules. Instead, we will here derive the path-integral formalism for fermions directly from the principles of quantum mechanics, as we did for bosons.[9]

As before, we will start with a general quantum mechanical system, with 'coordinates' Q_a and canonical conjugate 'momenta' P_a, but now satisfying anticommutation rather than commutation relations:

$$\left\{Q_a, P_b\right\} = i\,\delta_{ab}\,, \tag{9.5.1}$$

$$\left\{Q_a, Q_b\right\} = \left\{P_a, P_b\right\} = 0\,. \tag{9.5.2}$$

(These are Schrödinger-picture operators, or in other words Heisenberg-picture operators at time $t = 0$.) Later we will replace the discrete index a with a spatial position $\mathbf{x}$ and a field index m.

We wish first to construct a complete basis for the states on which the Qs and Ps act. Note that for any given a, we have

$$Q_a^2 = P_a^2 = 0\,. \tag{9.5.3}$$

It follows that there will always be a 'ket' state $|0\rangle$ annihilated by all Q_a:

$$Q_a|0\rangle = 0\,, \tag{9.5.4}$$

and a 'bra' state $\langle 0|$ annihilated (from the right) by all P_a:

$$\langle 0|\,P_a = 0\,. \tag{9.5.5}$$

For instance, we can take

$$|0\rangle \propto \left(\prod_a Q_a\right)|f\rangle\,, \qquad \langle 0| \propto \langle g|\left(\prod_a P_a\right)\,,$$

where $|f\rangle$ and $\langle g|$ are any kets and bras for which these expressions do not vanish. (They cannot vanish for all $|f\rangle$ and $\langle g|$ unless the operators $\prod_a Q_a$ and $\prod_a P_a$ vanish, which we assume not to be the case.) These states satisfy Eqs. (9.5.4) and (9.5.5) by virtue of Eq. (9.5.3). They are not in general unique, because there may be other bosonic degrees of freedom that distinguish the various possible $|0\rangle$ and $\langle 0|$, but for simplicity we will limit ourselves here to the case where the only degrees of freedom are those described by the fermionic operators Q_a and P_a, and will assume that the states satisfying Eqs. (9.5.4) and (9.5.5) are unique up to constant factors, which we choose so that

$$\langle 0|0\rangle = 1\,. \tag{9.5.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

(Note that this normalization convention could not be imposed if we had defined $\langle 0|$ as the left-eigenstate of the Q_a with eigenvalue zero, because in this case $\langle 0|\{Q_a, P_b\}|0\rangle$ would vanish, which with Eq. (9.5.1) would imply that $\langle 0|0\rangle = 0$.)

As we saw in Section 7.5, in the Dirac theory Q_a is not Hermitian, but instead has an adjoint $-iP_a$, in which case $\langle 0|$ *can* be regarded as simply the adjoint of $|0\rangle$. However, there are fermionic operators (such as the 'ghost' fields to be introduced in Volume II) for which P_a is unrelated to the adjoint of Q_a. In what follows we will not need to assume anything about the adjoints of Q_a or P_a, or about any relation between $|0\rangle$ and $\langle 0|$.

A complete basis for the states of this system is provided by $|0\rangle$ and the states (antisymmetric in indices $a, b, \cdots$)

$$|a, b, \cdots\rangle \equiv P_a P_b \cdots |0\rangle \tag{9.5.7}$$

with any number of *different* Ps acting on $|0\rangle$. That is, the result of acting on these states with any operator function of the Ps and Qs can be written as a linear combination of the same set of states. In particular, if an index a is unequal to any of the indices appearing in $|b, c, \cdots\rangle$, then

$$Q_a|b, c, \cdots\rangle = 0\,, \tag{9.5.8}$$

$$P_a|b, c, \cdots\rangle = |a, b, c \cdots\rangle\,. \tag{9.5.9}$$

On the other hand, if a *is* equal to one of the indices in the sequence, $b, c, \cdots$, we can always rewrite the state (possibly changing its sign) so that a is the first of these indices, in which case we have

$$Q_a|a, b, c, \cdots\rangle = i|b, c \cdots\rangle\,, \tag{9.5.10}$$

$$P_a|a, b, c, \cdots\rangle = 0\,. \tag{9.5.11}$$

Similarly, we may define a complete dual basis, consisting of $\langle 0|$ and the states (also antisymmetric in the indices)

$$\langle a, b, \cdots| \equiv \langle 0|\cdots(-iQ_b)(-iQ_a)\,. \tag{9.5.12}$$

Using Eqs. (9.5.4)–(9.5.6) and the anticommutation relation (9.5.1), we see that the scalar products of these states take the values

$$\begin{aligned}\langle c, d, \cdots|a, b, \cdots\rangle &= \langle 0|\cdots(-iQ_d)(-iQ_c)\,P_aP_b\cdots|0\rangle \\ &= \begin{cases} 0 & \text{if } \{c, d, \cdots\} \neq \{a, b, \cdots\} \\ 1 & \text{if } c = a,\ d = b, \text{ etc.} \end{cases}\end{aligned} \tag{9.5.13}$$

where $\{\cdots\}$ here denotes the set of indices within the brackets, irrespective of order.

In deriving the Feynman rules, we would like to be able to rewrite sums over intermediate states like (9.5.7) as integrals over eigenstates of the Q_a or the P_a. However, it is not possible for these operators to have

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

eigenvalues (other than zero) in the usual sense. Suppose we try to find a state $|q\rangle$ that satisfies (for all a)

$$Q_a|q\rangle = q_a|q\rangle \,. \tag{9.5.14}$$

From Eq. (9.5.2) we see that

$$q_a q_b + q_b q_a = 0 \tag{9.5.15}$$

which is impossible for ordinary numbers. However, nothing can stop us from introducing an algebra of 'variables' (known as *Grassmann variables*) q_a, which act like c-numbers as far as the physical Hilbert space is concerned, but which still satisfy the anticommutation relations (9.5.15). We will require further that

$$\{q_a, q'_b\} = \{q_a, Q_b\} = \{q_a, P_b\} = 0, \tag{9.5.16}$$

where q and q' denote any two 'values' of these variables. We can now construct eigenstates $|q\rangle$ satisfying Eq. (9.5.14):

$$|q\rangle = \exp\left(-i\sum_a P_a q_a\right)|0\rangle \tag{9.5.17}$$

with the exponential defined as usual by its power series expansion. (To verify Eq. (9.5.14), use the fact that all $P_a q_a$ commute with one another and have zero square, so that

$$\begin{aligned}
[Q_a - q_a]\,|q\rangle &= [Q_a - q_a]\,\exp(-iP_a q_a)\,\exp\left(-i\sum_{b\neq a} P_b q_b\right)|0\rangle \\
&= [Q_a - q_a]\,[1 - iP_a q_a]\,\exp\left(-i\sum_{b\neq a} P_b q_b\right)|0\rangle \\
&= [-i\{Q_a, P_a\}q_a - q_a]\,\exp\left(-i\sum_{b\neq a} P_b q_b\right)|0\rangle = 0
\end{aligned}$$

as required by Eq. (9.5.14).) We can also define left-eigenstates $\langle q|$ (*not* the adjoints of $|q\rangle$), as

$$\langle q| \equiv \langle 0|\left(\prod_a Q_a\right)\exp\left(-i\sum_a q_a P_a\right) = \langle 0|\left(\prod_a Q_a\right)\exp\left(+i\sum_a P_a q_a\right), \tag{9.5.18}$$

where $\prod_a$ is the product in whatever order we take as standard. By the same argument as for Eq. (9.5.14), we see that

$$\langle q|\,Q_a \;=\; \langle q|q_a \,. \tag{9.5.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

These eigenstates have the scalar product

$$\langle q'|q\rangle = \langle 0|\left(\prod_a Q_a\right)\exp\left(i\sum_b P_b(q'_b - q_b)\right)|0\rangle$$
$$= \langle 0|\left(\prod_a Q_a\right)\left(\prod_b (1 + iP_b(q'_b - q_b))\right)|0\rangle .$$

Moving each Q_a to the right (starting with the rightmost) yields factors $i^2(q'_a - q_a)$, which we move to the right out of the scalar product, so

$$\langle q'|q\rangle = \prod_a (q_a - q'_a) . \tag{9.5.20}$$

We shall see that Eq. (9.5.20) plays the role of a delta function in integrals over the qs.

In the same way, we can construct right- and left-eigenstates of the P_a:

$$P_a|p\rangle = p_a|p\rangle , \tag{9.5.21}$$
$$\langle p|P_a = \langle p|p_a , \tag{9.5.22}$$

where the p_a are like q_a anticommuting c-numbers (taken for convenience to anticommute with the q_a and all fermionic operators as well as each other), and

$$|p\rangle = \exp\left(-i\sum_a Q_a p_a\right)\left(\prod_b P_b\right)|0\rangle , \tag{9.5.23}$$

$$\langle p| = \langle 0|\exp\left(-i\sum_a p_a Q_a\right) \tag{9.5.24}$$

with scalar product (now derived by moving the Ps to the left)

$$\langle p'|p\rangle = \prod_a (p'_a - p_a) . \tag{9.5.25}$$

The scalar products of these two sorts of eigenstate with each other are

$$\langle q|p\rangle = \langle q|\exp\left(-i\sum_a Q_a p_a\right)\left(\prod_a P_a\right)|0\rangle$$
$$= \left(\prod_a \exp(-iq_a p_a)\right)\langle q|\left(\prod_a P_a\right)|0\rangle$$
$$= \left(\prod_a \exp(-iq_a p_a)\right)\langle 0|\left(\prod_a Q_a\right)\left(\prod_a P_a\right)|0\rangle$$

and so

$$\langle q|p\rangle = \chi_N \exp\left(-i\sum_a q_a p_a\right) = \chi_N \exp\left(i\sum_a p_a q_a\right) , \tag{9.5.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

where χ_N is a phase that depends only on the number N of Q_a operators:

$$\chi_N \equiv \langle 0| \left(\prod_a Q_a \right) \left(\prod_a P_a \right) |0\rangle = i^N (-1)^{N(N-1)/2} .$$

Somewhat more simply, we also find

$$\langle p|q\rangle = \prod_a \exp(-ip_a q_a) . \tag{9.5.27}$$

It is easy to see that the states $|q\rangle$ are in a sense a complete set (and so also are the $|p\rangle$.) From the definitions (9.5.17), we see that the state $|a, b, \cdots\rangle$ in the general basis is (up to a phase) just the coefficient of the product $q_a q_b \cdots$ in an expansion of $|q\rangle$ in a sum of products of qs. Therefore we can write any state $|f\rangle$ in the form

$$|f\rangle = f_0 |q\rangle_0 + \sum_a f_a |q\rangle_a + \sum_{a \neq b} f_{ab} |q\rangle_{ab} + \cdots ,$$

where the fs are numerical coefficients, and a subscript $a, b, \cdots$ on $|q\rangle$ denotes the coefficient of $q_a q_b \cdots$ in $|q\rangle$.

In summing over states, it will be very convenient to introduce a sort of integration over fermionic variables, known as *Berezin integration*,[10] that is designed to pick out the coefficients of such products of anticommuting c-numbers. For any set of such variables ξ_n (either ps or qs or both together), the most general function $f(\xi)$ (either a c-number or a state-vector like $|q\rangle$) can be put in the form

$$f(\xi) = \left(\prod_n \xi_n \right) c \; + \; \text{terms with fewer } \xi \text{ factors} \tag{9.5.28}$$

and the integral over the ξs is defined simply by

$$\int \left(\widetilde{\prod_n} d\xi_n \right) f(\xi) \equiv c \tag{9.5.29}$$

with the tilde in Eq. (9.5.29) indicating that we use the convenient convention that the differentials are written in an order *opposite* to that of the product of integration variables in Eq. (9.5.28). Since this product is antisymmetric under the interchange of any two ξs, the integral is likewise antisymmetric under the interchange of any two $d\xi$s, so these 'differentials' effectively anticommute

$$d\xi_n d\xi_m + d\xi_m d\xi_n = 0 . \tag{9.5.30}$$

Also, the coefficient c may itself depend on other unintegrated c-number variables that anticommute with the ξs over which we integrate, in which case it is important to standardize the definition of c by moving all ξs to the left of c before integrating over them, as we have done in Eq. (9.5.28).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

For instance, the most general function of a pair of anticommuting c-numbers ξ_1 and ξ_2 takes the form

$$f(\xi_1,\xi_2) = \xi_1\xi_2 c_{12} + \xi_1 c_1 + \xi_2 c_2 + d$$

because the squares and all higher powers of ξ_1 and ξ_2 vanish. This function has the integrals

$$\int d\xi_1\, f(\xi_1,\xi_2) = \xi_2 c_{12} + c_1\,, \quad \int d\xi_2\, f(\xi_1,\xi_2) = -\xi_1 c_{12} + c_2\,,$$

$$\int d\xi_2\, d\xi_1\, f(\xi_1,\xi_2) = c_{12}\,.$$

Note that the multiple integral is the same as a repeated integral:

$$\int d\xi_2\, d\xi_1\, f(\xi_1,\xi_2) = \int d\xi_2 \Big[\int d\xi_1\, f(\xi_1,\xi_2)\Big]\,,$$

a result that can easily be extended to integrals over any number of fermionic variables. (It was in order to obtain this result without extra sign factors that we took the product of differentials in Eq. (9.5.29) to be in the opposite order to the product of variables in Eq. (9.5.28).) Indeed, we could have first defined the integral over a single anticommuting c-number ξ_1, and then defined multiple integrals in the usual way by iteration. The most general function of anticommuting c-numbers is linear in any one of them

$$f(\xi_1,\xi_2,\cdots) = b(\xi_2\cdots) + \xi_1 c(\xi_2\cdots)$$

(because $\xi_1^2 = 0$), and its integral over ξ_1 is defined as

$$\int d\xi_1 f(\xi_1,\xi_2,\cdots) = c(\xi_2,\cdots)\,.$$

Repeating this process leads to the same multiple integral as defined by Eqs. (9.5.28) and (9.5.29).

This definition of integration shares some other properties with multiple integrals (from $-\infty$ to $+\infty$) over ordinary real variables, but there are significant differences.

Obviously, Berezin integration is linear, in the sense that

$$\int \left(\tilde{\prod_n} d\xi_n\right) \big[f(\xi)+g(\xi)\big] = \int \left(\tilde{\prod_n} d\xi_n\right) f(\xi) + \int \left(\tilde{\prod_n} d\xi_n\right) g(\xi) \quad (9.5.31)$$

and also

$$\int \left(\tilde{\prod_n} d\xi_n\right) \big[f(\xi)a(\xi')\big] = \left[\int \left(\tilde{\prod_n} d\xi_n\right) f(\xi)\right] a(\xi')\,, \quad (9.5.32)$$

where $a(\xi')$ is any function (including a constant) of any anticommuting c-numbers ξ'_m over which we are *not* integrating. However, linearity with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

respect to left-multiplication is not so obvious. If we are integrating over ν variables, then since ξ'_m is assumed to anticommute with all ξ_n, we have

$$a\left((-)^\nu \xi'\right)\left(\prod_n \xi_n\right) = \left(\prod_n \xi_n\right) a(\xi')$$

and so

$$\int \left(\prod_n d\xi_n\right)\left[a\left((-)^\nu \xi'\right) f(\xi)\right] = a(\xi') \int \left(\prod_n d\xi_n\right) f(\xi)\,. \tag{9.5.33}$$

It is therefore very convenient (though not strictly necessary) to take the differentials $d\xi_n$ to anticommute with all anticommuting variables (including the ξ_n):

$$(d\xi_n)\xi'_m + \xi'_m(d\xi_n) = 0 \tag{9.5.34}$$

in which case Eq. (9.5.33) reads more simply

$$\int a(\xi')\left(\prod_n d\xi_n\right) f(\xi) = a(\xi') \int \left(\prod_n d\xi_n\right) f(\xi)\,. \tag{9.5.35}$$

Another similarity with ordinary integration is that, for an arbitrary anticommuting c-number ξ' independent of ξ,

$$\int \left(\prod_n d\xi_n\right) f(\xi + \xi') = \int \left(\prod_n d\xi_n\right) f(\xi) \tag{9.5.36}$$

since shifting ξ by a constant only affects the terms in f with fewer than the total number of ξ-variables.

On the other hand, consider a change of variables

$$\xi_n \to \xi'_n = \sum_m \mathscr{S}_{nm}\xi_m\,, \tag{9.5.37}$$

where $\mathscr{S}$ is an arbitrary non-singular matrix of ordinary numbers. The product of the new variables is

$$\prod_n \xi'_n = \sum_{m_1 m_2 \cdots} \left(\prod_n \mathscr{S}_{nm_n}\xi_{m_n}\right).$$

But $\prod_n \xi_{m_n}$ here is just the same as the product (in the original order) $\prod_n \xi_n$, except for a sign $\epsilon[m]$ which is $+1$ or -1 according to whether the permutation $n \to m_n$ is an even or odd permutation of the original order:

$$\prod_n \xi'_n = \left[\sum_{m_1 m_2 \cdots} \left(\prod_n \mathscr{S}_{nm_n}\right)\epsilon[m]\right] \prod_n \xi_n = (\mathrm{Det}\ \mathscr{S}) \prod_n \xi_n\,.$$

This applies whatever order we take for the ξ_n, as long as we take the ξ'_n in the same order. It follows that the coefficient of $\prod_n \xi'_n$ in any function

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$f(\xi)$ is just $(\mathrm{Det}\,\mathscr{S})^{-1}$ times the coefficient of $\prod_n \xi_n$, a statement we write as

$$\int \left(\tilde{\prod_n} d\xi'_n\right) f = (\mathrm{Det}\,\mathscr{S})^{-1} \int \left(\tilde{\prod_n} d\xi_n\right) f . \tag{9.5.38}$$

This is the usual rule for changing variables of integration, except that $(\mathrm{Det}\,\mathscr{S})$ appears to the power -1 instead of $+1$. We shall use Eq. (9.5.38) and the linearity properties (9.5.31), (9.5.32), and (9.5.35) later to evaluate the integrals encountered in deriving the Feynman rules for theories with fermions.

We can now use this definition of integration to write the completeness condition as a formula for an integral over eigenvalues. As already mentioned, any state $|f\rangle$ can be expanded in a series of the states $|0\rangle$, $|a\rangle$, $|a,b\rangle$, etc. and these states are (up to a phase) the coefficients of the products 1, q_a, $q_a q_b$, etc. in the Q-eigenstate $|q\rangle$. According to the definition of integration here, we can pick out the coefficient of any product $q_b q_c q_d \cdots$ in the state $|q\rangle$ by integrating the product of $|q\rangle$ with all q_a with a *not* equal to $b, c, d, \cdots$. Thus, by choosing a function $f(q)$ as a suitable sum of such products of qs, we can write any state $|f\rangle$ as an integral:

$$|f\rangle = \int \left(\tilde{\prod_a} dq_a\right) |q\rangle f(q) = \int |q\rangle \left(\tilde{\prod_a} dq_a\right) f(q) . \tag{9.5.39}$$

(We can move $|q\rangle$ to the left of the differentials without any sign changes because the exponential in Eq. (9.5.17) used to define $|q\rangle$ involves only even numbers of fermionic quantities.) To find the function $f(q)$ for a given state-vector $|f\rangle$, take the scalar product of Eq. (9.5.39) with some bra $\langle q'|$ (with q' any fixed Q-eigenvalue). According to Eqs. (9.5.35) and (9.5.20), this is

$$\langle q'|f\rangle = \int \left(\prod_a (q_a - q'_a)\right) \left(\tilde{\prod_b} dq_b\right) f(q) .$$

Moving every factor $(q_a - q'_a)$ to the right past every differential dq_b yields a sign factor $(-)^{N^2} = (-)^N$, where N is now the total number of q_a variables, so

$$\langle q'|f\rangle = (-)^N \int \left(\tilde{\prod_b} dq_b\right) \left(\prod_a (q_a - q'_a)\right) f(q) .$$

We can rewrite $f(q)$ as $f(q' + (q - q'))$ and expand in powers of $q - q'$. All terms beyond the lowest order vanish when multiplied with the product

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$\prod(q_a - q'_a)$, so

$$\left(\prod_a (q_a - q'_a)\right) f(q) = \left(\prod_a (q_a - q'_a)\right) f(q') , \tag{9.5.40}$$

which partly justifies our earlier remark that Eq. (9.5.20) plays the role of a delta function for integrals over the qs. Using Eq. (9.5.32), we now have

$$\langle q'|f\rangle = (-)^N \left[\int \left(\tilde{\prod_b} dq_b\right) \left(\prod_a (q_a - q'_a)\right) f(q')\right] .$$

The term in the integrand proportional to $\prod q_a$ has coefficient $f(q')$, so according to our definition of integration $\langle q'|f\rangle = (-)^N f(q')$. Inserting this back in Eq. (9.5.39) gives our completeness relation

$$|f\rangle = (-)^N \int |q\rangle \left(\tilde{\prod_b} dq_b\right) \langle q|f\rangle ,$$

or as an operator equation

$$1 = \int |q\rangle \left(\tilde{\prod_a} -dq_a\right) \langle q| . \tag{9.5.41}$$

In exactly the same way, we can also show that

$$1 = \int |p\rangle \left(\tilde{\prod_a} dp_a\right) \langle p| . \tag{9.5.42}$$

We are now in a position to calculate transition matrix elements. As before, we define time-dependent operators

$$Q_a(t) \equiv \exp(iHt)\, Q_a \exp(-iHt) \tag{9.5.43}$$

$$P_a(t) \equiv \exp(iHt)\, P_a \exp(-iHt) \tag{9.5.44}$$

and their right- and left-eigenstates

$$|q;t\rangle \equiv \exp(iHt)|q\rangle , \qquad |p;t\rangle \equiv \exp(iHt)|p\rangle , \tag{9.5.45}$$

$$\langle q;t| \equiv \langle q|\exp(-iHt) , \qquad \langle p;t| \equiv \langle p|\exp(-iHt) . \tag{9.5.46}$$

The scalar product between q-eigenstates defined at infinitesimally close times is then

$$\langle q';\tau + d\tau|q;\tau\rangle = \langle q'|\exp(-iH d\tau)|q\rangle .$$

Now insert Eq. (9.5.42) to the left of the operator $\exp(-iH d\tau)$. It is convenient here to define the Hamiltonian operator $H(P,Q)$ with all Ps to the left of all Qs, so that (for $d\tau$ infinitesimal)

$$\langle p|\exp\Big(-iH(P,Q)d\tau\Big)|q\rangle = \langle p|q\rangle \exp\Big(-iH(p,q)d\tau\Big) .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

(We could move the c-number $H(p,q)$ to either side of the matrix element without any sign changes because each term in the Hamiltonian is assumed to contain an even number of fermionic operators.) This gives

$$\begin{aligned}\langle q';\tau+d\tau|q;\tau\rangle &= \int\langle q'|p\rangle\left(\tilde{\prod_a}\,dp_a\right)\langle p|\exp(-iHd\tau)|q\rangle \\ &= \int\langle q'|p\rangle\left(\tilde{\prod_a}\,dp_a\right)\langle p|q\rangle\exp\Big(-iH(p,q)d\tau\Big)\,.\end{aligned}$$

Using Eqs. (9.5.26) and (9.5.27), and noting that the products p_aq_a and $p_aq'_a$ commute with all anticommuting c-numbers, we find

$$\langle q';\tau+d\tau|q;\tau\rangle = \int\left(\tilde{\prod_a}\,i\,dp_a\right)\exp\left[i\sum_a{}'\,p_a(q'_a-q_a)-iH(p,q)d\tau\right]\,. \tag{9.5.47}$$

The rest of the derivation follows the same lines as in Section 9.1. To calculate the matrix element $\langle q';t'|O_A(P(t_A),Q(t_A))\,O_B(P(t_B),Q(t_B))\cdots|q;t\rangle$ of a product of operators (with $t'>t_A>t_B>\cdots>t$), divide the time-interval from t to t' into a large number of very close time steps; at each time step insert the completeness relation (9.5.41); use Eq. (9.5.47) to evaluate the resulting matrix elements (with $\mathcal{O}_A,\mathcal{O}_B$, etc. inserted where appropriate); move all differentials to the left (this introduces no sign changes, because at each step we have an equal number of dps and dqs); and then introduce functions $q_a(t)$ and $p_a(t)$ that interpolate between the values of q_a and p_a at each step. We then find

$$\begin{aligned}&\langle q';t'|T\left\{\mathcal{O}_A\Big(P(t_A),Q(t_A)\Big),\ \mathcal{O}_B\Big(P(t_B),\ Q(t_B)\Big),\cdots\right\}|q;t\rangle \\ &\quad=(-i)^N\chi_N\int_{q_a(t)=q_a,q_a(t')=q'_a}\left(\tilde{\prod_{a\tau}}\,dq_a(\tau)dp_a(\tau)\right) \\ &\quad\times\ \mathcal{O}_A\Big(p(t_A),q(t_A)\Big)\,\mathcal{O}_B\Big(p(t_B),q(t_B)\Big)\cdots \\ &\quad\times\ \exp\left[i\int_t^{t'}d\tau\left\{\sum_a p_a(\tau)\dot{q}_a(\tau)-H\Big(p(\tau),q(\tau)\Big)\right\}\right]\,.\end{aligned} \tag{9.5.48}$$

The symbol T here denotes the ordinary product if the times are in the order originally assumed, $t_A>t_B>\cdots$. However, the right-hand side is totally symmetric in the $\mathcal{O}_A,\mathcal{O}_B,\cdots$ (except for minus signs where anticommuting c-numbers are interchanged) so this formula holds for general times (between t and t'), provided T is interpreted as the time-ordered product, with an overall minus sign if time-ordering the operators involves an odd number of permutations of fermionic operators.

Up to this point we have kept track of the overall phase factor $(-i)^N\chi_N$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

But in fact these phases contribute only to the vacuum–vacuum transition amplitude, and hence will not be of importance to us.

The transition to quantum field theory follows along the same lines as described for bosonic fields in Section 9.2. The vacuum expectation value of a time-ordered product of operators is given by a formula just like Eq. (9.2.17):

$$\begin{aligned}&\left\langle \text{VAC, out}|T\left\{\mathcal{O}_A\left[P(t_A),Q(t_A)\right],\ \mathcal{O}_B\left[P(t_B),Q(t_B)\right],\cdots\right\}|\text{VAC, in}\right\rangle\\ &\propto \int\left[\prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau)\right]\left[\prod_{\tau,\mathbf{x},m} dp_m(\mathbf{x},\tau)\right]\mathcal{O}_A\left[p(t_A),q(t_A)\right]\\ &\times \mathcal{O}_B\left[p(t_B),q(t_B)\right]\cdots\exp\left[i\int_{-\infty}^{\infty} d\tau\left\{\int d^3x\ \sum_m p_m(\mathbf{x},\tau)\,\dot{q}_m(\mathbf{x},\tau)\right.\right.\\ &\qquad\left.\left.-H\left[q(\tau),p(\tau)\right]+\ i\epsilon\text{ terms}\right\}\right]\end{aligned}\tag{9.5.49}$$

where the proportionality constant is the same for all operators $\mathcal{O}_A$, $\mathcal{O}_B$, etc., and the '$i\epsilon$ terms' again arise from the wave function of the vacuum. As before, we have replaced each discrete index like a with a space position $\mathbf{x}$ and a field index m. We are also dropping the tilde on the product of differentials, since it only affects the constant phase in the path integral.

A major difference between the fermionic and bosonic cases is that here we will not want to integrate out the ps before the qs. Indeed, in the standard model of electroweak interactions (and in other theories, such as the older Fermi theory of beta decay) the canonical conjugates p_m are auxiliary fields unrelated to the $\dot{q}_m$, and the Lagrangian is linear in the $\dot{q}_m$, so that the quantity $\int d^3x\ \sum_m p_m\dot{q}_m - H$ in Eq. (9.5.49) as it stands *is* the Lagrangian L. Each term in the Hamiltonian for a fermionic field that carries a non-vanishing quantum number (like the electron field in quantum electrodynamics) generally contains an equal number of ps (proportional to $q^\dagger$) and qs. In particular, the free-particle term H_0 in the Hamiltonian is bilinear in p and q, so that

$$\begin{aligned}&\int_{-\infty}^{\infty} d\tau\left\{\int d^3x\ \sum_m p_m(\mathbf{x},\tau)\dot{q}_m(\mathbf{x},\tau) - H_0\left[q(\tau),p(\tau)\right]+\ i\epsilon\text{ terms}\right\}\\ &=-\sum_{mn}\int d^4x\,d^4y\ \mathscr{D}_{mx,ny}\,p_m(x)\,q_n(y)\end{aligned}\tag{9.5.50}$$

with $\mathscr{D}$ some numerical 'matrix'. The interaction Hamiltonian $V \equiv H - H_0$ is a sum of products of equal numbers of fermionic qs and ps (with coefficients that may depend on bosonic fields) so when we expand Eq. (9.5.49) in powers of the V we encounter a sum of fermionic integrals

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

of the form

$$\mathscr{I}_{n_1 m_1 n_2 m_2 \cdots n_N m_N}(x_1, y_1, x_2, y_2 \cdots, x_N, y_N) \equiv \int \Big[\prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau) \Big]$$
$$\times \Big[\prod_{\tau,\mathbf{x},m} dp_m(\mathbf{x},\tau) \Big] q_{m_1}(x_1)\, p_{n_1}(y_1)\, q_{m_2}(x_2)\, p_{n_2}(y_2) \cdots q_{m_N}(x_N)\, p_{n_N}(y_N)$$
$$\times \exp\Big(-i \sum_{mn} \int d^4x\, d^4y\, \mathscr{D}_{mx,ny}\, p_m(x)\, q_n(y) \Big) , \tag{9.5.51}$$

one such term for each possible set of vertices in the Feynman diagram, with coefficients contributed by each vertex given by i times the coefficient of the product of fields in the corresponding term in the interaction.

To calculate this sort of integral, first consider a generating function for all these integrals:

$$\mathscr{I}(f,g) \equiv \int \Big[\prod_{\mathbf{x},\tau,m} dq_m(\mathbf{x},\tau)\, dp_m(\mathbf{x},\tau) \Big]$$
$$\times \exp\Big(-i \sum_{mn} \int d^4x\, d^4y\, \mathscr{D}_{mx,ny}\, p_m(x)\, q_n(y)$$
$$-i \sum_m \int d^4x\, p_m(x)\, f_m(x) - i \sum_n \int d^4y\, g_n(y)\, q_n(y) \Big) , \tag{9.5.52}$$

where $f_m(x)$ and $g_n(y)$ are arbitrary anticommuting c-number functions. We shift to new variables of integration

$$p'_m(x) = p_m(x) + \sum_n \int d^4y\, g_n(y)\, (\mathscr{D}^{-1})_{ny,mx} ,$$

$$q'_n(y) = q_n(y) + \sum_m \int d^4x\, (\mathscr{D}^{-1})_{ny,mx}\, f_m(x) .$$

Using the translation invariance condition (9.5.36), we then find

$$\mathscr{I}(f,g) = \exp\Big(i \sum_{mn} \int d^4x\, d^4y\, (\mathscr{D}^{-1})_{ny,mx}\, g_n(y)\, f_m(x) \Big)$$
$$\times \int \Big[\prod_{\mathbf{x},\tau,m} dq'_m(\mathbf{x},\tau)\, dp'_m(\mathbf{x},\tau) \Big]$$
$$\times \exp\Big(-i \sum_{mn} \int d^4x\, d^4y\, \mathscr{D}_{mx,ny}\, p'_m(x)\, q'_n(y) \Big) . \tag{9.5.53}$$

The integral is a constant (i.e., independent of the functions f and g) which can be shown using Eq. (9.5.38) to be proportional to Det $\mathscr{D}$. Of more importance to us is the first factor. Expanding this factor in powers

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

of gf and comparing with the direct expansion of Eq. (9.5.52), we see that

$$\mathscr{I}_{n_1 m_1 n_2 m_2 \cdots n_N m_N}(x_1, y_1, x_2, y_2 \cdots, x_N, y_N) \propto \sum_{\text{pairings}} \delta_{\text{pairing}} \prod_{\text{pairs}} \left(-i\mathscr{D}^{-1} \right)_{\text{paired } mx,ny} \tag{9.5.54}$$

with a proportionality constant that is independent of the x, y, m, or n, and also independent of the number of these variables. The sum is over all different ways of pairing ps with qs, not counting as different pairings that only differ in the order of the pairs. In other words, we sum over the $N!$ permutations either of the ps or the qs. The sign factor δ_{pairing} is $+1$ if this permutation is even; -1 if it is odd.

This sign factor and sum over pairings are just the same as we encountered in our earlier derivation of the Feynman rules, with the sum over pairings corresponding to the sum over ways of connecting the lines associated with vertices in the Feynman diagrams, and the factors $(\mathscr{D}^{-1})_{mx,ny}$ playing the role of the propagator for the pairing of $q_m(x)$ with $p_n(y)$. In the Dirac formalism for spin $\frac{1}{2}$, the free-particle action is

$$\int_{-\infty}^{\infty} d\tau \left\{ \int d^3x \sum_m p_m(\mathbf{x}, \tau) \dot{q}_m(\mathbf{x}, \tau) - H_0\left[q(\tau), p(\tau)\right] \right\} = -\int d^4x \, \bar{\psi}(x) \left[\gamma^\mu \partial_\mu + m\right] \psi(x) , \tag{9.5.55}$$

where in the usual notation the canonical variables here are

$$q_m(x) = \psi_m(x) , \qquad p_m(x) = -[\bar{\psi}(x)\gamma^0]_m = i\psi_m^\dagger(x) \tag{9.5.56}$$

with m a four-valued Dirac index. Comparing this with Eq. (9.5.50) , we find here

$$\mathscr{D}_{mx,ny} = \left[\gamma^0 \left(\gamma^\mu \frac{\partial}{\partial x^\mu} + m - i\epsilon\right)\right]_{mn} \delta^4(x-y) = \int \frac{d^4k}{(2\pi)^4} \left(\gamma^0 [i\gamma^\mu k_\mu + m - i\epsilon]\right)_{mn} e^{ik\cdot(x-y)} . \tag{9.5.57}$$

(Though we shall not work it out in detail, the $i\epsilon$ term here arises in much the same way as for the scalar field in Section 9.2.) The propagator is then

$$(\mathscr{D}^{-1})_{mx,ny} = \int \frac{d^4k}{(2\pi)^4} \left([i\gamma^\mu k_\mu + m - i\epsilon]^{-1}[-\gamma^0]\right)_{mn} e^{ik\cdot(x-y)} , \tag{9.5.58}$$

just as we found in the operator formalism. The extra factor $-\gamma^0$ arises because this propagator is the vacuum expectation value of $T\{\psi_m(x), -[\bar{\psi}(y)\gamma^0]_n\}$, not $T\{\psi_m(x), \bar{\psi}_n(y)\}$.

As one example of a problem that is easier to solve by path-integral than by operator methods, let us calculate the field dependence of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

vacuum→vacuum amplitude for a Dirac field that interacts only with an external field. Take the Lagrangian as

$$\mathscr{L} = -\bar{\psi}[\gamma^\mu \partial_\mu + m + \Gamma]\psi \,, \tag{9.5.59}$$

where $\Gamma(x)$ is an x-dependent matrix representing the interaction of the fermion with the external field. According to Eq. (9.5.49), the vacuum persistence amplitude in the presence of this external field is

$$\langle \text{VAC, out}|\text{VAC, in}\rangle_\Gamma \propto \int \Big[\prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau)\Big]\Big[\prod_{\tau,\mathbf{x},m} dp_m(\mathbf{x},\tau)\Big]$$
$$\times \exp\left\{-i\int d^4x\, p^T \gamma^0[\gamma^\mu\partial_\mu + m + \Gamma - i\epsilon]\, q\right\} \tag{9.5.60}$$

with a proportionality constant that is independent of $\Gamma(x)$. We write this as

$$\langle \text{VAC, out}|\text{VAC, in}\rangle_\Gamma \propto \int \Big[\prod_{\tau,\mathbf{x},m} dq_m(\mathbf{x},\tau)\Big]\Big[\prod_{\tau,\mathbf{x},m} dp_m(\mathbf{x},\tau)\Big]$$
$$\times \exp\left\{-i\sum_{mn}\int d^4x\, d^4y\; p_m(x)\, q_n(y)\, \mathscr{K}[\Gamma]_{mx,ny}\right\}, \tag{9.5.61}$$

where

$$\mathscr{K}[\Gamma]_{mx,ny} = \left(\gamma^0\left[\gamma^\mu\frac{\partial}{\partial x^\mu} + m + \Gamma(x) - i\epsilon\right]\right)_{mn}\delta^4(x-y)\,. \tag{9.5.62}$$

To evaluate this, we change the variables of integration $q_n(x)$ to

$$q'_m(x) \equiv \sum_n \int d^4y\; \mathscr{K}[\Gamma]_{mx,ny}\, q_n(y)\,. \tag{9.5.63}$$

The remaining integral is now Γ-independent, so the whole dependence of the vacuum persistence amplitude is contained in the determinant arising according to Eq. (9.5.38) from the change of variables:

$$\langle \text{VAC, out}|\text{VAC, in}\rangle_\Gamma \propto \text{Det}\,\mathscr{K}[\Gamma]\,. \tag{9.5.64}$$

To recover the results of perturbation theory, let us write

$$\mathscr{K}[\Gamma] \equiv \mathscr{D} + \mathscr{G}[\Gamma]\,, \tag{9.5.65}$$

$$\mathscr{G}[\Gamma]_{mx,ny} = \left(\gamma^0\,\Gamma(x)\right)_{mn}\delta^4(x-y)\,, \tag{9.5.66}$$

and expand in powers of $\mathscr{G}[\Gamma]$. Eq. (9.5.64) gives then

$$\langle \text{VAC, out}|\text{VAC, in}\rangle_\Gamma \propto \text{Det}\left(\mathscr{D}[1 + \mathscr{D}^{-1}\mathscr{G}[\Gamma]]\right)$$
$$= [\text{Det}\,\mathscr{D}]\;\exp\left(\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\text{Tr}\,(\mathscr{D}^{-1}\mathscr{G}[\Gamma])^n\right)\,. \tag{9.5.67}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

This is just what we should expect from the Feynman rules: the contributions from internal lines and vertices in this theory are $-i\mathscr{D}^{-1}$ and $-i\mathscr{G}[\Gamma]$; the trace of the product of n factors of $-\mathscr{D}^{-1}\mathscr{G}[\Gamma]$ thus corresponds to a loop with n vertices connected by n internal lines; $1/n$ is the usual combinatoric factor associated with such loops (see Section 6.1); the sign factor is $(-1)^{n+1}$ rather than $(-1)^n$ because an extra minus sign is associated with fermion loops; and the sum over n appears as the argument of an exponential because the vacuum persistence amplitude receives contributions from graphs with any number of disconnected loops. The Γ-independent factor Det $\mathscr{D}$ is less easy to derive from the Feynman rules; it represents the contribution of any number of fermion loops that carry no vertices.

More to the point, a formula like Eq. (9.5.64) allows us to derive nonperturbative results by using topological theorems to derive information about the eigenvalues of kernels like $\mathscr{K}[\Gamma]$. This will be pursued further in Volume II.

9.6 Path-Integral Formulation of Quantum Electrodynamics

The path-integral approach to quantum field theory really comes into its own when applied to gauge theories of massless spin one particles, such as quantum electrodynamics. The derivation of the Feynman rules for quantum electrodynamics in the previous chapter involved a fair amount of hand-waving, in arguing that the terms in the photon propagator $\Delta^{\mu\nu}(q)$ proportional to q^μ or q^ν could be dropped, and that the purely time-like terms would just cancel the Coulomb term in the Hamiltonian, so that the effective photon propagator could be taken as $\eta^{\mu\nu}/q^2$. To give a real justification of this result using the methods of Chapter 8 would involve us in a complicated analysis of Feynman diagrams. But as we shall now see, the path-integral approach yields the desired form of the photon propagator, without ever having to think about the details of Feynman diagrams.

In Chapter 8 we found that in Coulomb gauge, the Hamiltonian for the interaction of photons with charged particles takes the form

$$H[\mathbf{A},\mathbf{\Pi}_\perp,\cdots] = H_{\rm M} + \int d^3x \left[\tfrac{1}{2}\mathbf{\Pi}_\perp{}^2 + \tfrac{1}{2}(\nabla\times\mathbf{A})^2 - \mathbf{A}\cdot\mathbf{J}\right] + V_{\rm Coul}\,. \tag{9.6.1}$$

Here $\mathbf{A}$ is the vector potential, subject to the Coulomb gauge condition

$$\nabla\cdot\mathbf{A} = 0\,, \tag{9.6.2}$$

while $\mathbf{\Pi}_\perp$ is the solenoidal part of its canonical conjugate, satisfying the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

same constraint

$$\nabla \cdot \mathbf{\Pi}_\perp = 0 \,. \tag{9.6.3}$$

Also, $H_{\rm M}$ is the matter Hamiltonian and $V_{\rm Coul}$ is the Coulomb energy

$$V_{\rm Coul}(t) = \tfrac{1}{2} \int d^3x\, d^3y\, J^0(\mathbf{x},t) J^0(\mathbf{y},t) \Big/ 4\pi |\mathbf{x}-\mathbf{y}| \,. \tag{9.6.4}$$

Just as for any other Hamiltonian system, we can calculate vacuum expectation values of time-ordered products as path integrals*

$$\begin{aligned} \langle T\{\mathcal{O}_A \mathcal{O}_B \cdots\}\rangle_{\rm VAC} = \int & \left[\prod_{x,i} da_i(x) \prod_{x,i} d\pi_i(x) \prod_{x,\ell} d\psi_\ell(x)\right] \mathcal{O}_A \mathcal{O}_B \cdots \\ \times \exp & \left\{ i \int d^4x \left[\boldsymbol{\pi}\cdot\dot{\mathbf{a}} - \tfrac{1}{2}\boldsymbol{\pi}^2 - \tfrac{1}{2}(\nabla\times\mathbf{a})^2 + \mathbf{a}\cdot\mathbf{J} + \mathscr{L}_{\rm M}\right] - i\int dt\, V_{\rm Coul} \right\} \\ \times & \left[\prod_x \delta\Big(\nabla\cdot\mathbf{a}(x)\Big)\right] \left[\prod_x \delta\Big(\nabla\cdot\boldsymbol{\pi}(x)\Big)\right] , \end{aligned} \tag{9.6.5}$$

where $\psi_\ell(x)$ are generic matter fields. In writing Eq. (9.6.5) in terms of a matter Lagrangian density, we are assuming that $H_{\rm M}$ is local and either linear in the matter πs (as in spinor electrodynamics) or quadratic with field-independent coefficients (as in scalar electrodynamics). We have inserted delta functions** in Eq. (9.6.5) to enforce the constraints (9.6.2) and (9.6.3).

The argument of the exponential in Eq. (9.6.5) is evidently quadratic in the independent components of $\boldsymbol{\pi}$ (say, π_1 and π_2), with field-independent coefficients in the term of second order in $\boldsymbol{\pi}$. Thus, according to Eq. (9.A.9), the integral over $\boldsymbol{\pi}$ can be done (up to a constant factor) by setting $\boldsymbol{\pi}$ equal to the stationary point of the argument of the exponential,

* Note that $\pi(x)$ is the interpolating c-number field for the quantum operator $\mathbf{\Pi}_\perp$, whose commutation relations with each other and with $\mathbf{A}$ are the same as those of $\mathbf{\Pi}$, but which unlike $\mathbf{\Pi}$ commutes with all canonical matter variables.

** This is not strictly accurate. If we take the canonical variables to be, say, a_1, a_2 and π_1, π_2, with a_3 and π_3 regarded as functionals of these variables given by Eqs. (9.6.2) and (9.6.3), then we should insert the delta functions

$$\prod_x \delta\left(a_3(x) + \partial_3^{-1}\Big(\partial_1 a_1(x) + \partial_2 a_2(x)\Big)\right) \delta\left(\pi_3(x) + \partial_3^{-1}\Big(\partial_1 \pi_1(x) + \partial_2 \pi_2(x)\Big)\right) .$$

However, this differs from the product of delta functions in Eq. (9.6.5) only by a factor $\mathrm{Det}\,\partial_3^2$, which although infinite is field-independent and hence cancels in ratios like (9.4.1).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

$\boldsymbol{\pi} = \dot{\mathbf{a}}$:

$$\langle T\{\mathcal{O}_A\mathcal{O}_B\cdots\}\rangle_{\text{VAC}} = \int \left[\prod_{x,i} da_i(x) \prod_{x,\ell} d\psi_\ell(x)\right]$$
$$\times\, \mathcal{O}_A\mathcal{O}_B\cdots \exp\left\{i\int d^4x \left[\frac{1}{2}\dot{\mathbf{a}}^2 - \frac{1}{2}(\nabla\times\mathbf{a})^2 + \mathbf{a}\cdot\mathbf{j} + \mathscr{L}_{\text{M}}\right]\right.$$
$$\left. -i\int dt\, V_{\text{Coul}} + i\epsilon \text{ terms}\right\} \left[\prod_x \delta\big(\nabla\cdot\mathbf{a}(x)\big)\right] . \qquad (9.6.6)$$

To bring out the essential covariance of this result, we use a trick. Introduce a new variable of integration $a^0(x)$, and replace the Coulomb term $-\int dt\, V_{Coul}$ in the action with

$$\int d^4x \left[-a^0(x)j^0(x) + \tfrac{1}{2}\big(\nabla a^0(x)\big)^2\right] . \qquad (9.6.7)$$

Since (9.6.7) is quadratic in a^0, the integral over a^0 can be done (up to a constant factor) by setting $a^0(x)$ equal to the stationary point of (9.6.7), i.e., to the solution of

$$-j^0(x) - \nabla^2 a^0(x) = 0$$

or in other words, to

$$a^0(\mathbf{x},t) = \int d^3y\, \frac{j^0(\mathbf{y},t)}{4\pi|\mathbf{x}-\mathbf{y}|} . \qquad (9.6.8)$$

Using this in Eq. (9.6.7) just gives the Coulomb action $-\int dt\, V_{\text{Coul}}$. Hence we can rewrite the argument of the exponential in Eq. (9.6.6) as

$$\tfrac{1}{2}\dot{\mathbf{a}}^2 - \tfrac{1}{2}(\nabla\times\mathbf{a})^2 + \mathbf{a}\cdot\mathbf{j} + \mathscr{L}_{\text{M}} - a^0 j^0 + \tfrac{1}{2}(\nabla a^0)^2$$
$$= -\tfrac{1}{4} f_{\mu\nu}f^{\mu\nu} + a_\mu j^\mu + \mathscr{L}_{\text{M}} + \text{total derivatives}$$

with $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$, and integrate over a^0 as well as over $\mathbf{a}$ and matter fields. That is, the path integral (9.6.6) is now

$$\langle T\{\mathcal{O}_A\mathcal{O}_B\cdots\}\rangle_{\text{VAC}} \propto \int \left[\prod_{x,\mu} da_\mu(x)\right] \left[\prod_{x,\ell} d\psi_\ell(x)\right]$$
$$\times\, \mathcal{O}_A\mathcal{O}_B\cdots \exp\big(iI[a,\psi]\big) \prod_x \delta\big(\nabla\cdot\mathbf{a}(x)\big) , \qquad (9.6.9)$$

where I is the original action

$$I[a,\psi] = \int d^4x \left[-\tfrac{1}{4} f_{\mu\nu}f^{\mu\nu} + a_\mu j^\mu + \mathscr{L}_{\text{M}}\right] + i\epsilon \text{ terms} . \qquad (9.6.10)$$

Now everything is manifestly Lorentz- and gauge-invariant, except for the final product of delta functions which enforce the Coulomb gauge

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

condition.** To make further progress, we shall use a simple version of a trick[4,5] that in Volume II will be used to treat the more difficult case of non-Abelian gauge theories. For simplicity, we shall deal here with the case where the operators $\mathcal{O}_A[A,\Psi]$, $\mathcal{O}_B[A,\Psi],\cdots$ as well as the action $I[a,\psi]$ and measure $[\prod da][\prod d\psi]$ are gauge-invariant.

First, replace the field variables of integration $a_\mu(x)$ and $\psi(x)$ everywhere in Eq. (9.6.9) with the new variables

$$a_{\mu\Lambda}(x) \equiv a_\mu(x) + \partial_\mu \Lambda(x)\,, \tag{9.6.11}$$

$$\psi_{\ell\Lambda}(x) \equiv \exp\left(i q_\ell \Lambda(x)\right)\psi_\ell(x) \tag{9.6.12}$$

with arbitrary finite $\Lambda(x)$. This step is a mathematical triviality, like changing an integral $\int_{-\infty}^{\infty} f(x)dx$ to read $\int_{-\infty}^{\infty} f(y)dy$, and does not require use of the postulated gauge invariance of the theory. Next, use gauge invariance to replace $a_{\mu\Lambda}(x)$ and $\psi_{\ell\Lambda}(x)$ in the action, measure, and $\mathcal{O}$-functions with the original fields $a_\mu(x)$ and $\psi_\ell(x)$, respectively. Eq. (9.6.9) then becomes

$$\begin{aligned}\left\langle T\left\{\mathcal{O}_A[A,\Psi],\,\mathcal{O}_B[A,\Psi],\cdots\right\}\right\rangle_{\text{VAC}} \\ \propto \int \left[\prod_{x,\mu} da_\mu(x)\right]\left[\prod_{x,\ell} d\psi_\ell(x)\right]\,\mathcal{O}_A[a,\psi]\mathcal{O}_B[a,\psi]\cdots \\ \times \exp\left(iI[a,\psi]\right)\prod_x \delta\left(\nabla\cdot\mathbf{a}(x) + \nabla^2\Lambda(x)\right)\,.\end{aligned} \tag{9.6.13}$$

Now, the function $\Lambda(x)$ was chosen at random, so despite appearances the right-hand side of Eq. (9.6.13) cannot depend on this function. We shall exploit this fact to put the path integral in a much more convenient form. Multiply Eq. (9.6.13) by the functional

$$B[\Lambda,a] = \exp\left(-\tfrac{1}{2}i\alpha\int d^4x\left(\partial_0 a^0 - \nabla^2\Lambda\right)^2\right) \tag{9.6.14}$$

(where α is an arbitrary constant), and integrate over $\Lambda(x)$. By shifting the integration variable $\Lambda(x)$, and noting the actual Λ-independence of (9.6.13), we see that the effect is simply to multiply Eq. (9.6.13) with the field-independent constant

$$\int\left[\prod_x d\Lambda(x)\right]\,\exp\left(-\tfrac{1}{2}i\alpha\int d^4x\left(\nabla^2\Lambda\right)^2\right)\,. \tag{9.6.15}$$

** Note that now $a^0(x)$ is *not* equal to the value (9.6.8), but is an independent variable of integration. We will not integrate over $a^0(x)$ first, which would lead back to Eq. (9.6.6), but instead will treat it in tandem with $\mathbf{a}(x)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

This factor cancels out in the connected part of the vacuum expectation value, and thus has no physical effect. But (9.6.13) is only Λ-independent *after* we integrate over $a^\mu(x)$ and $\psi(x)$. We can just as well integrate over $\Lambda(x)$ *before* we integrate over $a^\mu(x)$ and $\psi(x)$, in which case the factor $\prod_x \delta\left(\nabla \cdot \mathbf{a}(x) + \nabla^2\Lambda\right)$ in Eq. (9.6.13) is replaced with

$$\int \left[\prod_x d\Lambda(x)\right] \exp\left(-\tfrac{1}{2}i\alpha \int d^4x \left(\partial_0 a^0 - \nabla^2\Lambda\right)^2\right) \prod_x \delta\left(\nabla \cdot \mathbf{a}(x) + \nabla^2\Lambda\right)$$
$$\propto \exp\left(-\tfrac{1}{2}i\alpha \int d^4x \left(\partial_\mu a^\mu\right)^2\right), \tag{9.6.16}$$

where "$\propto$" again means proportional with a field-independent factor. Dropping constant factors, Eq. (9.6.9) now becomes

$$\langle T\{\mathcal{O}_A\mathcal{O}_B\cdots\}\rangle_{\text{VAC}} \propto \int \left[\prod_{x,\mu} da_\mu(x)\right] \left[\prod_{x,\ell} d\psi_\ell(x)\right] \mathcal{O}_A\mathcal{O}_B\cdots \exp\left(iI_{\text{eff}}[a,\psi]\right), \tag{9.6.17}$$

where

$$I_{\text{eff}}[a,\psi] = I[a,\psi] - \tfrac{1}{2}\alpha \int (\partial_\mu a^\mu)^2 d^4x\,. \tag{9.6.18}$$

This is now manifestly Lorentz-invariant.

We consider the new term in (9.6.18) as a contribution to the unperturbed part of the action, whose photonic part now reads

$$I_0[a] = \int d^4x \left[-\tfrac{1}{4}(\partial_\mu a_\nu - \partial_\nu a_\mu)(\partial^\mu a^\nu - \partial^\nu a^\mu) - \tfrac{1}{2}\alpha(\partial_\mu a^\mu)^2 + i\epsilon \text{ terms}\right]$$
$$= -\tfrac{1}{2}\int d^4x\, d^4y\, a^\mu(x)\, a^\nu(y)\, D_{\mu x,\nu y}, \tag{9.6.19}$$

where

$$\mathcal{D}_{\mu x,\nu y} = \left[\eta_{\mu\nu}\frac{\partial^2}{\partial x^\rho \partial y_\rho} - (1-\alpha)\frac{\partial^2}{\partial x^\mu \partial y^\nu}\right] \delta^4(x-y) + i\epsilon \text{ terms}$$
$$= (2\pi)^{-4}\int d^4q \left[\eta_{\mu\nu}\, q^2 - (1-\alpha)q_\mu q_\nu - i\epsilon\eta_{\mu\nu}\right] e^{iq\cdot(x-y)}\,. \tag{9.6.20}$$

The photon propagator is then found immediately by inverting the 4×4 matrix in the integrand of Eq. (9.6.20)

$$\Delta_{\mu x,\nu y} = (2\pi)^{-4}\int d^4q \left[\frac{\eta^{\mu\nu}}{q^2 - i\epsilon} + \frac{(1-\alpha)}{\alpha}\frac{q^\mu q^\nu}{(q^2 - i\epsilon)^2}\right] e^{iq\cdot(x-y)}\,. \tag{9.6.21}$$

We are free to choose α as seems most convenient. Two common choices are $\alpha = 1$, which yields the propagator in *Feynman gauge*:

$$\Delta^{\text{Feynman}}_{\mu x,\nu y} = (2\pi)^{-4}\int d^4q \left[\frac{\eta^{\mu\nu}}{q^2 - i\epsilon}\right] e^{iq\cdot(x-y)} \tag{9.6.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

or $\alpha = \infty$, in which case the factor (9.6.14) acts as a delta function, and we obtain the propagator in *Landau gauge* (often also called *Lorentz gauge*):

$$\Delta^{\text{Landau}}_{\mu x,\nu y} = (2\pi)^{-4} \int d^4q \left[\frac{\eta^{\mu\nu}}{q^2 - i\epsilon} - \frac{q^\mu q^\nu}{(q^2 - i\epsilon)^2} \right] e^{iq\cdot(x-y)} . \tag{9.6.23}$$

Practical calculations are made far more convenient by working with such manifestly Lorentz-invariant interactions and propagators.

9.7 Varieties of Statistics*

We can now take up a question raised in Chapter 4: what are the possibilities for the change of state-vectors when we interchange identical particles?

For this purpose, we will consider the preparation of the initial or final states in a scattering process. Suppose that a set of indistinguishable particles in either of these states is brought to a particular configuration with momenta $\mathbf{p}_1$, $\mathbf{p}_2$, etc. from a standard configuration with momenta $\mathbf{k}_1$, $\mathbf{k}_2$, etc., by some sort of slowly varying external fields, keeping the particles far enough apart in the process to justify the use of non-relativistic quantum mechanics. (Spin indices are not shown explicitly here; they should be understood to accompany momentum labels.) To calculate the amplitude for this process we can use the path-integral method,** taking the qs and ps of Section 9.1 as particle positions and momenta, rather than fields and their canonical conjugates. These always satisfy canonical commutation rather than anticommutation relations, whether or not the particles are bosons or fermions or something else, so at this point we are not committing ourselves to any particular statistics. The path-integral formula (9.1.34) gives an amplitude $\langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\text{D}}$ as an integral over paths in which one particle is brought continuously from momentum $\mathbf{k}_1$ to momentum $\mathbf{p}_1$, another identical particle is brought continuously from momentum $\mathbf{k}_2$ to momentum $\mathbf{p}_2$, and so on. The subscript 'D' indicates that this is the amplitude we would calculate for distinguishable particles. In particular, this amplitude is symmetric under permutations of the $\mathbf{p}$s *and* simultaneous permutations of the $\mathbf{k}$s, but has no particular symmetry under separate permutations of the $\mathbf{p}$s *or* $\mathbf{k}$s. But

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** Here I am following the discussion of Laidlaw and C. DeWitt,[11] except that they apply the path-integral method to the whole scattering process, rather than just the preparation of initial or final states. In a relativistic theory the possibility of particle creation and annihilation makes it necessary to apply the path-integral method to fields rather than particle orbits. For us, this is not a problem, because we limit such calculations to sufficiently early or late times, when the particles participating in a scattering process are all far apart.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

if the particles are really indistinguishable, then there are other paths that are topologically distinct, but that yield the same final configuration. For *space* dimensionality $d \geq 3$, the only such paths† are those that take $\mathbf{k}_1, \mathbf{k}_2, \cdots$ into some non-trivial permutation $\mathscr{P}$ of $\mathbf{p}_1, \mathbf{p}_2, \cdots$. Hence the true amplitude should be written

$$\langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle = \sum_{\mathscr{P}} C_{\mathscr{P}} \langle \mathbf{p}_{\mathscr{P}1}, \mathbf{p}_{\mathscr{P}2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} , \tag{9.7.1}$$

the sum running over all $N!$ permutations of the N indistinguishable particles in the state, and $C_{\mathscr{P}}$ a set of complex constants. These amplitudes must satisfy a composition rule appropriate for indistinguishable particles:

$$\langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle = \frac{1}{N!} \int d^3q_1 \, d^3q_2 \cdots \langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{q}_1, \mathbf{q}_2, \cdots \rangle \times \langle \mathbf{q}_1, \mathbf{q}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle . \tag{9.7.2}$$

Using Eq. (9.7.1), this is the requirement

$$\sum_{\mathscr{P}} C_{\mathscr{P}} \langle \mathbf{p}_{\mathscr{P}1}, \mathbf{p}_{\mathscr{P}2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} = \frac{1}{N!} \sum_{\mathscr{P}', \mathscr{P}''} C_{\mathscr{P}'} C_{\mathscr{P}''} \int d^3q_1 \, d^3q_2 \cdots \times \langle \mathbf{p}_{\mathscr{P}'1}, \mathbf{p}_{\mathscr{P}'2}, \cdots | \mathbf{q}_1, \mathbf{q}_2, \cdots \rangle_{\mathrm{D}} \langle \mathbf{q}_{\mathscr{P}''1}, \mathbf{q}_{\mathscr{P}''2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} .$$

Applying a permutation $\mathscr{P}''$ to both the initial and final states in the first amplitude on the right, this is

$$\sum_{\mathscr{P}} C_{\mathscr{P}} \langle \mathbf{p}_{\mathscr{P}1}, \mathbf{p}_{\mathscr{P}2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} = \frac{1}{N!} \sum_{\mathscr{P}', \mathscr{P}''} C_{\mathscr{P}'} C_{\mathscr{P}''} \int d^3q_1 \, d^3q_2 \cdots \times \langle \mathbf{p}_{\mathscr{P}''\mathscr{P}'1}, \mathbf{p}_{\mathscr{P}''\mathscr{P}'2}, \cdots | \mathbf{q}_{\mathscr{P}''1}, \mathbf{q}_{\mathscr{P}''2}, \cdots \rangle_{\mathrm{D}} \langle \mathbf{q}_{\mathscr{P}''1}, \mathbf{q}_{\mathscr{P}''2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} .$$

But the amplitudes $\langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}}$ satisfy the composition rule for *distinguishable* particles

$$\langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} = \int d^3q_1 \, d^3q_2 \cdots \langle \mathbf{p}_1, \mathbf{p}_2, \cdots | \mathbf{q}_1, \mathbf{q}_2, \cdots \rangle_{\mathrm{D}} \times \langle \mathbf{q}_1, \mathbf{q}_2, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} , \tag{9.7.3}$$

so the composition rule for the physical amplitudes may be written

$$\sum_{\mathscr{P}} C_{\mathscr{P}} \langle \mathbf{p}_{\mathscr{P}1}, \mathbf{p}_{\mathscr{P}2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} = \frac{1}{N!} \sum_{\mathscr{P}', \mathscr{P}''} C_{\mathscr{P}'} C_{\mathscr{P}''} \times \langle \mathbf{p}_{\mathscr{P}''\mathscr{P}'1}, \mathbf{p}_{\mathscr{P}''\mathscr{P}'2}, \cdots | \mathbf{k}_1, \mathbf{k}_2, \cdots \rangle_{\mathrm{D}} ,$$

† This is expressed formally in the statement that the first homotopy group of configuration space in $d \geq 3$ is the permutation group.[12] By 'configuration space' for N indistinguishable particles is meant the space of N d-vectors, excluding d-vectors that coincide with (or are within an arbitrary limiting distance of) each other, and identifying configurations that differ only by a permutation of the vectors.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

which will be satisfied if and only if

$$C_{\mathscr{P}'\mathscr{P}''} = C_{\mathscr{P}'} C_{\mathscr{P}''} . \tag{9.7.4}$$

That is, the coefficients $C_{\mathscr{P}}$ must furnish a one-dimensional representation of the permutation group. But the permutation group has only two such representations: one is the identity, with $C_{\mathscr{P}} = +1$ for all permutations, and the other is the alternating representation, with $C_{\mathscr{P}} = +1$ or $C_{\mathscr{P}} = -1$ according to whether $\mathscr{P}$ is an even or odd permutation. These two possibilities correspond to Bose or Fermi statistics, respectively.‡

The nice feature of this argument is that it makes it clear why the case of two space dimensions is an exception. In this case there is a much richer variety of topologically distinct paths.¶ For instance, a path in which one particle circles another a definite number of times cannot be deformed into a path where it does not. In consequence, in two space dimensions it is possible to have *anyons*,[15] particles with more general permutation properties than just Fermi or Bose statistics.[8a]

Appendix Gaussian Multiple Integrals

We wish first to calculate the multiple integral, over a finite number of real variables ξ_r, of the exponential of a general quadratic function of ξ:

$$\mathscr{I} \equiv \int_{-\infty}^{\infty} \prod_r d\xi_r \exp\Big\{ -Q(\xi) \Big\} , \tag{9.A.1}$$

$$Q(\xi) = \tfrac{1}{2} \sum_{rs} K_{rs} \xi_r \xi_s + \sum_r L_r \xi_r + M , \tag{9.A.2}$$

where K_{rs}, L_r, and M are arbitrary constants, except that the matrix K is required to be symmetric and non-singular. For this purpose, we begin by considering the case where K_{rs}, L_r, and M are all real, with K_{rs} also positive. The result in the general case can then be obtained by analytic continuation.

Any real symmetric matrix can be diagonalized by an orthogonal matrix. Therefore, there is a matrix $\mathscr{S}$ with transpose $\mathscr{S}^{\mathrm{T}} = \mathscr{S}^{-1}$ such that

$$\left(\mathscr{S}^{\mathrm{T}} K \mathscr{S} \right)_{rs} = \delta_{rs} \kappa_r . \tag{9.A.3}$$

‡ There has been much discussion in the literature of possibilities other than Bose or Fermi statistics, often under the label *parastatistics*. It has been shown[13] that parastatistics theories in $d \geq 3$ space dimensions are equivalent to theories in which all particles are ordinary fermions or bosons, but carrying an extra quantum number, so that wave functions could have unusual properties under permutations of momenta and spins.

¶ This is expressed in the statement that the first homotopy group of configuration space in two space dimensions is not the permutation group, but a larger group known as the *braid group*.[14]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

Because K is assumed positive and non-singular, the eigenvalues κ_r are positive-definite. We can use the matrix $\mathscr{S}$ to perform a change of variables:

$$\xi_r = \sum_s \mathscr{S}_{rs}\xi'_s \,. \tag{9.A.4}$$

The Jacobian $|\mathrm{Det}\,\mathscr{S}|$ of this transformation is unity, so the multiple integral (9.A.1) is now given by a product of ordinary integrals:

$$\begin{aligned}\mathscr{I} &= e^{-M}\prod_r \int_{-\infty}^{\infty} d\xi' \exp\left\{-\frac{\kappa_r}{2}\xi'^2 - (\mathscr{S}^{\mathrm{T}}L)_r\xi'\right\} \\ &= e^{-M}\prod_r \sqrt{\frac{2\pi}{\kappa_r}}\exp\left\{\frac{1}{2\kappa_r}(\mathscr{S}^{\mathrm{T}}L)_r^2\right\} \,. \end{aligned} \tag{9.A.5}$$

But the determinant and the reciprocal of Eq. (9.A.3) give

$$\mathrm{Det}\,K = \prod_r \kappa_r \,, \qquad K^{-1}_{rs} = \sum_\ell \mathscr{S}_{r\ell}\mathscr{S}_{s\ell}\kappa_\ell^{-1} \,,$$

so Eq. (9.A.5) may be written

$$\mathscr{I} = \left(\mathrm{Det}\left(\frac{K}{2\pi}\right)\right)^{-1/2}\exp\left\{\tfrac{1}{2}\sum_{rs} L_r L_s K^{-1}_{rs} - M\right\} \,. \tag{9.A.6}$$

Eq. (9.A.1) defines a function of K_{rs}, L_r, and M that is analytic in K_{rs} in a finite region around the surface where K_{rs} is real and positive, where the integral converges, and for such K_{rs} is analytic everywhere in L_r and M. Since (9.A.6) equals (9.A.1) for K_{rs}, L_r, and M all real, with K_{rs} also positive, Eq. (9.A.6) provides an analytic continuation of Eq. (9.A.1) to the whole complex plane, with a cut required by the square root. The sign of the square root is fixed by this analytic continuation. In field theory K_{rs} is actually imaginary, except for a small real part due to the '$i\epsilon$ term'.

It is useful to express Eq. (9.A.6) in terms of the stationary point of the function (9.A.2):

$$\bar{\xi}_r = -\textstyle\sum_s K^{-1}_{rs} L_s \,, \tag{9.A.7}$$

$$\partial Q(\xi)/\partial\xi_r = 0 \qquad \text{at} \quad \xi = \bar{\xi} \,, \tag{9.A.8}$$

as

$$\mathscr{I} = \left(\mathrm{Det}\left(\frac{K}{2\pi}\right)\right)^{-1/2}\exp\left\{-Q(\bar{\xi})\right\} \,. \tag{9.A.9}$$

This is the result to remember: *Gaussian integrals can be evaluated up to a determinant factor by setting the integration variable equal to the point where the argument of the exponential is stationary.*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

We next wish to use this result to calculate the integrals

$$I_{r_1\cdots r_{2N}} \equiv \int \left(\prod_r d\xi_r\right) \xi_{r_1}\xi_{r_2}\cdots\xi_{r_{2N}} \exp\left\{-\tfrac{1}{2}\sum_{rs} K_{rs}\xi_r\xi_s\right\} . \quad (9.A.10)$$

(Integrals of this sort with an odd number of ξ-factors in the integrand obviously vanish.) From the power-series expression of $\exp\left(-\sum_r L_r\xi_r\right)$ in Eq. (9.A.1), we have the sum rule

$$\begin{aligned}
&\sum_{N=0}^{\infty} \sum_{r_1 r_2\cdots r_{2N}} \frac{1}{(2N)!} I_{r_1 r_2\cdots r_{2N}} L_{r_1} L_{r_2}\cdots L_{r_{2N}} \\
&\quad = \int \left(\prod_r d\xi_r\right) \exp\left\{-\sum_r L_r\xi_r - \frac{1}{2}\sum_{rs} K_{rs}\xi_r\xi_s\right\} \\
&\quad = \left[\mathrm{Det}\left(\frac{K}{2\pi}\right)\right]^{-1/2} \exp\left\{\frac{1}{2}\sum_{rs} L_r L_s K^{-1}_{rs}\right\} \\
&\quad = \left[\mathrm{Det}\left(\frac{K}{2\pi}\right)\right]^{-1/2} \sum_{N=0}^{\infty} \frac{1}{N!2^N}\left(\sum_{rs} L_r L_s K^{-1}_{rs}\right)^N . \quad (9.A.11)
\end{aligned}$$

Comparing the coefficients of $L_{r_1}L_{r_2}\cdots L_{r_{2N}}$ on both sides, we see that $I_{r_1 r_2\cdots r_{2N}}$ must be proportional to a sum of products of elements of K^{-1}, which symmetry requires to take the form

$$I_{r_1 r_2\cdots r_{2N}} = c_N \sum_{\substack{\text{pairings} \\ \text{of } r_1\cdots r_{2N}}} \prod_{\text{pairs}} (K^{-1})_{\text{paired indices}} . \quad (9.A.12)$$

Here the sum is over all ways of pairing the indices $r_1\cdots r_{2N}$, with two pairings being considered the same if they differ only by the order of the pairs, or by the order of indices within a pair. To calculate the constant factor c_N, we note that the number v_N of terms in the sum over pairings in Eq. (9.A.12) is equal to the number $(2N)!$ of permutations of the indices, divided by the number of ways $N!$ of permuting index pairs and by the number 2^N of permutations within index pairs

$$v_N = \frac{(2N)!}{N!2^N} . \quad (9.A.13)$$

Therefore, Eq. (9.A.12) gives

$$\sum_{r_1 r_2\cdots r_{2N}} L_{r_1}L_{r_2}\cdots L_{r_{2N}} I_{r_1 r_2\cdots r_{2N}} = v_N c_N \left(\sum_{rs} L_r L_s K^{-1}_{rs}\right)^N . \quad (9.A.14)$$

Comparing this with Eq. (9.A.11) shows that the factors $(2N)!$ and $N!2^N$ are cancelled by v_N, leaving us with

$$c_N = \left[\mathrm{Det}\left(\frac{K}{2\pi}\right)\right]^{-1/2} . \quad (9.A.15)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

For instance,

$$I_{r_1r_2} = I_0(K^{-1})_{r_1r_2}\,, \tag{9.A.16}$$

$$I_{r_1r_2r_3r_4} = I_0\Big[(K^{-1})_{r_1r_2}(K^{-1})_{r_3r_4} + (K^{-1})_{r_1r_3}(K^{-1})_{r_2r_4} + (K^{-1})_{r_1r_4}(K^{-1})_{r_2r_3}\Big]\,, \tag{9.A.17}$$

and so on, where I_0 is the integral with no indices

$$I_0 \equiv \int \left(\prod_r d\xi_r\right) \exp\left\{-\tfrac{1}{2}\sum_{rs} K_{rs}\xi_r\xi_s\right\} = \left[\mathrm{Det}\left(\frac{K}{2\pi}\right)\right]^{-1/2}\,. \tag{9.A.18}$$

Problems

1. Consider a non-relativistic particle of mass m, moving along the x-axis in a potential $V(x) = m\omega^2x^2/2$. Use path-integral methods to find the probability that if the particle is at x_1 at time t_1, then it is between x and $x + dx$ at time t.

2. Find the wave function in field space of a state consisting of a single spinless particle of mass $m \neq 0$. Use the result to derive the Feynman rules for emission or absorption of such a particle.

3. Find the wave function in field space of the vacuum in the theory of a neutral vector field of mass $m \neq 0$. Use the result to derive the form of the $i\epsilon$ terms in the propagator of this field.

4. The Lagrangian density of the free spin 3/2 Rarita–Schwinger field ψ^μ is

   $$\mathscr{L} = -\bar\psi^\mu(\gamma^\nu\partial_\nu + m)\psi_\mu - \tfrac{1}{3}\bar\psi^\mu(\gamma_\mu\partial_\nu + \gamma_\nu\partial_\mu)\psi^\nu + \tfrac{1}{3}\bar\psi^\mu\gamma_\mu(\gamma^\sigma\partial_\sigma - m)\gamma^\nu\psi_\nu.$$

   Use path-integral methods to find the propagator of this field.

References

1. R. P. Feynman, *The Principle of Least Action in Quantum Mechanics* (Princeton University, 1942; University Microfilms Publication No. 2948, Ann Arbor). Also see R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965). For a general reference, see J. Glimm and A. Jaffe, *Quantum Physics–A Functional Integral Point of View*, 2nd edn (Springer-Verlag, New York, 1987).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

2. P. A. M. Dirac, *Phys. Zeits. Sowjetunion* **3**, 62 (1933).

3. R. P. Feynman, *Rev. Mod. Phys.* **20**, 367 (1948); *Phys. Rev.*, **74**, 939, 1430 (1948); **76**, 749, 769 (1949); **80**, 440 (1950).

4. L. D. Faddeev and V. N. Popov, *Phys. Lett.* **B25**, 29 (1967). Also see R. P. Feynman, *Acta Phys. Pol.* **24**, 697 (1963); S. Mandelstam, *Phys. Rev.* **175**, 1580, 1604 (1968).

5. B. De Witt, *Phys. Rev. Lett.* **12**, 742 (1964).

6. G. 't Hooft, *Nucl. Phys.* **B35**, 167 (1971).

7. I. S. Gerstein, R. Jackiw, B. W. Lee, and S. Weinberg, *Phys. Rev.* **D3**, 2486 (1971).

8. L. D. Faddeev, *Teor. Mat. Fizika*, **1**, 3 (1969); translation in *Theor. Math. Phys.* **1**, 1 (1970).

8*a*. J. Schwinger, *Proc. Nat. Acad. Sci.* **44**, 956 (1958).

8*b*. K. Osterwalder and R. Schrader, *Phys. Rev. Lett.* **29**, 1423 (1972); *Commun. Math. Phys.* **31**, 83 (1973); *Commun. Math. Phys.* **42**, 281 (1975). The Osterwalder-Schrader axioms require smoothness, Euclidean covariance, 'reflection positivity,' permutation symmetry, and cluster decomposition.

9. This section was largely inspired by discussions with J. Polchinski.

10. F. A. Berezin, *The Method of Second Quantization* (Academic Press, New York, 1966).

11. M. G. G. Laidlaw and C. M. De Witt, *Phys. Rev.* D **3**, 1375 (1970).

12. J. M. Leinaas and J. Myrheim, *Nuovo Cimento* **37 B**, 1 (1977).

13. Y. Ohnuki and S. Kamefuchi, *Phys. Rev.* **170**, 1279 (1968); *Ann. Phys.* **51**, 337 (1969); K. Drühl, R. Haag, and J. E. Roberts, *Commun. Math. Phys.* **18**, 204 (1970).

14. The braid group was introduced by E. Artin. See *The Collected Papers of E. Artin*, ed. by S. Lang and J. E. Tate (Addison-Wesley, Reading, MA, 1965).

15. F. Wilczek, *Phys. Rev. Lett.* **49**, 957 (1982); K. Fredenhagen, M. R. Gaberdiel, and S. M. Rüger, Cambridge preprint DAMTP-94-90 (1994). Also see J. M. Leinaas and J. Myrheim, Ref. 12.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.011

10
Non-Perturbative Methods

We are now going to begin our study of higher-order contributions to physical processes, corresponding to Feynman diagrams involving one or more loops. It will be very useful in this work to have available a method of deriving results valid to all orders in perturbation theory (and in some cases beyond perturbation theory). In this chapter we will exploit the field equations and commutation relations of the interacting fields in the Heisenberg picture for this purpose. The essential bridge between the Heisenberg picture and the Feynman diagrams of perturbation theory is provided by the theorem proved in Section 6.4: the sum of all diagrams for a process $\alpha \to \beta$ with extra vertices inserted corresponding to operators $o_a(x)$, $o_b(y)$, etc. is given by the matrix element of the time-ordered product of the corresponding Heisenberg-picture operators

$$\left(\Psi_\beta{}^-, T\left\{-iO_a(x), -iO_b(y)\cdots\right\}\Psi_\alpha{}^+\right) .$$

As a special case, where the operators $O_a(x)$, $O_b(x)$, etc. are elementary particle fields, this matrix element equals the sum of all Feynman diagrams with incoming lines on the mass shell corresponding to the state α, outgoing lines on the mass shell corresponding to the state β, and lines off the mass shell (including propagators) corresponding to the operators $O_a(x)$, $O_b(x)$, etc. After exploring some of the non-perturbative results that can be obtained in this way we will be in a good position to take up the perturbative calculation of radiative corrections.

10.1 Symmetries

One obvious but important use of the theorem quoted above is to extend the application of symmetry principles from S-matrix elements, where all external lines have four-momenta on the mass shell, to parts of Feynman diagrams, with some or all external lines off the mass shell.

For instance, consider the symmetry of spacetime translational invariance. This symmetry has as a consequence the existence of a Hermitian

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

four-vector operator P^μ, with the property that, for any local function $O(x)$ of field operators and their canonical conjugates,

$$\left[P_\mu, O(x)\right] = i\frac{\partial}{\partial x^\mu}O(x)\,. \tag{10.1.1}$$

(See Eqs. (7.3.28) and (7.3.29).) Also, the states α and β are usually chosen to be eigenstates of the four-momentum:

$$P^\mu \Psi_\alpha{}^+ = p_\alpha^\mu \Psi_\alpha{}^+\,, \qquad P^\mu \Psi_\beta{}^- = p_\beta^\mu \Psi_\beta{}^-\,. \tag{10.1.2}$$

It follows that for any set of local functions $O_a(x)$, $O_b(x)$, etc. of fields and/or field derivatives

$$\begin{aligned}(p_{\beta\mu} - p_{\alpha\mu})&\left(\Psi_\beta{}^-,\, T\left\{O_a(x_1),\, O_b(x_2)\cdots\right\}\Psi_\alpha{}^+\right)\\ &= \left(\Psi_\beta{}^-, \left[P_\mu,\, T\left\{O_a(x_1),\, O_b(x_2)\cdots\right\}\right]\Psi_\alpha{}^+\right)\\ &= i\left(\frac{\partial}{\partial x_1^\mu} + \frac{\partial}{\partial x_2^\mu} + \cdots\right)\left(\Psi_\beta{}^- T\left\{O_a(x_1), O_b(x_2), \cdots\right\}\Psi_\alpha{}^+\right)\,.\end{aligned} \tag{10.1.3}$$

This has the solution

$$\begin{aligned}&\left(\Psi_\beta{}^-,\, T\left\{O_a(x_1), O_b(x_2), \cdots\right\}\Psi_\alpha{}^+\right)\\ &= \exp\left(i(p_\alpha - p_\beta)\cdot x\right) F_{ab\cdots}(x_1 - x_2, \cdots)\,,\end{aligned} \tag{10.1.4}$$

where x is any sort of average spacetime coordinate

$$x^\mu = c_1 x_1^\mu + c_2 x_2^\mu + \cdots\,, \qquad c_1 + c_2 + \cdots = 1 \tag{10.1.5}$$

and F depends only on differences among the xs. (In particular, a vacuum expectation value can depend only on the coordinate differences.) We can Fourier transform Eq. (10.1.4) by integrating separately over x^μ and the coordinate differences, with the result that

$$\begin{aligned}&\int d^4x_1\, d^4x_2 \cdots \left(\Psi_\beta{}^-,\, T\left\{O_a(x_1), O_b(x_2), \cdots\right\}\Psi_\alpha{}^+\right)\\ &\times \exp(-ik_1\cdot x_1 - ik_2\cdot x_2 - \cdots) \propto \delta^4(p_\alpha - p_\beta - k_1 - k_2 - \cdots)\,.\end{aligned} \tag{10.1.6}$$

We saw in Section 6.4 that the matrix element of the time-ordered product is given by applying the usual coordinate-space Feynman rules to the sum of all graphs with incoming particles corresponding to particles in α, outgoing particles in β, and external lines that simply terminate in vertices at $x_1, x_2, \cdots$. The Fourier transform (10.1.6) is correspondingly given by applying the momentum-space Feynman rules to the same sum of Feynman diagrams, with off-shell external lines carrying four-momenta $k_1, k_2, \cdots$ into the diagrams. Eq. (10.1.6) is then just the statement that this sum of Feynman graphs conserves four-momentum. The result is obvious in perturbation theory, because four-momentum is conserved at

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

every vertex, so it is not surprising to see the same result emerging without having to rely on perturbation theory.

With somewhat more effort, one can use the Lorentz transformation properties of the Heisenberg-picture fields and the 'in' and 'out' states to show that the sum of all graphs with a given set of on- and off-shell lines satisfies the same Lorentz transformation conditions as the lowest-order terms.

Similar arguments apply to the conservation of internal quantum numbers, like electric charge. As shown in Section 7.3, a field or other operator $O_a(x)$ that destroys a charge q_a (or creates a charge $-q_a$) will satisfy

$$[Q, O_a(x)] = -q_a O_a(x)$$

in the Heisenberg and interaction pictures alike. Also, if the free-particle states α and β have charges q_α and q_β, then so do the corresponding 'in' and 'out' states. We then have

$$\begin{aligned}(q_\beta - q_\alpha) &\left(\Psi_\beta{}^-, T\left\{O_a(x), O_b(y), \cdots\right\}\Psi_\alpha{}^+\right) \\ &= \left(\Psi_\beta{}^-, \left[Q, T\left\{O_a(x), O_b(y), \cdots\right\}\right]\Psi_\alpha{}^+\right) \\ &= -(q_a + q_b + \cdots)\left(\Psi_\beta{}^-, T\left\{O_a(x), O_b(y), \cdots\right\}\Psi_\alpha{}^+\right) .\end{aligned}$$

Thus the amplitude $\left(\Psi_\beta{}^-, T\left\{O_a(x), O_b(y), \cdots\right\}\Psi_\alpha{}^+\right)$ vanishes unless charge is conserved

$$q_\beta = q_\alpha - q_a - q_b - \cdots \quad . \tag{10.1.7}$$

A somewhat less trivial example is provided by the symmetry of charge-conjugation invariance. As we saw in Chapter 5, there is an operator C that interchanges electron and positron operators

$$\begin{aligned}\mathsf{C}\, a(\mathbf{p}, \sigma, e^-)\mathsf{C}^{-1} &= \xi^*\, a(\mathbf{p}, \sigma, e^+) , \\ \mathsf{C}\, a(\mathbf{p}, \sigma, e^+)\mathsf{C}^{-1} &= \xi\, a(\mathbf{p}, \sigma, e^-) ,\end{aligned}$$

with ξ a phase factor. For the free electron field $\psi(x)$, this gives

$$\mathsf{C}\psi(x)\,\mathsf{C}^{-1} = -\xi^* \beta \mathscr{C} \psi(x)^* ,$$

where βC is a 4×4 matrix, which (for the Dirac matrix representation we have been using, with γ_5 diagonal) takes the form

$$\beta\, C = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad .$$

Applied to the free-particle electric current in spinor electrodynamics, this

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

gives

$$\mathsf{C}(\bar{\psi}\gamma^\mu\psi)\mathsf{C}^{-1} = -\bar{\psi}C\gamma^{\mu T}C\psi = -\bar{\psi}\gamma^\mu\psi \ .$$

If C is to be conserved in electrodynamics, it must then also be defined to anticommute with the free photon field

$$\mathsf{C}(a^\mu)\mathsf{C}^{-1} = -a^\mu \ .$$

In theories like electrodynamics for which C commutes with the interaction as well as H_0, it also commutes with the similarity transformation $\Omega(t)$ between the Heisenberg and interaction pictures, and so it anticommutes with the electric current of the interacting fields

$$\mathsf{C}(\bar{\Psi}\gamma^\mu\Psi)\mathsf{C}^{-1} = -\bar{\Psi}\gamma^\mu\Psi \tag{10.1.8}$$

and the electromagnetic field in the Heisenberg-picture

$$\mathsf{C}(A^\mu)\mathsf{C}^{-1} = -A^\mu \ . \tag{10.1.9}$$

It follows then that the vacuum expectation value of the time-ordered product of any odd number of electromagnetic currents and/or fields vanishes. Therefore the sum of all Feynman graphs with an odd number of external photon lines (off or on the photon mass shell) and no other external lines vanishes.

This result is known as *Furry's theorem.*[1] It can be proved perturbatively by noting that a graph consisting of electron loops ℓ, to each of which are attached n_ℓ photon lines, must have numbers I and E of internal and external photon lines related by an analog of Eq. (6.3.11):

$$2I + E = \sum_\ell n_\ell \ .$$

Hence if E is odd at least one of the loops must have attached an odd number of photon lines. For any such loop there is a cancellation between the two diagrams in which the electron arrows circulate around the loop in opposite directions. Hence Furry's theorem is a somewhat less trivial consequence of a symmetry principle than translation or Lorentz invariance; it is not true of individual diagrams, but rather of certain sums of diagrams. Figure 10.1 illustrates the application of Furry's theorem that was historically most important, its use to show that the scattering of a photon by an external electromagnetic field receives no contributions of first order (or any odd order) in the external field.

10.2 Pole Structure

One of the most important uses of the non-perturbative methods described in this chapter is to clarify the pole structure of Feynman amplitudes as

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

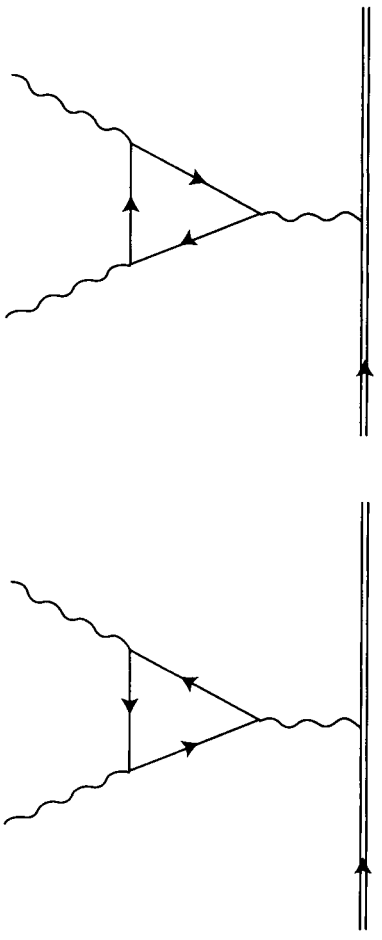

Figure 10.1. The lowest-order diagrams for the scattering of a photon by an electromagnetic field. Here straight lines represent virtual electrons; wavy lines represent real and virtual photons; and the double line represents a heavy particle like an atomic nucleus that serves as a source of an electromagnetic field. The contributions of these two diagrams cancel, as required by charge-conjugation invariance.

functions of the momenta carried by external lines. Often the S-matrix for a physical process can be well approximated by the contribution of a single pole. Also, an understanding of this pole structure will help us later in dealing with radiative corrections to particle propagators.

Consider the momentum-space amplitude

$$\int d^4x_1 \cdots d^4x_n \, e^{-iq_1\cdot x_1} \cdots e^{-iq_n\cdot x_n} \left\langle T\left\{A_1(x_1)\cdots A_n(x_n)\right\}\right\rangle_0 \equiv G(q_1 \cdots q_n) \,. \tag{10.2.1}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

The As are Heisenberg-picture operators of arbitrary Lorentz type, and $\langle\cdots\rangle_0$ denotes the expectation value in the true vacuum $\Psi_0{}^+ = \Psi_0{}^- \equiv \Psi_0$. As discussed in Section 6.4, if $A_1, \cdots A_n$ are ordinary fields appearing in the Lagrangian, then (10.2.1) is a sum of the terms calculated using the ordinary Feynman rules, for all graphs with external lines corresponding to the fields $A_1, \cdots A_n$, carrying off-shell four-momenta $q_1 \cdots q_n$ into the graph. However, we will not be limited to this case; the A_i may be arbitrary local functions of fields and field derivatives.

We are interested in poles of G at certain values of the invariant squares of the total four-momenta carried by various subsets of the external lines. To be definite, let's consider G as a function of q^2, where

$$q \equiv q_1 + \cdots + q_r = -q_{r+1} - \cdots - q_n \tag{10.2.2}$$

with $1 \leq r \leq n-1$. We will show that G has a pole at $q^2 = -m^2$, where m is the mass of any one-particle state that has non-vanishing matrix elements with the states $A_1^\dagger \cdots A_r^\dagger \Psi_0$ and $A_{r+1} \cdots A_n \Psi_0$, and that the residue at this pole is given by

$$\begin{aligned} G \to & \frac{-2i\sqrt{\mathbf{q}^2 + m^2}}{q^2 + m^2 - i\epsilon} (2\pi)^7 \delta^4(q_1 + \cdots + q_n) \\ & \times \sum_\sigma M_{0|\mathbf{q},\sigma}(q_2 \cdots q_r) M_{\mathbf{q},\sigma|0}(q_{r+2} \cdots q_n) \end{aligned} \tag{10.2.3}$$

where the Ms are defined by*

$$\begin{aligned} & \int d^4x_1 \cdots d^4x_r \, e^{-iq_1\cdot x_1} \cdots e^{-iq_r\cdot x_r} \left(\Psi_0, T\left\{A_1(x_1) \cdots A_r(x_r)\right\}\Psi_{\mathbf{p},\sigma}\right) \\ & = (2\pi)^4 \delta^4(q_1 + \cdots + q_r - p) M_{0|\mathbf{p},\sigma}(q_2 \cdots q_r) \,, \end{aligned} \tag{10.2.4}$$

$$\begin{aligned} & \int d^4x_{r+1} \cdots d^4x_n \, e^{-iq_{r+1}\cdot x_{r+1}} \cdots e^{-iq_n\cdot x_n} \\ & \qquad \times \left(\Psi_{\mathbf{p},\sigma}, T\left\{A_{r+1}(x_{r+1}) \cdots A_n(x_n)\right\}\Psi_0\right) \\ & = (2\pi)^4 \delta^4(q_{r+1} + \cdots + q_n + p) M_{\mathbf{p},\sigma|0}(q_{r+2} \cdots q_n) \end{aligned} \tag{10.2.5}$$

(with $p^0 \equiv \sqrt{\mathbf{p}^2 + m^2}$), and the sum is over all spin (or other) states of the particle of mass m.

Before proceeding to the proof, it will help to clarify the significance of

* Recall that in the absence of time-varying external fields, there is no distinction between 'in' and 'out' one-particle states, so that $\Psi_{\mathbf{p},\sigma}{}^+ = \Psi_{\mathbf{p},\sigma}{}^- = \Psi_{\mathbf{p},\sigma}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

(10.2.3) if we write it in the somewhat long-winded form

$$G(q_1 \cdots q_n) \to \sum_\sigma \int d^4k$$
$$\times \left[(2\pi)^4\,\delta^4(q_1 + \cdots + q_r - k)\,(2\pi)^{3/2}\left(2\sqrt{\mathbf{k}^2+m^2}\right)^{1/2} M_{0|\mathbf{k},\sigma}(q_2 \cdots q_r)\right]$$
$$\times \left[\frac{-i}{(2\pi)^4}\,\frac{1}{k^2+m^2-i\epsilon}\right]$$
$$\times \Bigg[(2\pi)^4\,\delta^4(k + q_{r+1} + \cdots + q_n)\,(2\pi)^{3/2}$$
$$\times \left(2\sqrt{\mathbf{k}^2+m^2}\right)^{1/2} M_{\mathbf{k},\sigma|0}(q_{r+2} \cdots q_n)\Bigg] \quad . \tag{10.2.6}$$

This is just what we should expect from a Feynman diagram with a single internal line for a particle of mass m connecting the first r and the last $n-r$ external lines.** However, it is *not* necessary that the particle of mass m correspond to a field that appears in the Lagrangian of the theory. Eqs. (10.2.3) and (10.2.6) apply even if this particle is a bound state of the so-called elementary particles whose fields do appear in the Lagrangian. In this case, the pole arises not from single Feynman diagrams, like Figure 10.2, but rather from infinite sums of diagrams, such as the one shown in Figure 10.3. This is the first place where the methods of this chapter take us beyond results that could be derived as properties of each order of perturbation theory.

Now to the proof. Among the $n!$ possible orderings of the times $x_1^0 \cdots x_n^0$ in Eq. (10.2.1), there are $n!/r!(n-r)!$ for which the first r of the x_i^0 are *all* larger than the last $n-r$. Isolating the contribution of this part of the volume of integration in Eq. (10.2.1), we have

$$G(q_1 \cdots q_n) = \int d^4x_1 \cdots d^4x_n\, e^{-iq_1\cdot x_1} \cdots e^{-iq_n\cdot x_n}$$
$$\times\theta\left(\min\left[x_1^0 \cdots x_r^0\right] - \max\left[x_{r+1}^0 \cdots x_n^0\right]\right)$$
$$\times\left(\Psi_0,\ T\left\{A_1(x_1)\cdots A_r(x_r)\right\}\,T\left\{A_{r+1}(x_{r+1})\cdots A_n(x_n)\right\}\Psi_0\right)$$
$$+ \text{OT}, \tag{10.2.7}$$

where 'OT' denotes the other terms arising from different time-orderings. We can evaluate the matrix element here by inserting a complete set of

** See Figure 10.2. The factors $(2\pi)^{3/2}\left[2\sqrt{\mathbf{k}^2+m^2}\right]^{1/2}$ just serve to remove kinematic factors associated with the mass m external line in $M_{0|\mathbf{k},\sigma}$ and $M_{\mathbf{k},\sigma|0}$. Also, the sum over σ of the product of coefficient-function factors from these two matrix elements yields the numerator of the propagator associated with the internal line in Figure 10.2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

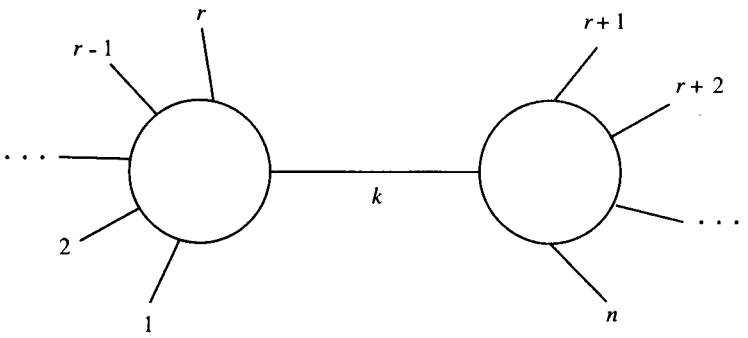

Figure 10.2. A Feynman diagram with the pole structure (10.2.6). Here the line carrying a momentum k represents an elementary particle, one whose field appears in the Lagrangian.

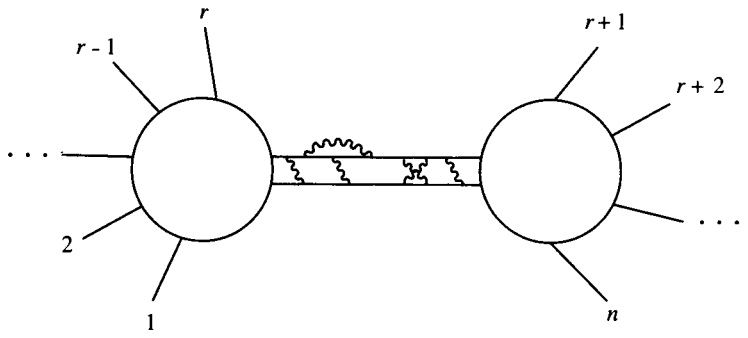

Figure 10.3. Figure 10.3. A Feynman diagram of the class whose sum has the pole structure (10.2.6). Here the pole is due to a composite particle, a bound state of two elementary particles. The elementary particles are represented by straight lines, and interact by the exchange of particles represented by wavy lines.

intermediate states between time-ordered products. Among these may be the single-particle state $\Psi_{\mathbf{p},\sigma}$ of a definite species of mass m. Further isolating the contribution of these one-particle intermediate states, we have

$$G(q_1 \cdots q_n) = \int d^4x_1 \cdots d^4x_n \, e^{-iq_1\cdot x_1} \cdots e^{-iq_n\cdot x_n}$$
$$\theta\left(\min\left[x_1^0 \cdots x_r^0\right] - \max\left[x_{r+1}^0 \cdots x_n^0\right]\right) \sum_\sigma \int d^3p$$
$$\left(\Psi_0,\, T\left\{A_1(x_1)\cdots A_r(x_r)\right\}\Psi_{\mathbf{p},\sigma}\right)\left(\Psi_{\mathbf{p},\sigma},\, T\left\{A_{r+1}(x_{r+1})\cdots A_n(x_n)\right\}\Psi_0\right)$$
$$+ \text{OT}, \tag{10.2.8}$$

where 'OT' now denotes other terms, here arising not only from other time-orderings, but also from other intermediate states. It will be convenient

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

to shift variables of integration, so that

$$\begin{aligned} x_i &= x_1 + y_i\,, \qquad && i = 2, 3, \cdots r\,, \\ x_i &= x_{r+1} + y_i\,, && i = r+2, \cdots n\,, \end{aligned}$$

and use the results of the previous section to write

$$\begin{aligned} &\left(\Psi_0,\, T\left\{A_1(x_1)\cdots A_r(x_r)\right\}\Psi_{\mathbf{p},\sigma}\right) \\ &\quad = e^{ip\cdot x_1}\left(\Psi_0,\, T\left\{A_1(0)A_2(y_2)\cdots A_r(y_r)\right\}\Psi_{\mathbf{p},\sigma}\right)\,, \end{aligned} \tag{10.2.9}$$

$$\begin{aligned} &\left(\Psi_{\mathbf{p},\sigma},\, T\left\{A_{r+1}(x_{r+1})\cdots A_n(x_n)\right\}\Psi_0\right) \\ &\quad = e^{-ip\cdot x_{r+1}}\left(\Psi_{\mathbf{p},\sigma},\, T\left\{A_{r+1}(0)\cdots A_n(y_n)\right\}\Psi_0\right)\,. \end{aligned} \tag{10.2.10}$$

Also, the argument of the theta function becomes

$$\begin{aligned} &\min\,[x_1^0 \cdots x_r^0] - \max\,[x_{r+1}^0 \cdots x_n^0] \\ &\quad = x_1^0 - x_{r+1}^0 + \min\,[0\, y_2^0 \cdots y_r^0] - \max\,[0\, y_{r+1}^0 \cdots y_n^0]\,. \end{aligned}$$

We also insert the Fourier representation (6.2.15) of the step function

$$\theta(\tau) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{d\omega\, e^{-i\omega\tau}}{\omega + i\epsilon}\,.$$

The integrals over x_1 and x_{r+1} now just yield delta functions:

$$\begin{aligned} G(q_1 \cdots q_n) &= \int d^4y_2 \cdots d^4y_r\, d^4y_{r+2} \cdots d^4y_n \\ &\times e^{-iq_2\cdot y_2} \cdots e^{-iq_r\cdot y_r} e^{-iq_{r+2}\cdot y_{r+2}} \cdots e^{-iq_n\cdot y_n} \\ &\times -\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{d\omega}{\omega + i\epsilon}\exp\left(-i\omega\left[\min[0\, y_2^0 \cdots y_r^0] - \max[0\, y_{r+1}^0 \cdots y_n^0]\right]\right) \\ &\times \sum_\sigma \int d^3p\left(\Psi_0,\, T\left\{A_1(0)\cdots A_r(y_r)\right\}\Psi_{\mathbf{p},\sigma}\right) \\ &\qquad\qquad \times \left(\Psi_{\mathbf{p},\sigma}, T\left\{A_{r+1}(0)\cdots A_n(y_n)\right\}\Psi_0\right) \\ &\times (2\pi)^4\delta^3(\mathbf{p} - \mathbf{q}_1 - \cdots - \mathbf{q}_r)\,\delta(\sqrt{\mathbf{p}^2 + m^2} + \omega - q_1^0 - \cdots - q_r^0) \\ &\times (2\pi)^4\,\delta^3(\mathbf{q}_{r+1} + \cdots + \mathbf{q}_n + \mathbf{p})\,\delta\left(q_{r+1}^0 + \cdots + q_n^0 + \sqrt{\mathbf{p}^2 + m^2} + \omega\right) \\ &\quad + \text{OT}\quad. \end{aligned} \tag{10.2.11}$$

We are interested here only in the pole that arises from the vanishing of the denominator $\omega + i\epsilon$, so for our present purposes we can set the factor $\exp(-i\omega[\min - \max])$ equal to unity. The integrals over both $\mathbf{p}$ and ω are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

now trivial, and yield the pole

$$G(q_1 \cdots q_n) \to i(2\pi)^7 \delta^4(q_1 + \cdots + q_n) \left[q^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon \right]^{-1} \times \sum_\sigma M_{0|\mathbf{q},\sigma}(q_2 \cdots q_n) M_{\mathbf{q},\sigma|0}(q_{r+2} \cdots q_n) + \cdots \tag{10.2.12}$$

where now

$$q \equiv q_1 + \cdots + q_r = -q_{r+1} - \cdots - q_n ,$$

$$M_{0|\mathbf{q},\sigma}(q_2 \cdots q_r) \equiv \int d^4y_2 \cdots d^4y_r \, e^{-iq_2 \cdot y_2} \cdots e^{-iq_r \cdot y_r} \times \left(\Psi_0, \; T\left\{ A_1(0) A_2(y_2) \cdots A_r(y_r) \right\} \Psi_{\mathbf{q},\sigma} \right) , \tag{10.2.13}$$

$$M_{\mathbf{q},\sigma|0}(q_{r+2} \cdots q_n) \equiv \int d^4y_{r+2} \cdots d^4y_n \, e^{-iq_{r+2} \cdot y_{r+2}} \cdots e^{-iq_n \cdot y_n} \times \left(\Psi_{\mathbf{q},\sigma}, \; T\left\{ A_{r+1}(0) A_{r+2}(y_{r+2}) \cdots A_n(y_n) \right\} \Psi_0 \right) , \tag{10.2.14}$$

and the final '$\cdots$' in Eq. (10.2.12) denotes terms that do not exhibit this particular pole. (The 'other terms' arising from other single-particle states produce poles in q at different positions, while those arising from multi-particle states produce branch points in q, and those arising from other time-orderings produce poles and branch cuts in other variables.) Using Eqs. (10.2.9) and (10.2.10), it is easy to see that these Ms are the same as defined by Eqs. (10.2.4) and (10.2.5). Also, near the pole we can write

$$\frac{1}{q^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon} = \frac{-q^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon}{-(q^0)^2 + (\sqrt{\mathbf{q}^2 + m^2} - i\epsilon)^2} \to \frac{-2\sqrt{\mathbf{q}^2 + m^2}}{q^2 + m^2 - i\epsilon} .$$

(We again redefine ϵ by a positive factor $2\sqrt{\mathbf{q}^2 + m^2}$, which is permissible since ϵ stands for any positive infinitesimal.) Eq. (10.2.12) is thus the same as the desired result (10.2.3).

This result has a classic application to the theory of nuclear forces. Let $\Phi_a(x)$ be any real field or combination of fields (for instance, proportional to a quark–antiquark bilinear $\bar{q}\gamma_5\tau_a q$) that has a non-vanishing matrix element between a one-pion state of isospin a and the vacuum, normalized so that

$$\langle \mathrm{VAC} | \, \Phi_a(0) \, | \pi_b, \mathbf{p} \rangle = (2\pi)^{-3/2} (2p^0)^{-1/2} \delta_{ab} . \tag{10.2.15}$$

The matrix element of Φ_a between one-nucleon states with four-momenta p,p' then has a pole at $(p - p')^2 \to -m_\pi^2$ which isospin and Lorentz invariance (including space inversion invariance) dictate must take the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

form**

$$\langle N', \sigma', \mathbf{p}' | \Phi_a(0) | N, \sigma, \mathbf{p} \rangle \to i(2\pi)^{-3} G_\pi \times \frac{\left(\bar{u}' \gamma_5 \tau_a u\right)}{(p-p')^2 + m_\pi^2}, \qquad (10.2.16)$$

where u and u' are the initial and final nucleon spinor coefficient functions, including the nucleon wave functions in isospin space, and τ_a with $a = 1, 2, 3$ are the 2×2 Pauli isospin matrices. The constant G_π is known as the *pion–nucleon coupling constant*. This pole is not actually in the physical region for the matrix element (10.2.16), for which $(p-p')^2 \geq 0$, but it can be reached by analytic extension of this matrix element, for instance by considering the off-shell matrix element

$$\int d^4x \, d^4x' \, e^{-ip\cdot x} e^{ip'\cdot x'} \langle T\{\Phi_a(0) \bar{N}(x) N'(x')\}\rangle_{\text{VAC}},$$

where N and N' are appropriate components of a field operator or product of field operators with non-vanishing matrix elements between one-nucleon states and the vacuum. The theorem proved above in this section shows then that exchange of a pion in the scattering of two nucleons with initial four-momenta p_1, p_2, and final four-momenta p_1', p_2' yields a pole at $(p_1 - p_1')^2 = (p_2 - p_2')^2 \to -m_\pi^2$:

$$S_{N_1' N_2', N_1 N_2} \to -i(2\pi)^4 \delta^4(p_1' + p_2' - p_1 - p_2) \frac{G_\pi^2}{(p_1 - p_1')^2 + m_\pi^2}$$
$$\times (2\pi)^{-3}\left(\bar{u}_{1'} \gamma_5 \tau_a u_1\right) \times (2\pi)^{-3}\left(\bar{u}_2' \gamma_5 \tau_a u_2\right). \qquad (10.2.17)$$

(The easiest way to get the phases and numerical factors right in such formulas is to use Feynman diagrams; our theorem just says that the pole structure is the same as would be found in a field theory in which the Lagrangian involved an elementary pion field.) Again, this pion pole is not actually in the physical region for scattering of nucleons on the mass shell, for which $(p_1 - p_1')^2 \geq 0$, but it can be reached by analytic extension of the S-matrix element, for instance by considering the off-shell matrix

** Lorentz and isospin invariance requires this matrix to take the form $(\bar{u}' \Gamma \tau_a u)$, where Γ is a 4×4 matrix for which the bilinear $(\bar{\psi}' \Gamma \psi)$ transforms as a pseudoscalar. Like any 4×4 matrix, Γ can be expanded as a sum of terms proportional to the Dirac matrices 1, γ_μ, $[\gamma_\mu, \gamma_\nu]$, $\gamma_5 \gamma_\mu$, and γ_5. The coefficients must be respectively pseudoscalar, pseudovector, pseudotensor, vector, and scalar. Out of the two momenta p and p' it is possible to construct no pseudoscalars or pseudovectors; just one pseudotensor, proportional to $\epsilon^{\mu\nu\rho\sigma} p_\rho p'_\sigma$; two independent vectors, proportional to p_μ or p'_μ; and a scalar proportional to unity, in each case with a proportionality factor depending on the only independent scalar variable, $(p-p')^2$. By using the momentum-space Dirac equations for u and u', it is easy to see that the tensor and pseudovector matrices in Γ give contributions proportional to γ_5.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

element

$$\int d^4x_1\, d^4x_2\, d^4x_1'\, d^4x_2'\, e^{-ip_1\cdot x_1} e^{-ip_2\cdot x_2} e^{ip_1'\cdot x_1'} e^{ip_2'\cdot x_2'}$$
$$\times \langle\, T\{\bar{N}_1(x_1), \bar{N}_2(x_2), N_1'(x_1'), N_2'(x_2')\}\rangle_{\text{VAC}}\,.$$

Although this pole is not in the physical region for nucleon–nucleon scattering, the pion mass is small enough so that the pole is quite near the physical region, and under some circumstances may dominate the scattering amplitude, as for instance for large ℓ in the partial wave expansion.

Interpreted in coordinate space, a pole like this at $(p_1 - p_1')^2 = (p_2 - p_2')^2 \to -m_\pi^2$ implies a force of range $1/m_\pi$. For instance, in Yukawa's original theory[2] of nuclear force the exchange of mesons (then assumed scalar rather than pseudoscalar) produced a local potential of the form $\exp(-m_\pi r)/4\pi r$, which in the first Born approximation yields an S-matrix for non-relativistic nucleon scattering proportional to the Fourier transform:

$$\int d^3x_1\, d^3x_2\, d^3x_1'\, d^3x_2'\, e^{-i\mathbf{x}_1\cdot\mathbf{p}_1} e^{-i\mathbf{x}_2\cdot\mathbf{p}_2} e^{i\mathbf{x}_1'\cdot\mathbf{p}_1'} e^{i\mathbf{x}_2'\cdot\mathbf{p}_2'}$$
$$\times \frac{\exp\left(-m_\pi|\mathbf{x}_1 - \mathbf{x}_2|\right)}{4\pi|\mathbf{x}_1 - \mathbf{x}_2|}\delta^3(\mathbf{x}_1 - \mathbf{x}_1')\delta^3(\mathbf{x}_2 - \mathbf{x}_2')$$
$$= -(2\pi)^3\delta^3(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_1' - \mathbf{p}_2')\frac{1}{(\mathbf{p}_1 - \mathbf{p}_1')^2 + m_\pi^2}\,.$$

The factor $1/[(\mathbf{p}_1 - \mathbf{p}_1')^2 + m_\pi^2]$ is just the non-relativistic limit of the propagator $1/[(p_1 - p_1')^2 + m_\pi^2]$ in (10.2.17). (In (10.2.17) the energy transfer $p_1^0 - p_1'^0$ for $|\mathbf{p}_1| \ll m_N$ and $|\mathbf{p}_1'| \ll m_N$ equals $[\mathbf{p}_1^2 - \mathbf{p}_1'^2]/2m_N$, which is negligible compared with the magnitude $|\mathbf{p}_1 - \mathbf{p}_1'|$ of the momentum transfer.) When Yukawa's theory was first proposed, it was generally supposed that this sort of momentum-dependence arises from the appearance of a meson field in the theory. It was not until the 1950s that it became generally understood that the existence of a pole at $(p_1 - p_1')^2 \to -m_\pi^2$ follows from the existence of a pion *particle* and has nothing to do with whether this is an elementary particle with its own field in the Lagrangian.

10.3 Field and Mass Renormalization

We will now use a special case of the result of the previous section to clarify the treatment of radiative corrections in the internal and external line of general processes.

The special case that concerns us here is the one in which the four-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

momentum of a single external line approaches the mass shell. (In the notation of the previous section, this corresponds to taking $r = 1$.) We will consider a function

$$G_\ell(q_1 q_2 \cdots) = \int d^4x_1\, d^4x_2 \cdots\; e^{-iq_1\cdot x_1}\; e^{-iq_2\cdot x_2} \cdots \times \left(\Psi_0,\, T\left\{\mathcal{O}_\ell(x_1),\, A_2(x_2), \cdots\right\}\Psi_0\right), \qquad (10.3.1)$$

where $\mathcal{O}_\ell(x)$ is a Heisenberg-picture operator, with the Lorentz transformation properties of some sort of free field ψ_ℓ belonging to an irreducible representation of the homogeneous Lorentz group (or the Lorentz group including space inversion for theories that conserve parity), as labelled by the subscript ℓ, and A_2, A_3, etc. are arbitrary Heisenberg-picture operators. Suppose there is a one-particle state $\Psi_{\mathbf{q}_1,\sigma}$ that has non-vanishing matrix elements with the states $\mathcal{O}_\ell^\dagger\Psi_0$ and with $A_2A_3\cdots\Psi_0$. Then according to the theorem proved in the previous section, G_ℓ has a pole at $q_1^2 = -m^2$, with

$$G_\ell(q_1 q_2 \cdots) \to \frac{-2i\sqrt{\mathbf{q}_1{}^2 + m^2}}{q_1^2 + m^2 - i\epsilon}\,(2\pi)^3 \sum_\sigma \left(\Psi_0, \mathcal{O}_\ell(0)\Psi_{\mathbf{q}_1,\sigma}\right) \times \int d^4x_2 \cdots e^{-iq_2\cdot x_2} \cdots \left(\Psi_{\mathbf{q}_1,\sigma}\, T\left\{A_2(x_2)\cdots\right\}\Psi_0\right), \qquad (10.3.2)$$

We use Lorentz invariance to write

$$\left(\Psi_0, \mathcal{O}_\ell(0)\Psi_{\mathbf{q}_1,\sigma}\right) = (2\pi)^{-3/2}\, N\, u_\ell(\mathbf{q}_1,\sigma), \qquad (10.3.3)$$

where $u_\ell(\mathbf{q},\sigma)$ is (aside from the factor $(2\pi)^{-3/2}$) the coefficient function* appearing in the free field ψ_ℓ with the same Lorentz transformation properties as $\mathcal{O}_\ell$, and N is a constant. (It was in order to obtain Eq. (10.3.3) with a single free constant N that we had to assume that $\mathcal{O}_\ell$ transforms irreducibly.) We also define a 'truncated' matrix element M_ℓ by

$$\int d^4x_2 \cdots e^{-iq_2\cdot x_2} \cdots \left(\Psi_{\mathbf{q}_1,\sigma}\, T\left\{A_2(x_2)\cdots\right\}\Psi_0\right) \equiv N^{-1}(2\pi)^{-3/2} \sum_\ell u_\ell^*(\mathbf{q}_1,\sigma)\, M_\ell(q_2\cdots). \qquad (10.3.4)$$

Eq. (10.3.2) then reads, for $q^2 \to -m_1^2$

$$G_\ell \to \frac{-2i\sqrt{\mathbf{q}_1{}^2 + m^2}}{q_1^2 + m^2 - i\epsilon} \sum_{\sigma,\ell'} u_\ell(\mathbf{q}_1,\sigma) u_{\ell'}^*(\mathbf{q}_1,\sigma) M_{\ell'}. \qquad (10.3.5)$$

According to Eqs. (6.2.2) and (6.2.18), the quantity multiplying $M_{\ell'}$ in (10.3.5) is the momentum space matrix propagator $-i\Delta_{\ell\ell'}(q_1)$ for the free

* For instance, for a conventionally normalized free scalar field, $u_\ell(\mathbf{q}_1,\sigma) = [2\sqrt{\mathbf{q}_1^2 + m^2}]^{-1/2}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

field with the Lorentz transformation properties of $\mathcal{O}_\ell$ (or at least its limiting behavior for $q_1^2 \to -m^2$), so (10.3.5) allows us to identify M_ℓ as the sum of all graphs with external lines carrying momenta $q_1, q_2 \cdots$ corresponding to the operators $\mathcal{O}_\ell, A_2, \cdots$, but with the final propagator for the $\mathcal{O}_\ell$ line stripped away. Eq. (10.3.4) is then just the usual prescription for how to calculate the matrix element for emission of a particle from the sum of Feynman diagrams: strip away the particle propagator, and contract with the usual external line factor $(2\pi)^{-3/2} u_\ell^*$. The only discrepancy with the usual Feynman rules is the factor N.

The above theorem is a famous result due to Lehmann, Symanzik, and Zimmerman,[3] known as the *reduction formula*, which we have proved here by a somewhat different method that has allowed us easily to generalize this result to the case of arbitrary spin. One important aspect of this result is that it applies to any sort of operator; $\mathcal{O}_\ell$ need not be some field that actually appears in the Lagrangian, and the particle it creates may be a bound state composed of those particles whose fields do occur in the Lagrangian. It provides an important lesson even where $\mathcal{O}_\ell$ is some field Ψ_ℓ in the Lagrangian: if we are to use the usual Feynman rules to calculate S-matrix elements, then we should first redefine the normalization of the fields by a factor $1/N$, so that (with apologies for the multiple use of the symbol Ψ):

$$\left(\Psi_0, \Psi_\ell(0)\Psi_{\mathbf{q},\sigma}\right) = (2\pi)^{-3/2}\, u_\ell(\mathbf{q},\sigma)\,. \tag{10.3.6}$$

A field normalized as in Eq. (10.3.6) is called a *renormalized field.*

The field renormalization constant N shows up in another place. Suppose that there is just one of the operators $A_2, A_3, \cdots$ in Eq. (10.3.1), and take it to be the adjoint of a member of the same field multiplet as $\mathcal{O}_\ell$. Then Eq. (10.3.2) reads

$$\begin{aligned}
&\int d^4x_1 \int d^4x_2\, e^{-iq_1\cdot x_1}\, e^{-iq_2\cdot x_2} \left(\Psi_0,\, T\left\{\mathcal{O}_\ell(x_1)\mathcal{O}_{\ell'}^\dagger(x_2)\right\}\Psi_0\right) \\
&\overset{q_1^2\to -m^2}{\to} \frac{-2i\sqrt{\mathbf{q}_1{}^2+m^2}\,(2\pi)^3}{q_1^2+m^2-i\epsilon} \sum_\sigma \left(\Psi_0, \mathcal{O}_\ell(0)\Psi_{\mathbf{q}_1,\sigma}\right) \\
&\qquad \times \int d^4x_2\, e^{-iq_2\cdot x_2} e^{-iq_1\cdot x_2} \left(\Psi_{\mathbf{q}_1,\sigma}, \mathcal{O}_{\ell'}^\dagger(0)\Psi_0\right) \\
&= \frac{-2i|N|^2\sqrt{\mathbf{q}_1{}^2+m^2}}{q_1^2+m^2-i\epsilon} \sum_\sigma u_\ell(\mathbf{q}_1,\sigma)\, u_{\ell'}^*(\mathbf{q}_1,\sigma)\,(2\pi)^4\,\delta^4(q_1+q_2)\,.
\end{aligned}$$

This is just the usual behavior of a propagator (the sum of all graphs with two external lines) near its pole, except for the factor $|N|^2$. According to Eq. (10.3.6), this factor is absent in the propagator of the renormalized field Ψ_ℓ. Thus *a renormalized field is one whose propagator has the same*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

behavior near its pole as for a free field, and the renormalized mass is defined by the position of the pole.

To see how this works in practice, consider the theory of a real self-interacting scalar field Φ_B, the subscript B being added here to remind us that so far this is a 'bare' (i.e., unrenormalized) field. The Lagrangian density is taken as usual as

$$\mathscr{L} = -\tfrac{1}{2}\partial_\mu\Phi_B\partial^\mu\Phi_B - \tfrac{1}{2}m_B^2\Phi_B^2 - V_B(\Phi_B)\,. \tag{10.3.7}$$

In general there would be no reason to expect that the field Φ_B would satisfy condition (10.3.6), nor that the pole in q^2 would be at $-m_B^2$, so let us introduce a renormalized field and mass

$$\Phi \equiv Z^{-1/2}\Phi_B\,, \tag{10.3.8}$$

$$m^2 \equiv m_B^2 + \delta m^2\,, \tag{10.3.9}$$

with Z to be chosen so that Φ does satisfy Eq. (10.3.6), and δm^2 chosen so that the pole of the propagator is at $q^2 = -m^2$. (The use of the symbol Z in this context has become conventional; there is a different Z for each field in the Lagrangian.) The Lagrangian density (10.3.7) may then be rewritten

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1, \tag{10.3.10}$$

$$\mathscr{L}_0 = -\tfrac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \tfrac{1}{2}m^2\Phi^2, \tag{10.3.11}$$

$$\mathscr{L}_1 = -\tfrac{1}{2}(Z-1)[\partial_\mu\Phi\partial^\mu\Phi + m^2\Phi^2] + \tfrac{1}{2}Z\,\delta m^2\Phi^2 - V(\Phi)\,, \tag{10.3.12}$$

where

$$V(\Phi) \equiv V_B(\sqrt{Z}\Phi)\,.$$

In calculating the corrections to the complete momentum space propagator of the renormalized scalar field, conventionally called $\Delta'(q)$, it is convenient to consider separately the *one-particle-irreducible* graphs: those connected graphs (excluding a graph consisting of a single scalar line) that cannot be disconnected by cutting through any one internal scalar line. An example is shown in Figure 10.4. It is conventional to write the sum of all such graphs, with the two external line propagator factors $-i(2\pi)^{-4}(q^2 + m^2 - i\epsilon)^{-1}$ omitted, as $i(2\pi)^4\Pi^*(q^2)$, with the asterisk to remind us that these are one-particle-irreducible graphs. Then the corrections to the complete propagator are given by a sum of chains of one, two, or more of these one-particle-irreducible subgraphs connected with the usual uncorrected

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

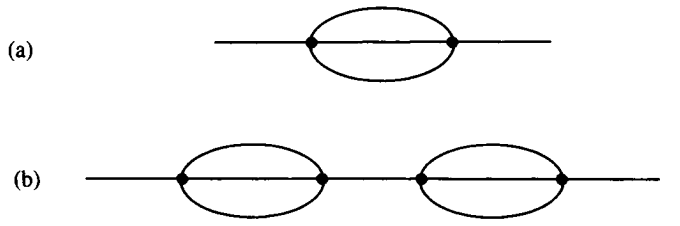

Figure 10.4. Diagrams that (a) are, or (b) are not, one-particle irreducible. These diagrams are drawn for a theory with some sort of quadrilinear interaction, like the theory of a scalar field ϕ with interaction proportional to ϕ^4.

propagator factors:

$$\begin{aligned}\frac{-i}{(2\pi)^4}\Delta'(q) &= \frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon} \\ &+\left[\frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon}\right]\left[i(2\pi)^4\Pi^*(q^2)\right]\left[\frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon}\right] \\ &+\left[\frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon}\right]\left[i(2\pi)^4\Pi^*(q^2)\right]\left[\frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon}\right] \\ &\times\left[i(2\pi)^4\Pi^*(q^2)\right]\left[\frac{-i}{(2\pi)^4}\frac{1}{q^2+m^2-i\epsilon}\right]+\cdots \end{aligned} \tag{10.3.13}$$

or more simply

$$\begin{aligned}\Delta'(q) &= [q^2+m^2-i\epsilon]^{-1}+[q^2+m^2-i\epsilon]^{-1}\Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1} \\ &+[q^2+m^2-i\epsilon]^{-1}\Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1}\Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1}+\cdots .\end{aligned} \tag{10.3.14}$$

Summing the geometric series, this gives

$$\Delta'(q) = \left[q^2+m^2-\Pi^*(q^2)-i\epsilon\right]^{-1} . \tag{10.3.15}$$

In calculating Π^*, we encounter a tree graph arising from a single insertion of vertices corresponding to the terms in Eq. (10.3.12) proportional to $\partial_\mu\Phi\partial^\mu\Phi$ and Φ^2, plus a term $\Pi^*_{\rm LOOP}$ arising from loop graphs like that in Figure 10.4(a):

$$\Pi^*(q^2) = -(Z-1)[q^2+m^2]+Z\,\delta m^2+\Pi^*_{\rm LOOP}(q^2) . \tag{10.3.16}$$

The condition that m^2 is the true mass of the particle is that the pole of the propagator should be at $q^2=-m^2$, so that

$$\Pi^*(-m^2) = 0 . \tag{10.3.17}$$

Also, the condition that the pole of the propagator at $q^2=-m^2$ should

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

have a unit residue (like the uncorrected propagator) is that

$$\left[\frac{d}{dq^2}\Pi^*(q^2)\right]_{q^2=-m^2} = 0\,. \tag{10.3.18}$$

These conditions allow us to evaluate Z and δm^2:

$$Z\,\delta m^2 = -\Pi^*_{\rm LOOP}(-m^2)\,, \tag{10.3.19}$$

$$Z = 1 + \left[\frac{d}{dq^2}\Pi^*_{\rm LOOP}(q^2)\right]_{q^2=-m^2}\,. \tag{10.3.20}$$

This incidentally shows that $Z\,\delta m^2$ and $Z-1$ are given by a series of terms containing one or more coupling constant factors, justifying the treatment of the first two terms in Eq. (10.3.12) as part of the interaction $\mathscr{L}_1$.

In actual calculations it is simplest just to say that from the loop terms $\Pi^*_{\rm LOOP}(q^2)$ we must subtract a first-order polynomial in q^2 with coefficients chosen so that the difference satisfies Eqs. (10.3.17) and (10.3.18). As we shall see, this subtraction procedure incidentally cancels the infinities that arise from the momentum space integrals in $\Pi^*_{\rm LOOP}$. However, as this discussion should make clear, *the renormalization of masses and fields has nothing directly to do with the presence of infinities, and would be necessary even in a theory in which all momentum space integrals were convergent.*

An important consequence of the conditions (10.3.17) and (10.3.18) is that it is not necessary to include radiative corrections in external lines on the mass shell. That is,

$$\Big[\Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1} + \Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1}\Pi^*(q^2)[q^2+m^2-i\epsilon]^{-1} + \cdots\Big]_{q^2\to -m^2} = 0\,. \tag{10.3.21}$$

Similar remarks apply to particles of arbitrary spin. For instance, for the 'bare' Dirac field the Lagrangian is

$$\mathscr{L} = -\bar{\Psi}_{\rm B}[\not\partial + m_{\rm B}]\Psi_{\rm B} - V_{\rm B}(\bar{\Psi}_{\rm B}\Psi_{\rm B})\,. \tag{10.3.22}$$

We introduce renormalized fields and masses

$$\Psi \equiv Z_2^{-1/2}\Psi_{\rm B}\,, \tag{10.3.23}$$

$$m = m_{\rm B} + \delta m\,. \tag{10.3.24}$$

(The subscript 2 on Z_2 is conventionally used to distinguish the renormalization constant of a fermion field.) The Lagrangian density is then rewritten

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1\,, \tag{10.3.25}$$

$$\mathscr{L}_0 = -\bar{\Psi}[\not\partial + m]\Psi\,, \tag{10.3.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

$$\mathscr{L}_1 = -(Z_2 - 1)[\bar{\Psi}[\partial\!\!\!/ + m]\Psi] + Z_2\delta m\bar{\Psi}\Psi - V_B(Z_2\bar{\Psi}\Psi)\,. \qquad (10.3.27)$$

Let $i(2\pi)^4\Sigma^*(\not{k})$ be the sum of all connected graphs, with one fermion line coming in with four-momentum k and one going out with the same four-momentum, that cannot be disconnected by cutting through any single internal fermion line, and with external line propagator factors $-i(2\pi)^{-4}$ and $[i\not{k} + m - i\epsilon]^{-1}$ omitted. (Lorentz invariance is being used to justify writing Σ^* as an ordinary function of the Lorentz scalar matrix $\not{k} \equiv k_\mu\gamma^\mu$.) Then the complete fermion propagator is

$$\begin{aligned} S'(k) &= [i\not{k} + m - i\epsilon]^{-1} + [i\not{k} + m - i\epsilon]^{-1}\Sigma^*(\not{k})[i\not{k} + m - i\epsilon]^{-1} \\ &\quad + [i\not{k} + m - i\epsilon]^{-1}\Sigma^*(\not{k})[i\not{k} + m - i\epsilon]^{-1}\Sigma^*(\not{k})[i\not{k} + m - i\epsilon]^{-1} + \cdots \\ &= [i\not{k} + m - \Sigma^*(\not{k}) - i\epsilon]^{-1}\,. \end{aligned} \qquad (10.3.28)$$

In calculating $\Sigma^*(\not{k})$ we take into account the tree graphs from the terms in Eq. (10.3.27) proportional to $\bar{\Psi}\,\partial\!\!\!/\Psi$ and $\bar{\Psi}\Psi$ as well as loop contributions:

$$\Sigma^*(\not{k}) = -(Z_2 - 1)[i\not{k} + m] + Z_2\delta m + \Sigma^*_{\text{LOOP}}(\not{k})\,. \qquad (10.3.29)$$

The condition that the complete propagator has a pole at $k^2 = -m^2$ with the same residue as the uncorrected propagator is then that

$$\Sigma^*(im) = 0\,, \qquad (10.3.30)$$

$$\left.\frac{\partial\Sigma^*(\not{k})}{\partial\not{k}}\right|_{k=im} = 0\,, \qquad (10.3.31)$$

and hence

$$Z_2\delta m = -\Sigma^*_{\text{LOOP}}(im)\,, \qquad (10.3.32)$$

$$Z_2 = 1 - i\left.\frac{\partial\Sigma^*_{\text{LOOP}}(\not{k})}{\partial\not{k}}\right|_{\not{k}=im}\,. \qquad (10.3.33)$$

Just as for scalars, the vanishing of $[i\not{k} + m]^{-1}\Sigma^*(\not{k})$ in the limit $\not{k} \to im$ tells us that radiative corrections may be ignored in external fermion lines. Corresponding results for the photon propagator will be derived in Section 10.5.

10.4 Renormalized Charge and Ward Identities

The use of the commutation and conservation relations of Heisenberg-picture operators allows us to make a connection between the charges (or other similar quantities) in the Lagrangian density and the properties of physical states. Recall that the invariance of the Lagrangian density

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

with respect to global gauge transformations $\Psi_\ell \to \exp(iq_\ell\alpha)\Psi_\ell$ (with α an arbitrary constant phase) implies the existence of a current

$$J^\mu = -i \sum_\ell \frac{\partial \mathscr{L}}{\partial(\partial_\mu \Psi_\ell)} q_\ell \Psi_\ell \,, \tag{10.4.1}$$

satisfying the conservation condition

$$\partial_\mu J^\mu = 0 \,. \tag{10.4.2}$$

This implies that the space-integral of the time component of J^μ is time-independent:

$$i \frac{d}{dt} Q = [Q, H] = 0 \,, \tag{10.4.3}$$

where

$$Q \equiv \int d^3x \, J^0 \,. \tag{10.4.4}$$

(There is a very important possible exception here, that the integral (10.4.4) may not exist if there are long-range forces due to massless scalars in the system. We will return to this point when we consider broken symmetries in Volume II.) Also, since it is a space-integral, Q is manifestly translation-invariant

$$[\mathbf{P}, Q] = 0 \tag{10.4.5}$$

and since J^μ is a four-vector, Q is invariant with respect to homogeneous Lorentz transformations

$$[J^{\mu\nu}, Q] = 0 \,. \tag{10.4.6}$$

It follows that Q acting on the true vacuum Ψ_0 must be another Lorentz-invariant state of zero energy and momentum, and hence (assuming no vacuum degeneracy) must be proportional to Ψ_0 itself. But the proportionality constant must vanish, because Lorentz invariance requires that $(\Psi_0, J_\mu \Psi_0)$ vanish. Hence

$$Q\Psi_0 = 0 \,. \tag{10.4.7}$$

Also, Q acting on any one-particle state $\Psi_{\mathbf{p},\sigma,n}$ must be another state with the same energy, momentum, and Lorentz transformation properties, and thus (assuming no degeneracy of one-particle states) must be proportional to the same one-particle state

$$Q\Psi_{\mathbf{p},\sigma,n} = q_{(n)} \Psi_{\mathbf{p},\sigma,n} \,. \tag{10.4.8}$$

The Lorentz invariance of Q ensures that the eigenvalue $q_{(n)}$ is independent of $\mathbf{p}$ and σ, depending only on the species of the particle. This eigenvalue is what is known as the electric charge (or whatever other quantum number

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

of which J^μ may be the current) of the one-particle state. To relate this to the q_ℓ parameters in the Lagrangian, we note that the canonical commutation relations give

$$\left[J^0(\mathbf{x},t),\ \Psi_\ell(\mathbf{y},t)\right] = -q_\ell \Psi_\ell(\mathbf{y},t)\delta^3(\mathbf{x}-\mathbf{y})\,, \tag{10.4.9}$$

or integrating over $\mathbf{x}$:

$$\left[Q, \Psi_\ell(y)\right] = -q_\ell \Psi_\ell(y)\,. \tag{10.4.10}$$

The same is true of any local function $F(y)$ of the fields and field derivatives and their adjoints, containing definite numbers of each:

$$\left[Q, F(y)\right] = -q_F F(y)\,, \tag{10.4.11}$$

where q_F is the sum of the q_ℓ for all fields and field derivatives in $F(y)$, minus the sum of the q_ℓ for all field adjoints and their derivatives. Taking the matrix element of this equation between a one-particle state and the vacuum, and using Eqs. (10.4.7) and (10.4.8), we have

$$\left(\Psi_0,\ F(y)\Psi_{\mathbf{p},\sigma,n}\right)(q_F - q_{(n)}) = 0\,. \tag{10.4.12}$$

Hence we must have

$$q_{(n)} = q_F \tag{10.4.13}$$

as long as

$$\left(\Psi_0,\ F(y)\,\Psi_{\mathbf{p},\sigma,n}\right) \neq 0\,. \tag{10.4.14}$$

As we saw in the previous section, Eq. (10.4.14) is the condition that assures that momentum space Green's functions involving F have poles corresponding to the one-particle state $\Psi_{\mathbf{p},\sigma,n}$. For a one-particle state corresponding to one of the fields in the Lagrangian we could take $F = \Psi_\ell$, in which case $q_F = q_\ell$, but our results here apply to general one-particle states, whether or not their fields appear in the Lagrangian.

This almost, but not quite, tells us that despite all the possible high-order graphs that affect the emission and absorption of photons by charged particles, the physical electric charge is just equal to a parameter q_ℓ appearing in the Lagrangian (or to a sum of such parameters, like q_F .) The qualification that has to be added here is that the requirement, that the Lagrangian be invariant under the transformations $\Psi_\ell \to \exp(iq_\ell\alpha)\Psi_\ell$, does nothing to fix the over-all scale of the quantities q_ℓ. The physical electric charges are those that determine the response of matter fields to a given *renormalized* electromagnetic field A^μ. That is, the scale of the q_ℓ is fixed by requiring that the renormalized electromagnetic field appears in the matter Lagrangian $\mathscr{L}_M$ in the linear combinations $[\partial_\mu - iq_\ell A_\mu]\Psi_\ell$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

so that the current J^μ is

$$J^\mu = \frac{\delta \mathscr{L}_M}{\delta A_\mu} . \tag{10.4.15}$$

But A^μ and q_ℓ are not the same as the 'bare electromagnetic field' A^μ_B and 'bare charges' $q_{B\ell}$ that appear in the Lagrangian when we write it in its simplest form

$$\mathscr{L} = -\tfrac{1}{4}(\partial_\mu A_{B\nu} - \partial_\nu A_{B\mu})(\partial^\mu A^\nu_B - \partial^\nu A^\mu_B) + \mathscr{L}_M\left(\Psi_\ell, [\partial_\mu - iq_{B\ell}A_{B\mu}]\Psi_\ell\right) . \tag{10.4.16}$$

The renormalized electromagnetic field (defined to have a complete propagator whose pole at $p^2 = 0$ has unit residue) is conventionally written in terms of A^μ_B as

$$A^\mu = Z_3^{-1/2} A^\mu_B , \tag{10.4.17}$$

so in order for the charge q_ℓ to characterize the response of the charged particles to a given renormalized electromagnetic field, we should define the renormalized charges by

$$q_\ell = \sqrt{Z_3}\, q_{B\ell} . \tag{10.4.18}$$

We see that the physical electric charge q of any particle is just proportional to a parameter q_B related to those appearing in the Lagrangian, with a proportionality constant $\sqrt{Z_3}$ that is the same for all particles. This helps us to understand how a particle like the proton, that is surrounded by a cloud of virtual mesons and other strongly interacting particles, can have the same charge as the positron, whose interactions are all much weaker. It is only necessary to assume that for some reason the charges $q_{B\ell}$ in the Lagrangian are equal and opposite for the electron and for those particles (two u quarks and one d quark) that make up the proton; the effect of higher-order corrections then appears solely in the *common* factor $\sqrt{Z_3}$.

In order for charge renormalization to arise only from radiative corrections to the photon propagator, there must be cancellations among the great variety of other radiative corrections to the propagators and electromagnetic vertices of the charged particles. We can see a little more deeply into the nature of these cancellations by making use of the celebrated relations between these charged particle propagators and vertices known as the *Ward identities.*

For instance, consider the Green's function for an electric current $J^\mu(x)$ together with a Heisenberg-picture Dirac field $\Psi_n(y)$ of charge q and its covariant adjoint $\bar{\Psi}_m(z)$. We define the electromagnetic vertex function Γ^μ

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

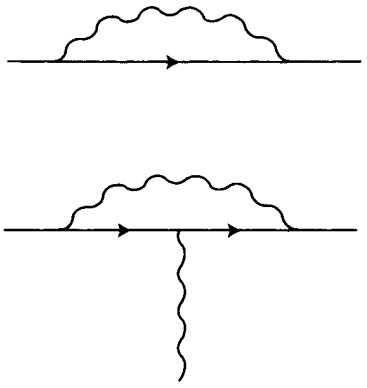

Figure 10.5. Diagrams for the first corrections to the electron propagator and vertex function in quantum electrodynamics. Here straight lines are electrons; wavy lines are photons.

of the charged particle by

$$\int d^4x\, d^4y\, d^4z\, e^{-ip\cdot x} e^{-ik\cdot y} e^{+i\ell\cdot z} \left(\Psi_0,\ T\left\{J^\mu(x)\,\Psi_n(y)\,\bar{\Psi}_m(z)\right\}\Psi_0\right)$$
$$\equiv -\,i(2\pi)^4 q S'_{nn'}(k)\Gamma^\mu_{n'm'}(k,\ell)S'_{m'm}(\ell)\,\delta^4(p+k-\ell)\,, \tag{10.4.19}$$

where

$$-\,i(2\pi)^4 S'_{nm}(k)\,\delta^4(k-\ell) \equiv \int d^4y\, d^4z\, \left(\Psi_0,\ T\left\{\Psi_n(y)\bar{\Psi}_m(z)\right\}\Psi_0\right) e^{-ik\cdot y} e^{+i\ell\cdot z}\,. \tag{10.4.20}$$

According to the theorem of Section 6.4, Eq. (10.4.20) gives the sum of all Feynman graphs with one incoming and one outgoing fermion line, i.e., the complete Dirac propagator. Also, Eq. (10.4.19) gives the sum of all such graphs with an extra photon line attached, so Γ^μ is the sum of "vertex" graphs with one incoming Dirac line, one outgoing Dirac line, and one photon line, but with the complete Dirac external line propagators and the bare photon external line propagator stripped away. To make the normalization of S' and Γ^μ perfectly clear, we mention that in the limit of no interactions, these functions take the values

$$S'(k) \to [i\gamma_\lambda k^\lambda + m - i\epsilon]^{-1}\,, \quad \Gamma^\mu(k,\ell) \to \gamma^\mu\,.$$

The one-loop diagrams that provide corrections to these limiting values are shown in Figure 10.5.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

We can derive a relation between Γ^μ and S' by use of the identity

$$\frac{\partial}{\partial x^\mu} T\left\{J^\mu(x)\Psi_n(y)\bar{\Psi}_m(z)\right\} = T\left\{\partial_\mu J^\mu(x)\Psi_n(y)\bar{\Psi}_m(z)\right\} + \delta(x^0-y^0)\, T\left\{\left[J^0(x),\Psi_n(y)\right]\bar{\Psi}_m(z)\right\} + \delta(x^0-z^0)\, T\left\{\Psi_n(y)\left[J^0(x),\bar{\Psi}_m(z)\right]\right\}, \tag{10.4.21}$$

where the delta functions arise from time-derivatives of step functions. The conservation condition (10.4.2) tells us that the first term vanishes, while the second and third terms can be calculated using the commutation relations (10.4.9), which here give

$$\left[J^0(\mathbf{x},t),\ \Psi_n(\mathbf{y},t)\right] = -q\Psi_n(\mathbf{y},t)\delta^3(\mathbf{x}-\mathbf{y}) \tag{10.4.22}$$

and its adjoint

$$\left[J^0(\mathbf{x},t),\ \bar{\Psi}_n(\mathbf{y},t)\right] = q\bar{\Psi}_n(\mathbf{y},t)\delta^3(\mathbf{x}-\mathbf{y})\,. \tag{10.4.23}$$

Eq. (10.4.21) then reads

$$\frac{\partial}{\partial x^\mu}\, T\left\{J^\mu(x)\Psi_n(y)\bar{\Psi}_m(z)\right\} = -q\,\delta^4(x-y)\, T\left\{\Psi_n(y)\bar{\Psi}_m(z)\right\} + q\,\delta^4(x-z)\, T\left\{\Psi_n(y)\bar{\Psi}_m(z)\right\}. \tag{10.4.24}$$

Inserting this in the Fourier transform (10.4.19) gives

$$(\ell-k)_\mu\, S'(k)\,\Gamma^\mu(k,\ell)\,S'(\ell) = i\,S'(\ell) \;-\; i\,S'(k)$$

or in other words

$$(\ell-k)_\mu\Gamma^\mu(k,\ell) = i\,S'^{-1}(k) - iS'^{-1}(\ell)\,. \tag{10.4.25}$$

This is known as the *generalized Ward identity*, first derived (by these methods) by Takahashi.[4] The original Ward identity, derived earlier by Ward[5] from a study of perturbation theory, can be obtained from Eq. (10.4.25) by letting ℓ approach k. In this limit, Eq. (10.4.25) gives

$$\Gamma^\mu(k,k) = -i\,\frac{\partial}{\partial k_\mu}\,S'^{-1}(k)\,. \tag{10.4.26}$$

The fermion propagator is related to the self-energy insertion $\Sigma^*(\not{k})$ by Eq. (10.3.28)

$$S'^{-1}(k) = i\not{k} + m - \Sigma^*(\not{k})\,,$$

so Eq. (10.4.26) may be written

$$\Gamma^\mu(k,k) = \gamma^\mu + i\,\frac{\partial}{\partial k_\mu}\,\Sigma^*(\not{k})\,. \tag{10.4.27}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

For a *renormalized* Dirac field, Eqs. (10.3.31) and (10.4.27) tell us that on the mass shell

$$\bar{u}'_k \, \Gamma^\mu(k,k) \, u_k = \bar{u}'_k \, \gamma^\mu \, u_k \, , \tag{10.4.28}$$

where $[i\gamma_\mu k^\mu + m]u_k = [i\gamma_\mu k^\mu + m]u'_k = 0$. Thus the renormalization of the fermion field ensures that the radiative corrections to the vertex function Γ_μ cancel when a fermion on the mass shell interacts with an electromagnetic field with zero momentum transfer, as is the case when we set out to measure the fermion's electric charge. If we had not used a renormalized fermion field then the corrections to the vertex function would have just cancelled the corrections due to radiative corrections to the external fermion lines, leaving the electric charge again unchanged.

10.5 Gauge Invariance

The conservation of electric charge may be used to prove a useful result for the quantities

$$M_{\beta\alpha}^{\mu\mu'\cdots}(q,q',\cdots) \equiv \int d^4x \int d^4x' \cdots e^{-iq\cdot x} e^{-iq'\cdot x'} \cdots \times \left(\Psi_\beta^-, \, T\left\{ J^\mu(x), J^{\mu'}(x') \cdots \right\} \Psi_\alpha^+ \right) . \tag{10.5.1}$$

In theories like spinor electrodynamics in which the electromagnetic interaction is linear in the field A^μ, this is the matrix element for emission (and/or absorption) of on- or off-shell photons having four-momenta q, q', etc. (and/or $-q$, $-q'$, etc.), with external line photon coefficient functions or propagators omitted, in an arbitrary transition $\alpha \to \beta$. Our result is that Eq. (10.5.1) vanishes when contracted with any one of the photon four-momenta:

$$q_\mu M_{\beta\alpha}^{\mu\mu'\cdots}(q,q',\cdots) = q'_{\mu'} M_{\beta\alpha}^{\mu\mu'\cdots}(q,q',\cdots) = \cdots = 0 \, . \tag{10.5.2}$$

Since M is defined symmetrically with respect to the photon lines, it will be sufficient to show the vanishing of the first of these quantities.

For this purpose, note that by an integration by parts

$$q_\mu \, M_{\beta\alpha}^{\mu\mu'\cdots}(q,q',\cdots) = -i \int d^4x \int d^4x' \cdots \times e^{-iq\cdot x} e^{-iq'\cdot x'} \cdots \left(\Psi_\beta^-, \frac{\partial}{\partial x^\mu} \, T\left\{ J^\mu(x), J^{\mu'}(x') \cdots \right\} \Psi_\alpha^+ \right) . \tag{10.5.3}$$

The electric current $J^\mu(x)$ is conserved, but this does not immediately imply that Eq. (10.5.3) vanishes, because we still have to take account

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

of the x^0-dependence contained in the theta functions that appear in the definition of the time-ordered product. For instance, for just two currents

$$T\left\{J^\mu(x)J^\nu(y)\right\} = \theta(x^0 - y^0)J^\mu(x)J^\nu(y) + \theta(y^0 - x^0)J^\nu(y)J^\mu(x)$$

so, taking account of the conservation of $J^\mu(x)$:

$$\begin{aligned}\frac{\partial}{\partial x^\mu}\, T\left\{J^\mu(x)\, J^\nu(y)\right\} &= \delta(x^0 - y^0)\, J^0(x)\, J^\nu(y) - \delta(y^0 - x^0)\, J^\nu(y)\, J^0(x) \\ &= \delta(x^0 - y^0)\left[J^0(x),\, J^\nu(y)\right] . \end{aligned} \tag{10.5.4}$$

With more than two currents, we get an equal-time commutator like this (inside the time-ordered product) for each current aside from $J^\mu(x)$ itself. To evaluate this commutator, we recall that (as shown in the previous section) for any product F of field operators and their adjoints and/or derivatives

$$\left[J^0(\vec{x}, t),\, F(\vec{y}, t)\right] = -q_F\, F(\vec{x}, t)\, \delta^3(\vec{x} - \vec{y}) ,$$

where q_F is the sum of the q_ℓs for the fields and field derivatives in F, minus the sum of the q_ℓs for the field adjoints and their derivatives. For the electric current, q_J is zero; $J^\nu(y)$ is itself an electrically neutral operator. It follows that

$$\left[J^0(\vec{x}, t),\, J^\nu(\vec{y}, t)\right] = 0 \tag{10.5.5}$$

and therefore Eq. (10.5.4) vanishes, so that Eq. (10.5.3) gives

$$q_\mu\, M^{\mu\mu'\cdots}_{\beta\alpha}(q, q', \cdots) = 0 \tag{10.5.6}$$

as was to be proved.

There is an important qualification here. In deriving Eq. (10.5.5) we should take into account the fact that a product of fields at the same spacetime point y like the current operator $J^\nu(y)$ can only be properly defined through some regularization procedure that deals with the infinities in such products. In many cases it turns out that there are non-vanishing contributions to the commutator of $J^0(\vec{x}, t)$ with the regulated current $J^i(\vec{y}, t)$, known as *Schwinger terms*.[6] Where the current includes terms arising from a charged scalar field Φ, there are additional regulator-independent Schwinger terms involving $\Phi^\dagger\Phi$. However, all these Schwinger terms are cancelled in multi-photon amplitudes by the contribution of additional interactions that are quadratic in the electromagnetic field, either arising from the regulator procedure (if gauge-invariant) or, as for charged scalars, directly from terms in the Lagrangian. We will be dealing mostly with charged spinor fields, and will use a regularization procedure (dimensional regularization) that does not lead to Schwinger terms, so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

in what follows we will ignore this issue and continue to use the naive commutation relation (10.5.5).

The same argument yields a result like Eq. (10.5.2) even if other particles besides photons are off the mass shell, provided that all *charged* particles are taken on the mass shell, i.e., kept in the states $\Psi_\beta{}^-$ and $\Psi_\alpha{}^+$. Otherwise, the left-hand side of Eq. (10.5.2) receives contributions from non-vanishing equal-time commutators, such as those we encountered in the derivation of the Ward identity in the previous section.

One consequence of Eq. (10.5.2) is that S-matrix elements are unaffected if we change any photon propagator $\Delta_{\mu\nu}(q)$ by

$$\Delta_{\mu\nu}(q) \to \Delta_{\mu\nu}(q) + \alpha_\mu q_\nu + q_\mu \beta_\nu \tag{10.5.7}$$

or if we change any photon polarization vector by

$$e_\rho(\mathbf{k},\lambda) \to e_\rho(\mathbf{k},\lambda) + c k_\rho \,, \tag{10.5.8}$$

where $k^0 \equiv |\mathbf{k}|$, and α_μ, β_ν, and c are entirely arbitrary (not necessarily constants, and not necessarily the same for all propagators or polarization vectors.) This is (somewhat loosely) called the gauge invariance of the S-matrix.

To prove this result it is only necessary to display the explicit dependence of the S-matrix on photon polarization vectors and propagators

$$\begin{aligned} S_{\beta\alpha} \propto \int & d^4q_1\, d^4q_2 \cdots \Delta_{\mu_1\nu_1}(q_1)\Delta_{\mu_2\nu_2}(q_2)\cdots \\ & \times e^*_{\rho_1}(\mathbf{k}'_1\lambda'_1) e^*_{\rho_2}(\mathbf{k}'_2\lambda'_2)\cdots e_{\sigma_1}(\mathbf{k}_1\lambda_1) e_{\sigma_2}(\mathbf{k}_2\lambda_2)\cdots \\ \times M_{ba}^{\mu_1\mu_2\cdots\nu_1\nu_2\cdots\rho_1\rho_2\cdots\sigma_1\sigma_2\cdots}(&-q_1,-q_2,\cdots,q_1,q_2,\cdots,-k'_1,-k'_2\cdots,k_1,k_2\cdots) \end{aligned} \tag{10.5.9}$$

where $M^{\rho\sigma\cdots}$ is the matrix element (10.5.1) calculated in the absence of electromagnetic interactions.* The invariance of Eq. (10.5.9) under the 'gauge transformations' (10.5.7) and (10.5.8) follows immediately from the conservation conditions (10.5.2). (In Section 9.6 we used the path-integral formalism to prove a special case of this theorem, that vacuum expectation values of time-ordered products of gauge-invariant operators are independent of the constant α in the propagator (9.6.21).) This result is not as elementary as it looks, as it applies not to individual diagrams, but only to sums of diagrams in which the current vertices are inserted in all possible places in the diagrams.

There is a particularly important application of Eq. (10.5.2) to the calculation of the photon propagator. The complete photon propagator,

* The states a and b are the same as α and β, but with photons deleted. Note that the arguments of M are all taken to be *incoming* four-momenta, which is why we have to insert various signs for some of the arguments of M in Eq. (10.5.9).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

conventionally called $\Delta'_{\mu\nu}(q)$, takes the form

$$\Delta'_{\mu\nu}(q) = \Delta_{\mu\nu}(q) + \Delta_{\mu\rho}(q)M^{\rho\sigma}(q)\Delta_{\sigma\nu}(q)\,, \tag{10.5.10}$$

where $M^{\rho\sigma}$ is proportional to the matrix element (10.5.1) with two currents and α and β both the vacuum state, and $\Delta_{\mu\nu}$ is the bare photon propagator, written here in a general Lorentz-invariant gauge as

$$\Delta_{\mu\nu}(q) \equiv \frac{\eta_{\mu\nu} - \xi(q^2)q_\mu q_\nu/q^2}{q^2 - i\epsilon}\,. \tag{10.5.11}$$

From Eq. (10.5.2) we have here $q^\mu M_{\mu\nu}(q) = 0$, so that

$$q^\mu \Delta'_{\mu\nu}(q) = q^\mu \Delta_{\mu\nu}(q) = \frac{q_\nu(1 - \xi(q^2))}{q^2 - i\epsilon}\,. \tag{10.5.12}$$

On the other hand, just as we did for scalar and spinor fields in Section 10.3, we may express the complete photon propagator in terms of a sum $\Pi^*(q)$ of graphs with two external photon lines that (unlike M) are one-photon-irreducible:

$$\begin{aligned}\Delta'(q) &= \Delta(q) + \Delta(q)\Pi^*(q)\Delta(q) + \Delta(q)\Pi^*(q)\Delta(q)\Pi^*(q)\Delta(q) + \cdots \\ &= [\Delta(q)^{-1} - \Pi^*(q)]^{-1}\end{aligned} \tag{10.5.13}$$

or in other words

$$\Delta'_{\mu\nu}(q) = \Delta_{\mu\nu}(q) + \Delta_{\mu\rho}(q)\Pi^{*\rho\sigma}(q)\Delta'_{\sigma\nu}(q)\,. \tag{10.5.14}$$

Then in order to satisfy Eq. (10.5.12), we must have

$$q_\rho \Pi^{*\rho\sigma}(q) = 0\,. \tag{10.5.15}$$

This together with Lorentz invariance tells us that $\Pi^*(q)$ must take the form

$$\Pi^{*\rho\sigma}(q) = (q^2\eta^{\rho\sigma} - q^\rho q^\sigma)\pi(q^2)\,. \tag{10.5.16}$$

Then Eq. (10.5.13) yields a complete propagator of the form

$$\Delta'_{\mu\nu}(q) = \frac{\eta_{\mu\nu} - \tilde{\xi}(q^2)q_\mu q_\nu/q^2}{[q^2 - i\epsilon][1 - \pi(q^2)]}\,, \tag{10.5.17}$$

where

$$\tilde{\xi}(q^2) = \xi(q^2)[1 - \pi(q^2)] + \pi(q^2)\,. \tag{10.5.18}$$

Now, because $\Pi^*_{\mu\nu}(q)$ receives contributions only from one-photon-irreducible graphs, it is expected not to have any pole at $q^2 = 0$. (There is an important exception in the case of broken gauge symmetry, to be discussed in Volume II.) In particular, the absence of poles at $q^2 = 0$ in the $q_\mu q_\nu$ term in $\Pi^*_{\mu\nu}(q)$ tells us that the function $\pi(q^2)$ in Eq. (10.5.16) also has no such pole, and so the pole in the complete propagator (10.5.17) is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

still at $q^2 = 0$, indicating that *radiative corrections do not give the photon a mass.*

For a renormalized electromagnetic field, radiative corrections should also not alter the gauge-invariant part of the residue of the photon pole in Eq. (10.5.17), so

$$\pi(0) = 0 \,. \tag{10.5.19}$$

This condition leads to a determination of the electromagnetic field renormalization constant Z_3. Recall that when expressed in terms of the renormalized field (10.4.17), the electrodynamic Lagrangian takes the form

$$\mathscr{L} = -\tfrac{1}{4}Z_3(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \mathscr{L}_M\left(\Psi_\ell, [\partial_\mu - iq_\ell A_\mu]\Psi_\ell\right) \,.$$

The function $\pi(q^2)$ in the one-photon-irreducible amplitude is then

$$\pi(q^2) = 1 - Z_3 + \pi_{\text{LOOP}}(q^2) \,, \tag{10.5.20}$$

where π_{LOOP} is the contribution of loop diagrams. It follows that

$$Z_3 = 1 + \pi_{\text{LOOP}}(0) \,. \tag{10.5.21}$$

In practice, we just calculate the loop contributions and subtract a constant in order to make $\pi(0)$ vanish.

Incidentally, Eq. (10.5.18) shows that for $q^2 \neq 0$ the gauge term in the photon propagator *is* altered by radiative corrections. The one exception is the case of Landau gauge, for which $\tilde{\xi} = \xi = 1$ for all q^2.

10.6 Electromagnetic Form Factors and Magnetic Moment

Suppose that we want to calculate the scattering of a particle by an external electromagnetic field (or by the electromagnetic field of another particle), to first order in this electromagnetic field, but to all orders in all other interactions (including electromagnetic) of our particle. For this purpose, we need to know the sum of the contributions of all Feynman diagrams with one incoming and one outgoing particle line, both on the mass shell, plus a photon line, which may be on or off the mass shell. According to the theorem of Section 6.4, this sum is given by the one-particle matrix element of the electromagnetic current $J^\mu(x)$. Let us see what governs the general form of this matrix element.

According to spacetime translation invariance, the one-particle matrix element of the electromagnetic current takes the form

$$\left(\Psi_{\mathbf{p}',\sigma'},\ J^\mu(x)\Psi_{\mathbf{p},\sigma}\right) = \exp(i(p - p')\cdot x)\left(\Psi_{\mathbf{p}',\sigma'}, J^\mu(0)\Psi_{\mathbf{p},\sigma}\right) \,. \tag{10.6.1}$$

The current conservation condition $\partial_\mu J^\mu = 0$ then requires

$$(p' - p)_\mu \left(\Psi_{\mathbf{p}',\sigma'},\ J^\mu(0)\Psi_{\mathbf{p},\sigma}\right) = 0 \,. \tag{10.6.2}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

Also, setting $\mu = 0$ and integrating over all $\mathbf{x}$ gives

$$\left(\Psi_{\mathbf{p}',\sigma'},\ Q\Psi_{\mathbf{p},\sigma}\right) = (2\pi)^3\delta^3(\mathbf{p}-\mathbf{p}')\left(\Psi_{\mathbf{p}',\sigma'},\ J^0(0)\Psi_{\mathbf{p},\sigma}\right) .$$

Using Eq. (10.4.8), this gives

$$\left(\Psi_{\mathbf{p},\sigma'},\ J^0(0)\Psi_{\mathbf{p},\sigma}\right) = (2\pi)^{-3}q\,\delta_{\sigma'\sigma} , \tag{10.6.3}$$

where q is the particle charge.

We also have at our disposal the constraints on the current matrix elements imposed by Lorentz invariance. To explore these, we will limit ourselves to the simplest cases: spin zero and spin $\frac{1}{2}$. The analysis presented here provides an example of techniques that are useful for other currents, such as those of the semi-leptonic weak interactions.

Spin Zero

For spin zero, Lorentz invariance requires the one-particle matrix element of the current to take the general form

$$\left(\Psi_{\mathbf{p}'},\ J^\mu(0)\Psi_{\mathbf{p}}\right) = q(2\pi)^{-3}(2p'^0)^{-1/2}(2p^0)^{-1/2}\mathcal{J}^\mu(p',p), \tag{10.6.4}$$

where p^0 and p'^0 are the mass-shell energies ($p^0 = \sqrt{\mathbf{p}^2+m^2}$), and $\mathcal{J}^\mu(p',p)$ is a four-vector function of the two four-vectors p'^μ and p^μ. (We have extracted a factor of the charge q of the particle from $\mathcal{J}$ for future convenience.) Obviously, the most general such four-vector function takes the form of a linear combination of p'^μ and p^μ, or equivalently of $p'^\mu + p^\mu$ and $p'^\mu - p^\mu$, with scalar coefficients. But the scalars p^2 and p'^2 are fixed at the values $p^2 = p'^2 = -m^2$, so the scalar variables that can be formed from p^μ and p'^μ can be taken as functions only of $p\cdot p'$, or equivalently of

$$k^2 \equiv (p-p')^2 = -2m^2 - 2p\cdot p' . \tag{10.6.5}$$

Thus the function $\mathcal{J}^\mu(p',p)$ must take the form

$$\mathcal{J}^\mu(p',p) = (p'+p)^\mu F(k^2) + i(p'-p)^\mu H(k^2) . \tag{10.6.6}$$

The fact that J^μ is Hermitian implies that $\mathcal{J}^\mu(p',p)^* = \mathcal{J}^\mu(p,p')$, so that both $F(k^2)$ and $H(k^2)$ are real.

Now $(p'-p)\cdot(p'+p)$ vanishes, while $(p'-p)^2 = k^2$ is not generally zero, so the condition of current conservation is simply

$$H(k^2) = 0 . \tag{10.6.7}$$

Also, setting $\mathbf{p}' = \mathbf{p}$ and $\mu = 0$ in Eq. (10.6.4), and comparing with Eq. (10.6.3), we find that

$$F(0) = 1 . \tag{10.6.8}$$

The function $F(k^2)$ is called the *electromagnetic form factor* of the particle.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

Spin $\frac{1}{2}$

For spin $\frac{1}{2}$, Lorentz invariance requires the one-particle matrix element of the current to take the general form

$$\left(\Psi_{\mathbf{p}',\sigma'},\, J^{\mu}(0)\Psi_{\mathbf{p},\sigma}\right) = iq\,(2\pi)^{-3}\,\bar{u}(\mathbf{p}',\sigma')\Gamma^{\mu}(p',p)u(\mathbf{p},\sigma) \tag{10.6.9}$$

where Γ^{μ} is a four-vector 4×4 matrix function of p^{ν}, p'^{ν}, and γ^{ν}, and u is the usual Dirac coefficient function. We have extracted a factor iq to make the normalization of Γ^{μ} the same as in the previous section.

Just as for any 4×4 matrix, we may expand Γ^{μ} in the 16 covariant matrices 1, γ_{ρ}, $[\gamma_{\rho},\gamma_{\sigma}]$, $\gamma_5\gamma_{\rho}$, and γ_5. The most general four-vector Γ^{μ} can therefore be written as a linear combination of

$$\begin{aligned}
1: &\quad p^{\mu},\ p'^{\mu}\\
\gamma_{\rho}: &\quad \gamma^{\mu},\ p^{\mu}\not{p},\ p'^{\mu}\not{p},\ p^{\mu}\not{p}',\ p'^{\mu}\not{p}'\\
[\gamma_{\rho},\gamma_{\sigma}]: &\quad [\gamma^{\mu},\not{p}],\ [\gamma^{\mu},\not{p}'],\ [\not{p},\not{p}']p^{\mu},\ [\not{p},\not{p}']p'^{\mu}\\
\gamma_5\gamma_{\rho}: &\quad \gamma_5\gamma_{\rho}\,\epsilon^{\rho\mu\nu\sigma}p_{\nu}p'_{\sigma}\\
\gamma_5: &\quad \text{NONE}
\end{aligned}$$

with the coefficient of each term a function of the only scalar variable in the problem, the quantity (10.6.5). This can be greatly simplified by using the Dirac equations satisfied by u and $\bar{u}$:

$$\bar{u}(\mathbf{p}',\sigma')\,(i\not{p}'+m)=0\,,\qquad (i\not{p}+m)\,u(\mathbf{p},\sigma)=0\,.$$

In consequence, we can drop* all but the first three entries: p^{μ}, p'^{μ}, and γ^{μ}. We conclude that, on the fermion mass shell, Γ^{μ} may be expressed as a linear combination of γ^{μ}, p^{μ}, and p'^{μ}, which we choose to write as

$$\begin{aligned}
\bar{u}(\mathbf{p}',\sigma')\Gamma^{\mu}(p',p)u(\mathbf{p},\sigma) = \bar{u}(\mathbf{p}',\sigma')\Big[&\gamma^{\mu}F(k^2)\\
&-\frac{i}{2m}\,(p+p')^{\mu}G(k^2)+\frac{(p-p')^{\mu}}{2m}H(k^2)\Big]u(\mathbf{p},\sigma).
\end{aligned}\tag{10.6.10}$$

---

* This is obvious for the terms $p^{\mu}\not{p}$, $p'^{\mu}\not{p}$, $p^{\mu}\not{p}'$, and $p'^{\mu}\not{p}'$, which may be replaced respectively with imp^{μ}, imp'^{μ}, imp^{μ}, and imp'^{μ}, which are the same as terms already on our list. Also, we can write

$$[\gamma^{\mu},\not{p}] = 2\gamma^{\mu}\not{p} - \{\gamma^{\mu},\not{p}\} = 2\gamma^{\mu}\not{p} - 2p^{\mu}\,,$$

which may be replaced with $2im\gamma^{\mu} - 2p^{\mu}$, a linear combination of terms already on our list. The same applies to $[\gamma^{\mu},\not{p}']$. Also,

$$[\not{p},\not{p}'] = -2\not{p}'\not{p} + \{\not{p},\not{p}'\} = -2\not{p}'\not{p} + 2p\cdot p'\,,$$

which may be replaced with $2m^2 + 2p\cdot p' = -k^2$. Hence the terms $[\not{p},\not{p}']p^{\mu}$ and $[\not{p},\not{p}']p'^{\mu}$ give nothing new. Finally, to deal with the last term we may use the relation

$$\gamma_5\gamma_{\rho}\,\epsilon^{\rho\mu\nu\sigma} = \tfrac{1}{6}i\left(\gamma^{\mu}\gamma^{\nu}\gamma^{\sigma}+\gamma^{\sigma}\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\sigma}\gamma^{\mu}-\gamma^{\nu}\gamma^{\mu}\gamma^{\sigma}-\gamma^{\mu}\gamma^{\sigma}\gamma^{\nu}-\gamma^{\sigma}\gamma^{\nu}\gamma^{\mu}\right).$$

Contracting this with p_{ν} and p'_{σ} and then moving all $\not{p}$ factors to the right and $\not{p}'$ factors to the left again gives a linear combination of p^{μ}, p'^{μ}, and γ^{μ}.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

The hermiticity of $J^\mu(0)$ implies that

$$\beta\,\Gamma^{\mu\dagger}(p',p)\beta = -\Gamma^\mu(p,p')\,, \tag{10.6.11}$$

so that $F(k^2)$, $G(k^2)$, and $H(k^2)$ must all be *real* functions of k^2.

The conservation condition (10.6.2) is automatically satisfied by the first two terms in Eq. (10.6.10), because

$$(p'-p)_\mu\gamma^\mu = -i\left[(i\not{p}'+m)-(i\not{p}+m)\right]$$

and

$$(p'-p)\cdot(p'+p) = p'^2 - p^2\,.$$

On the other hand, $(p'-p)^2$ does not in general vanish, so current conservation requires the third term to vanish

$$H(k^2) = 0\,. \tag{10.6.12}$$

Also, letting $\mathbf{p}' \to \mathbf{p}$ in Eqs. (10.6.9) and (10.6.10), we find

$$\left(\Psi_{\mathbf{p},\sigma'}J^\mu(0)\Psi_{\mathbf{p},\sigma}\right) = i\,q(2\pi)^{-3}\bar{u}(\mathbf{p},\sigma')\left[\gamma^\mu F(0) - \frac{i}{m}\,p^\mu G(0)\right]u(\mathbf{p},\sigma)\,.$$

Using the identity $\{\gamma^\mu, i\not{p}+m\} = 2m\gamma^\mu + 2ip^\mu$, we also have

$$\bar{u}(\mathbf{p},\sigma')\gamma^\mu u(\mathbf{p},\sigma) = -\frac{ip^\mu}{m}\,\bar{u}(\mathbf{p},\sigma')u(\mathbf{p},\sigma)\,.$$

Recall also that

$$\bar{u}(\mathbf{p},\sigma')u(\mathbf{p},\sigma) = \delta_{\sigma'\sigma}m/p^0$$

and therefore

$$\left(\Psi_{\mathbf{p},\sigma'},\, J^\mu(0)\Psi_{\mathbf{p},\sigma}\right) = q(2\pi)^{-3}(p^\mu/p^0)\delta_{\sigma'\sigma}\left[F(0)+G(0)\right]\,. \tag{10.6.13}$$

Comparing this with Eq. (10.6.3) yields the normalization condition

$$F(0)+G(0) = 1\,. \tag{10.6.14}$$

It may be useful to note that the electromagnetic vertex matrix Γ^μ is commonly written in terms of two other matrices, as

$$\begin{aligned}\bar{u}(\mathbf{p}',\sigma')\Gamma^\mu(p',p)u(\mathbf{p},\sigma) = \bar{u}(\mathbf{p}',\sigma')\,&\Big[\gamma^\mu F_1(k^2)\\ &+\ \tfrac{1}{2}i\,[\gamma^\mu,\gamma^\nu]\,(p'-p)_\nu F_2(k^2)\Big]\,u(\mathbf{p},\sigma)\,.\end{aligned} \tag{10.6.15}$$

As already mentioned, we may rewrite the matrix appearing in the second term in terms of those used in defining $F(k^2)$ and $G(k^2)$:

$$\begin{aligned}&\bar{u}(\mathbf{p}',\sigma')\ \tfrac{1}{2}i\,[\gamma^\mu,\gamma^\nu]\,(p'-p)_\nu\,u(\mathbf{p},\sigma)\\ &\quad= \bar{u}(\mathbf{p}',\sigma')\left[-i\not{p}'\gamma^\mu + \tfrac{1}{2}i\{\gamma^\mu,\not{p}'\} - i\gamma^\mu\not{p} + \tfrac{1}{2}i\{\gamma^\mu,\not{p}\}\right]u(\mathbf{p},\sigma)\\ &\quad= \bar{u}(\mathbf{p}',\sigma')\left[i(p'^\mu+p^\mu)+2m\gamma^\mu\right]u(\mathbf{p},\sigma)\,.\end{aligned} \tag{10.6.16}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

Comparing Eq. (10.6.15) with Eq. (10.6.10), we find

$$F(k^2) = F_1(k^2) + 2m F_2(k^2) \tag{10.6.17}$$

$$G(k^2) = -2m F_2(k^2) . \tag{10.6.18}$$

The normalization condition (10.6.14) now reads

$$F_1(0) = 1 .$$

In order to evaluate the magnetic moment of our particle in terms of its form factors, let us consider the spatial part of the vertex function in the case of small momenta, $|\mathbf{p}|, |\mathbf{p}'| \ll m$. For this purpose, it is useful to use Eq. (10.6.16) to rewrite Eq. (10.6.10) (with $H = 0$) in a third form:

$$\bar{u}(\mathbf{p}', \sigma')\Gamma^\mu(p', p)u(\mathbf{p}, \sigma) = \frac{-i}{2m}\bar{u}(\mathbf{p}', \sigma') \left[(p + p')^\mu\{F(k^2) + G(k^2)\} - \frac{1}{2}[\gamma^\mu, \gamma^\nu](p' - p)_\nu F(k^2)\right] u(\mathbf{p}, \sigma) . \tag{10.6.19}$$

For zero momenta the matrix elements of the commutators of Dirac matrices are given by (5.4.19) and (5.4.20) as

$$\bar{u}(0, \sigma')[\gamma^i, \gamma^j]u(0, \sigma) = 4i\epsilon_{ijk}\left(J_k^{(\frac{1}{2})}\right)_{\sigma', \sigma} , \qquad \bar{u}(0, \sigma')[\gamma^i, \gamma^0]u(0, \sigma) = 0 ,$$

where $\mathbf{J}^{(\frac{1}{2})} = \frac{1}{2}\boldsymbol{\sigma}$ is the angular momentum matrix for spin $\frac{1}{2}$. Hence to first order in the small momenta,

$$\bar{u}(\mathbf{p}', \sigma')\boldsymbol{\Gamma}(p', p)u(\mathbf{p}, \sigma) \to \frac{-i}{2m}(\mathbf{p} + \mathbf{p}')\delta_{\sigma', \sigma} + \frac{1}{m}[(\mathbf{p} - \mathbf{p}') \times \mathbf{J}^{(\frac{1}{2})}]_{\sigma'\sigma} F(0) . \tag{10.6.20}$$

In a very weak time-independent external vector potential $A(\mathbf{x})$ the matrix element of the interaction Hamiltonian $H' = -\int d^3x\, \mathbf{J}(\mathbf{x}) \cdot A(\mathbf{x})$ between one-particle states of small momentum is therefore

$$(\Psi_{\mathbf{p}', \sigma'}, H'\Psi_{\mathbf{p}, \sigma}) = \frac{-iqF(0)}{m(2\pi)^3} \int d^3x\, e^{i(\mathbf{p} - \mathbf{p}')\cdot\mathbf{x}} A(\mathbf{x}) \cdot [(\mathbf{p} - \mathbf{p}') \times \mathbf{J}^{(\frac{1}{2})}]_{\sigma'\sigma}$$

$$= -\frac{qF(0)}{m(2\pi)^3} \int d^3x\, e^{i(\mathbf{p} - \mathbf{p}')\cdot\mathbf{x}} (\mathbf{J}^{(\frac{1}{2})})_{\sigma'\sigma} \cdot \mathbf{B}(\mathbf{x}) , \tag{10.6.21}$$

where $\mathbf{B} = \nabla \times A$ is the magnetic field. Hence in the limit of a slowly varying weak magnetic field, the matrix element of the interaction Hamiltonian is

$$(\Psi_{\mathbf{p}', \sigma'}, H'\Psi_{\mathbf{p}, \sigma}) = -\frac{qF(0)}{m}(\mathbf{J}^{(\frac{1}{2})})_{\sigma'\sigma} \cdot \mathbf{B}\, \delta^3(\mathbf{p} - \mathbf{p}') . \tag{10.6.22}$$

The magnetic moment μ for an arbitrary particle of general spin j is defined by the statement that the matrix element of the interaction of the particle with a weak static slowly varying magnetic field is

$$(\Psi_{\mathbf{p}', \sigma'}, H'\Psi_{\mathbf{p}, \sigma}) = -\frac{\mu}{j}(\mathbf{J}^{(j)})_{\sigma'\sigma} \cdot \mathbf{B}\, \delta^3(\mathbf{p} - \mathbf{p}') . \tag{10.6.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

Hence Eq. (10.6.22) gives the magnetic moment of a particle of charge q, mass m, and spin $\frac{1}{2}$ as:

$$\mu = \frac{qF(0)}{2m} . \tag{10.6.24}$$

This contains as a special case the celebrated Dirac result[7] $\mu = q/2m$ for a spin $\frac{1}{2}$ particle without radiative corrections.

We mention without proof that the form factors $F(k^2)$ and $G(k^2)$ of the proton may be measured for $k^2 > 0$ by comparison of experimental data for electron–proton scattering with the Rosenbluth formula[8] for the laboratory frame differential cross-section:

$$\frac{d\sigma}{d\Omega} = \frac{e^4}{4(4\pi)^2 E_0^2} \frac{\cos^2(\theta/2)}{\sin^4(\theta/2)} \left[1 + \frac{2E_0}{m} \sin^2(\theta/2)\right]^{-1} \times \left\{ \left(F(k^2) + G(k^2)\right)^2 + \frac{k^2}{4m^2} \left(2F^2(k^2) \tan^2(\theta/2) + G^2(k^2)\right) \right\} ,$$

where E_0 is the energy of the incident electron (taken here with $E_0 \gg m_e$); θ is the scattering angle; and

$$k^2 = \frac{4E_0^2 \sin^2(\theta/2)}{1 + (2E_0/m) \sin^2(\theta/2)} .$$

10.7 The Källen–Lehmann Representation*

We saw in Section 10.2 that the presence of one-particle intermediate states leads to poles in Fourier transforms of matrix elements of time-ordered products, like (10.2.1). Multi-particle intermediate states lead to more complicated singularities, which are difficult to describe in general. But in the special case of a vacuum expectation value involving just two operators, we have a convenient representation that explicitly displays the analytic structure of the Fourier transform. In particular, this representation may be used for propagators, where the two operators are the fields of elementary particles. When combined with the positivity requirements of quantum mechanics, this representation yields interesting bounds on the asymptotic behavior of propagators and the magnitude of renormalization constants.

Consider a complex scalar Heisenberg-picture operator $\Phi(x)$, which may or may not be an elementary particle field. The vacuum expectation

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

value of a product $\Phi(x)\Phi^\dagger(y)$ may be expressed as

$$\langle\Phi(x)\Phi^\dagger(y)\rangle_0 = \sum_n \langle 0|\Phi(x)|n\rangle\ \langle n|\Phi^\dagger(y)|0\rangle\ , \tag{10.7.1}$$

where the sum runs over any complete set of states. (Here the sum over n includes integrals over continuous labels as well as sums over discrete labels.) Choosing these states as eigenstates of the momentum four-vector P^μ, translational invariance tells us that

$$\begin{aligned}\langle 0|\Phi(x)|n\rangle &= \exp(ip_n\cdot x)\langle 0|\Phi(0)|n\rangle\ ,\\ \langle n|\Phi^\dagger(y)|0\rangle &= \exp(-ip_n\cdot y)\langle n|\Phi^\dagger(0)|0\rangle\end{aligned} \tag{10.7.2}$$

and so

$$\langle\Phi(x)\Phi^\dagger(y)\rangle_0 = \sum_n \exp(ip_n\cdot(x-y))\,|\langle 0|\Phi(0)|n\rangle|^2\ . \tag{10.7.3}$$

It is convenient to rewrite this in terms of a *spectral function*. Note that the sum $\sum_n \delta^4(p-p_n)\,|\langle 0|\Phi(0)|n\rangle|^2$ is a scalar function of the four-vector p^μ, and therefore may depend only on p^2 and (for $p^2 \le 0$) on the step function $\theta(p^0)$. In fact, the intermediate states in Eq. (10.7.3) all have $p^2 \le 0$ and $p^0 > 0$, so this sum takes the form

$$\sum_n \delta^4(p-p_n)\,|\langle 0|\Phi(0)|n\rangle|^2 = (2\pi)^{-3}\,\theta(p^0)\,\rho(-p^2) \tag{10.7.4}$$

with $\rho(-p^2) = 0$ for $p^2 > 0$. (The factor $(2\pi)^{-3}$ is extracted from ρ for future convenience.) The spectral function $\rho(-p^2)$ is clearly real and positive. With this definition, we can rewrite Eq. (10.7.3) as

$$\begin{aligned}\langle\Phi(x)\Phi^\dagger(y)\rangle_0 &= (2\pi)^{-3}\int d^4p\ \exp[ip\cdot(x-y)]\ \theta(p^0)\,\rho(-p^2)\\ &= (2\pi)^{-3}\int d^4p\int_0^\infty d\mu^2\ \exp[ip\cdot(x-y)]\ \theta(p^0)\\ &\qquad\times\rho(\mu^2)\,\delta(p^2+\mu^2)\ .\end{aligned} \tag{10.7.5}$$

Interchanging the order of integration over p^μ and μ^2, this may be expressed as

$$\langle\Phi(x)\Phi^\dagger(y)\rangle_0 = \int_0^\infty d\mu^2\ \rho(\mu^2)\,\Delta_+(x-y;\mu^2)\ , \tag{10.7.6}$$

where Δ_+ is the familiar function

$$\Delta_+(x-y;\mu^2) \equiv (2\pi)^{-3}\int d^4p\ \exp[ip\cdot(x-y)]\ \theta(p^0)\,\delta(p^2+\mu^2)\ . \tag{10.7.7}$$

In just the same way, we can show that

$$\langle\Phi^\dagger(y)\Phi(x)\rangle_0 = \int_0^\infty d\mu^2\ \bar\rho(\mu^2)\,\Delta_+(y-x;\mu^2) \tag{10.7.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

with a second spectral function $\bar{\rho}(\mu^2)$ defined by

$$\sum_n \delta^4(p - p_n) \, |\langle n|\Phi(0)|0\rangle|^2 = (2\pi)^{-3} \, \theta(p^0) \, \bar{\rho}(-p^2) \, . \tag{10.7.9}$$

We now make use of the causality requirement, that the commutator $[\Phi(x), \Phi^\dagger(y)]$ must vanish for space-like separations $x - y$. The vacuum expectation value of the commutator is

$$\langle [\Phi(x), \Phi^\dagger(y)] \rangle_0 = \int_0^\infty d\mu^2 \, \left(\rho(\mu^2) \, \Delta_+(x - y; \mu^2) - \bar{\rho}(\mu^2) \, \Delta_+(y - x; \mu^2) \right) \, . \tag{10.7.10}$$

As noted in Section 5.2, for $x - y$ space-like the function $\Delta_+(x - y)$ does not vanish, but it does become *even*. In order for (10.7.10) to vanish for arbitrary space-like separations, it is thus necessary that

$$\rho(\mu^2) = \bar{\rho}(\mu^2) \, . \tag{10.7.11}$$

This is a special case of the CPT theorem, proved here without the use of perturbation theory; for whatever states with $p^2 = -\mu^2$ have the quantum numbers of the operator Φ, there must be corresponding states with $p^2 = -\mu^2$ that have the quantum numbers of the operator $\Phi^\dagger$.

Using Eq. (10.7.11), the vacuum expectation of the time-ordered product is

$$\langle T \left\{ \Phi(x) \Phi^\dagger(y) \right\} \rangle_0 = -i \int_0^\infty d\mu^2 \, \rho(\mu^2) \, \Delta_F(x - y; \mu^2) \, , \tag{10.7.12}$$

where $\Delta_F(x - y; \mu^2)$ is the Feynman propagator for a spinless particle of mass μ:

$$-i\Delta_F(x - y; \mu^2) \equiv \theta(x^0 - y^0) \Delta_+(x - y; \mu^2) - \theta(y^0 - x^0) \Delta_+(y - x; \mu^2) \, . \tag{10.7.13}$$

Borrowing the notation introduced in Section 10.3 for complete propagators, we introduce the momentum space function

$$-i\Delta'(p) \equiv \int d^4x \, \exp[-ip \cdot (x - y)] \, \langle T \left\{ \Phi(x) \Phi^\dagger(y) \right\} \rangle_0 \, . \tag{10.7.14}$$

Recall that

$$\int d^4x \, \exp[-ip \cdot (x - y)] \, \Delta_F(x - y; \mu^2) = \frac{1}{p^2 + \mu^2 - i\epsilon} \, . \tag{10.7.15}$$

This yields our spectral representation:[9]

$$\Delta'(p) = \int_0^\infty \rho(\mu^2) \, \frac{d\mu^2}{p^2 + \mu^2 - i\epsilon} \, . \tag{10.7.16}$$

One immediate consequence of this result and the positivity of $\rho(\mu^2)$ is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

that $\Delta'(p)$ cannot vanish for $|p^2| \to \infty$ faster** than the bare propagator $1/(p^2 + m^2 - i\epsilon)$. From time to time the suggestion is made to include higher derivative terms in the unperturbed Lagrangian, which would make the propagator vanish faster than $1/p^2$ for $|p^2| \to \infty$, but the spectral representation shows that this would necessarily entail a departure from the positivity postulates of quantum mechanics.

We can use the spectral representation together with equal-time commutation relations to derive an interesting sum rule for the spectral function. If $\Phi(x)$ is a conventionally normalized (not renormalized) canonical field operator, then

$$\left[\frac{\partial\Phi(\mathbf{x},t)}{\partial t}, \Phi^\dagger(\mathbf{y},t)\right] = -i\delta^3(\mathbf{x}-\mathbf{y}) . \tag{10.7.17}$$

We note that

$$\frac{\partial}{\partial x^0}\Delta_+(x-y)\Big|_{x^0=y^0} = -i\delta^3(\mathbf{x}-\mathbf{y})$$

so the spectral representation (10.7.10) and the commutation relations (10.7.17) together tell us that

$$\int_0^\infty \rho(\mu^2)\, d\mu^2 = 1 . \tag{10.7.18}$$

This implies that for $|p^2| \to \infty$, the momentum space propagator (10.7.16) of the unrenormalized fields has the free-field asymptotic behavior

$$\Delta'(p) \to \frac{1}{p^2} .$$

This result is only meaningful within a suitable scheme for regulating ultraviolet divergences; in perturbation theory the unrenormalized fields have infinite matrix elements, and their propagator is ill-defined.

Now consider the possibility that there is a one-particle state $|\mathbf{k}\rangle$ of mass m with a non-vanishing matrix element with the state $\langle 0|\Phi(0)$. Lorentz

** In fact, it is not even certain that $\Delta'(p)$ vanishes for $|p^2| \to \infty$ at all, even though this would seem to follow from the spectral representation. The problem arises from the interchange of the integrals over p^μ and μ^2. What is certain is that $\Delta'(p)$ is an analytic function of $-p^2$ with a discontinuity across the positive real axis $-p^2 = \mu^2$ given by $\pi\rho(\mu^2)$, as can be shown by the methods of the next section. From this, it follows that $\Delta'(p)$ is given by a dispersion relation with spectral function $\rho(\mu^2)$ and possible subtractions:

$$\Delta'(p) = P(p^2) + (-p^2 + \mu_0^2)^n \int_0^\infty \frac{\rho(\mu^2)}{(\mu^2+\mu_0^2)^n} \frac{d\mu^2}{p^2+\mu^2-i\epsilon} ,$$

where n is a positive integer, μ_0^2 is an arbitrary positive constant, and $P(p^2)$ is a μ_0^2-dependent polynomial in p^2 of order $n-1$ that is absent for $n = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

invariance requires this matrix element to take the form

$$\langle 0|\Phi(0)|\mathbf{k}\rangle = (2\pi)^{-3/2}\left(2\sqrt{\mathbf{k}^2+m^2}\right)^{-1/2} N\,, \tag{10.7.19}$$

where N is a constant. According to the general results of Section 10.3, the propagator $\Delta'(p)$ of the unrenormalized fields should have a pole at $p^2 \to -m^2$ with residue $Z \equiv |N|^2 > 0$. That is,

$$\rho(\mu^2) = Z\delta(\mu^2 - m^2) + \sigma(\mu^2)\,, \tag{10.7.20}$$

where $\sigma(\mu^2) \geq 0$ is the contribution of multi-particle states. Together with Eq. (10.7.18), this has the consequence that

$$1 = Z + \int_0^\infty \sigma(\mu^2)\, d\mu^2 \tag{10.7.21}$$

and so

$$Z \leq 1 \tag{10.7.22}$$

with the equality reached only for a free particle, for which $\langle 0|\Phi(x)$ has no matrix elements with multi-particle states.

Because Z is positive, Eq. (10.7.21) can also be regarded as providing an upper bound on the coupling of the field Φ to multi-particle states:

$$\int_0^\infty \sigma(\mu^2)\, d\mu^2 \leq 1 \tag{10.7.23}$$

with the equality reached for $Z = 0$. The limit $Z = 0$ has an interesting interpretation as a condition for a particle to be composite rather than elementary.[10] In this context, a 'composite' particle may be understood to be one whose field does not appear in the Lagrangian. Consider such a particle, say a neutral particle of spin zero, and suppose that its quantum numbers allow it to be destroyed by an operator $F(\Psi)$ constructed out of other fields. We can freely introduce a field Φ for this particle by adding a term to the Lagrangian density of the form[†] $\Delta\mathscr{L} = (\Phi - F(\Psi))^2$, because the path integral over Φ can be done by setting it equal to the stationary point $\Phi = F(\Psi)$, at which $\Delta\mathscr{L} = 0$. But suppose instead we write $\Delta\mathscr{L} = \Delta\mathscr{L}_0 + \Delta\mathscr{L}_1$, where $\Delta\mathscr{L}_0 \equiv -\frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}m^2\Phi^2$ is the usual free-field Lagrangian, and treat $\Delta\mathscr{L}_1 \equiv \Delta\mathscr{L} - \Delta\mathscr{L}_0$ as an interaction. A term $\frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi$ in the interaction is nothing new. We encountered such a term in Eq. (10.3.12), multiplied by a factor $(1 - Z)$; the only new thing is that here $Z = 0$. Instead of adjusting Z to satisfy the field renormalization condition $\Pi^{*\prime}(0) = 0$, here we must regard this as a condition on the

[†] This is known in condensed matter physics as a 'Hubbard–Stratonovich transformation'.[11] It will be used to introduce fields for pairs of electrons in our discussion of superconductivity in Volume II.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

coupling constants of the composite particle. Unfortunately, it has not been possible to implement this procedure in quantum field theories, because as we have seen $Z = 0$ means that the particle couples as strongly as possible to its constituents, and this rules out the use of perturbation theory. The condition $Z = 0$ does prove useful in non-relativistic quantum mechanics; for instance, it fixes the coupling of the deuteron to the neutron and proton.[12]

Although the spectral representation has been derived here only for a spinless field, it is easy to generalize these results to other fields. Indeed, in the next chapter we shall show that to order e^2, the Z-factor for the electromagnetic field (conventionally called Z_3) is given by

$$Z_3 = 1 - \frac{e^2}{12\pi^2} \ln\left(\frac{\Lambda^2}{m_e^2}\right)$$

(where $\Lambda \gg m_e$ is an ultraviolet cutoff), in agreement with the bound (10.7.22).

10.8 Dispersion Relations*

The failure of early attempts to apply perturbative quantum field theory to the strong and weak nuclear forces had led theorists by the late 1950s to attempt the use of the analyticity and unitarity of scattering amplitudes as a way of deriving general non-perturbative results that would not depend on any particular field theory. This started with a revival of interest in dispersion relations. In its original form,[13] a dispersion relation was a formula giving the real part of the index of refraction in terms of an integral over its imaginary part. It was derived from an analyticity property of the index of refraction as a function of frequency, which followed from the condition that electromagnetic signals in a medium cannot travel faster than light in a vacuum. By expressing the index of refraction in terms of the forward photon scattering amplitude, the dispersion relation could be rewritten as a formula for the real part of the forward scattering amplitude as an integral of its imaginary part, and hence via unitarity in terms of the total cross-section. One of the exciting things about this relation was that it provided an alternative to conventional perturbation theory; given the scattering amplitude to order e^2, one could calculate the cross-section and the imaginary part of the scattering amplitude to order e^4, and then use the dispersion relation to

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

calculate the real part of the forward scattering amplitude to this order, without having ever to calculate a loop graph.

The modern approach to dispersion relations began in 1954 with the work of Gell-Mann, Goldberger, and Thirring.[14] Instead of considering the propagation of light in a medium, they derived the analyticity of the scattering amplitude directly from the condition of microscopic causality, which states that commutators of field operators vanish when the points at which the operators are evaluated are separated by a space-like interval. This approach allowed Goldberger[15] soon thereafter to derive a very useful dispersion relation for the forward pion–nucleon scattering amplitude.

To see how to use the principle of microscopic causality, consider the forward scattering in the laboratory frame of a massless boson of any spin on an arbitrary target α of mass $m_\alpha > 0$ and $\mathbf{p}_\alpha = 0$. (This has important applications to the scattering not only of photons but also pions in the limit $m_\pi = 0$, to be discussed in Volume II.) By a repeated use of Eq. (10.3.4) or the Lehmann–Symanzik–Zimmerman theorem,[3] the S-matrix element here is

$$S = \frac{1}{(2\pi)^3\sqrt{4\omega\omega'|N|^2}} \lim_{k^2\to 0} \lim_{k'^2\to 0} \times \int d^4x \int d^4y\, e^{-ik'\cdot y} e^{ik\cdot x}(i\Box_y)(i\Box_x)\langle\alpha|T\{A^\dagger(y), A(x)\}|\alpha\rangle\,. \quad (10.8.1)$$

Here k and k' are the initial and final boson four-momenta, with $\omega = k^0$, $\omega' = k'^0$; $A(x)$ is any Heisenberg-picture operator with a non-vanishing matrix element $\langle \mathrm{VAC}|A(x)|k\rangle = (2\pi)^{-3/2}(2\omega)^{-1/2}Ne^{ik\cdot x}$ between the one-boson state $|k\rangle$ and the vacuum; and N is the constant in this matrix element. In photon scattering $A(x)$ would be one of the transverse components of the electromagnetic field, while for massless pion scattering it would be a pseudoscalar function of hadron fields. The differential operators $-i\Box_x$ and $-i\Box_y$ are inserted to supply factors of ik'^2 and ik^2 that are needed to cancel the external line boson propagators. Letting these operators act on $A^\dagger(y)$ and $A(x)$, we have

$$S = \frac{-1}{(2\pi)^3\sqrt{4\omega\omega'|N|^2}} \lim_{k^2\to 0} \lim_{k'^2\to 0} \times \int d^4x \int d^4y\, e^{-ik'\cdot y} e^{ik\cdot x}\langle\alpha|T\{J^\dagger(y), J(x)\}|\alpha\rangle + \mathrm{ETC}\,, \quad (10.8.2)$$

where $J(x) \equiv \Box_x A(x)$, and 'ETC' denotes the Fourier transform of equal time commutator terms arising from the derivative acting on the step functions in the time-ordered product. The commutators of operators like $A(x)$ and $A^\dagger(y)$ (or their derivatives) vanish for $x^0 = y^0$ unless $\mathbf{x} = \mathbf{y}$, so the 'ETC' term is the Fourier transform of a differential operator acting on $\delta^4(x - y)$, and is hence a polynomial function of the boson

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

four-momenta. We are concerned here with the analytic properties of the S-matrix element, so the details of this polynomial will be irrelevant.

Using translation invariance, Eq. (10.8.2) gives the S-matrix element as $S = -2\pi i\delta^4(k' - k)M(\omega)$, where

$$M(\omega) = \frac{-i}{2\omega|N|^2} F(\omega)\,, \tag{10.8.3}$$

$$F(\omega) \equiv \int d^4x\, e^{i\omega\ell\cdot x}\langle\alpha|T\{J^\dagger(0)\,,\, J(x)\}|\alpha\rangle \;+\; \text{ETC}\,, \tag{10.8.4}$$

it now being understood that $k^\mu = \omega\ell^\mu$, where ℓ is a fixed four-vector with $\ell^\mu\ell_\mu = 0$ and $\ell^0 = 1$.

The time-ordered product can be rewritten in terms of commutators in two different ways:

$$\begin{aligned} T\{J^\dagger(0), J(x)\} &= \theta(-x^0)[J^\dagger(0), J(x)] + J(x)J^\dagger(0) \\ &= -\theta(x^0)[J^\dagger(0), J(x)] + J^\dagger(0)J(x)\,. \end{aligned} \tag{10.8.5}$$

Correspondingly, we can write

$$F(\omega) = F_A(\omega) + F_+(\omega) = F_R(\omega) + F_-(\omega)\,, \tag{10.8.6}$$

where

$$F_A(\omega) \equiv \int d^4x\, \theta(-x^0)\, \langle\alpha|[J^\dagger(0)\,,\, J(x)]|\alpha\rangle\, e^{i\omega\ell\cdot x} \;+\; \text{ETC}\,, \tag{10.8.7}$$

$$F_R(\omega) \equiv -\int d^4x\, \theta(x^0)\, \langle\alpha|[J^\dagger(0)\,,\, J(x)]|\alpha\rangle\, e^{i\omega\ell\cdot x} \;+\; \text{ETC}\,, \tag{10.8.8}$$

$$F_+(\omega) \equiv \int d^4x\, \langle\alpha|J(x)\, J^\dagger(0)|\alpha\rangle\, e^{i\omega\ell\cdot x}\,, \tag{10.8.9}$$

$$F_-(\omega) \equiv \int d^4x\, \langle\alpha|J^\dagger(0)\, J(x)|\alpha\rangle\, e^{i\omega\ell\cdot x}\,. \tag{10.8.10}$$

Microscopic causality tells us that the integrands in (10.8.7) and (10.8.8) vanish unless x^μ is within the light cone, and the step functions then require that x^μ is in the backward light cone in (10.8.7), so that $x\cdot\ell > 0$, and in the forward light cone in Eq. (10.8.8), so that $x\cdot\ell < 0$. We conclude that $F_A(\omega)$ is analytic for $\text{Im}\,\omega > 0$ and $F_R(\omega)$ is analytic for $\text{Im}\,\omega < 0$, because in both cases the factor $e^{i\omega\ell\cdot x}$ provides a cutoff for the integral over x^μ. (Recall that the 'ETC' term is a polynomial, and hence analytic at all finite points.) We may then define a function

$$\mathscr{F}(\omega) \equiv \begin{cases} F_A(\omega) & \text{Im}\,\omega > 0 \\ F_R(\omega) & \text{Im}\,\omega < 0 \end{cases} \tag{10.8.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

which is analytic in the whole complex ω plane, except for a cut on the real axis.

We can now derive the dispersion relation. According to Eq. (10.8.6), the discontinuity of $\mathscr{F}(\omega)$ across the cut at any real E is

$$\mathscr{F}(E+i\epsilon)-\mathscr{F}(E-i\epsilon)=F_A(E)-F_R(E)=F_-(E)-F_+(E)\,. \quad (10.8.12)$$

If $\mathscr{F}(\omega)/\omega^n$ vanishes as $|\omega|\to\infty$ in the upper or lower half-plane, then by dividing by any polynomial $P(\omega)$ of order n we obtain a function that vanishes for $|\omega|\to\infty$ and is analytic except for the cut on the real axis and poles at the zeroes ω_ν of $P(\omega)$. (Where $\mathscr{F}(\omega)$ itself vanishes as $|\omega|\to\infty$, we can take $P(\omega)=1$.) According to the method of residues, we then have

$$\frac{\mathscr{F}(\omega)}{P(\omega)}+\sum_\nu\frac{\mathscr{F}(\omega_\nu)}{(\omega_\nu-\omega)\,P'(\omega_\nu)}=\frac{1}{2\pi i}\oint_C\frac{\mathscr{F}(z)\,dz}{(z-\omega)\,P(z)}\,, \quad (10.8.13)$$

where ω is any point off the real axis, and C is a contour consisting of two segments: one running just above the real axis from $-\infty+i\epsilon$ to $+\infty+i\epsilon$ and then around a large semi-circle back to $-\infty+i\epsilon$, and the other just below the real axis from $+\infty-i\epsilon$ to $-\infty-i\epsilon$ and then around a large semi-circle back to $+\infty-i\epsilon$. Because the function $\mathscr{F}(z)/P(z)$ vanishes for $|z|\to\infty$, we can neglect the contribution from the large semi-circles. Using Eq. (10.8.12), Eq. (10.8.13) becomes

$$\mathscr{F}(\omega)=Q(\omega)+\frac{P(\omega)}{2\pi i}\int_{-\infty}^{+\infty}\frac{F_-(E)-F_+(E)}{(E-\omega)\,P(E)}\,dE\,, \quad (10.8.14)$$

where $Q(\omega)$ is the $(n-1)$th-order polynomial

$$Q(\omega)\equiv -P(\omega)\sum_\nu\frac{\mathscr{F}(\omega_\nu)}{(\omega_\nu-\omega)\,P'(\omega_\nu)}\,.$$

A dispersion relation of this form, with $P(\omega)$ and $Q(\omega)$ of order n and $n-1$ respectively, is said to have n *subtractions*. If we can take $P=1$ then $Q=0$, and the dispersion relation is said to be unsubtracted.

If we now let ω approach the real axis from above, Eq. (10.8.14) gives

$$F_A(\omega)=Q(\omega)+\frac{P(\omega)}{2\pi i}\int_{-\infty}^{+\infty}\frac{F_-(E)-F_+(E)}{(E-\omega-i\epsilon)\,P(E)}\,dE\,. \quad (10.8.15)$$

Recalling Eqs. (10.8.6) and (3.1.25), this is

$$F(\omega)=Q(\omega)+\frac{1}{2}F_-(\omega)+\frac{1}{2}F_+(\omega)+\frac{P(\omega)}{2\pi i}\int_{-\infty}^{+\infty}\frac{F_-(E)-F_+(E)}{(E-\omega)\,P(E)}\,dE \quad (10.8.16)$$

with $1/(E-\omega)$ now interpreted as the principal value function $\mathscr{P}/(E-\omega)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

This result is useful because the functions $F_{\pm}(E)$ may be expressed in terms of measurable cross-sections. Summing over a complete set of multi-particle intermediate states β in Eqs. (10.8.9) and (10.8.10) (including integrations over the momenta of the particles in β) and using translation invariance again, we have

$$F_{+}(E) = (2\pi)^4 \sum_{\beta} \left| \langle \beta | J(0)^{\dagger} | \alpha \rangle \right|^2 \delta^4(-p_\alpha + E\ell + p_\beta) , \tag{10.8.17}$$

$$F_{-}(E) = (2\pi)^4 \sum_{\beta} |\langle \beta | J(0) | \alpha \rangle|^2 \delta^4(p_\alpha + E\ell - p_\beta) . \tag{10.8.18}$$

But the matrix elements for the absorption of the massless scalar boson B in $B + \alpha \to \beta$ or its antiparticle B^c in $B^c + \alpha \to \beta$ are

$$-2i\pi M_{B^c+\alpha\to\beta} = \frac{(2\pi)^4}{(2\pi)^{3/2}\sqrt{2E_{B^c}N}} \langle \beta | J^{\dagger}(0) | \alpha \rangle , \tag{10.8.19}$$

$$-2i\pi M_{B+\alpha\to\beta} = \frac{(2\pi)^4}{(2\pi)^{3/2}\sqrt{2E_{B}N}} \langle \beta | J(0) | \alpha \rangle . \tag{10.8.20}$$

Comparing with Eq. (3.4.15), we see that $F_{\pm}(E)$ may be expressed in terms of total cross-sections** at energies $\mp E$:

$$F_{+}(E) = \theta(-E) \frac{2|E||N|^2}{(2\pi)^3} \sigma_{\alpha+B^c}(|E|) , \tag{10.8.21}$$

$$F_{-}(E) = \theta(E) \frac{2E|N|^2}{(2\pi)^3} \sigma_{\alpha+B}(E) . \tag{10.8.22}$$

The scattering amplitude (10.8.3) is now, for real $\omega > 0$,

$$M(\omega) = \frac{-iQ(\omega)}{2\omega|N|^2} - \frac{i}{2(2\pi)^3}\sigma_{\alpha+B}(\omega) - \frac{P(\omega)}{\omega(2\pi)^4} \int_0^{\infty} \left[\frac{\sigma_{\alpha+B}(E)}{(E-\omega)P(E)} + \frac{\sigma_{\alpha+B^c}(E)}{(E+\omega)P(-E)} \right] E\, dE . \tag{10.8.23}$$

It is more usual to express this dispersion relation in terms of the amplitude $f(\omega)$ for forward scattering in the laboratory frame, defined so that the laboratory frame differential cross-section in the forward direction is $|f(\omega)|^2$. This amplitude is given in terms of $M(\omega)$ by

** In some cases where selection rules allow the transition $\alpha \to \alpha + B$ and $\alpha \to \alpha + B^c$, the functions $F_{\pm}(E)$ also contain terms proportional to $\delta(E)$ arising from the contribution of the one-particle state α in the sum over intermediate states β. This does not occur for transversely polarized photons, or for pseudoscalar pions in the limit $m_\pi \to 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

$f(\omega) = -4\pi^2\omega\, M(\omega) = 2\pi^2 i F(\omega)/|N|^2$, so Eq. (10.8.23) now reads

$$f(\omega) = R(\omega) + \frac{i\omega}{4\pi}\sigma_{\alpha+B}(\omega) + \frac{P(\omega)}{4\pi^2}\int_0^\infty \left[\frac{\sigma_{\alpha+B}(E)}{(E-\omega)P(E)} + \frac{\sigma_{\alpha+B^c}(E)}{(E+\omega)P(-E)}\right] E\, dE\,,$$

where $R(\omega) \equiv 2i\pi^2\, Q(\omega)/|N|^2$. The optical theorem (3.6.4) tells us that the second term on the right-hand side equals $i\mathrm{Im}\, f(\omega)$, so this can just as well be written in the more conventional form

$$\mathrm{Re}\, f(\omega) = R(\omega) + \frac{P(\omega)}{4\pi^2}\int_0^\infty \left[\frac{\sigma_{\alpha+B}(E)}{(E-\omega)P(E)} + \frac{\sigma_{\alpha+B^c}(E)}{(E+\omega)P(-E)}\right] E\, dE\,, \quad (10.8.24)$$

In particular, we see that $R(\omega)$ is real if we choose $P(\omega)$ real.

The forward scattering amplitude also satisfies an important symmetry condition. By changing the integration variable in Eqs. (10.8.7) and (10.8.8) from x to $-x$ and then using the translation-invariance property

$$\langle\alpha|[J^\dagger(0)\,,\, J(-x)]|\alpha\rangle = \langle\alpha|[J^\dagger(x)\,,\, J(0)]|\alpha\rangle$$

we see that for $\mathrm{Im}\,\omega \leq 0$, $F_A(-\omega)$ is the same as $F_R(\omega)$, except for an interchange of J with $J^\dagger$. That is,

$$F_A(-\omega) = F_R^c(\omega) \quad \text{for} \quad \mathrm{Im}\,\omega \leq 0\,,$$

where a superscript c indicates that the amplitude refers to the scattering of the antiparticle B^c on α. (We leave it to the reader to show that this relation is not upset by the equal-time commutator terms in Eqs. (10.8.7) and (10.8.8).) In the same way, we find

$$F_R(-\omega) = F_A^c(\omega) \quad \text{for} \quad \mathrm{Im}\,\omega \geq 0\,,$$

and for real ω

$$F_\pm(-\omega) = F_\mp(\omega)\,.$$

Using these relations in (10.8.6), and recalling that $f(\omega)$ is proportional to $F(\omega)$, we find the *crossing symmetry* relation, that for real ω

$$f(-\omega) = f^c(\omega). \quad (10.8.25)$$

We are free to take $P(\omega)$ as any polynomial of sufficiently high order, but $R(\omega)$ then depends not only on $P(\omega)$ but also on the values of $\mathscr{F}(\omega)$ at the zeroes of $P(\omega)$. For $P(\omega)$ real and of nth order, the only free parameters in Eq. (10.8.16) are the n real coefficients in the real $(n-1)$th order polynomial $R(\omega)$. Hence Eq. (10.8.16) contains just n unknown real independent constants, the coefficients in the polynomial $R(\omega)$ for a given $P(\omega)$. We therefore wish to take the order n of the otherwise arbitrary polyomial $P(\omega)$ to be as small as possible.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

We might try taking $P(\omega) = 1$, but this doesn't work. The analysis of Section 3.7 suggests that the forward scattering amplitude should grow like ω or perhaps as fast as $\omega \ln^2 \omega$. In this case for $f(\omega)/P(\omega)$ to vanish as $\omega \to \infty$, it is sufficient to take $P(\omega)$ as a second order polynomial, so that $R(\omega)$ is linear in ω. Choosing $P(E) = E^2$ for convenience, Eq. (10.8.24) then becomes

$$\begin{aligned} \mathrm{Re}\, f(\omega) &= a + b\omega \\ &\quad + \frac{\omega^2}{4\pi^2} \int_0^\infty \left[\frac{\sigma_{\alpha+B}(E)}{(E-\omega)} + \frac{\sigma_{\alpha+B^c}(E)}{(E+\omega)} \right] \frac{dE}{E} , \end{aligned} \tag{10.8.26}$$

with a and b unknown real constants. The crossing symmetry condition (10.8.25) tells us that the corresponding constants in the dispersion relation for the antiparticle scattering amplitude $f^c(\omega)$ are

$$a^c = a , \qquad b^c = -b . \tag{10.8.27}$$

If we assume for instance that the cross-sections $\sigma_{\alpha+B}(E)$ and $\sigma_{\alpha+B^c}(E)$ behave for $E \to \infty$ as different constants times $(\ln E)^r$, then (10.8.26) would give

$$\mathrm{Re}\, f(\omega) \sim [\sigma_{\alpha+B}(\omega) - \sigma_{\alpha+B^c}(\omega)]\omega \ln \omega \sim \omega(\ln \omega)^{r+1} \tag{10.8.28}$$

so the real part of the scattering amplitude would grow faster than the imaginary part by a factor $\ln \omega$. This is implausible; we saw in Section 3.7 that the real part of the forward scattering amplitude is expected to become much smaller than the imaginary part for $\omega \to \infty$, as confirmed by experiment. We conclude that if $\sigma_{\alpha+B}(E)$ and $\sigma_{\alpha+B^c}(E)$ do behave for $E \to \infty$ as constants times $(\ln E)^r$ then the constants must be the same. Because we are concerned here with the high-energy limit, this result does not depend on the assumption that B is a massless boson, so in the same sense, *the ratio of the cross-sections of any particle and its antiparticle on a fixed target should approach unity at high energy.* This result is a somewhat generalized version of what is known as Pomeranchuk's theorem.[16] (Pomeranchuk considered only the case $r = 0$, while Section 3.7 and the observed behavior of cross-sections both suggest that $r = 2$ is more likely.)

Although Pomeranchuk took his estimates of the asymptotic behavior of scattering amplitudes from arguments like those of Section 3.7, today high energy behavior is usually inferred from Regge pole theory.[17] It would take us too far from our subject to go into details about this; suffice it to say that for hadronic processes the asymptotic behavior of $f(\omega)$ as ω goes to infinity is a sum over terms proportional to $\omega^{\alpha_n(0)}$, where $\alpha_n(t)$ are a set of 'Regge trajectories', each representing the exchange of an infinite family of different one-hadron states in the collision process. The leading trajectory (actually, a complex of many trajectories) in hadron–

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

hadron scattering is the 'Pomeron,' for which $\alpha(0)$ is close to unity. It is this trajectory that gives cross-sections that are approximately constant for $E \to \infty$. According to Pomeranchuk's theorem, the Pomeron couples equally to any hadron and its antiparticle. We can estimate $\alpha_n(0)$ for the lower Regge trajectories from the spectrum of hadronic states. A necessary though not sufficient condition[18] for a mesonic resonance of spin j to occur at a mass m is that m^2 equals the value of t where one of trajectories $\alpha_n(t)$ equals j. Apart from the Pomeron, the leading trajectory in pion–nucleon scattering is that on which we find the $j = 1$ ρ meson at $m = 770$ MeV, the $j = 3$ g meson at $m = 1690$ MeV, and a $j = 5$ meson at $m = 2350$ MeV. Extrapolating these values of $\alpha(t)$ down to $t = 0$, we can estimate that this trajectory has $\alpha(0) \approx 0.5$. This trajectory couples with opposite sign to π^+ and π^-, so for pion–nucleon scattering we expect $f(\omega) - f^c(\omega)$ to behave roughly like $\sqrt{\omega}$.

For photon scattering there is no distinction between B and B^c, so here Eq. (10.8.27) gives $b = 0$, and Eq. (10.8.26) reads

$$f(\omega) = a + \frac{\omega^2}{2\pi^2} \int_0^\infty \frac{\sigma(E)}{E^2 - \omega^2} dE \,. \tag{10.8.29}$$

This is essentially the original Kramers–Kronig[13] relation. As we shall see in Section 13.5, for a target of charge e and mass m the constant a has the known value $\mathrm{Re}\, f(0) = -e^2/m$.

Problems

1. Consider a neutral vector field $v_\mu(x)$. What conditions must be imposed on the sum $\Pi^*_{\mu\nu}(k)$ of one-particle-irreducible graphs with two external vector field lines in order that the field should be properly renormalized and describe a particle of renormalized mass m? How do we split the free-field and interacting terms in the Lagrangian to achieve this?

2. Derive the generalized Ward identity that governs the electromagnetic vertex function of a charged scalar field.

3. What is the most general form of the matrix element $\langle \mathbf{p}_2 \sigma_2 | J^\mu(x) | \mathbf{p}_1 \sigma_1 \rangle$ of the electromagnetic current $J^\mu(x)$ between two spin $\frac{1}{2}$ one-particle states of *different* masses m_1 and m_2 and equal parity? What if the parities were opposite? (Assume parity conservation throughout.)

4. Derive the spectral (Källen–Lehmann) representation for the vacuum expectation value $\langle T\{J^\mu(x)\ J^\nu(y)^\dagger\}\rangle_0$, where $J^\mu(x)$ is a complex conserved current.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

5. Derive the spectral (Källen–Lehmann) representation for the vacuum expectation value $\langle T\{\psi_n(x)\,\bar{\psi}_m(y)\}\rangle_0$, where $\psi(x)$ is a Dirac field.

6. Without using any assumptions about the asymptotic behavior of the scattering amplitude or cross-sections, show that it is impossible for forward photon scattering amplitudes to satisfy unsubtracted dispersion relations.

7. Derive the spectral (Källen–Lehmann) representation for a complex scalar field by using the methods of dispersion theory.

8. Use dispersion theory and the results of Section 8.7 to calculate the amplitude for forward photon–electron scattering in the electron rest frame to order e^4.

References

1. W. H. Furry, *Phys. Rev.* **51**, 125 (1937).

2. H. Yukawa, *Proc. Phys.-Math. Soc. Japan* **17**, 48 (1935).

3. H. Lehmann, K. Symanzik, and W. Zimmerman, *Nuovo Cimento* **1**, 205 (1955).

4. Y. Takahashi, *Nuovo Cimento*, Ser. 10, **6**, 370 (1957).

5. J. C. Ward, *Phys. Rev.* **78**, 182 (1950).

6. J. Schwinger, *Phys. Rev. Lett.* **3**, 296 (1959).

7. P. A. M. Dirac, *Proc. Roy. Soc. (London)* **A117**, 610 (1928).

8. M. N. Rosenbluth, *Phys. Rev.* **79**, 615 (1950).

9. G. Källen, *Helv. Phys. Acta* **25**, 417 (1952); *Quantum Electrodynamics* (Springer-Verlag, Berlin, 1972); H. Lehmann, *Nuovo Cimento* **11**, 342 (1954).

10. J. C. Howard and B. Jouvet, *Nuovo Cimento* **18**, 466 (1960); M. J. Vaughan, R. Aaron, and R. D. Amado, *Phys. Rev.* **1254**, 1258 (1961); S. Weinberg, in *Proceedings of the 1962 High-Energy Conference at CERN* (CERN, Geneva, 1962): p. 683.

11. R. L. Stratonovich, *Sov. Phys. Dokl.* **2**, 416 (1957); J. Hubbard, *Phys. Rev. Lett.* **3**, 77 (1959).

12. S. Weinberg, *Phys. Rev.* **137**, B672 (1965).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

13. H. A. Kramers, *Atti Congr. Intern. Fisici, Como* (Nicolo Zanichelli, Bologna, 1927); reprinted in H. A. Kramers, *Collected Scientific Papers* (North-Holland, Amsterdam, 1956); R. de Kronig, *Ned. Tyd. Nat. Kunde* **9**, 402 (1942); *Physica* **12**, 543 (1946); J. S. Toll, The Dispersion Relation for Light and its Application to Problems Involving Electron Pairs (Princeton University Ph. D. Thesis 1952). For historical reviews, see J. D. Jackson in *Dispersion Relations*, ed. by G. R. Screaton (Oliver and Boyd, Edinburgh, 1961); M. L. Goldberger in *Dispersion Relations and Elementary Particles*, ed. by C. De Witt and R. Omnes (Hermann, Paris, 1960).

14. M. Gell-Mann, M. L. Goldberger, and W. Thirring, *Phys. Rev.* **95**, 1612 (1954). The non-perturbative nature of this result was shown by M. L. Goldberger, *Phys. Rev.* **97**, 508 (1955).

15. M. L. Goldberger, *Phys. Rev.* **99**, 979 (1955).

16. I. Ia. Pomeranchuk, *J. Expt. Theor. Phys. (USSR.)* **34**, 725 (1958). English version: *Soviet Physics - JETP* **34**(7), 499 (1958). For a generalization, see S. Weinberg, *Phys. Rev.* **124**, 2049 (1961).

17. See, e.g., P. D. B. Collins, *An Introduction to Regge Theory and High Energy Physics* (Cambridge University Press, Cambridge, 1977). The original references are T. Regge, *Nuovo Cimento* **14**, 951 (1959): **18**, 947 (1960).

18. The graph of spin versus squared mass is known as a Chew–Frautschi plot; see G. F. Chew and S. C. Frautschi, *Phys. Rev. Lett.* **8**, 41 (1962).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.012

11

One-Loop Radiative Corrections in Quantum Electrodynamics

In this chapter we shall proceed to carry out some of the classic one-loop calculations in the theory of charged leptons — massive spin $\frac{1}{2}$ particles that interact only with the electromagnetic field. There are three known species or 'flavors' of leptons: the electron and muon, and the heavier, more recently discovered tauon. For definiteness we shall refer to the charged particles in our calculations here as 'electrons,' though most of our calculations will apply equally to muons and tauons. After some generalities in Section 11.1, we will move on to the calculation of the vacuum polarization in Section 11.2, the anomalous magnetic moment of the electron in Section 11.3, and the electron self-energy in Section 11.4. Along the way, we will introduce a number of the mathematical techniques that prove useful in such calculations, including the use of Feynman parameters, Wick rotation, and both the dimensional regularization of 't Hooft and Veltman and the older regularization method of Pauli and Villars. Although we shall encounter infinities, it will be seen that the final results are finite if expressed in terms of the renormalized charge and mass. In the next chapter we shall extend what we have learned here about renormalization to general theories in arbitrary orders of perturbation theory.

11.1 Counterterms

The Lagrangian density for electrons and photons is taken in the form*

$$\mathscr{L} = -\tfrac{1}{4}F_{\mathrm{B}}^{\mu\nu}F_{\mathrm{B}\,\mu\nu} - \bar{\psi}_{\mathrm{B}}\left[\gamma_\mu\left[\partial^\mu + ie_{\mathrm{B}}A_{\mathrm{B}}^\mu\right] + m_{\mathrm{B}}\right]\psi_{\mathrm{B}} \tag{11.1.1}$$

where $F_{\mathrm{B}}^{\mu\nu} \equiv \partial^\mu A_{\mathrm{B}}^\nu - \partial^\nu A_{\mathrm{B}}^\mu$; A_{B}^μ and ψ_{B} are the bare (i.e., unrenormalized) fields of the photon and electron, and $-e_{\mathrm{B}}$ and m_{B} are the bare charge and

* In this chapter we will not be making transformations between Heisenberg- and interaction-picture operators, so we shall return to a conventional notation, in which an upper case A and a lower case ψ are used to denote the photon and charged particle fields, respectively.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

mass of the electron. As described in the previous chapter, we introduce renormalized fields and charge and mass:

$$\psi \equiv Z_2^{-1/2}\psi_B\,, \tag{11.1.2}$$

$$A^\mu \equiv Z_3^{-1/2}A_B^\mu\,, \tag{11.1.3}$$

$$e \equiv Z_3^{+1/2}e_B\,, \tag{11.1.4}$$

$$m \equiv m_B + \delta m\,, \tag{11.1.5}$$

with the constants Z_2, Z_3, and δm adjusted so that the propagators of the renormalized fields have poles in the same position and with the same residues as the propagators of the free fields in the absence of interactions. The Lagrangian may then be written in terms of renormalized quantities, as

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1 + \mathscr{L}_2\,, \tag{11.1.6}$$

where

$$\mathscr{L}_0 = -\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} - \bar{\psi}\left[\gamma_\mu\,\partial^\mu + m\right]\psi\,, \tag{11.1.7}$$

$$\mathscr{L}_1 = -ieA_\mu\bar{\psi}\,\gamma^\mu\psi\,, \tag{11.1.8}$$

and $\mathscr{L}_2$ is a sum of 'counterterms'

$$\begin{aligned}\mathscr{L}_2 = &-\tfrac{1}{4}(Z_3-1)F^{\mu\nu}F_{\mu\nu} - (Z_2-1)\bar{\psi}\left[\gamma_\mu\,\partial^\mu + m\right]\psi \\ &+Z_2\delta m\bar{\psi}\,\psi - ie(Z_2-1)A_\mu\bar{\psi}\,\gamma^\mu\psi\,.\end{aligned} \tag{11.1.9}$$

It will turn out that all of the terms in $\mathscr{L}_2$ are of second order and higher order in e, and that these terms just suffice to cancel the ultraviolet divergences that arise from loop graphs.

11.2 Vacuum Polarization

We now begin our first calculation of a radiative correction involving loop graphs, the so-called vacuum polarization effect, consisting of the corrections to the propagator associated with an internal photon line. Vacuum polarization produces measurable shifts in the energy levels of hydrogen, and makes an important correction to the energies of muons bound in atomic orbits around heavy nuclei. Also, as we shall see in Volume II, the calculation of the vacuum polarization provides a key

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

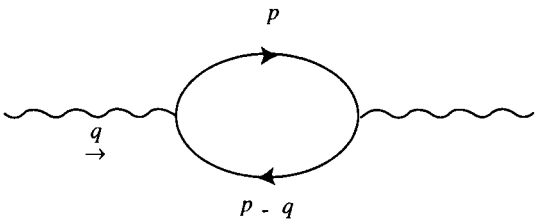

Figure 11.1. The one-loop diagram for the vacuum polarization in quantum electrodynamics. Here wavy lines represent photons; lines carrying arrows represent electrons.

element in the calculation of the high energy behavior of electrodynamics and other gauge theories.

As in Section 10.5, we define $i(2\pi)^4\Pi^{*\mu\nu}(q)$ as the sum of all connected graphs with two external photon lines with polarization indices μ and ν and carrying four-momentum q into and out of the diagram, not including photon propagators for the two external lines, and with the asterisk indicating that we exclude diagrams that can be disconnected by cutting through some internal photon line. The complete photon propagator $\Delta'^{\mu\nu}(q)$ is given by Eq. (10.5.13):

$$\Delta' = \Delta\,[1 - \Pi^*\Delta]^{-1}\,, \tag{11.2.1}$$

where $\Delta^{\mu\nu}(q)$ is the photon propagator without radiative corrections. Our task here is to calculate the leading contributions to $\Pi^{*\rho\sigma}(q)$.

In lowest order there is a one-loop contribution to Π^*, corresponding to the diagram in Figure 11.1:

$$i(2\pi)^4\Pi^{*\rho\sigma}_{1\text{ loop}}(q) = -\int d^4p\,\mathrm{Tr}\left\{\left[\frac{-i}{(2\pi)^4}\,\frac{-i\,\not{p}+m}{p^2+m^2-i\epsilon}\right]\right.$$
$$\left.\times\left[(2\pi)^4 e\gamma^\rho\right]\left[\frac{-i}{(2\pi)^4}\,\frac{-i(\not{p}-\not{q})+m}{(p-q)^2+m^2-i\epsilon}\right]\left[(2\pi)^4 e\gamma^\sigma\right]\right\} \tag{11.2.2}$$

with the first minus sign on the right required by the presence of a fermion loop. More simply, this is

$$\Pi^{*\rho\sigma}_{1\text{ loop}}(q) = \frac{-ie^2}{(2\pi)^4}\int d^4p\,\frac{\mathrm{Tr}\left\{[-i\,\not{p}+m]\gamma^\rho\,[-i(\not{p}-\not{q})+m]\,\gamma^\sigma\right\}}{(p^2+m^2-i\epsilon)\,((p-q)^2+m^2-i\epsilon)}\,. \tag{11.2.3}$$

The first step in doing this integral is to use a trick introduced by Feynman.[1] We use the elementary formula

$$\frac{1}{AB} = \int_0^1 \frac{dx}{[(1-x)A+xB]^2} \tag{11.2.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

to write the product of scalar propagators in Eq. (11.2.3) as

$$\begin{aligned}\frac{1}{(p^2+m^2-i\epsilon)((p-q)^2+m^2-i\epsilon)} &= \int_0^1 \Big[(p^2+m^2-i\epsilon)(1-x)\\ &\quad + \Big((p-q)^2+m^2-i\epsilon\Big)x\Big]^{-2}dx\\ &= \int_0^1 \Big[p^2+m^2-i\epsilon-2p\cdot qx+q^2x\Big]^{-2}dx\\ &= \int_0^1 \Big[(p-qx)^2+m^2-i\epsilon+q^2x(1-x)\Big]^{-2}dx\,.\end{aligned}$$

(This is a special case of a class of integrals given in the Appendix to this chapter.) We can now shift the variable of integration in momentum space*

$$p \to p+qx\,,$$

so that Eq. (11.2.3) becomes

$$\begin{aligned}\Pi_{1\,\text{Loop}}^{*\rho\sigma}(q) &= \frac{-ie^2}{(2\pi)^4}\int_0^1 dx \int d^4p\Big[p^2+m^2-i\epsilon+q^2x(1-x)\Big]^{-2}\\ &\times \text{Tr}\left\{\left[-i(\not{p}+\not{q}x)+m\right]\gamma^\rho\left[-i(\not{p}-\not{q}(1-x))+m\right]\gamma^\sigma\right\}\,. \quad (11.2.5)\end{aligned}$$

Using the results of the Appendix to Chapter 8, the trace here can easily be calculated as

$$\begin{aligned}&\text{Tr}\left\{\left[-i(\not{p}+\not{q}x)+m\right]\gamma^\rho\left[-i(\not{p}-\not{q}(1-x))+m\right]\gamma^\sigma\right\}\\ &= 4\Big[-(p+qx)^\rho(p-q(1-x))^\sigma+(p+qx)\cdot(p-q(1-x))\eta^{\rho\sigma}\\ &\quad -(p+qx)^\sigma(p-q(1-x))^\rho+m^2\eta^{\rho\sigma}\Big]\,. \quad (11.2.6)\end{aligned}$$

Our next step is called a *Wick rotation*.[2] As long as $-q^2 < 4m^2$, the quantity $m^2+q^2x(1-x)$ is positive for all x between 0 and 1, so the poles in the integrand of Eq. (11.2.5) are at $p^0 = \pm\sqrt{\mathbf{p}^2+m^2+q^2x(1-x)}-i\epsilon$, i.e., just above the negative real axis and just below the positive real axis. (See Figure 11.2.) We can rotate the contour of integrations of p^0 counterclockwise without crossing either of these poles, so that instead of integrating p^0 on the real axis from $-\infty$ to $+\infty$, we integrate it on the imaginary axis from $-i\infty$ to $+i\infty$. That is, we can write $p^0 = ip^4$, with p^4 integrated over real values from $-\infty$ to $+\infty$. (If an $i\epsilon$ instead of $-i\epsilon$ had appeared in the denominator of the propagator, then we would have been setting $p^0 = -ip^4$, with p^4 again integrated over real values from $-\infty$ to

* Strictly speaking, this step is only valid in convergent integrals. In principle, in order to justify the shift of variables, we should introduce some regulator scheme to make all integrals converge, such as the dimensional regularization scheme discussed below.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

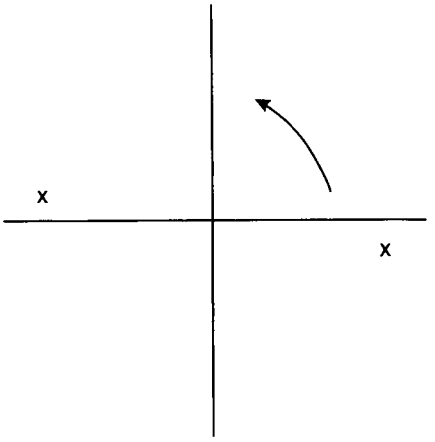

Figure 11.2. Wick rotation of the p^0 contour of integration. Small x's mark the poles in the p^0 complex plane; the arrow indicates the direction of rotation of the contour of integration, from the real to the imaginary p^0-axes.

$+\infty$. The effect would be a change of sign of $\Pi^{*\rho\sigma}_{1\text{ loop}}(q)$.) Eq. (11.2.5) now becomes

$$\Pi^{*\rho\sigma}_{1\text{ loop}}(q) = \frac{4e^2}{(2\pi)^4}\int_0^1 dx \int (d^4p)_E \left[p^2 + m^2 + q^2x(1-x)\right]^{-2}$$
$$\times\Big[-(p+qx)^\rho(p-q(1-x))^\sigma + (p+qx)\cdot(p-q(1-x))\,\eta^{\rho\sigma}$$
$$-(p+qx)^\sigma(p-q(1-x))^\rho + m^2\eta^{\rho\sigma}\Big], \qquad (11.2.7)$$

where

$$(d^4p)_E = dp^1dp^2dp^3dp^4$$

and all scalar products are evaluated using the Euclidean norm

$$a\cdot b = a^1b^1 + a^2b^2 + a^3b^3 + a^4b^4$$

with the understanding that $q^4 \equiv -iq^0$. Also, $\eta^{\rho\sigma}$ can be taken as either the Kronecker delta, with the indices running over 1, 2, 3, 4, or as the usual Minkowski tensor, with the indices running over 1, 2, 3, 0.

The integral (11.2.7) is badly divergent. Eventually all infinities will cancel, but to see this it is necessary at intermediate stages of the calculation to use some sort of regularization technique that makes the integrals finite. It would not do simply to cut off the integrals at some maximum momentum Λ, integrating only over p^μ with $p^2 < \Lambda^2$, because this would amount to introducing a step function $\theta(\Lambda^2 - p^2)$ into the electron propagator, and the Ward identity (10.4.25) shows that in order to maintain

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

gauge invariance, any modification of the electron propagator must be accompanied with a modification of the electron–photon vertex. In fact, with an ordinary cutoff Λ, radiative corrections would induce a photon mass, a clear violation of the requirements of gauge invariance.

Experience has shown that the most convenient method for regulating divergent integrals without impairing gauge invariance is the dimensional regularization technique introduced by 't Hooft and Veltman[3] in 1972, based on a continuation from four to an arbitrary number d of spacetime dimensions. This amounts to carrying out angular averages in integrals like (11.2.7) by dropping all terms that are odd in p, and replacing the terms that have even numbers of p-factors with**

$$p^\mu p^\nu \to p^2\eta^{\mu\nu}\Big/d\,, \tag{11.2.8}$$

$$p^\mu p^\nu p^\rho p^\sigma \to (p^2)^2[\eta^{\mu\nu}\eta^{\rho\sigma}+\eta^{\mu\rho}\eta^{\nu\sigma}+\eta^{\mu\sigma}\eta^{\nu\rho}]\Big/d(d+2)\,, \tag{11.2.9}$$

Also, after writing the integrand in this way as a function only of p^2, the volume element d^4p_E is to be replaced with $\Omega_d\kappa^{d-1}d\kappa$, where $\kappa\equiv\sqrt{p^2}$, and Ω_d is the area of a unit sphere in d dimensions

$$\Omega_d = 2\pi^{d/2}\Big/\Gamma(d/2)\,. \tag{11.2.10}$$

The integral (11.2.7) now converges for complex spacetime dimensionality d. We can continue the integral through complex d-values to $d=4$, the infinities then reappearing as factors $(d-4)^{-1}$.

For the integral (11.2.7), dimensional regularization gives

$$\Pi^{*\rho\sigma}_{1\text{ loop}}(q) = \frac{4e^2\Omega_d}{(2\pi)^4}\int_0^1 dx\int_0^\infty \kappa^{d-1}d\kappa\left[\kappa^2+m^2+q^2x(1-x)\right]^{-2}$$
$$\times\left[\frac{-2\kappa^2}{d}\eta^{\rho\sigma}+2q^\rho q^\sigma x(1-x)+\Big(\kappa^2-q^2x(1-x)\Big)\eta^{\rho\sigma}+m^2\eta^{\rho\sigma}\right].$$

The integrals over κ can be carried out for any complex d (or for any real d, aside from the even integers). We use the well-known formulas (given in greater generality in the Appendix to this chapter):

$$\int_0^\infty \kappa^{d-1}[\kappa^2+v^2]^{-2}d\kappa = \tfrac{1}{2}(v^2)^{\frac{d}{2}-2}\,\Gamma(d/2)\,\Gamma(2-d/2)\,, \tag{11.2.11}$$

$$\int_0^\infty \kappa^{d+1}[\kappa^2+v^2]^{-2}d\kappa = \tfrac{1}{2}(v^2)^{\frac{d}{2}-1}\Gamma(1+d/2)\,\Gamma(1-d/2)\,, \tag{11.2.12}$$

** These expressions may most easily be derived by noting that their form is dictated by Lorentz invariance and the symmetry among the indices μ,ν,ρ, etc., while the factors may be found by requiring that both sides give the same result when contracted with ηs.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

and find

$$\Pi^{*\rho\sigma}_{1\,\text{loop}}(q) = \frac{2e^2\Omega_d}{(2\pi)^4}$$
$$\times \int_0^1 dx \Big[(1-2/d)\,\eta^{\rho\sigma}\left(m^2+q^2x(1-x)\right)^{\frac{d}{2}-1}\Gamma(1+d/2)\,\Gamma(1-d/2)$$
$$+ \left(2q^\rho q^\sigma x(1-x) - q^2\eta^{\rho\sigma}x(1-x) + m^2\eta^{\rho\sigma}\right)\left(m^2+q^2x(1-x)\right)^{\frac{d}{2}-2}$$
$$\times \Gamma(d/2)\,\Gamma(2-d/2)\Big].$$

The two terms in the integrand can be combined, using

$$(1-2/d)\,\Gamma(1+d/2)\,\Gamma(1-d/2) = -\Gamma(d/2)\,\Gamma(2-d/2)\,.$$

We find

$$\Pi^{*\rho\sigma}_{1\,\text{loop}}(q) = \frac{4e^2\Omega_d}{(2\pi)^4}\,\Gamma(d/2)\,\Gamma(2-d/2)\,(q^\rho q^\sigma - q^2\eta^{\rho\sigma})$$
$$\times \int_0^1 dx\; x(1-x)(m^2+q^2x(1-x))^{\frac{d}{2}-2}\,. \qquad (11.2.13)$$

We note the very important result that this contribution to $\Pi^{*\rho\sigma}$ satisfies the relation

$$q_\rho \Pi^{*\rho\sigma}_{1\,\text{loop}}(q) = 0 \qquad (11.2.14)$$

that was derived in Section 10.5 on the basis of the conservation and neutrality of the electric current. It was precisely to achieve this result that we adopted the dimensional regularization scheme. The reason that dimensional regularization gives this result is that the conservation of current does not depend on the dimensionality of spacetime.

The gamma function $\Gamma(2-d/2)$ in Eq. (11.2.13) blows up for $d \to 4$. Fortunately, as we saw in Section 11.1, there is another term that must be added to $\Pi^{*\rho\sigma}(q)$, arising from the term $-\frac{1}{4}(Z_3-1)F_{\mu\nu}F^{\mu\nu}$ in the interaction Lagrangian. This term has a structure like Eq. (11.2.13)

$$\Pi^{*\rho\sigma}_{\mathscr{L}_2}(q) = -(Z_3-1)(q^2\eta^{\rho\sigma} - q^\rho q^\sigma)\,, \qquad (11.2.15)$$

so to order e^2, the complete Π^* has the form

$$\Pi^{*\rho\sigma}(q) = (q^2\eta^{\rho\sigma} - q^\rho q^\sigma)\pi(q^2)\,, \qquad (11.2.16)$$

with

$$\pi(q^2) = -\frac{4e^2\Omega_d}{(2\pi)^4}\,\Gamma\left(\tfrac{d}{2}\right)\Gamma\left(2-\tfrac{d}{2}\right)\int_0^1 dx\; x(1-x)(m^2+q^2x(1-x))^{\frac{d}{2}-2}$$
$$-(Z_3-1)\,. \qquad (11.2.17)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

As we saw in Section 10.5, the definition of the renormalized electromagnetic field requires that $\pi(0) = 0$ (in order that the residue of the pole in the complete photon propagator at $q^2 = 0$ should be the same as for the bare propagator, aside from gauge-dependent terms). Therefore, to order e^2,

$$Z_3 = 1 - \frac{4e^2\Omega_d}{(2\pi)^4}\,\Gamma\left(\tfrac{d}{2}\right)\Gamma\left(2-\tfrac{d}{2}\right)(m^2)^{\frac{d}{2}-2}\int_0^1 x(1-x)dx\,, \qquad (11.2.18)$$

so that, to order e^2,

$$\pi(q^2) = -\frac{4e^2\Omega_d}{(2\pi)^4}\,\Gamma\left(\tfrac{d}{2}\right)\Gamma\left(2-\tfrac{d}{2}\right)\int_0^1 dx\; x(1-x) \times\left[\Big(m^2+q^2x(1-x)\Big)^{\frac{d}{2}-2} - (m^2)^{\frac{d}{2}-2}\right]. \qquad (11.2.19)$$

Now we can remove the regularization, allowing d to approach its physical value $d = 4$. As mentioned before, there is an infinity in the one-loop contribution, arising from the limiting behavior of the Gamma function

$$\Gamma\left(2-\tfrac{d}{2}\right) \to \frac{1}{(2-d/2)} \quad -\gamma\,,$$

where γ is the Euler constant, $\gamma = 0.5772157$. The infinite part of $Z_3 - 1$ is given by using $1/(2-d/2)$ for $\Gamma(2-d/2)$, and replacing d everywhere else by 4:

$$(Z_3-1)_\infty = -\frac{4e^2\cdot 2\pi^2}{6(2\pi)^4}\,\frac{1}{2-d/2} = \frac{e^2}{6\pi^2}\,\frac{1}{d-4}\,. \qquad (11.2.20)$$

We shall see in Volume II that this result may be used to derive the leading term in the renormalization group equation for the electric charge.

The poles at $d = 4$ obviously cancel in $\pi(q^2)$, because for $d = 4$ both $(m^2+q^2x(1-x))^{\frac{d}{2}-2}$ and $(m^2)^{\frac{d}{2}-2}$ have the same limit, unity. For the same reason, the term $-\gamma$ in $\Gamma(2-d/2)$ cancels in the total $\pi(q^2)$, though it does make a finite contribution to Z_3-1. There are other finite contributions to Z_3-1, that arise from the product of the pole in $\Gamma(2-d/2)$ with the linear terms in the expansion of $\Omega_d\Gamma(d/2)$ around $d = 4$, but these also cancel in the total $\pi(q^2)$. Indeed, in carrying out our dimensional regularization, we might have replaced $(2\pi)^{-4}$ with $(2\pi)^{-d}$, and the factor $\mathrm{Tr}\,1 = 4$ might have been replaced with the dimensionality $2^{d/2}$ of gamma matrices in arbitrary even spacetime dimensionalities d, and these too would have contributed to the finite part of $Z_3 - 1$, but not of $\pi(q^2)$. Moreover, e^2 cannot be supposed to be d-independent, because as shown by inspection of Eq. (11.2.13), it has the d-dependent dimensionality $[\text{mass}]^{4-d}$. If we take $e^2 \propto \mu^{4-d}$, where μ is some quantity with the units of mass, then

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

there are additional finite terms in $Z_3 - 1$, arising from the product of the pole in $\Gamma(2 - d/2)$ with the term $(4 - d)\ln\mu$ in the expansion of μ^{4-d} in powers of $4 - d$, but again, these cancel between $Z_3 - 1$ and the one-loop contributions to $\pi(q^2)$.

The only terms that *do* contribute to $\pi(q^2)$ in the limit $d \to 4$ are those arising from the product of the pole in $\Gamma(2 - d/2)$ with the linear terms in the expansion of $(m^2 + q^2x(1-x))^{\frac{d}{2}-2}$ and $(m^2)^{\frac{d}{2}-2}$ in powers of $d - 4$:

$$(m^2+q^2x(1-x))^{\frac{d}{2}-2}-(m^2)^{\frac{d}{2}-2} \to (\tfrac{d}{2}-2)\ln\left(1+\frac{q^2x(1-x)}{m^2}\right). \quad (11.2.21)$$

This gives at last

$$\pi(q^2) = \frac{e^2}{2\pi^2}\int_0^1 x(1-x)\ln\left(1+\frac{q^2x(1-x)}{m^2}\right)dx. \quad (11.2.22)$$

The physical significance of the vacuum polarization can be explored by considering its effect on the scattering of two charged particles of spin $\frac{1}{2}$. The Feynman diagrams of Figure 11.3 make contributions to the scattering S-matrix element of the form

$$S_a(1,2 \to 1',2') = (2\pi)^{-12/2}\delta^4(p_{1'} + p_{2'} - p_1 - p_2)\left[e_1(2\pi)^4\bar{u}_{1'}\gamma^\mu u_1\right]$$
$$\times\left[-i(2\pi)^{-4}\frac{1}{q^2}\right]\left[e_2(2\pi)^4\bar{u}_{2'}\gamma_\mu u_2\right],$$

$$S_b(1,2 \to 1',2') = (2\pi)^{-12/2}\delta^4(p_{1'} + p_{2'} - p_1 - p_2)\left[e_1(2\pi)^4\bar{u}_{1'}\gamma^\mu u_1\right]$$
$$\times\left[-i(2\pi)^{-4}\frac{1}{q^2}\right]^2\left[i(2\pi)^4(q^2\eta_{\mu\nu} - q_\mu q_\nu)\pi(q^2)\right]\left[e_2(2\pi)^4\bar{u}_{2'}\gamma^\nu u_2\right],$$

where e_1 and e_2 are the charges of the two particles being scattered; $\pi(q^2)$ is calculated using for e in Eq. (11.2.22) the magnitude of the charge of the particle circulating in the loop in Figure 11.3; and q^μ is the momentum transfer $q \equiv p_1 - p_{1'} = p_{2'} - p_2$. Using the conservation property $q_\mu\bar{u}_{1'}\gamma^\mu u_1 = 0$ the two diagrams together yield an S-matrix element:

$$S_{a+b}(1,2 \to 1',2') = \frac{-ie_1e_2}{4\pi^2q^2}\left[1+\pi(q^2)\right]\delta^4(p_{1'} + p_{2'} - p_1 - p_2)$$
$$\times\left[\bar{u}_{1'}\gamma^\mu u_1\right]\left[\bar{u}_{2'}\gamma_\mu u_2\right]. \quad (11.2.23)$$

In the non-relativistic limit, $\bar{u}_{1'}\gamma^0 u_1 \simeq -i\delta_{\sigma_1'\sigma_1}$ while $\bar{u}_{1'}\gamma^i u_1 \simeq 0$, and likewise for particle 2. Also, in this limit q^0 is negligible compared with $|\mathbf{q}|$. Eq. (11.2.23) in this limit becomes

$$S_{a+b}(1,2 \to 1',2') = \frac{-ie_1e_2}{4\pi^2\mathbf{q}^2}\left[1+\pi(\mathbf{q}^2)\right]\delta^4(p_{1'} + p_{2'} - p_1 - p_2)\delta_{\sigma_1'\sigma_1}\delta_{\sigma_2'\sigma_2}. \quad (11.2.24)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

This may be compared with the S-matrix in the Born approximation due to a local spin-independent central potential $V(r)$:

$$S_{\text{Born}}(1,2 \to 1',2') = -2\pi i\delta(E_{1'} + E_{2'} - E_1 - E_2)T_{\text{Born}}(1,2 \to 1',2') . \tag{11.2.25}$$

$$T_{\text{Born}}(1,2 \to 1',2') = \delta_{\sigma_1'\sigma_1}\delta_{\sigma_2'\sigma_2}\int d^3x_1 \int d^3x_2\, V\left(|\mathbf{x}_1 - \mathbf{x}_2|\right) \times (2\pi)^{-12/2} e^{-i\mathbf{p}_{1'}\cdot\mathbf{x}_1} e^{-i\mathbf{p}_{2'}\cdot\mathbf{x}_2} e^{i\mathbf{p}_1\cdot\mathbf{x}_1} e^{i\mathbf{p}_2\cdot\mathbf{x}_2} . \tag{11.2.26}$$

Setting $\mathbf{x}_1 = \mathbf{x}_2 + \mathbf{r}$, this gives

$$S_{\text{Born}} = \frac{-i}{4\pi^2}\delta^4(p_{1'} + p_{2'} - p_1 - p_2)\delta_{\sigma_1'\sigma_1}\delta_{\sigma_2'\sigma_2} \times \int d^3r\, V(r)\, e^{-i\mathbf{q}\cdot\mathbf{r}} . \tag{11.2.27}$$

Comparing this with Eq. (11.2.23) shows that in the non-relativistic limit the diagrams of Figure 11.3 yield the same S-matrix element as a potential $V(r)$ such that

$$\int d^3r\, V(r)\, e^{-i\mathbf{q}\cdot\mathbf{r}} = e_1 e_2 \frac{1 + \pi(\mathbf{q}^2)}{\mathbf{q}^2}$$

or, inverting the Fourier transform,

$$V(r) = \frac{e_1 e_2}{(2\pi)^3}\int d^3q\, e^{i\mathbf{q}\cdot\mathbf{r}} \left[\frac{1 + \pi(\mathbf{q}^2)}{\mathbf{q}^2}\right] . \tag{11.2.28}$$

Eq. (11.2.28) is to first order in the radiative correction the same potential energy that would be produced by the electrostatic interaction of two extended charge distributions $e_1\eta(\mathbf{x})$ and $e_2\eta(\mathbf{y})$ at a distance r:

$$V(|\mathbf{r}|) = e_1 e_2 \int d^3x \int d^3y\, \frac{\eta(\mathbf{x})\eta(\mathbf{y})}{4\pi|\mathbf{x} - \mathbf{y} + \mathbf{r}|} , \tag{11.2.29}$$

where

$$\eta(\mathbf{r}) = \delta^3(\mathbf{r}) + \frac{1}{2(2\pi)^3}\int d^3q\, \pi(\mathbf{q}^2)\, e^{i\mathbf{q}\cdot\mathbf{r}} . \tag{11.2.30}$$

Note that

$$\int d^3r\, \eta(\mathbf{r}) = 1 + \tfrac{1}{2}\pi(0) = 1 , \tag{11.2.31}$$

so the total charges of particles 1 and 2, as determined from the long-range part of the Coulomb potential, are the same constants e_1 and e_2 that govern the interactions of the renormalized electromagnetic field.

For $|\mathbf{r}| \neq 0$ the integral (11.2.30) can be carried out by a straightforward

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

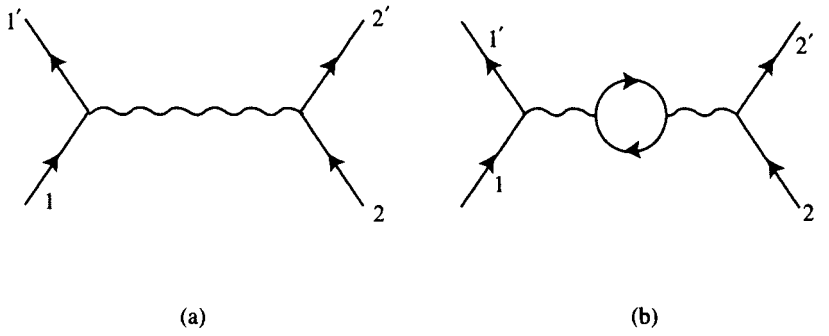

Figure 11.3. Two diagrams for the scattering of charged particles. Here lines carrying arrows are charged particles; wavy lines are photons. Diagram (b) represents the lowest-order vacuum polarization correction to the tree approximation graph (a).

contour integration:

$$\eta(\mathbf{r}) = -\frac{e^2}{8\pi^3 r^3}\int_0^1 x(1-x)\,dx\left[1+\frac{mr}{\sqrt{x(1-x)}}\right]\exp\left(\frac{-mr}{\sqrt{x(1-x)}}\right).$$

This expression is negative everywhere. However, we have seen that the integral of $\eta(\mathbf{r})$ over all $\mathbf{r}$ equals $+1$. Therefore, $\eta(\mathbf{r})$ must contain a term $(1+L)\delta^3(\mathbf{r})$ that is singular at $\mathbf{r}=0$, with L chosen to satisfy Eq. (11.2.31):

$$L = \frac{e^2}{8\pi^3}\int\frac{d^3r}{r^3}\int_0^1 x(1-x)\,dx\left[1+\frac{mr}{\sqrt{x(1-x)}}\right]\exp\left(\frac{-mr}{\sqrt{x(1-x)}}\right). \tag{11.2.32}$$

The complete expression for the charge distribution function is then

$$\eta(\mathbf{r}) = (1+L)\delta^3(\mathbf{r}) - \frac{e^2}{8\pi^3 r^3}\int_0^1 x(1-x)\,dx \times\left[1+\frac{mr}{\sqrt{x(1-x)}}\right]\exp\left(\frac{-mr}{\sqrt{x(1-x)}}\right). \tag{11.2.33}$$

The physical interpretation of this result is that a bare point charge attracts particles of charge of opposite sign out of the vacuum, repelling their antiparticles to infinity, so that the bare charge is partially shielded, yielding a renormalized charge smaller by a factor $1/(1+L)$. As a check, we may note that if we cut off the divergent integral (11.2.32) by taking the integral to extend only over $r \geq a$, we find that the part that is divergent for $a \to 0$ is

$$L_\infty = \frac{e^2}{12\pi^2}\ln a^{-1}. \tag{11.2.34}$$

Hence if we identify the momentum space cutoff Λ with a^{-1}, the divergent

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

part of L is related to the divergent part of $Z_3 - 1$ by

$$(Z_3 - 1)_\infty = -2L_\infty \,, \tag{11.2.35}$$

because to order e^2 the renormalized charge (10.4.18) is given by

$$e_\ell = Z_3^{1/2} e_{B\ell} \simeq (1 + \tfrac{1}{2}(Z_3 - 1))\, e_{B\ell} \simeq (1 + L)^{-1} e_{B\ell} \,. \tag{11.2.36}$$

Eq. (11.2.35) is confirmed below.

Vacuum polarization has a measurable effect on muonic atomic energy levels. As we shall see in Chapter 14, the effect of Feynman graph (b) in Figure 11.3 is to shift the energy of an atomic state with wave function $\psi(\mathbf{r})$ by

$$\Delta E = \int d^3r\, \Delta V(\mathbf{r})\, |\psi(\mathbf{r})|^2 \,, \tag{11.2.37}$$

where $\Delta V(\mathbf{r})$ is the perturbation in the potential (11.2.28):

$$\Delta V(r) = \frac{e_1 e_2}{(2\pi)^3} \int d^3q\, e^{i\mathbf{q}\cdot\mathbf{r}} \left[\frac{\pi(\mathbf{q}^2)}{\mathbf{q}^2}\right] \,. \tag{11.2.38}$$

This perturbation falls off exponentially for $r \gg m^{-1}$. On the other hand, the wave function of electrons in ordinary atoms will generally be confined within a much larger radius $a \gg m^{-1}$; for instance, for hydrogenic orbits of electrons around a nucleus of charge Ze we have $a = 137/Zm$ (where here $m = m_e$). The energy shift will then depend only on the behavior of the wave function for $r \ll a$. For orbital angular momentum ℓ, the wave function behaves like r^ℓ for $r \ll a$, so Eq. (11.2.37) gives ΔE proportional to a factor $(ma)^{-(2\ell+1)}$. The effect of vacuum polarization is therefore very much larger for $\ell = 0$ than for higher orbital angular momenta. For $\ell = 0$ the wave function is approximately equal to the constant $\psi(0)$ for r less than or of the order of m^{-1}, so Eq. (11.2.37) becomes

$$\Delta E = |\psi(0)|^2 \int d^3r\, \Delta V(\mathbf{r}) \,. \tag{11.2.39}$$

Using Eqs. (11.2.38) and (11.2.22), the integral of the shift in the potential (for $e_1 e_2 = -Ze^2$) is

$$\int d^3r\, \Delta V(r) = -Ze^2 \pi'(0) = -\frac{4Z\alpha^2}{15m^2} \,. \tag{11.2.40}$$

Also, in states of hydrogenic atoms with $\ell = 0$ and principal quantum number n the wave function at the origin is

$$\psi(0) = \frac{2}{\sqrt{4\pi}} \left(\frac{Z\alpha m}{n}\right)^{3/2} \,, \tag{11.2.41}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

so the energy shift (11.2.39) is

$$\Delta E = -\frac{4Z^4\alpha^5 m}{15\pi n^3} . \tag{11.2.42}$$

For instance, in the $2s$ state of hydrogen this energy shift is -1.122×10^{-7} eV, corresponding to a frequency shift $\Delta E/2\pi\hbar$ of -27.13 MHz. This is sometimes called the *Uehling effect*.[4] As discussed in Chapter 1, such tiny energy shifts became measurable because in the absence of various radiative corrections the pure Dirac theory would predict exact degeneracy of the $2s$ and $2p$ states of hydrogen. As we shall see in Chapter 14, most of the +1058 MHz 'Lamb shift' between the $2s$ and the $2p$ states comes from other radiative corrections, but the agreement between theory and experiment is good enough to verify the presence of the -37.13 MHz shift due to vacuum polarization.

Although vacuum polarization contributes only a small part of the radiative corrections in ordinary atoms, it dominates the radiative corrections in muonic atoms, in which a muon takes the place of the orbiting electron. This is because most radiative corrections give energy shifts in muonic atoms that on dimensional grounds are proportional to m_μ, while the integrated vacuum polarization energy $\int d^3r\, \Delta V$ due to an *electron* loop is still proportional to m_e^{-2} as in Eq. (11.2.40), giving an energy shift proportional to $m_\mu^3 m_e^{-2} = (210)^2 m_\mu$. However, in this case the muonic atomic radius is not much larger than the electron Compton wavelength, so the approximate result (11.2.39) only gives the order of magnitude of the energy shift due to vacuum polarization.

* * *

For the purposes of comparison with later calculations, note that if we had cut off the integral (11.2.7) at $\kappa = \Lambda$, then in place of Eq. (11.2.20) we would have encountered an integral of the form

$$(Z_3 - 1)_\infty = -\frac{e^2}{6\pi^2}\int_\mu^\Lambda \kappa^{d-5}\, d\kappa = \frac{e^2}{6\pi^2}\frac{\mu^{d-4} - \Lambda^{d-4}}{d-4} ,$$

where μ is an infrared effective cutoff of the order of the mass of the charged particle circulating in the loop of Figure 11.1. (The easiest way to find the constant factor here is to require that the limit of this expression for $d < 3$ and $\Lambda \to \infty$ matches Eq. (11.2.20).) With such an ultraviolet cutoff in place, we can pass to the limit $d \to 4$, and obtain

$$(Z_3 - 1)_\infty = -\frac{e^2}{6\pi^2}\ln(\Lambda/\mu) . \tag{11.2.43}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

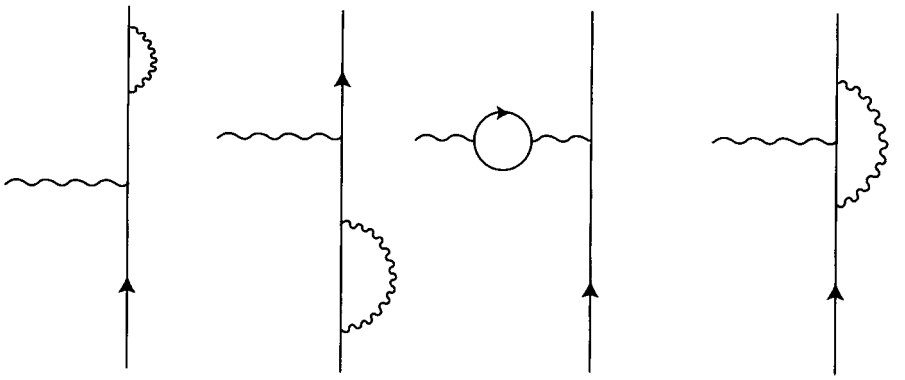

Figure 11.4. One-loop diagrams for the photon–lepton vertex function Γ^μ. Here wavy lines represent photons; other lines represent electrons or muons. Diagrams (a) and (b) are cancelled by lepton field renormalization terms; diagram (c) arises from the vacuum polarization calculated in Section 11.2; and (d) is the term calculated in Section 11.3.

11.3 Anomalous Magnetic Moments and Charge Radii

For our next example, we shall calculate the shift in the magnetic moment and the charge radius of an electron or muon due to lowest-order radiative corrections. The one-loop graphs and renormalization corrections for the photon–lepton vertex are shown in Figure 11.4. Of these graphs, those involving insertions in incoming or outgoing lepton lines vanish because the lepton is on the mass shell, as discussed in Section 10.3. The graph involving an insertion in the external photon line is the vacuum polarization effect, discussed in the previous section. This leaves one one-loop graph (the last in Figure 11.4) that needs to be calculated here:

$$\Gamma^\mu_{1\text{ loop}}(p',p) = \int d^4k \left[e\gamma^\rho(2\pi)^4\right]\left[\frac{-i}{(2\pi)^4}\,\frac{-i(\not{p}'-\not{k})+m}{(p'-k)^2+m^2-i\epsilon}\right][\gamma^\mu] \times\left[\frac{-i}{(2\pi)^4}\,\frac{-i(\not{p}-\not{k})+m}{(p-k)^2+m^2-i\epsilon}\right]\left[e\gamma_\rho(2\pi)^4\right]\left[\frac{-i}{(2\pi)^4}\,\frac{1}{k^2-i\epsilon}\right], \tag{11.3.1}$$

where p' and p are the final and initial lepton four-momenta, respectively. (The contribution of the vertex connecting the external photon line and the internal lepton line is taken as γ^μ, because a factor $e(2\pi)^4$ was extracted in defining Γ^μ.)

This integral has an obvious ultraviolet divergence, roughly like $\int d^4k/(k^2)^2$. Unlike the case of the vacuum polarization, here we do

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

not need a fancy regularization procedure like dimensional regularization to maintain the structure required by gauge invariance, because the photon is a neutral particle and so the integral may be rendered finite by suitable modifications of the photon propagator (for instance by including a factor $M^2/(k^2+M^2)$ with a large cutoff mass M), without having to introduce modifications elsewhere to maintain gauge invariance. In any case, as we shall see the anomalous magnetic moment and charge radii can be calculated without encountering any ultraviolet divergences at all. In what follows we shall leave the integrals for the vertex function in their infinite form, with it being understood that if necessary any divergent integrals can be expressed in terms of a cutoff mass M.

We start by combining denominators, using a repeated version of the Feynman trick described in the Appendix to this chapter

$$\frac{1}{ABC} = 2\int_0^1 dx \int_0^x dy \left[Ay + B(x-y) + C(1-x)\right]^{-3}. \tag{11.3.2}$$

Applied to the denominators in Eq. (11.3.1), this gives

$$\begin{aligned}
&\frac{1}{(p'-k)^2+m^2-i\epsilon}\,\frac{1}{(p-k)^2+m^2-i\epsilon}\,\frac{1}{k^2-i\epsilon} \\
&= 2\int_0^1 dx \int_0^x dy \left[\Big((p'-k)^2+m^2-i\epsilon\Big)y + \Big((p-k)^2+m^2-i\epsilon\Big)(x-y)\right. \\
&\qquad \left.+\Big(k^2-i\epsilon\Big)(1-x)\right]^{-3} \\
&= 2\int_0^1 dx \int_0^x dy \left[\Big(k-p'y-p(x-y)\Big)^2 + m^2x^2 + q^2y(x-y) - i\epsilon\right]^{-3},
\end{aligned} \tag{11.3.3}$$

where $q \equiv p - p'$ is the momentum transferred to the photon. Shifting the variable of integration

$$k \to k + p'y + p(x-y)$$

the integral (11.3.1) becomes

$$\begin{aligned}
\Gamma^{\mu}_{1\text{ loop}}(p',p) = \frac{2ie^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int \frac{d^4k}{\left[k^2 + m^2x^2 + q^2y(x-y) - i\epsilon\right]^3} \\
\times\gamma^{\rho}\Big[-i\Big(\not{p}'(1-y) - \not{k} - \not{p}(x-y)\Big) + m\Big]\gamma^{\mu} \\
\times\Big[-i\Big(\not{p}(1-x+y) - \not{k} - \not{p}'y\Big) + m\Big]\gamma_{\rho}.
\end{aligned} \tag{11.3.4}$$

Our next step is a Wick rotation. As explained in the previous section,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

the $-i\epsilon$ in the denominator dictates that when we rotate the k^0 contour of integration to the imaginary axis we must rotate counterclockwise, so that the integral over k^0 from $-\infty$ to $+\infty$ is replaced with an integral over imaginary values from $-i\infty$ to $+i\infty$, or equivalently over real values of $k^4 \equiv -ik^0$ from $-\infty$ to $+\infty$. We also exploit the rotational symmetry of the denominator in Eq. (11.3.4); we drop terms in the numerator of odd order in k, replace $k^\lambda k^\sigma$ with $\eta^{\lambda\sigma}k^2/4$, and replace the volume element $d^4k = idk^1dk^2dk^3dk^4$ with $2i\pi^2\kappa^3 d\kappa$, where κ is the Euclidean length of the four-vector k. Putting this all together, Eq. (11.3.4) now becomes

$$\begin{aligned}\Gamma^\mu_{1\text{ loop}}(p',p) = \frac{-4\pi^2e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa \Big\{ &-\kappa^2\gamma^\rho\gamma^\sigma\gamma^\mu\gamma_\sigma\gamma_\rho/4 \\ &+\gamma^\rho\Big[-i\Big(\not{p}'(1-y)-\not{p}(x-y)\Big)+m\Big]\gamma^\mu \\ &\times\Big[-i\Big(\not{p}(1-x+y)-\not{p}'y\Big)+m\Big]\gamma_\rho\Big\} \\ &\times\Big[\kappa^2+m^2x^2+q^2y(x-y)\Big]^{-3}. \end{aligned} \tag{11.3.5}$$

We are interested here only in the matrix element $\bar{u}'\Gamma^\mu u$ of the vertex function between Dirac spinors that satisfy the relations

$$\bar{u}'[i\not{p}'+m]=0\,, \qquad [i\not{p}+m]u=0\,.$$

We can therefore simplify this expression by using the anticommutation relations of the Dirac matrices to move all factors $\not{p}'$ to the left and all factors $\not{p}$ to the right, replacing them when they arrive on the right or left with im. After a straightforward but tedious calculation, Eq. (11.3.5) then becomes

$$\begin{aligned}\bar{u}'\Gamma^\mu_{\text{one loop}}(p',p)u = \frac{-4\pi^2e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa \\ \bar{u}'\Big\{\gamma^\mu\Big[-\kappa^2+2m^2(x^2-4x+2)+2q^2(y(x-y)+1-x)\Big] \\ +4im\,p'^\mu(y-x+xy)+4im\,p^\mu(x^2-xy-y)\Big\}u \\ \times\Big[\kappa^2+m^2x^2+q^2y(x-y)\Big]^{-3}. \end{aligned} \tag{11.3.6}$$

We next exploit the symmetry of the final factor under the reflection $y \to x-y$. Under this reflection, the functions $y-x+xy$ and x^2-xy-y that multiply p'^μ and p^μ are interchanged, so both may be replaced with their average:

$$\tfrac{1}{2}(y-x+xy)+\tfrac{1}{2}(x^2-xy-y) = -\tfrac{1}{2}x(1-x)\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

This gives finally

$$\bar{u}'\Gamma^{\mu}_{\text{one loop}}(p',p)u = \frac{-4\pi^2e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa$$
$$\times\, \bar{u}'\Big\{\gamma^\mu\Big[-\kappa^2 + 2m^2(x^2-4x+2) + 2q^2(y(x-y)+1-x)\Big]$$
$$-2im(p'^\mu + p^\mu)x(1-x)\Big\}u$$
$$\times\, \Big[\kappa^2 + m^2x^2 + q^2y(x-y)\Big]^{-3}. \tag{11.3.7}$$

Note that p^μ and p'^μ now enter only in the combination $p^\mu + p'^\mu$, as required by current conservation.

There are other diagrams that need to be taken into account. Of course, there is the zeroth-order term γ^μ in Γ^μ. The term proportional to $Z_2 - 1$ in the correction term (11.1.9) yields a term in Γ^μ

$$\Gamma^\mu_{\mathscr{L}_2} = (Z_2 - 1)\gamma^\mu. \tag{11.3.8}$$

Also, the effect of insertions of corrections to the external photon propagator is a term:

$$\Gamma^\mu_{\text{vac pol}}(p',p) = \frac{1}{(p'-p)^2 - i\epsilon}\Pi^{\mu\nu}(p'-p)\gamma_\nu. \tag{11.3.9}$$

The form of each of these terms is in agreement with the general result (10.6.10) (with $H(q^2) = 0$)

$$\bar{u}'\Gamma^\mu(p',p)u = \bar{u}'\left[\gamma^\mu F(q^2) - \frac{i}{2m}(p+p')^\mu G(q^2)\right]u. \tag{11.3.10}$$

To order e^2, the form factors are

$$F(q^2) = Z_2 + \pi(q^2) + \frac{4\pi^2e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa$$
$$\times \frac{\Big[\kappa^2 - 2m^2(x^2-4x+2) - 2q^2(y(x-y)+1-x)\Big]}{\Big[\kappa^2 + m^2x^2 + q^2y(x-y)\Big]^3}, \tag{11.3.11}$$

$$G(q^2) = \frac{-4\pi^2e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \frac{4m^2x(1-x)\kappa^3\, d\kappa}{\Big[\kappa^2 + m^2x^2 + q^2y(x-y)\Big]^3}, \tag{11.3.12}$$

where $\pi(q^2)$ is the vacuum polarization function (11.2.22).

The integral for the form factor $G(q^2)$ is finite as it stands:

$$G(q^2) = \frac{-e^2m^2}{4\pi^2}\int_0^1 dx \int_0^x dy \frac{x(1-x)}{m^2x^2 + q^2y(x-y)}. \tag{11.3.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

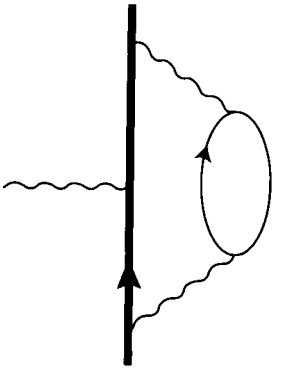

Figure 11.5. A two-loop diagram for the muon magnetic moment. Here the heavy straight line represents a muon; the light wavy lines are photons; and the other light lines are electrons. This diagram makes a relatively large contribution to the fourth-order muon gyromagnetic ratio, proportional to $\ln(m_\mu/m_e)$.

This makes it easy to calculate the anomalous magnetic moment. As noted in Section 10.6, it is only the γ^μ term that contributes to the magnetic moment, so the effect of radiative corrections is to multiply the Dirac value $e/2m$ of the magnetic moment by a factor $F(0)$. But the definition of e as the true lepton charge requires that

$$F(0) + G(0) = 1 , \tag{11.3.14}$$

so the magnetic moment may be expressed as

$$\mu = \frac{e}{2m}\Big(1 - G(0)\Big) . \tag{11.3.15}$$

From Eq. (11.3.13), we find

$$-G(0) = \frac{e^2}{8\pi^2} = 0.001161 . \tag{11.3.16}$$

This is the famous $\alpha/2\pi$ correction first calculated by Schwinger.[5]

Of course, this is only the first term in the radiative corrections to the magnetic moment. Even in just the next order, fourth order in e, there are so many terms that the calculations become quite complicated. However, because of the large muon–electron mass ratio, there is one fourth-order term in the magnetic moment of the *muon* that is somewhat larger than any of the others. It arises from the insertion of an *electron* loop in the virtual photon line of the second-order diagram, as shown in Figure 11.5. The effect of this electron loop is to change the photon propagator $1/k^2$ in Eq. (11.3.1) to $(1 + \pi_e(k^2))/k^2$, where $\pi_e(k^2)$ is given by Eq. (11.2.22),

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

but with the mass m taken as the *electron* mass:

$$\pi_e(k^2) = \frac{e^2}{2\pi^2} \int_0^1 x(1-x) \ln\left(1 + \frac{k^2 x(1-x)}{m_e^2}\right) dx .$$

Inspection of Eq. (11.3.12) shows that in calculating the muon magnetic moment the effective cutoff on the virtual photon momentum k is m_μ. The ratio m_μ/m_e is so large that for k^2 of order m_μ^2 we may approximate

$$\pi_e(k^2) \simeq \frac{e^2}{2\pi^2} \int_0^1 dx\, x(1-x) \ln(m_\mu^2/m_e^2) = \frac{e^2}{12\pi^2} \ln(m_\mu^2/m_e^2) \qquad (11.3.17)$$

with the neglected terms having coefficients of order unity in place of $\ln(m_\mu^2/m_e^2)$. Since this is a constant, the change in $-G(0)$ produced by adding an electron loop in the virtual photon line is simply given by multiplying our previous result (11.3.16) for $-G(0)$ by Eq. (11.3.17), so that now

$$\mu_\mu = \frac{e}{2m_\mu}\left(1 + \frac{e^2}{8\pi^2} + \frac{e^4}{96\pi^4}\left[\ln\frac{m_\mu^2}{m_e^2} + O(1)\right]\right) . \qquad (11.3.18)$$

(As we shall see in Volume II, this argument is a primitive version of the method of the renormalization group.) The result (11.3.18) may be compared with the full fourth-order result:[6]

$$\mu_\mu = \frac{e}{2m_\mu}\left(1 + \frac{e^2}{8\pi^2} + \frac{e^4}{96\pi^4}\left[\ln\frac{m_\mu^2}{m_e^2} - \frac{25}{6} + \frac{197}{24} + \frac{\pi^2}{2} + \frac{9\zeta(3)}{2} - 3\pi^2 \ln 2 + O\left(\frac{m_e}{m_\mu}\right)\right]\right) . \qquad (11.3.19)$$

It turns out that the '$O(1)$' terms multiplying $e^4/96\pi^4$ add up to -6.137, which is not very much smaller than $\ln(m_\mu^2/m_e^2) = 10.663$, so the approximation (11.3.18) gives the fourth-order terms only to a factor of order 2. The correct fourth-order result (11.3.19) gives $\mu_\mu = 1.00116546\, e/2m_\mu$, in comparison with the second-order result $\mu_\mu = 1.001161\, e/2m_\mu$ and the current experimental value,[7] $\mu_\mu = 1.001165923(8) e/2m_\mu$.

Now let us turn to the other form factor. The integral in Eq. (11.3.11) for $F(q^2)$ has an ultraviolet divergence. However, in order to satisfy the charge-non-renormalization condition (11.3.14), it is necessary that Z_2 take the value

$$Z_2 = 1 + \frac{e^2}{8\pi^2} - \frac{4\pi^2 e^2}{(2\pi)^4} \int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa \times \frac{\kappa^2 - 2m^2(x^2 - 4x + 2)}{\left[\kappa^2 + m^2 x^2\right]^3} . \qquad (11.3.20)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

(Recall that $\pi(0) = 0$.) This is itself ultraviolet divergent, with an infinite part

$$(Z_2 - 1)_\infty = -\frac{e^2}{8\pi^2}\int^\infty \frac{d\kappa}{\kappa}\,. \tag{11.3.21}$$

Inserting Eq. (11.3.20) back into Eq. (11.3.11) gives

$$F(q^2) = 1 + \frac{e^2}{8\pi^2} + \pi(q^2) + \frac{4\pi^2 e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \int_0^\infty \kappa^3 d\kappa \times \left\{ \frac{\left[\kappa^2 - 2m^2(x^2 - 4x + 2) - 2q^2(y(x-y) + 1 - x)\right]}{\left[\kappa^2 + m^2x^2 + q^2y(x-y)\right]^3} - \frac{\left[\kappa^2 - 2m^2(x^2 - 4x + 2)\right]}{\left[\kappa^2 + m^2x^2\right]^3} \right\}\,. \tag{11.3.22}$$

The integral over κ is now convergent:

$$F(q^2) = 1 + \frac{e^2}{8\pi^2} + \pi(q^2) + \frac{2\pi^2 e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \times \left\{ \frac{-m^2[x^2 - 4x + 2] - q^2[y(x-y) + 1 - x]}{m^2x^2 + q^2y(x-y)} + \frac{x^2 - 4x + 2}{x^2} - \ln\left[\frac{m^2x^2 + q^2y(x-y)}{m^2x^2}\right] \right\}\,. \tag{11.3.23}$$

However, we see that the integral over x and y now diverges logarithmically at $x = 0$ and $y = 0$, because there are two powers of x and/or y in the denominators, and just two differentials $dx\,dy$ in the numerator. This divergence can be traced to the vanishing of the denominator $[\kappa^2 + m^2x^2 + q^2y(x-y)]^3$ in Eq. (11.3.11) at $x = 0$, $y = 0$, and $\kappa = 0$. Because this infinity comes from the region of small rather than large κ, it is termed an *infrared divergence* rather than an ultraviolet divergence.

We shall give a comprehensive treatment of the infrared divergences in Chapter 13. It will be shown there that infrared divergences in the cross-section for processes like electron–electron scattering, such as those that are introduced by the infrared divergence in the electron form factor $F(q^2)$, are cancelled when we include the emission of low-energy photons as well as elastic scattering. Also, as we shall see in Chapter 14, when we calculate radiative corrections to atomic energy levels the infrared divergence in $F(q^2)$ is cut off because the bound electron is not exactly on the free-particle mass shell. For the present we shall continue our calculation by simply introducing a fictitious photon mass μ to cut off the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

infrared divergence in $F(q^2)$, leaving it for Chapter 14 to see how to use this result.

With a photon mass μ, the denominator $k^2 - i\epsilon$ in Eq. (11.3.1) would be replaced with $k^2 + \mu^2 - i\epsilon$. The effect would then be to add a term $\mu^2(1-x)$ to the cubed quantity in the denominators of Eqs. (11.3.3)–(11.3.7), (11.3.11), (11.3.20), and (11.3.22). Eq. (11.3.23) then is replaced with

$$F(q^2) = 1 + \frac{e^2}{8\pi^2} + \pi(q^2) + \frac{2\pi^2 e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \times \left\{ \frac{-m^2[x^2-4x+2] - q^2[y(x-y)+1-x]}{m^2x^2 + q^2y(x-y) + \mu^2(1-x)} + \frac{m^2[x^2-4x+2]}{m^2x^2+\mu^2(1-x)} - \ln\left[\frac{m^2x^2 + q^2y(x-y) + \mu^2(1-x)}{m^2x^2+\mu^2(1-x)}\right]\right\} . \tag{11.3.24}$$

This integral is now completely convergent. It can be expressed in terms of Spence functions, but the result is not particularly illuminating. For our purposes in Chapter 14, it will be sufficient to calculate the behavior of $F(q^2)$ for small q^2. We already know from the Ward identity that $F(0) = 1 - G(0) = 1 + e^2/8\pi^2$, so let us consider the first derivative $F'(q^2)$ at $q^2 = 0$. According to Eq. (11.3.24), this is

$$F'(0) = \pi'(0) + \frac{2\pi^2 e^2}{(2\pi)^4}\int_0^1 dx \int_0^x dy \times \left\{ -\frac{2y(x-y)+1-x}{m^2x^2+\mu^2(1-x)} + \frac{m^2[x^2-4x+2]y(x-y)}{[m^2x^2+\mu^2(1-x)]^2}\right\} . \tag{11.3.25}$$

The vacuum polarization contribution is given by Eq. (11.2.22) as

$$\pi'(0) = \frac{e^2}{60\pi^2 m^2} . \tag{11.3.26}$$

Dropping all terms proportional to powers of μ/m in Eq. (11.3.25), we then have*

$$F'(0) = \frac{e^2}{24\pi^2 m^2}\left[\ln\left(\frac{\mu^2}{m^2}\right) + \frac{2}{5} + \frac{1}{4}\right] \tag{11.3.27}$$

with the term $\frac{2}{5}$ the contribution of vacuum polarization. On the other hand, Eq. (11.3.13) shows that $G(q^2)$ has a finite derivative at $q^2 = 0$,

$$G'(0) = \frac{e^2}{48\pi^2 m^2} . \tag{11.3.28}$$

* The y-integral is trivial. The x-integral is most easily calculated in the limit $\mu \ll m$ by dividing the range of integration into two parts, one from 0 to s, where $\mu/m \ll s \ll 1$, and the second from s to 1.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

These results are most conveniently expressed in terms of the charge form factor $F_1(q^2)$, defined by the alternative representation (10.6.15) of the vertex function

$$\bar{u}(\mathbf{p}',\sigma')\Gamma^\mu(p',p)u(\mathbf{p},\sigma) = \bar{u}(\mathbf{p}',\sigma')\left[\gamma^\mu F_1(q^2) + \tfrac{1}{2}i[\gamma^\mu,\gamma^\nu](p'-p)_\nu F_2(q^2)\right]u(\mathbf{p},\sigma)\,. \quad (11.3.29)$$

According to Eqs. (10.6.17) and (10.6.18),

$$F_1(q^2) = F(q^2) + G(q^2)\,. \quad (11.3.30)$$

For $|q^2| \ll m^2$, this form factor is approximately

$$F_1(q^2) \simeq 1 + \frac{e^2}{24\pi^2}\left(\frac{q^2}{m^2}\right)\left[\ln\left(\frac{\mu^2}{m^2}\right) + \frac{2}{5} + \frac{3}{4}\right]\,. \quad (11.3.31)$$

This may be expressed in terms of a *charge radius* a, defined by the limiting behavior of the charge form factor for $q^2 \to 0$:

$$F_1(q^2) \to 1 - q^2a^2/6\,. \quad (11.3.32)$$

(This definition is motivated by the fact that the average of $\exp(i\mathbf{q}\cdot\mathbf{x})$ over a spherical shell of radius a goes as $1 - \mathbf{q}^2a^2/6$ for $\mathbf{q}^2a^2 \ll 1$.) We see that the charge radius of the electron is given by

$$a^2 = -\frac{e^2}{4\pi^2m^2}\left[\ln\left(\frac{\mu^2}{m^2}\right) + \frac{2}{5} + \frac{3}{4}\right]\,. \quad (11.3.33)$$

We will see in Chapter 14 that for electrons in atoms the role of the photon mass is played by an effective infrared cutoff that is much less than m, so the logarithm here is large and negative, yielding a positive value for a^2.

11.4 Electron Self-Energy

We conclude this chapter with a calculation of the electron self-energy function. This by itself does not have any direct experimental implications, but some of the results here will be useful in Chapter 14 and Volume II.

As in Section 10.3, we define $i(2\pi)^4[\Sigma^*(p)]_{\beta,\alpha}$ as the sum of all graphs with one incoming and one outgoing electron line carrying momenta p and Dirac indices α and β respectively, with the asterisk indicating that we exclude diagrams that can be disconnected by cutting through some internal electron line, and with propagators omitted for the two external lines. The complete electron propagator is then given by the sum

$$[-i(2\pi)^{-4}S'(p)] = [-i(2\pi)^{-4}S(p)] + [-i(2\pi)^{-4}S(p)][i(2\pi)^4\Sigma^*(p)][-i(2\pi)^{-4}S(p)] + \ldots\,, \quad (11.4.1)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

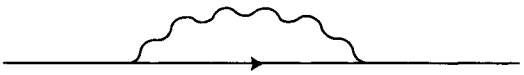

Figure 11.6. The one-loop diagram for the electron self-energy function. As usual, the straight line represents an electron; the wavy line is a photon.

where

$$S(p) \equiv \frac{-i\not{p} + m_e}{p^2 + m_e^2 - i\epsilon} . \tag{11.4.2}$$

The sum is trivial, and gives

$$S'(p) = [i\not{p} + m_e - \Sigma^*(p) - i\epsilon]^{-1} . \tag{11.4.3}$$

In lowest order there is a one-loop contribution to Σ^*, given by Figure 11.6:

$$i(2\pi)^4 \Sigma^*_{1\text{ loop}}(p) = \int d^4k \left[\frac{-i}{(2\pi)^4} \frac{\eta_{\rho\sigma}}{k^2 - i\epsilon}\right] \times [(2\pi)^4 e\gamma^\rho] \left[\frac{-i}{(2\pi)^4} \frac{-i\not{p} + i\not{k} + m_e}{(p-k)^2 + m_e^2 - i\epsilon}\right] [(2\pi)^4 e\gamma^\sigma]$$

or more simply

$$\Sigma^*_{1\text{ loop}}(p) = \frac{ie^2}{(2\pi)^4} \int d^4k \left[\frac{1}{k^2 - i\epsilon}\right] \times \left[\frac{\gamma^\rho(-i\not{p} + i\not{k} + m_e)\gamma_\rho}{(p-k)^2 + m_e^2 - i\epsilon}\right] . \tag{11.4.4}$$

(This is in Feynman gauge; amplitudes with charged particles off the mass shell are not gauge-invariant.) For use in our calculation of the Lamb shift, it will be convenient to use a method of regularization introduced by Pauli and Villars.[8] We replace the photon propagator $(k^2 - i\epsilon)^{-1}$ with

$$\frac{1}{k^2 - i\epsilon} - \frac{1}{k^2 + \mu^2 - i\epsilon} ,$$

so that the electron self-energy function is

$$\Sigma^*_{1\text{ loop}}(p) = \frac{ie^2}{(2\pi)^4} \int d^4k \left[\frac{1}{k^2 - i\epsilon} - \frac{1}{k^2 + \mu^2 - i\epsilon}\right] \times \left[\frac{\gamma^\rho(-i\not{p} + i\not{k} + m_e)\gamma_\rho}{(p-k)^2 + m_e^2 - i\epsilon}\right] . \tag{11.4.5}$$

Later we can drop the regulator by letting the regulator mass μ go to infinity. In Chapter 14 we will also be interested in the case where $\mu \ll m_e$.

We again use the Feynman trick to combine denominators, and recall

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

that $\gamma^\rho\gamma^\kappa\gamma_\rho = -2\gamma^\kappa$ and $\gamma^\rho\gamma_\rho = 4$. This gives

$$\Sigma^*_{1\text{ loop}}(p) = \frac{ie^2}{(2\pi)^4}\int d^4k\ [2i(\not{p} - \not{k}) + 4m_e] \times \int_0^1 dx \left[\frac{1}{((k-px)^2 + p^2x(1-x) + m_e^2x - i\epsilon)^2} - \frac{1}{((k-px)^2 + p^2x(1-x) + m_e^2x + \mu^2(1-x) - i\epsilon)^2}\right] \quad (11.4.6)$$

Shifting the variable of integration $k \to k + px$ and rotating the contour of integration gives

$$\Sigma^*_{1\text{ loop}}(p) = \frac{-2\pi^2e^2}{(2\pi)^4}\int_0^1 dx\ [2i(1-x)\not{p} + 4m_e]\int_0^\infty d\kappa\ \kappa^3 \times \left[\frac{1}{(\kappa^2 + p^2x(1-x) + m_e^2x)^2} - \frac{1}{(\kappa^2 + p^2x(1-x) + m_e^2x + \mu^2(1-x))^2}\right]. \quad (11.4.7)$$

The κ-integral is trivial:

$$\Sigma^*_{1\text{ loop}}(p) = \frac{-\pi^2e^2}{(2\pi)^4}\int_0^1 dx\ [2i(1-x)\not{p} + 4m_e] \times \ln\left(\frac{p^2x(1-x) + m_e^2x + \mu^2(1-x)}{p^2x(1-x) + m_e^2x}\right). \quad (11.4.8)$$

The interaction (11.1.9) also contributes a renormalization counterterm $-(Z_2 - 1)(i\not{p} + m_e) + Z_2\delta m_e$ in $\Sigma^*(p)$, with Z_2 and δm_e determined by the condition that the complete propagator $S'(p)$ regarded as a function of $i\not{p}$ should have a pole at $i\not{p} = -m_e$ with residue unity. (As we shall see in the next chapter, this makes Σ^* finite as $\mu \to \infty$ to all orders in e.) In lowest order, this gives

$$\delta m_e = -\Sigma^*_{1\text{ loop}}\Big|_{i\not{p}=-m_e} = \frac{2m_e\pi^2e^2}{(2\pi)^4}\int_0^1 dx\ [1+x]\ \ln\left(\frac{m_e^2x^2 + \mu^2(1-x)}{m_e^2x^2}\right), \quad (11.4.9)$$

$$Z_2 - 1 = -i\left.\frac{\partial\Sigma^*_{1\text{ loop}}}{\partial\not{p}}\right|_{i\not{p}=-m_e} = -\frac{2\pi^2e^2}{(2\pi)^4}\int_0^1 dx\left\{(1-x)\ln\left(\frac{m_e^2x^2 + \mu^2(1-x)}{m_e^2x^2}\right) - \frac{2\mu^2(1-x)^2(1+x)}{x(m_e^2x^2 + \mu^2(1-x))}\right\}. \quad (11.4.10)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

(To this order, we do not distinguish between δm_e and $Z_2 \delta m_e$.) Dropping terms that vanish for $\mu^2 \to \infty$, Eqs. (11.4.8)–(11.4.10) yield

$$\Sigma^*_{1\text{ loop}}(p) = \frac{-\pi^2 e^2}{(2\pi)^4} \int_0^1 dx\, [2i(1-x)\not{p} + 4m_e] \ln\left(\frac{\mu^2(1-x)}{p^2 x(1-x) + m_e^2 x}\right), \tag{11.4.11}$$

$$\delta m_e = \frac{2m_e \pi^2 e^2}{(2\pi)^4} \int_0^1 dx\, [1+x] \ln\left(\frac{\mu^2(1-x)}{m_e^2 x^2}\right), \tag{11.4.12}$$

$$Z_2 - 1 = \frac{-2\pi^2 e^2}{(2\pi)^4} \int_0^1 dx \left\{(1-x)\ln\left(\frac{\mu^2(1-x)}{m_e^2 x^2}\right) - \frac{2(1-x^2)}{x}\right\}. \tag{11.4.13}$$

Inspection then shows that in the complete self-energy function the $\ln \mu^2$ terms cancel, leaving us with

$$\begin{aligned}\Sigma^*_{\text{order } e^2}(p) &= \Sigma^*_{1\text{ loop}}(p) - (Z_2 - 1)(i\not{p} + m_e) + Z_2 \delta m_e \\ &= \frac{-2\pi^2 e^2}{(2\pi)^4} \int_0^1 dx \left\{[i(1-x)\not{p} + 2m_e] \ln\left(\frac{m_e^2(1-x)}{p^2 x(1-x) + m_e^2 x}\right)\right. \\ &\quad -m_e[1+x]\ln\left(\frac{1-x}{x^2}\right) \\ &\quad \left.-(i\not{p} + m_e)\left[(1-x)\ln\left(\frac{1-x}{x^2}\right) - \frac{2(1-x^2)}{x}\right]\right\}. \end{aligned} \tag{11.4.14}$$

There is still a divergence from the behavior of the last term as $x \to 0$, which can be traced to the singular behavior of the integral over the photon momentum k in Eq. (11.4.5) at $k^2 = 0$, when we take p^2 at the point $p^2 = -m_e^2$ where we evaluated $Z_2 - 1$. Such infrared divergences will be discussed in detail in Chapter 13. For the present, the point that concerns us is that the ultraviolet divergence has cancelled.

* * *

The result (11.4.9) for δm_e is of some interest in itself. Note that $\delta m_e / m_e > 0$, as we would expect for the electromagnetic self energy due to the interaction of a charge with its own field. But unlike the classical estimates of electromagnetic self-energy by Poincaré, Abraham, and others,[9] Eq. (11.4.9) is only logarithmically divergent in the limit $\mu \to \infty$ where the cutoff is removed. In this limit:

$$\delta m_e \to \frac{6m_e \pi^2 e^2}{(2\pi)^4} \ln\left(\frac{\mu}{m_e}\right). \tag{11.4.15}$$

In our calculation of the Lamb shift in Section 14.3 we will be interested

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

in the opposite limit, $\mu \ll m_e$. Here Eq. (11.4.9) gives

$$\delta m_e \to \frac{e^2\mu}{8\pi}\left[1 - \frac{3\mu}{2\pi m_e} + \ldots\right] . \tag{11.4.16}$$

Appendix Assorted Integrals

In order to combine the denominators of N propagators, we need to replace a product like $D_1^{-1}D_2^{-1}\ldots D_N^{-1}$ with an integral of a function that involves a linear combination of $D_1, D_2, \ldots D_N$. For this purpose it is often convenient to make use of the formula

$$\frac{1}{D_1 D_2 \ldots D_N} = (N-1)! \int_0^1 dx_1 \int_0^{x_1} dx_2 \cdots \int_0^{x_{N-2}} dx_{N-1}$$
$$\times \left[D_1 x_{N-1} + D_2(x_{N-2} - x_{N-1}) + \cdots + D_N(1 - x_1)\right]^{-N} . \tag{11.A.1}$$

In this chapter we have used special cases of this formula for $N = 2$ and $N = 3$.

After combining denominators, shifting the four-momentum variable of integration, Wick rotating, and using four-dimensional rotational invariance, we commonly encounter integrals of the form

$$\int d^4k \, \frac{(k^2)^n}{(k^2 + v^2)^m}$$

with $(k^2 + v^2)^m$ coming from the combined propagator denominators, and $(k^2)^n$ coming from the propagator numerators and vertex momentum factors. This is divergent for $2n + 4 \geq 2m$, but the integral can be given a finite value by analytically continuing the spacetime dimensionality from 4 to a complex value d. To evaluate the resulting integral, we use the well-known formula

$$\int_0^\infty d\kappa \frac{\kappa^{\ell-1}}{(\kappa^2 + v^2)^m} = v^{\ell-2m} \frac{\Gamma(\ell/2)\,\Gamma(m - \ell/2)}{2\,\Gamma(m)} , \tag{11.A.2}$$

where $\ell = d + 2n$. We used this formula in the special cases $n = 0, m = 2$ and $n = 1, m = 2$ in Section 11.2.

Ultraviolet divergences manifest themselves in Eq. (11.A.2) as poles in the factor $\Gamma(m - \ell/2) = \Gamma(m - n - d/2)$ as $d \to 4$ with fixed integer n. For $2 + n = m$, this factor goes as

$$\Gamma\left(\frac{4-d}{2}\right) \to \frac{2}{d-4} + \gamma , \tag{11.A.3}$$

where $\gamma = 0.5772157\cdots$ is the Euler constant. The limiting behavior for $2 + n > m$ can be obtained from (11.A.3) and the recursion relation for Gamma functions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

Problems

1. Calculate the contributions to the vacuum polarization function $\pi(q^2)$ and to Z_3 of one-loop graphs involving a charged spinless particle of mass m_s. What effect does this have on the energy shift of the $2s$ state of hydrogen, if $m_s >> Z\alpha m_e$?
2. Suppose that a neutral scalar field ϕ of mass m_ϕ has an interaction $g\phi\bar{\psi}\psi$ with the electron field. To one-loop order, what effect does this have on the magnetic moment of the electron? On Z_2?
3. Consider a neutral scalar field ϕ with mass m_ϕ and self-interaction $g\phi^3/6$. To one-loop order, calculate the S-matrix element for scalar–scalar scattering.
4. To one-loop order, calculate the effect of the neutral scalar field of Problem 2 on the mass shift δm_e of the electron.

References

1. R. P. Feynman, *Phys. Rev.* **76**, 769 (1949).
2. G. C. Wick, *Phys. Rev.* **96**, 1124 (1954).
3. G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44**, 189 (1972).
4. E. A. Uehling, *Phys. Rev.* **48**, 55 (1935). The one-loop function $\pi(q^2)$ was first given for $q^2 \neq 0$ by J. Schwinger, *Phys. Rev.* **75**, 651 (1949).
5. J. Schwinger, *Phys. Rev.* **73**, 416 (1948).
6. This is calculated (including terms that vanish for $m_e \ll m_\mu$) by H. Suura and E. Wichmann, *Phys. Rev.* **105**, 1930 (1957); A. Petermann, *Phys. Rev.* **105**, 1931 (1957); H. H. Elend, *Phys. Lett.* **20**, 682 (1966); **21**, 720 (1966); G. W. Erickson and H. H. T. Liu, UCD-CNL-81 report (1968).
7. J. Bailey *et al.* (CERN–Mainz–Daresbury Collaboration), *Nucl. Phys.* **B150**, 1 (1979). These experiments are done by observing the precession of the muon spin in a storage ring.
8. W. Pauli and F. Villars, *Rev. Mod. Phys.* **21**, 434 (1949). Also see J. Rayski, *Phys. Rev.* **75**, 1961 (1949).
9. See, e.g., A. I. Miller, *Theory of Relativity — Emergence (1905) and Early Interpretation (1905–1911)* (Addison-Wesley, Reading, MA, 1981): Chapter 1.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.013

12
General Renormalization Theory

We saw in the previous chapter that calculations in quantum electrodynamics involving one-loop graphs yield divergent integrals over momentum space, but that these infinities cancel when we express all parameters of the theory in terms of 'renormalized' quantities, such as the masses and charges that are actually measured. In 1949 Dyson[1] sketched a proof that this cancellation would take place to all orders in quantum electrodynamics. It was immediately apparent (and will be shown here in Sections 12.1 and 12.2) that Dyson's arguments apply to a larger class of theories with finite numbers of relatively simple interactions, the so-called *renormalizable* theories, of which quantum electrodynamics is just one simple example.

For some years it was widely thought that any sensible physical theory would have to take the form of a renormalizable quantum field theory. The requirement of renormalizability played a crucial role in the development of the modern 'standard model' of weak, electromagnetic, and strong interactions. But as we shall see here, the cancellation of ultraviolet divergences does not really depend on renormalizability; as long as we include every one of the infinite number of interactions allowed by symmetries, the so-called non-renormalizable theories are actually just as renormalizable as renormalizable theories.

It is generally believed today that the realistic theories that we use to describe physics at accessible energies are what are known as 'effective field theories.' As discussed in Section 12.3, these are low-energy approximations to a more fundamental theory that may not be a field theory at all. Any effective field theory necessarily includes an infinite number of non-renormalizable interactions. Nevertheless, as discussed in Sections 12.3 and 12.4, we expect that at sufficiently low energy all the non-renormalizable interactions in such effective field theories are highly suppressed. Renormalizable theories like quantum electrodynamics and the standard model thus retain their special status in physics, though for reasons that are somewhat different from those that originally motivated the assumption of renormalizability in these theories.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

12.1 Degrees of Divergence

Let us consider a very general sort of theory, containing interactions of varying types labelled i. Each interaction may be characterized by the number n_{if} of fields of each type f, and by the number d_i of derivatives acting on these fields.

We will start by calculating the 'superficial degree of divergence' D of an arbitrary connected one-particle irreducible Feynman diagram in such a theory. This is the number of factors of momentum in the numerator minus the number in the denominator of the integrand, plus four for every independent four-momentum over which we integrate. The superficial divergence is the actual degree of divergence of the integration over the region of momentum space in which the momenta of all internal lines go to infinity together. That is, if $D > 0$, then the part of the amplitude where all internal momenta go to infinity with a common factor $\kappa \to \infty$ will diverge like

$$\int^{\infty} \kappa^{D-1} d\kappa \,. \tag{12.1.1}$$

In the same sense, an integral with degree of divergence $D = 0$ is logarithmically divergent, and an integral with $D < 0$ is convergent, at least as far as this region of momentum space is concerned. We will come back later to the problem posed by subintegrations that behave worse than the integral over this region.

To calculate D, we will need to know the following about the diagram:

$$\begin{aligned} I_f &\equiv \text{number of internal lines of field type } f \,, \\ E_f &\equiv \text{number of external lines of field type } f \,, \\ N_i &\equiv \text{number of vertices of interaction type } i \,. \end{aligned}$$

We will write the asymptotic behavior of the propagator $\Delta_f(k)$ of a field of type f in the form

$$\Delta_f(k) \sim k^{-2+2s_f} \,. \tag{12.1.2}$$

Looking back at Chapter 6, we see that $s_f = 0$ for scalar fields, $s_f = \frac{1}{2}$ for Dirac fields, and $s_f = 1$ for massive vector fields. More generally, it can be shown that for massive fields of Lorentz transformation type (A, B), we have $s_f = A + B$. Speaking loosely, we may call s_f the 'spin.' However, dropping terms that because of gauge invariance have no effect, the effective photon propagator $\eta_{\mu\nu}/k^2$ has $s_f = 0$. A similar result applies to a massive vector field coupled to a conserved current, provided the current does not depend on the vector field. It can also be shown that, in the same sense, the graviton field $g_{\mu\nu}$ has a propagator also with $s_f = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

According to (12.1.2), the propagators make a total contribution to D equal to

$$\sum_f I_f(2s_f - 2)\,. \tag{12.1.3}$$

Also, the derivatives in each interaction of type i introduce d_i momentum factors into the integrand, yielding a total contribution to D equal to

$$\sum_i N_i\, d_i\,. \tag{12.1.4}$$

Finally, we need the total number of independent momentum variables of integration. Each internal line can be labelled with a four-momentum, but these are not all independent; the delta function associated with each vertex imposes a linear relation among these internal momenta, except that one delta function only serves to enforce conservation of the external momenta. Thus, the momentum space integration volume elements contribute to D a term

$$4\left[\sum_f I_f - \left(\sum_i N_i - 1\right)\right]\,, \tag{12.1.5}$$

which, of course, is just four times the number of independent loops in the diagram. Adding the contributions (12.1.3), (12.1.4), and (12.1.5), we find

$$D = \sum_f I_f(2s_f + 2) + \sum_i N_i(d_i - 4) + 4\,. \tag{12.1.6}$$

Eq. (12.1.6) is not very convenient as it stands, because it gives a value for D that seems to depend on the internal details of the Feynman diagram. Fortunately, it can be simplified by using the topological identities

$$2I_f + E_f = \sum_i N_i\, n_{if}\,. \tag{12.1.7}$$

(Each internal line contributes two of the lines attached to vertices, while each external line contributes only one.) Using Eq. (12.1.7) to eliminate I_f, we see that Eq. (12.1.6) becomes

$$D = 4 - \sum_f E_f(s_f + 1) - \sum_i N_i\Delta_i\,, \tag{12.1.8}$$

where Δ_i is a parameter characterizing interactions of type i:

$$\Delta_i \equiv 4 - d_i - \sum_f n_{if}(s_f + 1)\,. \tag{12.1.9}$$

This result could have been obtained by simple dimensional analysis, without considering the structure of Feynman diagrams. The propagator

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

of a field is a four-dimensional Fourier transform of the vacuum expectation value of a time-ordered product of a pair of free fields, so a conventionally normalized field f whose dimensionality* in powers of momentum is $\mathscr{D}_f$ will have a propagator of dimensionality $-4+2\mathscr{D}_f$. Hence if the propagator behaves like k^{-2+2s_f} when k is much larger than the mass, then the field must have a dimensionality with $-4+2\mathscr{D}_f = -2+2s_f$, or $\mathscr{D}_f = 1+s_f$. An interaction i with n_{if} such fields and d_i derivatives will then have dimensionality $d_i+\sum_f n_{if}(1+s_f)$. But the action must be dimensionless, so each term in the Lagrangian density must have dimensionality $+4$ to cancel the dimensionality -4 of d^4x. Hence the interaction must have a coupling constant of dimensionality $4-d_i-\sum_f n_{if}(1+s_f)$, which is just the parameter Δ_i. The momentum space amplitude corresponding to a connected Feynman graph with E_f external lines of type f is the Fourier transform over $4\sum_f E_f$ coordinates of a vacuum expectation value of the time-ordered product of fields with a total dimensionality $\sum_f E_f(1+s_f)$, so it has dimensionality $\sum_f E_f(-3+s_f)$. Of this dimensionality, -4 comes from a momentum space delta function, and $\sum_f E_f(-2+2s_f)$ is the dimensionality of the propagators for the external lines, so the momentum space integral itself together with all coupling constant factors has dimensionality

$$\sum_f E_f(-3+s_f) - (-4) - \sum_f E_f(-2+2s_f) = 4 - \sum_f E_f(s_f+1)\,.$$

The coupling constants for a given Feynman graph have total dimensionality $\sum_i N_i\Delta_i$, leaving the momentum space integral with dimensionality $4-\sum_f E_f(s_f+1)-\sum_i N_i\Delta_i$. As long as we are interested in the region of integration where all momenta go to infinity together, the degree of divergence of the momentum space integral is its dimensionality, thus justifying Eq. (12.1.8).

If all interactions have $\Delta_i \geq 0$, then Eq. (12.1.8) provides an upper bound on D that depends only on the numbers of external lines of each type, i.e., on the physical process whose amplitude is being calculated

$$D \leq 4 - \sum_f E_f(s_f+1)\,. \tag{12.1.10}$$

For example, in the simple version of quantum electrodynamics studied in the previous chapter, the Lagrangian included terms of the types shown in Table 12.1. All interactions here have $\Delta_i \geq 0$, and hence a Feynman diagram with E_γ external photon lines and E_e external Dirac lines will

* In this chapter, 'dimensionality' will always refer to the dimensionality in powers of mass or momentum, in units with $\hbar = c = 1$. We are using fields that are conventionally normalized, in the sense that the term in the free-field Lagrangian with the largest number of derivatives (which determines the asymptotic behavior of the propagator) has a dimensionless coefficient.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

Table 12.1. Terms in the Lagrangian density for quantum electrodynamics. Here d_i, $n_{i\gamma}$, and n_{ie} are the numbers of derivatives, photon fields, and electron fields in the interaction, and Δ_i is the dimensionality of the corresponding coefficient. (Recall that $s_\gamma = 0$, $s_e = \frac{1}{2}$.)

Interaction	d_i	$n_{i\gamma}$	n_{ie}	Δ_i
$-ie\bar{\psi}\not{A}\psi$	0	1	2	$4-1-3=0$
$-\frac{1}{4}(Z_3-1)F_{\mu\nu}F^{\mu\nu}$	2	2	0	$4-2-2=0$
$-(Z_2-1)\bar{\psi}\not{\partial}\psi$	1	0	2	$4-1-3=0$
$[-(Z_2-1)m+Z_2\delta m]\bar{\psi}\psi$	0	0	2	$4-3=1$

have superficial degree of divergence bounded by Eq. (12.1.10):

$$D \leq 4 - \tfrac{3}{2}E_e - E_\gamma \,. \tag{12.1.11}$$

Only a finite number of sets of external lines can yield superficially divergent integrals; these will be enumerated in Section 12.2. We are going to show that the limited number of divergences that appear in theories with $\Delta_i \geq 0$ for all interactions are automatically removed by a redefinition of a finite number of physical constants and a renormalization of fields. For this reason, such theories are called *renormalizable*. In Section 12.3 we will catalog all the renormalizable theories, and discuss the significance of renormalizability as a criterion for physical theories.

The term 'renormalizable' is also applied to individual interactions. Renormalizable interactions are those with $\Delta_i \geq 0$, whose coupling constants have positive or zero dimensionality. Sometimes one distinguishes between interactions with $\Delta_i = 0$, called simply renormalizable, and those with $\Delta_i > 0$, called *superrenormalizable*. Since adding additional fields or derivatives always lowers Δ_i, there can only be a finite number of renormalizable interactions involving fields of any given types. We have seen that all the interactions in the simplest version of quantum electrodynamics are renormalizable, with the $\bar{\psi}\psi$ terms superrenormalizable.

On the other hand, if any interaction has $\Delta_i < 0$, the degree of divergence (12.1.8) becomes larger and larger the more such vertices we include. No matter how large we take the various E_f, eventually with enough vertices of type i for which $\Delta_i < 0$, Eq. (12.1.8) will become positive (or zero), and the integral will diverge. Such interactions, whose couplings have negative-definite dimensionality, are called *non-renormalizable*;** the-

** In perturbative statistical mechanics, non-renormalizable interactions are called *irrelevant*, because they become less important in the limit of low energies. Renormalizable and super-renormalizable interactions are called *marginal* and *relevant*, respectively.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

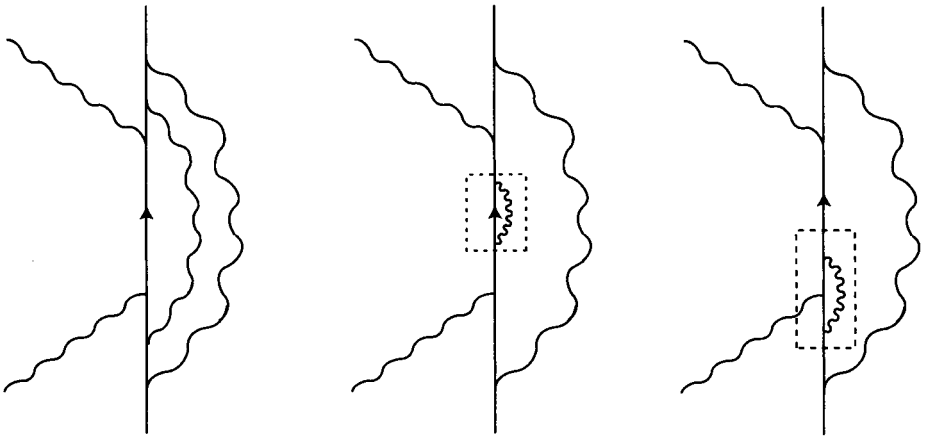

Figure 12.1. Some two-loop graphs for Compton scattering. Here straight lines are electrons; wavy lines are photons. The momentum space integral for diagram (a) is convergent, while for (b) and (c) it is divergent, due to the subintegration associated with the subgraphs surrounded by dotted lines.

ories with any non-renormalizable interactions are also known as non-renormalizable. But this does not mean that such theories are hopeless; we shall see that these divergences may also be absorbed into a redefinition of the parameters of the theory, but here we need an infinite number of couplings.

It should be kept in mind that we have here calculated the degree of divergence of Feynman diagrams arising only from regions of momentum space in which all internal four-momenta go to infinity together. Divergences can also arise from regions in which only the four-momenta of lines belonging to some subgraph go to infinity. For instance, in quantum electrodynamics Eq. (12.1.11) gives $D \leq -1$ for Compton scattering (where $E_e = 2$, $E_\gamma = 2$), and indeed graphs like Figure 12.1(a) are convergent, but a graph like Figure 12.1(b) or 12.1(c) is logarithmically divergent, because these graphs contain subgraphs (indicated by dotted boxes) with $D \geq 0$. We can think of the divergence of these graphs as being due to an anomalously bad asymptotic behavior that occurs when the eight components of the two independent internal four-momenta of these graphs go to infinity on a particular four-dimensional subspace, namely, that subspace in which the only four-momentum actually going to infinity is the one circulating in the loops that are inserted in the internal lines or at an electron–photon vertex.

It has been shown[2] that the requirement for the actual convergence of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

the amplitude corresponding to any graph is that power-counting should give $D < 0$ not only for the complete multiple integral for the whole amplitude, but also for any subintegration defined by holding any one or more linear combinations of the loop momenta fixed. (The graphs shown in Figures 12.1(b) and 12.1(c) fail this test because $D \geq 0$ for the subintegrations in which only the momenta for the loops within the dotted squares are integrated.) We will not repeat the rather long proof here, because it is well treated in earlier books,[3] and in any case the method of proof has little to do with how we actually do calculations. The next section will describe how this requirement is fulfilled.

12.2 Cancellation of Divergences

Consider a Feynman diagram, or part of a Feynman diagram, with positive superficial degree of divergence, $D \geq 0$. The part of the momentum space integral where all internal momenta go to infinity together will then diverge, like $\int^{\infty} k^{D-1}\, dk$. If we differentiate $D+1$ times with respect to any external momentum, we lower the net number of momentum factors in the integrand by $D+1$,* and hence render this part of the momentum space integral convergent. There may still be divergences arising from subgraphs, like those in Figures 12.1(b) and 12.1(c); for the moment we will ignore this possibility, returning to it later in this section. Since differentiation $D+1$ times renders the integral finite, it follows that the contribution of such a graph or subgraph can be written as a polynomial of order D in external momenta, with divergent coefficients, plus a finite remainder.

To see how this works without irrelevant complications, consider the logarithmically divergent one-dimensional integral

$$\mathscr{I}(q) \equiv \int_0^{\infty} \frac{dk}{k+q}$$

with $D = 1 - 1 = 0$. Differentiating once gives

$$\mathscr{I}'(q) \equiv -\int_0^{\infty} \frac{dk}{(k+q)^2} = -\frac{1}{q}\,,$$

so

$$\mathscr{I}(q) = -\ln q + c\,.$$

* For instance, if an internal scalar field line carries a momentum $k+p$, where p is a linear combination of external four-momenta and k is a four-momentum variable of integration, then the derivative of the propagator $[(k+p)^2+m^2]^{-1}$ with respect to p^{μ} gives $-2(k_{\mu}+p_{\mu})[(k+p)^2+m^2]^{-2}$, which goes as k^{-3} rather than k^{-2} for $k \to \infty$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

The constant c is obviously divergent, but the rest of the integral is perfectly finite. In exactly the same way, we can evaluate the $D = 1$ integral

$$\int_0^\infty \frac{k\,dk}{k+q} = a + bq + q \ln q$$

with divergent constants a and b.

Now, a polynomial term in external momenta is just what would be produced by adding suitable terms to the Lagrangian: if a graph with E_f external lines of type f has degree of divergence $D \geq 0$, then the ultraviolet divergent polynomial is the same as would be produced by adding various interactions i with $n_{if} = E_f$ fields of type f and $d_i \leq D$ derivatives. If there already are such interactions in the Lagrangian, then the ultraviolet divergences simply add corrections to the coupling constants of these interactions. Hence these infinities can be cancelled by including suitable infinite terms in these coupling constants. All that we ever measure is the sum of the bare coupling constant and the corresponding coefficient from one of the divergent polynomials, so if we demand that the sum equals the (presumably finite) measured value, then the bare coupling must automatically contain an infinity that cancels the infinity from the divergent integral over internal momenta. (One qualification: where the divergence occurs in a graph or subgraph with just two external lines, which appears as a radiative correction to a particle propagator, we must demand not that some effective coupling constant equals its measured value, but rather that the complete propagator has a pole at the same position and with the same residue as for free particles.) In this way, all infinities are absorbed into a redefinition of couplings constants, masses, and fields.

For this renormalization program to work, it is essential that the Lagrangian include *all* interactions that correspond to the ultraviolet divergent parts of Feynman amplitudes. The interactions in the Lagrangian are, of course, limited by various symmetry principles, such as Lorentz invariance, gauge invariance, etc., but these constrain the ultraviolet divergences in the same way. (It takes some work to prove that non-Abelian gauge symmetries constrain infinities in the same way that they constrain interactions. This will be shown in Volume II.) In the general case, there are no other limitations on the ultraviolet divergences, *so the Lagrangian must include every possible term consistent with symmetry principles.* (There are exceptions to this rule in supersymmetric theories.[4])

But there is an important class of theories with only a finite number of interactions, where the renormalization program also works. These are the so-called renormalizable theories, whose interactions all have $\Delta_i \geq 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

Eq. (12.1.8) then gives

$$D \leq 4 - \sum_f E_f(s_f + 1)\,,$$

so divergent polynomials arise in only a limited number of Feynman graphs or subgraphs: those with few enough external lines so that $D \geq 0$. The contribution of such divergent polynomials is just the same as would be produced by replacing the divergent graph or subgraph with a single vertex arising from a term in the Lagrangian with E_f fields of type f and $0,1,\cdots D$ derivatives. But, comparing with Eq. (12.1.9), we see that *these are precisely the same as the interactions that satisfy the renormalizability requirement* $\Delta_i \geq 0$, or in other words,

$$0 \leq d_i \leq 4 - \sum_f n_{if}(s_f + 1)\,.$$

In order for all infinities to cancel in a renormalizable theory, it is usually necessary that *all* renormalizable interactions that are allowed by symmetries must actually appear in the Lagrangian.** For instance, if there is a scalar (or pseudoscalar) field ϕ and fermion field ψ with interactions $\bar{\psi}\psi\phi$ (or $\bar{\psi}\gamma_5\psi\phi$) then we cannot exclude an interaction ϕ^4; otherwise there would be no counterterm to cancel the logarithmic divergence arising from fermion loops with four attached scalar or pseudoscalar lines.

Let's see in more detail how the cancellation of infinities works in the simplest version of quantum electrodynamics. Eq. (12.1.11) shows that the only graphs or subgraphs that could possibly yield divergent integrals are the following:

$\mathbf{E_e = 2}$, $\mathbf{E_\gamma = 1}$

This is the electron–photon vertex $\Gamma^{(\ell)}_\mu(p',p)$. (The superscript ℓ indicates that this includes only contributions from graphs with loops.) It has $D = 0$, so its divergent part is momentum-independent. Lorentz invariance then only allows this divergent constant to be proportional to γ_μ, so

$$\Gamma^{(\ell)}_\mu = L\gamma_\mu + \Gamma^{(f)}_\mu \tag{12.2.1}$$

with L a logarithmically divergent constant, and $\Gamma^{(f)}_\mu$ finite. This does not uniquely define the constant L, since we can always move a finite term

** In addition, interactions and mass terms that are not allowed by global symmetries may appear in the Lagrangian, as long as they are superrenormalizable, that is, with $\Delta_i > 0$. This is because the presence of a superrenormalizable coupling lowers the degree of divergence, so that the symmetry breaking does not affect those divergences that are cancelled by the strictly renormalizable couplings with $\Delta_i = 0$. Note that it is the *bare* strictly renormalizable couplings that would exhibit the symmetry; renormalized couplings that are defined in terms of mass-shell matrix elements generally show the effect of symmetry breaking.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

$\delta L\,\gamma_\mu$ from $\Gamma_\mu^{(f)}$ to $L\gamma_\mu$. To complete the definition, we may note that as shown in Section 10.4, the mass-shell matrix element of $\Gamma_\mu(p,p)$ and hence of $\Gamma_\mu^{(f)}(p,p)$ between mass-shell Dirac spinors is proportional to the same matrix element of γ^μ, so we may define L by the prescription that

$$\bar{u}(\mathbf{p},\sigma')\Gamma_\mu^{(f)}(p,p)u(\mathbf{p},\sigma)=0 \tag{12.2.2}$$

for $p^2+m_e^2=0$.

$\mathbf{E_e=2},\ \mathbf{E_\gamma=0}$

This is the electron self-energy insertion $\Sigma^*(p)$. It has $D=1$, so its divergent part is linear in the momentum p^μ carried by the incoming and outgoing fermion. Lorentz invariance (including parity conservation) will only allow it to be a function of $\not{p}$, so we may write the loop contribution as

$$\Sigma^{(\ell)}(p)=A-(i\not{p}+m)B+\Sigma^{(f)}(\not{p})\,, \tag{12.2.3}$$

where A and B are divergent constants, and $\Sigma^{(f)}$ is finite. Again, this does not uniquely define the constants A and B, because we can always shift $\Sigma^{(f)}$ by a finite first-order polynomial in $\not{p}$. We will define A and B by the prescription that

$$\Sigma^{(f)}=\frac{\partial\Sigma^{(f)}}{\partial\not{p}}=0\quad\text{for}\quad i\not{p}=-m\,. \tag{12.2.4}$$

Actually, B is not a new divergent constant. As long as we use a regularization procedure that respects current conservation, Γ_μ and Σ will be related by the Ward identity (10.4.27)

$$\Gamma^\mu(p,p)=\gamma^\mu+i\frac{\partial}{\partial p_\mu}\,\Sigma(p)$$

and therefore

$$L\gamma_\mu+\Gamma_\mu^{(f)}(p,p)=B\gamma_\mu+i\frac{\partial\Sigma^{(f)}(p)}{\partial p^\mu}\,. \tag{12.2.5}$$

Taking the matrix element of this equation between $\bar{u}(\mathbf{p},\sigma')$ and $u(\mathbf{p},\sigma)$ and using Eqs. (12.2.2) and (12.2.4), we find

$$L=B\,. \tag{12.2.6}$$

$\mathbf{E_\gamma=2},\ \mathbf{E_e=0}$

This is the photon self-energy insertion $\Pi^*_{\mu\nu}(q)$. It has $D=2$, so its divergent part is a second-order polynomial in q. Lorentz invariance only allows $\Pi^*_{\mu\nu}$ to take the form of a linear combination of $\eta_{\mu\nu}$ and $q_\mu q_\nu$ with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

coefficients depending only on q^2, so the loop contributions take the form

$$\Pi^{(\prime)}_{\mu\nu}(q) = C_1\eta_{\mu\nu} + C_2\eta_{\mu\nu}q^2 + C_3 q_\mu q_\nu + \text{finite terms}\,,$$

where C_1, C_2, and C_3 are divergent constants. As long as we use a regularization technique that respects current conservation, we must have

$$q^\mu \Pi^{(\prime)}_{\mu\nu}(q) = 0\,.$$

The same must then be true for the divergent terms, so $C_1 q_\nu + (C_2+C_3)q^2 q_\nu$ must be finite for all q. It follows that C_1 and $C_2 + C_3$ must be finite, and can therefore be lumped into the finite part of $\Pi^{(\prime)}_{\mu\nu}(q)$. Thus

$$\Pi^{(\prime)}_{\mu\nu}(q) = (\eta_{\mu\nu}q^2 - q_\mu q_\nu)\left(C + \pi(q^2)\right)\,, \tag{12.2.7}$$

where $\pi(q^2)$ is finite and C is the sole remaining divergence in $\Pi^{(\prime)}_{\mu\nu}$. To pin down the definition of C, we may move any finite constant $\pi(0)$ into C, so that

$$\pi(0) = 0\,. \tag{12.2.8}$$

$\mathbf{E_\gamma = 4}$, $\mathbf{E_e = 0}$

This is the amplitude $M_{\mu\nu\rho\sigma}$ for scattering of light by light. It has $D = 0$, so using Lorentz invariance and Bose statistics, it may be written (there is no non-loop contribution)

$$M_{\mu\nu\rho\sigma} = K(\eta_{\mu\nu}\eta_{\rho\sigma} + \eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho}) \;+\; \text{finite terms}$$

with K a potentially divergent constant. However, current conservation gives

$$q^\mu M_{\mu\nu\rho\sigma} = 0$$

and so $K(q_\nu\eta_{\rho\sigma} + q_\rho\eta_{\nu\sigma} + q_\sigma\eta_{\nu\rho})$ is finite. In order for this to be true for $q \neq 0$, K must itself be finite. This is a nice example of the role of symmetry principles in the renormalization program; if K had turned out to be infinite it could not be removed by renormalization of the coupling constant for an interaction $(A_\mu A^\mu)^2$, because no such interaction is allowed by gauge invariance, but K *is* finite because of current conservation conditions that are imposed by gauge invariance.

$\mathbf{E_\gamma = 1}$, $\mathbf{E_e = 0}$ and $\mathbf{E_e = 1}$, $\mathbf{E_\gamma = 0, 1, 2}$

These have $D = 3$ and $D = \frac{5}{2}, \frac{3}{2}$ and $\frac{1}{2}$, respectively, but Lorentz invariance makes all such graphs vanish.

$\mathbf{E_\gamma = 3}$, $\mathbf{E_e = 0}$

This has $D = 1$, but vanishes because of charge-conjugation invariance.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

The reader will perhaps have noticed that the independent divergent constants A, B, C are in one-to-one correspondence with the independent parameters Z_2, Z_3, and δm in the counterterm part (11.1.9) of the Lagrangian for quantum electrodynamics. These counterterms make a direct contribution $Z_2\delta m - (Z_2 - 1)(i\not{p} + m)$ to $\Sigma^*(p)$. The requirement that the position and residue of the one-particle pole be the same as in the free-field propagator means we must choose Z_2 and δm so that the total $\Sigma^*(p)$ satisfies Eq. (12.2.4), i.e.,

$$Z_2\,\delta m = -A, \tag{12.2.9}$$

$$Z_2 - 1 = -B, \tag{12.2.10}$$

so that the complete electron self-energy insertion is just the finite function $\Sigma^{(f)}(p)$:

$$\Sigma(p) = \Sigma^{(f)}(p)\,. \tag{12.2.11}$$

Also, $\mathscr{L}_2$ makes a direct contribution to Γ_μ equal to $(Z_2 - 1)\gamma_\mu$. Using Eq. (12.2.6), we see that the full vertex is

$$\Gamma_\mu = \gamma_\mu + (Z_2 - 1)\gamma_\mu + \Gamma_\mu^{(f)} = \gamma_\mu + \Gamma_\mu^{(f)}\,. \tag{12.2.12}$$

This is not only finite, but satisfies the condition

$$\bar{u}(\mathbf{p},\sigma')\Gamma_\mu(p,p)u(\mathbf{p},\sigma) = \bar{u}(\mathbf{p},\sigma')\gamma_\mu u(\mathbf{p},\sigma')\,, \tag{12.2.13}$$

as can also be seen from Eqs. (10.6.13) and (10.6.14). Finally, $\mathscr{L}_2$ makes a contribution $-(Z_3 - 1)(q^2\eta_{\mu\nu} - q_\mu q_\nu)$ to $\Pi^*_{\mu\nu}(q)$. In order that the photon propagator should have a pole with the same residue as for free fields we need the coefficient of $q^2\eta_{\mu\nu} - q_\mu q_\nu$ in the total $\Pi_{\mu\nu}(q)$ to vanish, so

$$Z_3 = 1 + C \tag{12.2.14}$$

and the photon propagator is then finite:

$$\Pi_{\mu\nu}(q) = (\eta_{\mu\nu}q^2 - q_\mu q_\nu)\pi(q^2)\,. \tag{12.2.15}$$

So far, we have only checked that the divergences, arising from the region of momentum space in which all internal momenta are large (and with generic ratios), are polynomials in external momenta that are cancelled by suitable counterterms. Such graphs are called *superficially convergent*. Before we conclude that all ultraviolet divergences actually are removed by renormalization, we need to consider the ultraviolet divergences arising in higher-order graphs when some subset of the momentum space integration variables rather than all of them go to infinity. For instance, in quantum electrodynamics the superficial divergences in subintegrations come from subgraphs that are either photon self-energy parts Π^*, or electron self-energy parts Σ^*, or electron–electron–photon vertices Γ^μ. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

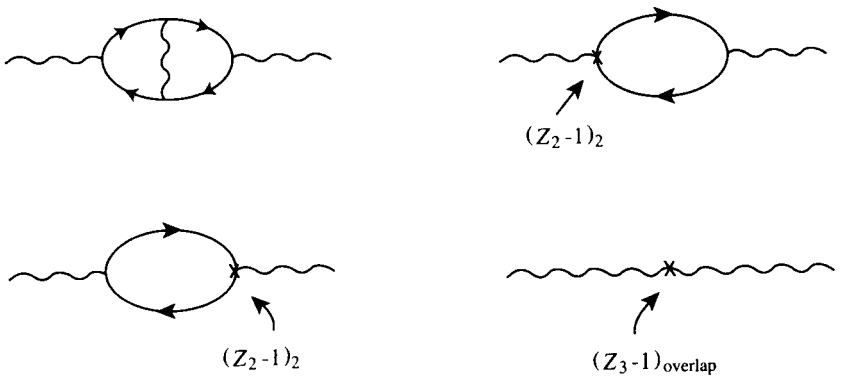

Figure 12.2. Some fourth-order graphs for the photon self-energy in quantum electrodynamics that involve overlapping divergences. Lines carrying arrows are electrons; wavy lines photons. The crosses mark the contribution of counterterms.

problem with such divergences is that they cannot be removed by differentiation with respect to external momenta; we are left with terms where the derivatives act only on internal lines in the parts of the graph which are *not* in the divergent subgraphs, and therefore do not reduce the degree of divergence of these subgraphs. As mentioned in the previous section, a graph or sum of graphs is actually convergent only if it and all its subintegrations are superficially convergent, in the sense of counting powers of momentum. But wherever such a divergent subgraph appears, it always comes accompanied with an infinite counterterm. In electrodynamics, these are the terms in Eq. (11.1.9): a term $-(Z_3-1)(q^2\eta_{\mu\nu}-q_\mu q_\nu)$ for each $\Pi^*_{\mu\nu}(q)$; a term $Z_2\delta m-(Z_2-1)(i\not{p}+m)$ for each $\Sigma^*(p)$; and a term $(Z_2-1)\gamma^\mu$ for each Γ^μ. Just as for the graph as a whole, these counterterms cancel the infinities from the divergent subgraphs.[1]

Unfortunately, there is a flaw in this simple argument — the possibility of overlapping divergences. That is, it is possible that two divergent subgraphs may share an internal line, so that we cannot regard them as independent divergent integrals. In quantum electrodynamics this happens only when two electron–electron–photon vertices overlap inside a photon or an electron self-energy insertion,† as shown in Figures 12.2 and 12.3.

A complete treatment of renormalization that takes account of overlap-

† The sharing of a line in two self-energy insertions or in a self-energy insertion and a vertex part would not leave enough external lines to attach such a subgraph to the rest of the diagram. Historically, the Ward identity (10.4.26) was used to by-pass the problem of overlapping divergences in the electron self-energy, by expressing the electron self-energy in terms of the vertex function, where overlapping divergences do not occur. This approach will not be followed here, as it is unnecessary, and in any case does not solve the problem for the self-energy of the photon or other neutral particles.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

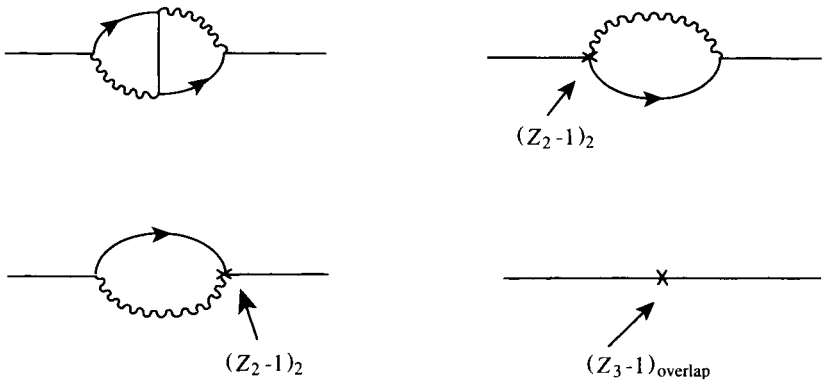

Figure 12.3. Some fourth-order graphs for the electron self-energy in quantum electrodynamics that involve overlapping divergences. Wavy lines are photons; other lines are electrons. The crosses mark the contribution of counterterms.

ping divergences should include a prescription for eliminating superficial ultraviolet divergences, not only in the overall integration but in all subintegrations as well, together with a proof that this prescription is (at least formally) implemented by renormalization of masses, fields, and coupling constants. The theorem of Ref. 2 then ensures that all Green's functions of renormalized fields are finite when expressed in terms of renormalized masses and couplings. The first proof that the renormalization of fields, couplings, and masses renders the whole integration and all its subintegrations superficially convergent was offered by Salam.[5] A more specific prescription for eliminating ultraviolet divergences was given by Bogoliubov and Parasiuk,[6] and corrected by Hepp,[7] and shown by them to be equivalent to a renormalization of fields, masses, and coupling constants. Finally, Zimmerman[8] proved that this prescription does eliminate all superficial divergences in the whole integration and all its subintegrations, and used the theorem of Ref. 2 to conclude from this that the renormalized Feynman momentum-space integrals are convergent.

Briefly, the 'BPHZ' prescription for eliminating superficial divergences requires that we consider all possible ways (called 'forests') of surrounding a whole graph and/or its subgraphs with boxes that may be nested within each other but do not overlap. (An example is given below.) For each forest we define a subtraction term by replacing the integrand for any subgraph of superficial divergence D within a box (starting with the innermost boxes and working outwards) with the first $D+1$ terms of its Taylor series expansion in the momenta flowing into or out of that box.‡

‡ As described here, this prescription applies to unrenormalizable as well as renormalizable theories. In renormalizable theories it implies that there is no subtraction unless the box contains one of the limited number of graphs corresponding to renormalizable terms in the Lagrangian.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

The subtracted Feynman amplitude is given by the original graph, minus all these subtraction terms, including the subtraction term for a forest consisting of a single box surrounding the whole graph.

It is fairly easy to see that the subtracted Feynman amplitude that is calculated in this way is the same as would be obtained by replacing all fields, coupling constants, and masses in the original Lagrangian by their renormalized counterparts. The difference between this procedure and the sort of renormalization we carried out in Chapter 11 is that the renormalized fields, coupling constants, and masses are defined in terms of amplitudes at an unconventional renormalization point, where all four-momenta vanish. (In this respect, the one-dimensional divergent integrals discussed at the start of this section provide an elementary example of the BPHZ method of separating divergent terms.) But there is nothing special about this renormalization point; once a Feynman amplitude is made convergent by expressing it in terms of these unconventional renormalized quantities, it can be rewritten in terms of conventionally renormalized fields, couplings, and masses without introducing new infinities.

It is not necessary to use the BPHZ subtraction prescription in practice. Replacing fields, masses, and couplings with their renormalized counterparts (defined using any convenient renormalization points) automatically provides counterterms that cancel all infinities. Instead of proving that the BPHZ subtraction prescription really does make all integrals converge, we shall just look at one example that shows how renormalization works, even in the presence of overlapping divergences.

Consider the fourth-order contribution to the photon self-energy insertion $\Pi^*_{\mu\nu}(q)$ shown in Figure 12.2. (The forests here consist of the whole integral over p and p', the subintegration over p alone, and the subintegration over p' alone.) Including the corresponding counterterms for the vertex parts and photon field renormalization, this has the value

$$\begin{aligned}[\Pi^*_{\mu\nu}(q)]_{\text{overlap}} = &-\frac{e^4}{(2\pi)^8}\int d^4p \int d^4p' \, \frac{1}{(p-p')^2 - i\epsilon} \\ &\times \operatorname{Tr}\{S(p')\gamma_\nu S(p'+q)\gamma^\rho S(p+q)\gamma_\mu S(p)\gamma_\rho\} \\ &-2(Z_2-1)_2 \frac{ie^2}{(2\pi)^4}\int d^4p \operatorname{Tr}\{\gamma_\nu S(p+q)\gamma_\mu S(p)\} \\ &-(Z_3-1)_{\text{overlap}}(q^2\eta_{\mu\nu} - q_\mu q_\nu), \end{aligned} \tag{12.2.16}$$

where $S(p) \equiv [-i\not{p} + m]/[p^2 + m^2 - i\epsilon]$; $(Z_2 - 1)_2$ is the term in $Z_2 - 1$ of second order in e; and $(Z_3 - 1)_{\text{overlap}}$ is a logarithmically divergent constant of fourth order in e that cancels the terms in $[\Pi^*_{\mu\nu}(q)]_{\text{overlap}}$ of second order in q^2. The factor 2 in the second term arises because there is a renormalization counterterm $Z_2 - 1$ for each of the two vertices in the second-order photon self-energy. Note, however, that the first term here

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

can be thought of *either* as the insertion of a vertex correction given by the p'-integral into a photon self-energy given by the p-integral, *or* as the insertion of a vertex correction given by the p-integral into a photon self-energy given by the p'-integral, but *not* as the insertion of two independent vertex corrections, because there is only one photon propagator.

To see how to handle the infinities in Eq. (12.2.16), note that

$$[(Z_2-1)_2+R_2]\gamma_\mu = \frac{ie^2}{(2\pi)^4}\int \frac{d^4p'}{p'^2-i\epsilon}\,\gamma_\rho\, S(p')\,\gamma_\mu\, S(p')\gamma^\rho\,, \qquad (12.2.17)$$

where R_2 is a finite remainder. (Lorentz invariance tells us that the integral on the right is proportional to γ_μ. The difference between this integral and $(Z_2-1)_2\gamma_\mu$ equals the complete renormalized electron–electron–photon vertex to second order in e at zero electron and photon momenta, and is therefore finite.) This allows us to rewrite Eq. (12.2.16) in the form

$$\begin{aligned}
[\Pi^*_{\mu\nu}(q)]_{\text{overlap}} = &-\frac{e^4}{(2\pi)^8}\int d^4p\int d^4p' \\
&\times\left[\frac{1}{(p-p')^2-i\epsilon}\mathrm{Tr}\,\{S(p')\,\gamma_\nu\,S(p'+q)\,\gamma^\rho\,S(p+q)\,\gamma_\mu\,S(p)\,\gamma_\rho\}\right. \\
&\quad -\frac{1}{p'^2-i\epsilon}\mathrm{Tr}\,\{S(p')\,\gamma_\nu\,S(p')\,\gamma^\rho\,S(p+q)\,\gamma_\mu\,S(p)\,\gamma_\rho\} \\
&\quad \left.-\frac{1}{p^2-i\epsilon}\mathrm{Tr}\,\{S(p')\,\gamma_\nu\,S(p'+q)\,\gamma^\rho\,S(p)\,\gamma_\mu\,S(p)\,\gamma_\rho\}\right] \\
&-2\,R_2\frac{ie^2}{(2\pi)^4}\int d^4p\;\mathrm{Tr}\,\{\gamma_\nu\;S(p+q)\,\gamma_\mu\,S(p)\} \\
&-(Z_3-1)_{\text{overlap}}(q^2\eta_{\mu\nu}-q_\mu q_\nu)\,. \qquad (12.2.18)
\end{aligned}$$

First consider integration over p' alone. Each of the first two terms is logarithmically divergent, but their difference is finite. The third term is also logarithmically divergent (with a gauge-invariant regulator), but the divergence in this term (unlike the first two terms) takes the form of a second-order polynomial in q, with the remainder finite. This remaining divergence is cancelled by the term $-(Z_3-1)(q^2\eta_{\mu\nu}-q_\mu q_\nu)$ that cancels all second-order terms in the expansion of $\Pi^*_{\mu\nu}(q)$. So the p'-subintegration gives a finite result. The symmetry of Eq. (12.2.16) shows that in exactly the same way the p subintegration also gives a finite result. Generic subintegrations over p and p' with $ap+bp'$ held fixed (where a and b are arbitrary non-zero constants) are manifestly convergent, and the integration over p and p' together is made finite by the counterterm $-(Z_3-1)_{\text{overlap}}(q^2\eta_{\mu\nu}-q_\mu q_\nu)$. Thus Eq. (12.2.18) *and* any of its subintegrations satisfy the power-counting requirements for convergence, and therefore

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

according to the theorem[2] quoted in the previous section, the whole expression actually converges.

* * *

In electrodynamics there is a natural definition of renormalized couplings as well as renormalized masses and fields. This is not always the case. For example, consider the theory of a single real scalar field $\phi(x)$ with Lagrangian density

$$\mathscr{L} = -\frac{1}{2}\,\partial_\lambda\phi\partial^\lambda\phi - \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{24}g\,\phi^4. \tag{12.2.19}$$

To one-loop order, the S-matrix for scalar–scalar scattering is given by the Feynman rules as

$$S(q_1q_2 \to q_1'q_2') = \frac{-i(2\pi)^4\delta^4(q_1' + q_2' - q_1 - q_2)}{(2\pi)^6(16E_1'E_2'E_1E_2)^{1/2}}F(q_1q_2 \to q_1'q_2')\,, \tag{12.2.20}$$

where

$$\begin{aligned} -i(2\pi)^4 F(q_1q_2 \to q_1'q_2') = -i(2\pi)^4 g + \frac{1}{2}\Big[-i(2\pi)^4 g\Big]^2 \left[\frac{-i}{(2\pi)^4}\right]^2 \\ \times \int d^4k \left[\frac{1}{\Big[(q_1+k)^2 + m^2 - i\epsilon\Big]\Big[(q_2-k)^2 + m^2 - i\epsilon\Big]}\right. \\ \left. + (q_2 \to -q_1') + (q_2 \to -q_2')\right], \end{aligned} \tag{12.2.21}$$

and q_1, q_2 and q_1', q_2' are the incoming and outgoing four-momenta. Combining denominators and rotating the k^0-integration contour as usual, this is

$$\begin{aligned} F = g - \frac{g^2}{16\pi^2}\int_0^\infty k^3dk \int_0^1 dx\Big\{\Big[k^2 + m^2 - sx(1-x)\Big]^{-2} \\ + \Big[k^2 + m^2 - tx(1-x)\Big]^{-2} + \Big[k^2 + m^2 - ux(1-x)\Big]^{-2}\Big\}, \end{aligned} \tag{12.2.22}$$

where s, t, and u are the *Mandelstam variables*

$$s = -(q_1 + q_2)^2\,, \qquad t = -(q_1 - q_1')^2\,, \qquad u = -(q_1 - q_2')^2\,, \tag{12.2.23}$$

related by $s + t + u = 4m^2$; also, x is the Feynman parameter introduced in combining denominators. With an ultraviolet cutoff at $k = \Lambda$, this gives

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

the result (for $\Lambda \gg m$)

$$F = g - \frac{g^2}{32\pi^2}\int_0^1 dx\left\{\ln\left(\frac{\Lambda^2}{m^2 - sx(1-x)}\right) + \ln\left(\frac{\Lambda^2}{m^2 - tx(1-x)}\right) + \ln\left(\frac{\Lambda^2}{m^2 - ux(1-x)}\right) - 3\right\}. \quad (12.2.24)$$

We can define the renormalized coupling g_R as the value of F at any point s, t, u we like, provided we stay in the region where F is real. For instance, suppose that in order to maintain the symmetry among the scalars, we choose to renormalize at the off-mass-shell point[¶] $q_1^2 = q_2^2 = q_1'^2 = q_2'^2 = \mu^2$, $s = t = u = -4\mu^2/3$. Defining the renormalized coupling g_R as the value of F at this point, we have

$$g = g_R + \frac{3g^2}{32\pi^2}\left[\ln\left(\frac{\Lambda^2}{\mu^2}\right) - 1 - \int_0^1 dx\ \ln\left(\frac{4x(1-x)}{3} + \frac{m^2}{\mu^2}\right)\right] + \cdots . \quad (12.2.25)$$

The cutoff dependence then cancels in Eq. (12.2.24) to order g_R^2, leaving a finite formula for F in terms of g_R:

$$F = g_R - \frac{g_R^2}{32\pi^2}\int_0^1 dx\left\{\ln\left(\frac{m^2 + 4x(1-x)\mu^2/3}{m^2 - sx(1-x)}\right) + \ln\left(\frac{m^2 + 4x(1-x)\mu^2/3}{m^2 - tx(1-x)}\right) + \ln\left(\frac{m^2 + 4x(1-x)\mu^2/3}{m^2 - ux(1-x)}\right)\right\} + \cdots . \quad (12.2.26)$$

Here μ^2 may be taken to be any real quantity greater than $-3m^2$, in which range g_R is real. The explicit μ-dependence in Eq. (12.2.26) is, of course, cancelled by the μ-dependence of the renormalized coupling. This freedom to change the renormalization prescription (which of course exists also in electrodynamics and other realistic theories) will be of great importance to us when we come to the renormalization group method in Volume II.

12.3 Is Renormalizability Necessary?

In the previous section we found a special class of theories having only a finite number of terms in the Lagrangian, to which the renormalization program is nevertheless applicable. These are theories in which all

[¶] Going back over the derivation of Eq. (12.2.25), one may check that in this derivation we have not used the conditions $q_1^2 = q_2^2 = q_1'^2 = q_2'^2 = -m^2$, so Eq. (12.2.24) is valid whatever we take for the external line masses.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

interactions satisfy the renormalizability condition

$$\Delta_i \equiv 4 - d_i - \sum_f n_{if}(s_f + 1) \geq 0,$$

where d_i and n_{if} are the numbers of derivatives and fields of type f in interactions of type i, and s_f is (with some qualifications) the spin of fields of type f. For renormalization to work in such theories, it is also usually necessary that all renormalizable interactions that are allowed by symmetry principles should actually appear in the Lagrangian.

It is important that there are only a limited number of such interaction types. Δ_i becomes negative if we have too many fields or derivatives, or fields of too high spin. Barring special cancellations, there are no renormalizable interactions at all involving fields with $s_f \geq 1$, because the only possible term in the Lagrangian with $\Delta_i \geq 0$ that involves such a field along with two or more other fields would involve a single $s_f = 1$ field along with two scalars and no derivatives, which would not be Lorentz-invariant. We shall see in Volume II that general, massless, spin one gauge fields in a suitable gauge effectively have $s_f = 0$, like the photon. Also, in Volume II we shall see that even massive gauge fields may effectively have $s_f = 0$, depending on where their mass comes from. Leaving aside these special cases, Table 12.2 gives a list of *all* renormalizable terms in the Lagrangian density that are allowed by Lorentz invariance and gauge invariance involving scalars ($s = 0$), photons ($s = 0$), and spin $\frac{1}{2}$ fermions ($s = \frac{1}{2}$).

We see that the requirement of renormalizability puts severe restrictions on the variety of physical theories that we may consider. Such restrictions provide a valuable key to the structure of physical theories. For instance, Lorentz and gauge invariance by themselves would allow the introduction of a 'Pauli' term proportional to $\bar{\psi}[\gamma_\mu, \gamma_\nu]\psi F^{\mu\nu}$ in the Lagrangian of quantum electrodynamics, which would make the magnetic moment of the electron an adjustable parameter, but we exclude such terms because they are not renormalizable. The successful predictions of quantum electrodynamics, such as the calculation of the magnetic moment of the electron outlined in Section 11.3, may be regarded as validations of the principle of renormalizability. The same applies to the standard model of weak, electromagnetic, and strong interactions, to be discussed in Volume II; there are any number of terms that might be added to this theory, such as four-fermion interactions among quarks and leptons, that would invalidate all the predictions of the standard model, and are excluded only because they are non-renormalizable.

Must we believe that the Lagrangian is restricted to contain only renormalizable interactions? As we saw in the previous section, if we include in the Lagrangian *all* of the infinite number of interactions allowed

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

Table 12.2. Allowed renormalizable terms in a Lagrangian density involving scalars ϕ, Dirac fields ψ, and photon fields A^μ. Here n_{if} and d_i are the number of fields of type f and the number of derivatives in an interaction of type i, and Δ_i is the dimensionality of the associated coefficient.

n_{if}			d_i	Δ_i	$\mathscr{H}_i$
Scalars	Photons	Spin $\frac{1}{2}$			
1	0	0	0	3	ϕ
2	0	0	0	2	ϕ^2
2	0	0	2	0	$\partial_\mu\phi\partial^\mu\phi$
3	0	0	0	1	ϕ^3
4	0	0	0	0	ϕ^4
2	1	0	1	0	$\phi\partial_\mu\phi A^\mu$
2	2	0	0	0	$\phi^2 A_\mu A^\mu$
1	0	2	0	0	$\phi\bar{\psi}\psi$
0	2	0	2	0	$F_{\mu\nu}F^{\mu\nu}$
0	0	2	0	1	$\bar{\psi}\psi$
0	0	2	1	0	$\bar{\psi}\gamma^\mu\partial_\mu\psi$
0	1	2	0	0	$\bar{\psi}\gamma^\mu A_\mu\psi$

by symmetries, then there will be a counterterm available to cancel every ultraviolet divergence. In this sense, as said earlier, non-renormalizable theories are just as renormalizable as renormalizable theories, as long as we include all possible terms in the Lagrangian.

In recent years it has become increasingly apparent that renormalizability is not a fundamental physical requirement, and that in fact any realistic quantum field theory will contain non-renormalizable as well as renormalizable terms. This change in point of view can be traced in part to the continued failure to find a renormalizable theory of gravitation. In the general class of metric theories of gravitation governed by Einstein's principle of equivalence there are no renormalizable interactions at all – generally covariant interactions must be constructed from the curvature tensor and its generally covariant derivatives, and hence, even in a 'gauge'

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

where the graviton propagator goes as k^{-2}, these interactions involve too many derivatives of the metric for renormalizability. In particular, we can easily see that general relativity is non-renormalizable from the fact that its coupling constant $8\pi G_N = (2.43 \times 10^{18}\ \text{GeV})^{-2}$ has negative dimensionality. Even if nothing else did, the cancellation of divergences due to virtual gravitons would require that the Lagrangian contain all interactions allowed by symmetries — not only interactions involving gravitons, but involving any particles.

But if renormalizability is not a fundamental physical principle, then how do we explain the success of renormalizable theories like quantum electrodynamics and the standard model? The answer can be seen by simple dimensional analysis. We have already noted that the coupling constant of an interaction of type i has dimensionality

$$[g_i] \sim [\text{mass}]^{\Delta_i}\ , \tag{12.3.1}$$

where Δ_i is the index (12.1.9). Non-renormalizable interactions are just those whose coupling constants have the dimensionality of *negative* powers of mass. Now, it is not unreasonable to guess from (12.3.1) that the coupling constants not only have dimensionalities governed by Δ_i, but are roughly of order

$$g_i \approx M^{\Delta_i}\ , \tag{12.3.2}$$

where M is some common mass. (This is found to be actually the case in the effective field theories discussed below and in more detail in Volume II.) In calculating physical processes at a characteristic momentum scale $k \ll M$, the inclusion of a non-renormalizable interaction of type i with $\Delta_i < 0$ will introduce a factor $g_i \approx M^{\Delta_i}$, which on dimensional grounds must be accompanied by a factor $k^{-\Delta_i}$, and so the effect of such an interaction is suppressed* for $k \ll M$ by a factor $(k/M)^{-\Delta_i} \ll 1$. (This argument will be made more carefully using the method of the renormalization group in Volume II.) The success of the renormalizable theories of electroweak and strong interactions shows only that M is very much larger than the energy scale at which these theories have been tested.

* It is essential at this point to assume that the ultraviolet divergences have been removed by renormalization, so that there are no factors of an ultraviolet cutoff Λ to mess up our dimensional analysis. Otherwise, dimensional analysis tells us that for $\Lambda \to \infty$, each additional non-renormalizable coupling constant factor g_i with $\Delta_i < 0$ would be accompanied with a growing factor $\Lambda^{-\Delta_i}$. This dimensional argument led Heisenberg[9] very early to classify interactions according to the dimensionality of their coupling constants, and to suggest[10] that new effects might arise at energies of order g_i^{1/Δ_i}, as for instance at the energy $G_F^{-1/2} \approx 300$ GeV, where G_F is the four-fermion coupling constant of the Fermi beta decay theory. After the development of renormalization theory it was noted by Sakata *et al.*[11] that the non-renormalizable theories are those whose coupling constants have negative dimensionality.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

For instance, the leading non-renormalizable corrections to the conventional electrodynamics of electrons or muons would be those interactions of dimension 5, which are suppressed by only one factor of $1/M$. There is just one such interaction allowed by Lorentz, gauge, and CP invariance, a Pauli term of order $(ie/2M)\bar{\psi}[\gamma_\mu, \gamma_\nu]\psi F^{\mu\nu}$. According to Eqs. (10.6.24), (10.6.17), and (10.6.19), such a term would contribute an amount of order $4e/M$ to the magnetic moment of the electron or muon. The calculated value of the magnetic moment of the electron agrees with experiment to within terms of order $10^{-10}e/2m_e$, so M must be greater than about $8 \times 10^{10} m_e = 4 \times 10^7$ GeV.

This limit may be weakened if other symmetries restrict the form of the non-renormalizable interactions. For instance, the conventional Lagrangian of quantum electrodynamics is invariant under a chiral transformation $\psi \to \gamma_5\psi$, except for a change of sign of the fermion mass term $-m\bar{\psi}\psi$. If we assume that the full Lagrangian is invariant under a formal symmetry $\psi \to \gamma_5\psi$, $m \to -m$, then a Pauli term in the Lagrangian would have to appear with an extra factor m/M, so that its contribution to the magnetic moment would be only of order $4em/M^2$. Because of the extra factor of m, here it is the muon rather than the electron that provides the most useful limit on M. The calculated value of the magnetic moment of the muon agrees with experiment to within terms of order $10^{-8}e/2m_\mu$, so M must be greater than about $\sqrt{8 \times 10^8} m_\mu = 3 \times 10^3$ GeV. In any case, if M is anywhere near as large as 10^{18} GeV, then we are certainly justified in neglecting any non-renormalizable interactions that might appear in quantum electrodynamics.

These considerations help us to cope with some of the puzzles associated with higher-derivative terms in the Lagrangian. For instance, in the general theory of a real scalar field ϕ, we would expect to find terms in the Lagrangian density of the form $\phi\Box^n\phi$. Any one such term would make a direct contribution to the scalar self-energy function $\Pi^*(q^2)$ proportional to $(q^2)^n$. If we were to include this contribution to all orders, but ignore all other effects of non-renormalizable interactions, then the propagator $\Delta'(q^2) = 1/(q^2 + m^2 - \Pi^*(q^2))$ would not have the simple pole in q^2 at negative q^2 expected from the general arguments of Section 10.7, but n such poles (some of which may coincide), generally at complex values of q^2. But if the non-renormalizable term $\phi\Box^n\phi$ has a coefficient of order $M^{-2(n-1)}$, where $M \gg m$, then the extra poles are at q^2 of order M^2, where it is illegitimate to ignore the infinite number of other non-renormalizable interactions that must also appear in the Lagrangian. Thus the appearance of higher-derivative terms in a general non-renormalizable Lagrangian is not in conflict with the general principles underlying quantum field theory that were used in Section 10.7. But by the same token, we also cannot use higher-derivative terms to avoid ultraviolet divergences altogether, as

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

has been repeatedly proposed. A term $M^{-2(n-1)}\phi\Box^n\phi$ in the Lagrangian density provides a cutoff at momenta $q^2 \approx M^2$, but at these momenta we cannot ignore all the other non-renormalizable interactions that must be present.

Although highly suppressed, non-renormalizable interactions may be detectable if they have effects that would otherwise be forbidden. For instance, we will see in Section 12.5 that the symmetries of charge-conjugation and space-inversion invariance are an automatic consequence of the structure of the electromagnetic interactions that is imposed by gauge invariance, Lorentz invariance, and renormalizability, but we can easily imagine non-renormalizable terms that would violate these symmetries, such as an electron electric dipole moment term $\bar{\psi}\gamma_5 [\gamma_\mu, \gamma_\nu] \psi F^{\mu\nu}$, or the Fermi interaction $\bar{\psi}\gamma_5\gamma_\mu\psi\bar{\psi}\gamma^\mu\psi$. It is widely believed today that the conservation of baryon and lepton number is violated by very small effects of highly suppressed non-renormalizable interactions. Another example of a detectable non-renormalizable interaction is provided by gravitation. As mentioned before, gravitons have no renormalizable interactions at all. But, of course, we detect gravitation, because it has the special property that the gravitational fields of all the particles in a macroscopic body add up coherently.

Although non-renormalizable theories involve an infinite number of free parameters, they retain considerable predictive power:[12] they allow us to calculate the non-analytic parts of Feynman amplitudes, like the $\ln q$ and $q \ln q$ terms in the one-dimensional examples at the beginning of the previous section. Such calculations just reproduce the results required by the axiom of S-matrix theory, that the S-matrix has only those singularities required by unitarity.

Paradoxically, it is just in the case where symmetry principles forbid renormalizable interactions that non-renormalizable quantum field theories prove the most useful. In such cases we can derive a useful perturbation theory by expanding in powers of k/M. This has been worked out in detail for the theory of low-energy pions,[12,13] to be discussed in detail in Volume II, and the theory of low-energy gravitons.[14] For a simpler example, consider the theory of a real scalar field, satisfying the principle of invariance under the field translation

$$\phi(x) \to \phi(x) + \epsilon$$

with ϵ an arbitrary constant. This symmetry forbids any renormalizable interactions or scalar mass, but it allows an infinite number of non-renormalizable derivative interactions

$$\mathscr{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{g}{4}(\partial_\mu\phi\partial^\mu\phi)^2 - \cdots,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

where $g \approx M^{-4}$, and '$\cdots$' denotes terms with more derivatives or fields. (For simplicity, it is assumed here that the theory also has a symmetry under the reflection $\phi \to -\phi$.) According to the above dimensional analysis, the graph for a general reaction in which all energies and momenta are of order $k \ll M$ is suppressed by a factor $(k/M)^v$, where

$$v = -\sum_i V_i \Delta_i = \sum_i V_i (d_i + n_i - 4) ,$$

with n_i and d_i the numbers of scalar fields and derivatives in an interaction of type i, and V_i the number of vertices for these interactions in our graph. For $k \ll M$, the dominant contributions to any process are those with the smallest value of v. The formula for v can be put in a more useful form by using the familiar topological identities for a connected graph:

$$\sum_i V_i = I - L + 1 , \qquad \sum_i V_i n_i = 2I + E ,$$

where I, E, and L are the numbers of internal lines, external lines, and loops in our graph. Combining these relations gives

$$v = 2E - 4 + 4L + \sum_i V_i (d_i - n_i) ,$$

Now, the field translation symmetry requires that every field must be accompanied with at least one derivative, so the quantity $d_i - n_i$ as well as L and V_i are non-negative for all interactions. Thus for a given process (that is, a fixed number E of external lines) the dominant terms will be those constructed solely from *tree* graphs (i.e., $L = 0$), and interactions with the minimum number $d_i = n_i$ of derivatives. That is, in leading order we can take the Lagrangian density to depend only on *first* derivatives of the field. Higher-order corrections may involve loops and/or interactions with more derivatives on some fields. But to any given order v in k/M, we need only consider a finite number of graphs, those with $L \leq (4-2E+v)/4$, and only a finite number of interaction types.

For instance, scalar–scalar scattering is given in leading order by the one-vertex tree graph calculated using the interaction $-g(\partial_\mu \phi \partial^\mu \phi)^2$ in first order. According to our formula for v, the leading correction, suppressed at low energy by a factor $(k/M)^2$, arises from another single-vertex tree graph, produced by an interaction with two additional derivatives of the form** $\partial_\mu \partial_\nu \phi \, \partial^\mu \partial^\nu \phi \, \partial_\lambda \phi \partial^\lambda \phi$. The next corrections, suppressed at low energy by two further factors of k/M, arise both from the one-loop diagram of Figure 12.4 (including permutations of external lines), calculated using only the interaction $-g(\partial_\mu \phi \partial^\mu \phi)^2$, and also from tree graphs with

** In accordance with the remarks of Section 7.7, we are excluding interactions involving $\Box\phi$, because the field equation for ϕ can be used to express such interactions in terms of the others.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

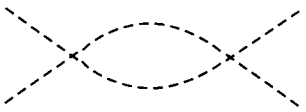

Figure 12.4. One-loop diagram for scalar–scalar scattering in the theory with derivative quadrilinear interactions.

a single vertex arising from a quartic interaction with eight derivatives, whose couplings contain infinite parts that cancel the ultraviolet divergence from the loop graph.† The loop graph also yields finite terms in the scattering amplitude proportional to terms like $s^4 \ln s + t^4 \ln t + u^4 \ln u$, $s^2t^2 \ln u + t^2u^2 \ln s + u^2s^2 \ln t$, etc., with calculable coefficients proportional to g^2. These finite terms simply represent the correction to the lowest-order scattering amplitude needed to ensure the unitarity of the S-matrix, but perturbative quantum field theory is by far the easiest way of calculating them.

Although non-renormalizable theories can provide useful expansions in powers of energy, they inevitably lose all predictive power at energies of the order of the common mass scale M that characterizes the various couplings. If we were to take these expansions literally, the results for S-matrix elements would violate unitarity bounds for $E \gg M$. There seem to be just two possibilities about what happens at such energies. One is that the growing strength of the effects of the non-renormalizable couplings somehow saturates, avoiding any conflict with unitarity.[15] The other is that new physics of some sort enters at the scale M. In this case, the non-renormalizable theories that describe nature at energies $E \ll M$ are just *effective field theories* rather than truly fundamental theories.

Probably the earliest example of an effective field theory was derived in the 1930s by Euler *et al.*,[16] as a theory of low-energy photon–photon interactions. (See Section 1.3.) In effect, they calculated the contribution to photon-photon scattering of Feynman diagrams such as Figure 12.5, and found that at energies much less than m_e the scattering of light by light was the same as would be calculated with an effective Lagrangian

$$\mathscr{L}_{\text{eff}} = \frac{2\alpha^2}{45m_e^4}\left[(\mathbf{E}^2 - \mathbf{B}^2)^2 + 7(\mathbf{E}\cdot\mathbf{B})^2\right] + \text{higher orders in } \frac{eE}{m_e^2} \;\&\; \frac{eB}{m_e^2}$$

† These are the only ultraviolet divergences encountered in one-loop graphs if we use dimensional regularization. For other methods of regularization there are also quartic and quadratic divergences, which are cancelled by counterterms in the four-scalar interactions with four or six derivatives.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

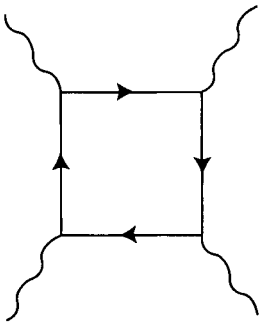

Figure 12.5. Diagram for photon–photon scattering, whose effect at low energy can be calculated from the effective Lagrangian of Euler *et al.*[16] Straight lines are electrons; wavy lines are photons.

Euler *et al.* used this effective Lagrangian only in the tree approximation, to calculate the leading terms in photon interaction matrix elements. It was not until much later that such Lagrangians, though non-renormalizable, were used beyond the tree approximation.[12,17]

In modern jargon, we say that in deriving this Lagrangian the electron is 'integrated out', because in the one-loop approximation we have

$$\exp\left(i\int \mathscr{L}_{\text{eff}}(\mathbf{E},\mathbf{B})d^4x\right) = \int\left[\prod_x d\psi_e(x)\right]\exp\left(i\int \mathscr{L}_{\text{QED}}(\psi_e,\mathbf{A})d^4x\right) .$$

A more general procedure is simply to write down the most general non-renormalizable effective Lagrangian, use it to calculate various amplitudes as an expansion in energies and momenta, and then choose the constants in the effective Lagrangian by matching the results it gives for these amplitudes to those derived from the underlying theory.

We will encounter effective field theories again, especially in considering broken symmetries in Volume II. As we shall see, effective field theories are useful even where they cannot be derived from an underlying theory, either because the theory is unknown, or because its interactions are too strong to allow the use of perturbation theory. Indeed, even if we knew nothing about the properties of charged particles, the scattering of photons at sufficiently low energy would have to be described by an effective Lagrangian consisting of the terms $(\mathbf{E}^2-\mathbf{B}^2)^2$ and $(\mathbf{E}\cdot\mathbf{B})^2$, because these are the unique quartic Lorentz- and gauge-invariant terms with no derivatives acting on $\mathbf{E}$ and $\mathbf{B}$. Terms with such derivatives would be suppressed at low photon energies E by additional factors of E/M, where M is some typical mass of the charged particles that are being integrated out. We can go further: we shall see that effective field theories are useful even where the light particles they describe are not present in the underlying theory at all, but composites of the heavy particles that are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

integrated out. The underlying theory might not even be a field theory at all — the problem of incorporating gravitation has led many theorists to believe that, in fact, it is a string theory. But wherever an effective field theory comes from, it is inevitably a non-renormalizable theory.

12.4 The Floating Cutoff*

Before closing this chapter, it is worth commenting on the relation between conventional renormalization theory and an approach pioneered by Wilson.[18] In Wilson's method one imposes a 'floating' finite ultraviolet cut-off (either sharp or smooth) at momenta with components of order Λ, and instead of taking $\Lambda \to \infty$, one requires that the bare constants of the theory (those appearing in the Lagrangian) depend on Λ in such a way that all observable quantities are Λ-independent.

It is convenient to work with dimensionless parameters. If a bare coupling or mass parameter $g_i(\Lambda)$ has dimensionality $[\text{mass}]^{\Delta_i}$, we define the corresponding dimensionless parameter $\mathcal{G}_i$ by

$$\mathcal{G}_i(\Lambda) \equiv \Lambda^{-\Delta_i} g_i(\Lambda) \,. \tag{12.4.1}$$

Ordinary dimensional analysis tells us that the value of $\mathcal{G}_i$ at one value Λ' of the cutoff can be expressed as a function of the values of the $\mathcal{G}_j$ at another value Λ of the cutoff, and the dimensionless ratio Λ'/Λ:

$$\mathcal{G}_i(\Lambda') = F_i\Big(\mathcal{G}(\Lambda), \Lambda'/\Lambda\Big) \,. \tag{12.4.2}$$

No dimensional parameters other than Λ' and Λ can appear in F, because no ultraviolet or infrared divergences can enter here; the difference between the constants at Λ and at Λ' arises from diagrams whose internal lines are restricted to have momenta between Λ and Λ'. Differentiating Eq. (12.4.2) with respect to Λ' and then setting Λ' equal to Λ yields a differential equation for $\mathcal{G}_i$:

$$\Lambda \frac{d}{d\Lambda} \mathcal{G}_i(\Lambda) = \beta_i(\mathcal{G}(\Lambda)) \,, \tag{12.4.3}$$

where $\beta_i(\mathcal{G}) \equiv [\partial/\partial z \, F_i(\mathcal{G}, z)]_{z=1}$. The functions $\beta_i(\mathcal{G})$ may be calculated for small couplings in perturbation theory. This is Wilson's version of the 'renormalization group' equation, which will be discussed in somewhat different terms in Volume II.

The Lagrangian for any finite value of the cutoff defines an effective field theory, in which instead of (or in addition to) integrating out 'heavy'

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

particles like the electron in the work of Euler *et al.*, one integrates out *all* particles with momenta greater than Λ. Even if one starts with a theory with a finite number of coupling parameters $\mathscr{G}_i^0$ at some cutoff Λ_0, at any other value of the cutoff the differential equation (12.4.3) will generally yield non-zero values for all couplings allowed by symmetry principles.**

We now distinguish between the renormalizable and unrenormalizable couplings, labelled $\mathscr{G}_a$ and $\mathscr{G}_n$, respectively, with a running over the finite number N of couplings (including masses) for which $\Delta_a \geq 0$, and n running over the infinite number of couplings with dimensionalities $\Delta_n < 0$. We want to show that if the couplings $\mathscr{G}_a(\Lambda_0)$ and $\mathscr{G}_n(\Lambda_0)$ at some initial cutoff value Λ_0 lie on a generic N-dimensional initial surface $\mathscr{S}_0$, then (with some qualifications) for $\Lambda \ll \Lambda_0$ they will approach a fixed surface $\mathscr{S}$ that is independent of both Λ_0 and the initial surface.† This fixed surface is stable, in the sense that from any point on the surface, the trajectory generated by Eq. (12.4.3) stays on the surface. Such a stable surface defines a finite-parameter set of theories whose physical content is cutoff-independent, which as argued in the previous section, is the essential property of renormalizable theories. Furthermore, this construction shows that a generic theory defined with cutoff Λ_0 will look for $\Lambda \ll \Lambda_0$ like a renormalizable theory.‡

To prove these results, consider any small perturbation $\delta\mathscr{G}_i(\Lambda)$ in the values of the $\mathscr{G}_i(\Lambda)$ satisfying Eq. (12.4.3). It will satisfy the differential equation

$$\Lambda \frac{d}{d\Lambda} \delta\mathscr{G}_i(\Lambda) = \sum_j M_{ij}(\mathscr{G}(\Lambda))\, \delta\mathscr{G}_j(\Lambda) \,, \tag{12.4.4}$$

where

$$M_{ij}(\mathscr{G}) \equiv \frac{\partial}{\partial \mathscr{G}_j} \beta_i(\mathscr{G}) \,. \tag{12.4.5}$$

This equation couples the renormalizable and unrenormalizable couplings, making it difficult to see the difference in their behavior. To decouple them, we introduce the linear combinations

$$\xi_n \equiv \delta\mathscr{G}_n - \sum_{ab} \frac{\partial \mathscr{G}_n}{\partial \mathscr{G}_a^0} \left(\frac{\partial \mathscr{G}}{\partial \mathscr{G}^0} \right)^{-1}_{ab} \delta\mathscr{G}_b \,, \tag{12.4.6}$$

** The only known exceptions to this rule are in theories based on supersymmetry.[4]

† This theorem is due to Polchinski.[19] What follows here is a shortened and less rigorous version. (In Polchinski's proof, the initial surface is taken to be that with all non-renormalizable couplings vanishing. As we shall see here, the couplings approach the same fixed surface for generic initial surfaces.)

‡ Of course some theories have symmetries and a field content that do not allow any renormalizable interactions. This is the case for theories containing only fermion fields, or only the gravitational field. Such theories for $\Lambda \ll \Lambda_0$ look like free-field theories.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

where $\mathcal{G}_a^0$ are the values of the renormalizable couplings at cutoff Λ_0, which we shall use as coordinates for the initial surface, and $\mathcal{G}_n$ are the values of the non-renormalizable couplings at cutoff Λ derived from the differential equation (12.4.3), with initial value for Λ_0 at the point on the initial surface with coordinates $\mathcal{G}_a^0$. To calculate the derivative of ξ_n with respect to Λ, we note that the derivatives $\partial\mathcal{G}_i/\partial\mathcal{G}_a^0$ satisfy the same differential equation (12.4.4) as the $\delta\mathcal{G}_i$. It is an elementary exercise then to show that

$$\Lambda\frac{d}{d\Lambda}\xi_n = \sum_m N_{nm}\,\xi_m\,, \tag{12.4.7}$$

where

$$N_{nm} \equiv M_{nm} - \sum_{ab}\frac{\partial\mathcal{G}_n}{\partial\mathcal{G}_a^0}\left(\frac{\partial\mathcal{G}}{\partial\mathcal{G}^0}\right)^{-1}_{ab} M_{bm}\,. \tag{12.4.8}$$

Now we must estimate the elements of N_{nm}. For a free-field theory no cutoff is needed, so for very small coupling all bare parameters $g_i(\Lambda)$ become Λ-independent. Hence for small couplings the dimensionless parameters $\mathcal{G}_i$ simply scale as $\Lambda^{-\Delta_i}$, and the matrix M_{ij} is given by

$$M_{ij} \approx -\Delta_i\delta_{ij}\,. \tag{12.4.9}$$

It follows that the matrix N_{nm} is given approximately by $-\Delta_n\delta_{nm}$. The defining characteristic of the non-renormalizable couplings is that $\Delta_n < 0$, so Eq. (12.4.7) tells us that, at least for couplings in some finite range, where N_{nm} is positive-definite, the ξ_n decay for $\Lambda \ll \Lambda_0$ like positive powers of Λ/Λ_0. In this limit, then, the perturbations are related by

$$\delta\mathcal{G}_n = \sum_{ab}\frac{\partial\mathcal{G}_n}{\partial\mathcal{G}_a^0}\left(\frac{\partial\mathcal{G}}{\partial\mathcal{G}^0}\right)^{-1}_{ab}\delta\mathcal{G}_b\,. \tag{12.4.10}$$

In particular, if we make a small change in the initial surface $\mathcal{S}_0$ and/or the starting point on that surface and/or the initial cutoff Λ_0, such that the perturbations $\delta\mathcal{G}_a$ in the renormalizable couplings vanish at some cutoff $\Lambda \ll \Lambda_0$, then the perturbations $\delta\mathcal{G}_n$ in all the other couplings at cutoff Λ also vanish. Thus the non-renormalizable couplings $\mathcal{G}_n(\Lambda)$ for $\Lambda \ll \Lambda_0$ can depend only on the renormalizable couplings $\mathcal{G}_a(\Lambda)$, not separately on the initial surface or the starting point on that surface or the initial cutoff Λ_0. At cutoff $\Lambda \ll \Lambda_0$ all the couplings therefore approach an N-dimensional surface $\mathcal{S}$, with coordinates $\mathcal{G}_a(\Lambda)$, which is independent of both the initial surface and of Λ_0. Note that the non-renormalizable couplings $\mathcal{G}_n$ are not generally small on $\mathcal{S}$; the important point is that they become functions of the renormalizable couplings. Changes in Λ with Λ remaining much less than Λ_0 will change the couplings, but the couplings will remain close to $\mathcal{S}$ (at least as long as the couplings do not

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

become so large that N_{nm} is no longer a positive-definite matrix). Hence $\mathscr{S}$ is a stable surface, as was to be proved.

We have seen that all physical quantities may be expressed in terms of Λ and the $\mathscr{G}_n(\Lambda)$, and are Λ-independent. This is true in particular of the N conventional renormalized couplings and masses, like e and m_e in quantum electrodynamics. But we can then invert this relation, and express the $\mathscr{G}_n(\Lambda)$ in terms of the conventional parameters and Λ. In this way we can justify the usual renormalization program: all physical quantities are expressed in a cutoff-independent way in terms of the conventional renormalized couplings and masses.

The Wilson approach has some advantages in practice. One does not have to worry about subintegrations and overlapping divergences; the momentum cutoff applies to all internal lines. Also, some of the non-renormalization theorems of supersymmetry theory, which tell us that certain couplings are not affected by radiative corrections, work only for the cutoff-dependent bare couplings.[20]

On the other hand, there are disadvantages to the Wilson approach. One must give up the special simplicities of working with renormalizable theories like quantum electrodynamics; once one starts integrating out particles with momenta above some scale Λ, the resulting effective field theory will contain all Lorentz- and gauge-invariant interactions, with Λ-dependent couplings. (Nevertheless, in physical processes at energies $E \ll \Lambda$, the dominant couplings will still be the renormalizable ones.) Also, the cutoff generally destroys *manifest* gauge invariance, and either manifest Lorentz invariance or unitarity. None of this is a problem in condensed matter physics, the original context of Wilson's approach, because no one would expect a realistic condensed matter theory to be strictly renormalizable, and there are no fundamental physical principles that are necessarily violated by a cutoff. In fact, in crystals there *is* a cutoff on phonon momenta, provided by the inverse lattice spacing.

At bottom, the difference between the conventional and the Wilson approach is one of mathematical convenience rather than of physical interpretation. Indeed, conventional renormalization already provides a sort of adjustable cutoff; when we express our answer in terms of coupling constants that are defined as the values of physical amplitudes at some momenta of order μ (as for the scalar field theory discussed in the previous section), the cancellations that make integrals converge begin to operate at virtual momenta of order μ. Conversely, the Λ-dependent coupling constants of the Wilson approach must ultimately be expressed in terms of observable masses and charges, and when this is done the results are, of course, the same as those obtained by conventional means.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

12.5 Accidental Symmetries*

In Section 12.3 we saw that there are good reasons to adopt renormalizable field theories as approximate descriptions of nature at sufficiently low energy. It often happens that the condition of renormalizability is so stringent that the effective Lagrangian automatically obeys one or more symmetries, which are not symmetries of the underlying theory, and may therefore be violated by the suppressed non-renormalizable terms in the effective Lagrangian. Indeed, most of the experimentally discovered symmetries of elementary particle physics are 'accidental symmetries' of this sort.

A classic example is provided by the inversions and flavor conservation in the electrodynamics of charged leptons. The most general renormalizable and gauge- and Lorentz-invariant Lagrangian density for photons and fields ψ_i of spin $\frac{1}{2}$ and charge $-e$ takes the form

$$\begin{aligned}\mathscr{L} = & -\tfrac{1}{4}Z_3 F_{\mu\nu}F^{\mu\nu} \\ & -\sum_{ij} Z_{Lij}\bar{\psi}_{Li}[\partial\!\!\!/ + ie A\!\!\!/]\psi_{Lj} - \sum_{ij} Z_{Rij}\bar{\psi}_{Ri}[\partial\!\!\!/ + ie A\!\!\!/]\psi_{Rj} \\ & -\sum_{ij} M_{ij}\bar{\psi}_{Li}\psi_{Rj} - \sum_{ij} M^{\dagger}_{ij}\bar{\psi}_{Ri}\psi_{Lj}\,, \end{aligned} \tag{12.5.1}$$

where i, j are summed over the three lepton flavors (e, μ, and τ), ψ_{iL} and ψ_{iR} are the left- and right-handed parts of the field ψ_i, defined by

$$\psi_{Li} = \tfrac{1}{2}(1+\gamma_5)\psi_i\,, \qquad \psi_{Ri} = \tfrac{1}{2}(1-\gamma_5)\psi_i\,, \tag{12.5.2}$$

and Z_L, Z_R, and M are numerical matrices. We are not assuming anything about lepton flavor conservation, so the matrices Z_{Lij}, Z_{Rij} and M_{ij} need not be diagonal. Also we are not assuming anything about invariance under P, C, or T invariance, so there is no necessary relation between Z_L and Z_R, or between M and $M^{\dagger}$. The only constraints on these matrices come from the reality of the Lagrangian density, which requires that Z_{Lij} and Z_{Rij} are Hermitian, and from the canonical anticommutation relations, which require that Z_{Lij} and Z_{Rij} are positive-definite.

Now suppose we replace the lepton fields ψ_L, ψ_R with new fields ψ'_L, ψ'_R defined by

$$\psi_L = S_L\psi'_L\,, \qquad \psi_R = S_R\psi'_R\,, \tag{12.5.3}$$

where $S_{L,R}$ are non-singular matrices that can be chosen as we like. The Lagrangian density when expressed in terms of these new fields then takes

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

the same form as in Eq. (12.5.1), but with new matrices

$$Z_L' = S_L^\dagger Z_L S_L\,, \qquad Z_R' = S_R^\dagger Z_R S_R\,, \qquad M' = S_L^\dagger M S_R\,. \tag{12.5.4}$$

We can choose S_L and S_R so that $Z_L' = Z_R' = 1$. (Take $S_{L,R} = U_{L,R}D_{L,R}$, where $U_{L,R}$ are the unitary matrices that diagonalize the positive-definite Hermitian matrices $Z_{L,R}$, and the $D_{L,R}$ are the diagonal matrices whose elements are the inverse square roots of the eigenvalues of $Z_{L,R}$.)

Now make another transformation, to lepton fields ψ_i'' defined by

$$\psi_L' = S_L'\psi_L''\,, \qquad \psi_R' = S_R'\psi_R''\,. \tag{12.5.5}$$

The Lagrangian density again takes the same form when expressed in terms of these new fields, with new matrices

$$Z_L'' = S_L^{'\dagger} S_L'\,, \qquad Z_R'' = S_R^{'\dagger} S_R'\,, \qquad M'' = S_L^{'\dagger} M' S_R'\,. \tag{12.5.6}$$

This time we take $S_{L,R}'$ unitary, so that again $Z_L'' = Z_R'' = 1$. We choose these unitary matrices so that M'' is real and diagonal. (By the polar decomposition principle, M' like any square matrix may be put in the form $M' = VH$, where V is unitary and H is Hermitian. Take $S_L' = S_R^{'\dagger} V^\dagger$ and choose S_R' as the unitary matrix that diagonalizes H.) Dropping primes, the Lagrangian density now takes the form

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}Z_3 F_{\mu\nu}F^{\mu\nu} - \sum_i \bar\psi_{Li}[\partial\!\!\!/ + ie A\!\!\!/]\psi_{Li} - \sum_i \bar\psi_{Ri}[\partial\!\!\!/ + ie A\!\!\!/]\psi_{Ri} \\ &- \sum_i m_i\bar\psi_{Li}\psi_{Ri} - \sum_i m_i\bar\psi_{Ri}\psi_{Li}\,,\end{aligned} \tag{12.5.7}$$

where m_i are real numbers, the eigenvalues of the Hermitian matrix H. Finally, this can be put in the more familiar form

$$\mathscr{L} = -\tfrac{1}{4}Z_3 F_{\mu\nu}F^{\mu\nu} - \sum_i \bar\psi_i[\partial\!\!\!/ + ie A\!\!\!/]\psi_i - \sum_i m_i\bar\psi_i\psi_i\,. \tag{12.5.8}$$

With the Lagrangian taking this form, it is now apparent that any renormalizable Lagrangian for lepton electrodynamics automatically conserves P, C, and T, as well as the numbers of leptons (minus the numbers of antileptons) of each flavor: electron, muon, and tauon.** In particular, despite the appearance of Eq. (12.5.1), this theory does not allow such processes as $\mu \to e+\gamma$. The reader may perhaps worry whether it is correct to identify the lepton fields as the ψ_i (previously called ψ_i'') appearing in Eq. (12.5.8), which obviously conserves lepton flavor, rather than the ψ_i appearing in Eq. (12.5.1), which seems to allow processes like $\mu \to e + \gamma$.

** This was first shown by Feinberg, Kabir, and myself.[21] Feinberg[22] had earlier noted that weak interaction effects in a theory with only one neutrino species would give rise to an observable rate for the process $\mu \to e + \gamma$, a difficulty that was only resolved by the discovery of a second neutrino species.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

Such worries may be put aside; as stressed in Section 10.3, there is no one field that can be identified as *the* field of the electron or muon. In fact, although Eq. (12.5.1) yields a non-vanishing matrix element for radiative decay of lepton 1 into lepton 2 *off* the lepton mass shell, by taking the lepton momenta on the mass shell we find a vanishing S-matrix for all such processes even when calculated using Eq. (12.5.1).

It was essential in deriving these results that the same electric charge appeared in Eq. (12.5.1) for both the left- and right-handed parts of the lepton fields or, in other words, that both left- and right-handed parts of the lepton fields transform in the same way under electromagnetic gauge transformations. As we shall see in Volume II, for similar reasons the modern renormalizable theory of strong interactions known as quantum chromodynamics automatically conserves C, and (aside from certain non-perturbative effects) P and T, as well as the numbers of quarks (minus the numbers of antiquarks) of each quark flavor. We shall also see in Volume II that the simplest version of the renormalizable standard model of weak and electromagnetic interactions automatically conserves lepton flavor (though not C and P) for reasons similar to those described here for electrodynamics. It remains an open possibility that non-renormalizable interactions arising from higher mass scales may violate any of these conservation laws.

Problems

1. List all the renormalizable (or superrenormalizable) Lorentz-invariant terms in the Lagrangian of a single scalar field for space-time dimensionalities 2, 3, and 6.

2. Show how the overlapping divergence in the electron self-energy is cancelled in quantum electrodynamics.

3. Consider the theory of a scalar field ϕ and spinor field ψ, with interaction Hamiltonian $g\phi\bar{\psi}\psi$. Write the one-loop part of the scalar self-energy function $\Pi^*(q)$ as a divergent polynomial in p^μ, plus an explicit convergent integral.

4. Suppose that the quantum electrodynamics of electrons and photons is actually an effective field theory, derived by integrating out unknown particles of mass $M \gg m_e$. Assume gauge invariance and Lorentz invariance, but not invariance under C, P, or T. What are the non-renormalizable terms in the Lagrangian of leading order in $1/M$? Of next to leading order?

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

References

1. F. J. Dyson, *Phys. Rev.* **75**, 486, 1736 (1949). A historical perspective is provided by *Renormalization*, ed. by L. M. Brown (Springer-Verlag, New York, 1993). For a comprehensive modern treatment, see J. Collins, *Renormalization* (Cambridge University Press, Cambridge, 1984).

2. S. Weinberg, *Phys. Rev.* **118**, 838 (1959). This proof relied only on general asymptotic properties of the integrands of Feynman graphs in Euclidean momentum space, obtained by Wick rotation of all integration contours. The proof was simplified through the use of more detailed properties of the integrand, by Y. Hahn and W. Zimmerman, *Commun. Math. Phys.* **10**, 330 (1968), and then extended to Minkowskian momentum space by W. Zimmerman, *Commun. Math. Phys.* **11**, 1 (1968).

3. J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965): Sections 19.10 and 19.11.

4. See, e.g., J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, Princeton, 1983), and original references quoted therein.

5. A. Salam, *Phys. Rev.* **82**, 217 (1951); *Phys. Rev.* **84**, 426 (1951).; P. T. Matthews and A. Salam, *Phys. Rev.* **94**, 185 (1954).

6. N. N. Bogoliubov and O. Parasiuk, *Acta Math.* **97**, 227 (1957)

7. K. Hepp, *Comm. Math. Phys.* **2**, 301 (1966). Hepp remarks that 'it is difficult to find two theorists whose understanding of the essential steps of the proof [of Bogoliubov and Parasiuk] is isomorphic,' but Hepp's paper is itself not easy to read.

8. W. Zimmerman, *Comm. Math. Phys.* **15**, 208 (1969). Also see W. Zimmerman, in *Lectures on Elementary Particles and Quantum Field Theory — Brandeis University Summer Institute in Theoretical Physics* (M.I.T. Press, Cambridge, 1970).

9. W. Heisenberg, *Z. Physik* **110**, 251 (1938).

10. W. Heisenberg, Ref. 6 and *Z. Physik* **113**, 61 (1939).

11. S. Sakata, H. Umezawa, and S. Kamefuchi, *Prog. Theor. Phys.* **7**, 327 (1952).

12. S. Weinberg, *Physica* **96A**, 327 (1979)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

13. J. Gasser and H. Leutwyler, *Ann. Phys.* (NY) **158**, 142 (1984); *Nucl. Phys.* **B250**, 465 (1985).

14. J. F. Donoghue, *Phys. Rev.* **D 50**, 3874 (1994).

15. One possible way that this can happen is through the phenomenon of 'asymptotic safety'; see S. Weinberg, in *General Relativity — An Einstein Centenary Survey*, ed. by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979): Section 16.3.

16. H. Euler and B. Kockel, *Naturwiss.* **23**, 246 (1935); W. Heisenberg and H. Euler, *Z. Physik* **98**, 714 (1936).

17. The Euler *et al.* effective Lagrangian has been used in one-loop calculations by J. Halter, *Phys. Lett.* **B 316**, 155 (1993).

18. K. G. Wilson, *Phys. Rev.* **B4**, 3174, 3184 (1971); *Rev. Mod. Phys.* **47**, 773 (1975).

19. J. Polchinski, *Nucl. Phys.* **B231**, 269 (1984); lecture in *Recent Directions in Particle Theory — Proceedings of the 1992 TASI Conference*, ed. by J. Harvey and J. Polchinski (World Scientific, Singapore, 1993): p. 235.

20. V. Novikov, M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Nucl. Phys.* **B229**, 381 (1983); M. A. Shifman and A. I. Vainshtein, *Nucl. Phys.* **B277**, 456 (1986); and references quoted therein. See also M. A. Shifman and A. I. Vainshtein, *Nucl. Phys.* **B359**, 571 (1991).

21. G. Feinberg, P. Kabir, and S. Weinberg, *Phys. Rev. Lett.* **3**, 527 (1959).

22. G. Feinberg, *Phys. Rev.* **110**, 1482 (1958).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.014

13
Infrared Effects

In the study of radiative corrections a special role is played by those corrections due to 'soft' photons: photons whose energy and momentum are much less than the masses and energies characteristic of the process in question. Not only are these corrections often so large that they must be summed to all orders of perturbation theory; they are so simple that this summation is not difficult. The contribution of photons of infinitely long wavelength takes the form of divergent integrals, but as we shall see these 'infrared divergences' all cancel.[1]

In most of this chapter we will deal with photons interacting with charged particles of arbitrary type and spin, including particles like atomic nuclei that have strong as well as electromagnetic interactions. But it is not difficult to adapt the calculations presented here to the infrared effects of other massless particles, such as the gluons of quantum chromodynamics. In Section 13.4 we shall explicitly consider very general theories of massless particles, and will show the cancellation of infrared divergences on general grounds.

After these generalities, we shall return to photons, and take up two topics of practical importance: the scattering of soft photons by charged particles with arbitrary non-electromagnetic interactions and arbitrary spin, and the treatment of heavy charged particles like atomic nuclei as a source of an external electromagnetic field.

13.1 Soft Photon Amplitudes

In this section we shall derive a universal formula that gives the amplitude for emission of any number of very-low-energy photons in a process $\alpha \to \beta$ involving any number of higher-energy charged particles of any types.

Let us start with the amplitude for emission of just one soft photon. If we attach the soft photon line with outgoing momentum q and polarization index μ to an outgoing charged-particle line that leaves some connected Feynman diagram for the process $\alpha \to \beta$, as in Figure 13.1(a), then we

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

must multiply the S-matrix element for $\alpha \to \beta$ with an additional charged-particle propagator carrying the momentum $p+q$ that the charged particle had before emitting the photon, together with the contribution of the new charged-particle–photon vertex. For charged particles of spin zero, mass m, and charge $+e$, these factors are

$$\left[i(2\pi)^4 e(2p^\mu + q^\mu)\right] \left[\frac{-i}{(2\pi)^4} \frac{1}{(p+q)^2 + m^2 - i\epsilon}\right] ,$$

which in the limit $q \to 0$ becomes

$$\frac{e\, p^\mu}{p \cdot q - i\epsilon} . \tag{13.1.1}$$

(We are freely redefining the scale of the positive infinitesimal ϵ, being careful only to keep track of its sign.) This result is actually true for charged particles of any spin. For instance, for a particle of spin $\frac{1}{2}$ and charge $+e$, we must replace the coefficient function $\bar{u}(\mathbf{p}, \sigma)$ for the outgoing charged particle with

$$\bar{u}(\mathbf{p}, \sigma) \left[-(2\pi)^4 e\gamma^\mu\right] \left[\frac{-i}{(2\pi)^4} \frac{-i(\not{p} + \not{q}) + m}{(p+q)^2 + m^2 - i\epsilon}\right] .$$

In the limit $q \to 0$ the numerator of the propagator is given by a sum of dyads:

$$-i\, \not{p} + m = 2p^0 \sum_{\sigma'} u(\mathbf{p}, \sigma')\, \bar{u}(\mathbf{p}, \sigma') ,$$

so we have a sum of equal-momentum matrix elements of γ^μ, given by

$$\bar{u}(\mathbf{p}, \sigma)\gamma^\mu u(\mathbf{p}, \sigma') = -i\delta_{\sigma,\sigma'} p^\mu / p^0 ,$$

and the effect again is to multiply the matrix element for the process $\alpha \to \beta$ by the factor (13.1.1). More generally, for any spin in the limit $q \to 0$ the four-momentum $p+q$ of the new internal charged particle line approaches the mass shell, so the numerator of the propagator approaches a sum of dyads of coefficient functions which convert the new vertex matrix into a factor proportional to p^μ and a unit matrix in helicity indices, leading again to the factor (13.1.1). Furthermore, as we saw in Chapter 10, higher-order corrections do not affect either the residue of the mass-shell poles in the propagators or the matrix element of the electric current between states of the same particle at equal momentum, so (13.1.1) gives the correct factor associated with the emission of a soft photon from an outgoing charged-particle line to all orders of perturbation theory.

The same reasoning applies to a photon emitted from an incoming charged-particle line of the process $\alpha \to \beta$, except that after the incoming particle emits a photon of four-momentum q the charged-particle line has

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

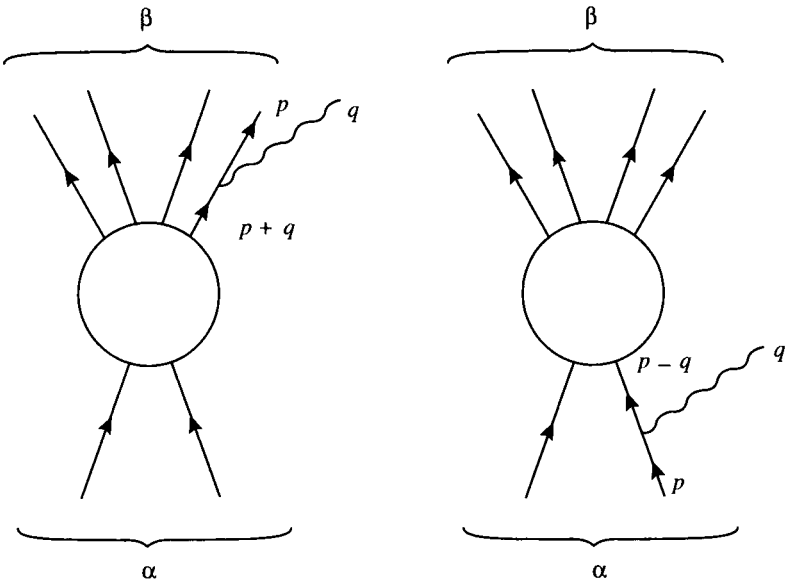

Figure 13.1. Dominant graphs for the emission of soft photons in an arbitrary process $\alpha \to \beta$. Straight lines are particles in the states α and β (including possible hard photons); wavy lines are soft photons.

four-momentum $p - q$, so in place of (13.1.1) we find a factor

$$\frac{e\,p^\mu}{-p\cdot q - i\epsilon}\,. \tag{13.1.2}$$

The photon can also, of course, be emitted from an internal line of the process $\alpha \to \beta$, but in this case there is no factor that goes as $(p\cdot q)^{-1}$ for $q \to 0$. The amplitude $M^\mu_{\beta\alpha}(q)$ (the S-matrix without the energy-momentum conservation delta function) for emitting a single soft photon with four-momentum q and polarization index μ in the process $\alpha \to \beta$ is therefore given in the limit $q \to 0$ by multiplying the matrix element $M_{\beta\alpha}$ for $\alpha \to \beta$ with a sum of terms like (13.1.1) and (13.1.2), one for each outgoing or incoming charged particle:

$$M^\mu_{\beta\alpha}(q) \to M_{\beta\alpha}\sum_n \frac{\eta_n e_n\, p_n^\mu}{p_n\cdot q - i\eta_n\epsilon}\,, \tag{13.1.3}$$

where p_n and $+e_n$ are the four-momentum and charge of the nth particle in the initial and final states, and η_n is a sign factor with the value $+1$ for particles in the final state β and -1 for particles in the initial state α.

Before going on to consider the emission of more than one soft photon, it is worth mentioning an important feature[2] of the formula (13.1.3). To

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

calculate the amplitude for emission of a photon of definite helicity, we must contract this expression with the corresponding photon polarization vector $e_\mu(\mathbf{q}, \pm)$. But as we saw in Section 5.9, $e^\mu(\mathbf{q}, \pm)$ is not a four-vector; under a Lorentz transformation $\Lambda^\mu_{\ \nu}$, the polarization vector is transformed into $\Lambda^\mu_{\ \nu} e^\nu(\mathbf{q}, \pm)$ plus a term proportional to q^μ. In order for this last term not to spoil Lorentz invariance, it is therefore necessary that $M^\mu_{\beta\alpha}(q)$ should vanish when contracted with q_μ. But for $q \to 0$, (13.1.3) gives

$$q_\mu M^\mu_{\beta\alpha}(q) \to M_{\beta\alpha} \sum_n \eta_n e_n \,. \tag{13.1.4}$$

The coefficient of $M_{\beta\alpha}$ on the right-hand side is just the total charge in the final state minus the total charge in the initial state, so the condition that it vanishes is just the condition that charge is conserved. Thus without any independent assumptions about gauge invariance, we see that *for particles of spin one and mass zero, Lorentz invariance requires the conservation of whatever coupling constant like electric charge governs the interaction of these particles at low energies.*

Incidentally, the amplitude for emitting a soft *graviton* of four-momentum q and tensor indices μ, ν in a process $\alpha \to \beta$ is given by a formula[3] analogous to (13.1.3):

$$M^{\mu\nu}_{\beta\alpha}(q) \to M_{\beta\alpha} \sum_n \frac{\eta_n f_n p_n^\mu p_n^\nu}{p_n \cdot q - i\eta_n \epsilon} \,, \tag{13.1.5}$$

where f_n is the coupling constant of the soft graviton to particles of type n. Lorentz invariance here requires that this vanish when contracted with q_μ. But

$$q_\mu M^{\mu\nu}_{\beta\alpha}(q) \to M_{\beta\alpha} \sum_n \eta_n f_n p_n^\nu \,, \tag{13.1.6}$$

so the sum $\sum f_n p_n^\nu$ is conserved. However, the only linear combination of the four-momenta that can be conserved without forbidding all non-trivial scattering processes is the total four-momentum, so in order for (13.1.6) to vanish, the f_n must all be equal. (The common value of all f_n may be identified as $\sqrt{8\pi G_N}$, where G_N is Newton's constant of gravitation.) Thus Lorentz invariance requires the result that low-energy massless particles of spin two couple in the same way to all forms of energy and momentum. This goes a long way toward showing that Einstein's principle of equivalence is a necessary consequence of Lorentz invariance as applied to massless particles of spin two. Likewise, the amplitude for emitting a soft massless particle of four-momentum q and spin $j \geq 3$ in a process $\alpha \to \beta$ is of the form

$$M^{\mu\nu\rho\cdots}_{\beta\alpha}(q) \to M_{\beta\alpha} \sum_n \frac{\eta_n g_n p_n^\mu p_n^\nu p_n^\rho \cdots}{p_n \cdot q - i\eta_n \epsilon} \,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

Lorentz invariance here requires that the sum $\sum g_n p_n^\nu p_n^\rho \cdots$ must be conserved. But no such quantity can be conserved without prohibiting all non-trivial scattering processes, so the g_n must all vanish. Massless particles of spin $j \geq 3$ may exist, but they cannot have couplings that survive in the limit of low energy, and in particular they cannot mediate inverse square law forces.

Now let us consider the emission of two soft photons. The contribution to the matrix element from a graph in which the two photons are emitted from different external lines of the process $\alpha \to \beta$ is given by multiplying the matrix element for $\alpha \to \beta$ by a product of factors like (13.1.1) or (13.1.2). Perhaps surprisingly, the same is true even if the two photons are emitted from the *same* external line. For example, if photon 1 is emitted from an external line of charge $+e$ and energy-momentum four-vector p after photon 2 we get a factor

$$\left[\frac{\eta e p^{\mu_1}}{p \cdot q_1 - i\eta\epsilon}\right]\left[\frac{\eta e p^{\mu_2}}{p \cdot (q_2 + q_1) - i\eta\epsilon}\right],$$

while if photon 2 is emitted after photon 1 the factor is

$$\left[\frac{\eta e p^{\mu_2}}{p \cdot q_2 - i\eta\epsilon}\right]\left[\frac{\eta e p^{\mu_1}}{p \cdot (q_1 + q_2) - i\eta\epsilon}\right].$$

(See Figure 13.2. Again, η is $+1$ or -1 according to whether the charged-particle line is outgoing or incoming.) These two factors add up to

$$\left[\frac{\eta e p^{\mu_1}}{p \cdot q_1 - i\eta\epsilon}\right]\left[\frac{\eta e p^{\mu_2}}{p \cdot q_2 - i\eta\epsilon}\right],$$

which is just a product of the same factors encountered in the emission of a single photon.

More generally, in emitting an arbitrary number of photons from a single external line we encounter a sum of the form*

---

* This identity may be proved by mathematical induction. We have already seen that it is true for two photons. Suppose that it is true for $N-1$ photons. For N photons we may then write the sum over permutations as a sum over the choice of the first photon to be emitted together with a sum over permutations of the remaining photons:

$$\begin{aligned}
&[p \cdot q_1 - i\eta\epsilon]^{-1}[p \cdot (q_1 + q_2) - i\eta\epsilon]^{-1} \cdots [p \cdot (q_1 + q_2 + \cdots + q_N) - i\eta\epsilon]^{-1} + \text{permutations} \\
&= \sum_{r=1}^{N}\left[p \cdot \left(\sum_{s=1}^{N} q_s\right) - i\eta\epsilon\right]^{-1} \prod_{s\neq r}[p \cdot q_s - i\eta\epsilon]^{-1} \\
&= \sum_{r=1}^{N}\left[p \cdot \left(\sum_{s=1}^{N} q_s\right) - i\eta\epsilon\right]^{-1} [p \cdot q_r - i\eta\epsilon]\prod_{s=1}^{N}[p \cdot q_s - i\eta\epsilon]^{-1} = \prod_{s=1}^{N}[p \cdot q_s - i\eta\epsilon]^{-1}
\end{aligned}$$

as was to be proved.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

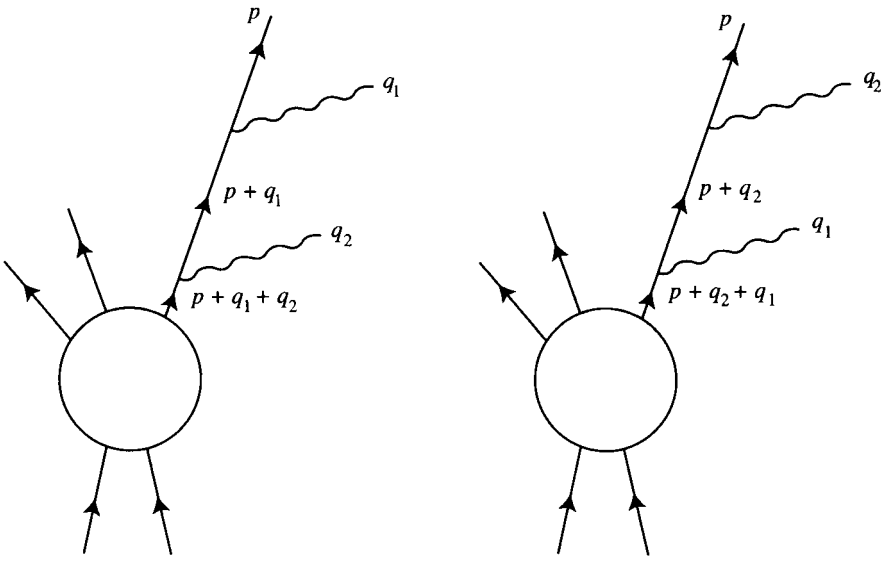

Figure 13.2. Graphs for the emission of two soft photons from the same outgoing charged particle. Straight lines are hard particles; wavy lines are soft photons.

$$[p\cdot q_1 - i\eta\epsilon]^{-1}[p\cdot(q_1+q_2) - i\eta\epsilon]^{-1}[p\cdot(q_1+q_2+q_3) - i\eta\epsilon]^{-1}\cdots$$
$$+ \text{ permutations}$$
$$= [p\cdot q_1 - i\eta\epsilon]^{-1}[p\cdot q_2 - i\eta\epsilon]^{-1}[p\cdot q_3 - i\eta\epsilon]^{-1}\cdots \quad . \tag{13.1.7}$$

It follows then that the amplitude $M_{\beta\alpha}^{\mu_1\cdots\mu_N}(q_1\cdots q_N)$ for emitting N very soft photons with polarization indices $\mu_1,\cdots\mu_N$ and four-momenta $q_1,\cdots q_N$ in the process $\alpha\to\beta$ is given in the limit $q\to 0$ by multiplying the matrix element $M_{\beta\alpha}$ for $\alpha\to\beta$ by a product of factors like that in (13.1.3), one for each photon:

$$M_{\beta\alpha}^{\mu_1\cdots\mu_N}(q_1\cdots q_N) \to M_{\beta\alpha}\prod_{r=1}^{N}\left(\sum_n \frac{\eta_n e_n p_n^{\mu_r}}{p_n\cdot q_r - i\eta_n\epsilon}\right) . \tag{13.1.8}$$

13.2 Virtual Soft Photons

We shall now use the results of the previous section to calculate the effect to all orders of radiative corrections involving virtual soft photons exchanged among the charged particle lines of a process $\alpha\to\beta$, as in Figure 13.3. By a 'soft' photon we mean one that carries momentum less than Λ, where Λ is some convenient dividing point chosen low enough to justify the approximations made in the previous section. We shall find that these soft photons introduce infrared divergences, so as a stop-gap

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

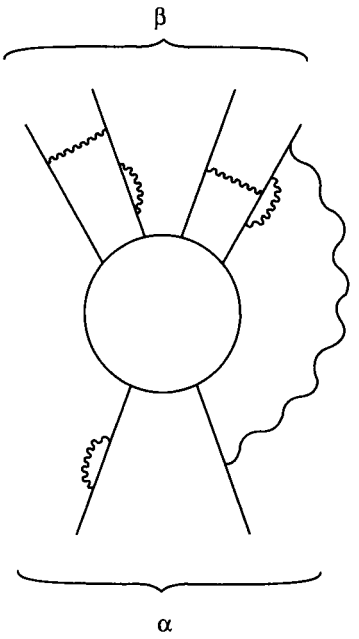

Figure 13.3. A typical dominant graph for the radiative corrections due to virtual soft photons to the S-matrix for the process $\alpha \to \beta$. Straight lines are particles in the states α and β (including possible hard photons); wavy lines are soft photons.

we will have to introduce also a lower bound λ on the photon momenta. It is important to recognize the difference between these two limits on the soft photon momenta. The upper cutoff Λ just serves to define what we mean by 'soft' photons; the Λ-dependence of the soft-photon radiative corrections is cancelled by the Λ-dependence of the rest of the amplitude, which is calculated including only virtual photons with momenta *greater* than Λ. On the other hand, the lower cutoff λ must eventually be removed by taking $\lambda \to 0$; as we shall see, the infrared divergences in this limit will be cancelled by the effects of real soft photon emission.

For each virtual soft photon we must supply a propagator factor

$$\frac{-i}{(2\pi)^4} \frac{\eta_{\mu\mu'}}{q^2 - i\epsilon} , \tag{13.2.1}$$

then multiply the amplitude (13.1.8) with the product of these propagators, contract photon polarization indices, and integrate over photon four-momenta. In addition for N virtual photons we must divide by a factor $2^N N!$, because the sum over all places to which we may attach the two ends of the soft photon lines includes spurious sums over the $N!$ permutations of the photon lines and over interchanges of the two ends of these lines. The effect of radiative corrections involving N soft photons

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

is then to multiply the matrix element $M_{\beta\alpha}$ for the process without such radiative corrections by a factor

$$\frac{1}{N!2^N}\left[\frac{1}{(2\pi)^4}\sum_{nm} e_n e_m \eta_n \eta_m J_{nm}\right]^N , \tag{13.2.2}$$

where

$$J_{nm} \equiv -i(p_n \cdot p_m)\int_{\lambda\le|\mathbf{q}|\le\Lambda} \frac{d^4q}{[q^2 - i\epsilon][p_n \cdot q - i\eta_n\epsilon][-p_m \cdot q - i\eta_m\epsilon]} . \tag{13.2.3}$$

Note that we have changed the sign of $p_m \cdot q$ in the denominator in (13.2.3), because if we define q as the momentum emitted by line n then $-q$ is the momentum emitted by line m.

Summing over N, we conclude that the matrix element for a process including radiative corrections due to any number of soft photons with momenta $|\mathbf{q}| \ge \lambda$ is given by

$$M^{\lambda}_{\beta\alpha} = M^{\Lambda}_{\beta\alpha} \exp\left[\frac{1}{2(2\pi)^4}\sum_{nm} e_n e_m \eta_n \eta_m J_{nm}\right] , \tag{13.2.4}$$

where $M^{\Lambda}_{\beta\alpha}$ is the amplitude including virtual photons only with momenta greater than Λ.

The integral over q^0 in (13.2.3) may be done by the method of residues. The integrand is analytic in q^0 except for four poles, at

$$\begin{aligned} q^0 &= |\mathbf{q}| - i\epsilon, & q^0 &= -|\mathbf{q}| + i\epsilon , \\ q^0 &= \mathbf{v}_n \cdot \mathbf{q} - i\eta_n\epsilon, & q^0 &= \mathbf{v}_m \cdot \mathbf{q} + i\eta_m\epsilon , \end{aligned}$$

where $\mathbf{v}_n \equiv \mathbf{p}_n/p_n^0$, and likewise for $\mathbf{v}_m$. If particle n is outgoing and particle m is incoming, then $\eta_n = +1$, $\eta_m = -1$, so by closing the q^0 contour in the upper half-plane we avoid the contributions from the poles at $q^0 = \mathbf{v}_n \cdot \mathbf{q} - i\eta_n\epsilon$ or $q^0 = \mathbf{v}_m \cdot \mathbf{q} + i\eta_m\epsilon$. Similarly, if n is incoming and m is outgoing we can avoid these two poles by closing the contour in the lower half-plane. In these two cases it is only one of the poles at $q^0 = \pm(|\mathbf{q}| - i\epsilon)$ that contributes, and we find a purely real integral:

$$J_{nm} = -\pi(p_n \cdot p_m)\int_{\lambda\le|\mathbf{q}|\le\Lambda} \frac{d^3q}{|\mathbf{q}|^3(E_n - \hat{\mathbf{q}} \cdot \mathbf{p}_n)(E_m - \hat{\mathbf{q}} \cdot \mathbf{p}_m)} \quad (\text{for } \eta_n = -\eta_m = \pm 1) \quad . \tag{13.2.5}$$

On the other hand, if particles n and m are both outgoing or both incoming, then the poles at $\mathbf{v}_n \cdot \mathbf{q} - i\eta_n\epsilon$ and $\mathbf{v}_m \cdot \mathbf{q} + i\eta_m\epsilon$ lie on opposite sides of the real q^0-axis, and we cannot avoid a contribution from one of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

them whichever way we close the contour:

$$J_{nm} = -\pi(p_n \cdot p_m) \int_{\lambda \le |\mathbf{q}| \le \Lambda} \frac{d^3q}{|\mathbf{q}|^3 (E_n - \hat{\mathbf{q}} \cdot \mathbf{p}_n)(E_m - \hat{\mathbf{q}} \cdot \mathbf{p}_m)} - \frac{4i\pi^3}{\beta_{nm}} \ln\left(\frac{\Lambda}{\lambda}\right) \quad (\text{for } \eta_n = \eta_m = \pm 1), \tag{13.2.6}$$

where β_{nm} is the relative velocity of particles n and m in the rest frame of either:

$$\beta_{nm} \equiv \sqrt{1 - \frac{m_n^2 m_m^2}{(p_n \cdot p_m)^2}} \,. \tag{13.2.7}$$

The imaginary term in Eq. (13.2.6) leads to an infrared-divergent phase factor[4] in Eq. (13.2.4), which drops out when we take the absolute value of the matrix element to calculate the rate for the process $\alpha \to \beta$. (This infinite phase factor is the relativistic counterpart of the well-known feature of non-relativistic Coulomb scattering, that the outgoing wave part of the Schrödinger wave function has a dependence on the radial coordinate r of the form $\exp(ipr - iv \ln r)/r$ instead of $\exp(ipr)/r$, where v is the product of charges divided by the relative velocity.[5]) The reaction rate *is* affected by the real part of J_{nm}, which for all η_n and η_m takes the value

$$\text{Re } J_{mn} = -\pi(p_n \cdot p_m) \int_{\lambda \le |\mathbf{q}| \le \Lambda} \frac{d^3q}{|\mathbf{q}|^3 (E_n - \hat{\mathbf{q}} \cdot \mathbf{p}_n)(E_m - \hat{\mathbf{q}} \cdot \mathbf{p}_m)} \,. \tag{13.2.8}$$

An elementary calculation gives

$$\text{Re } J_{mn} = \frac{2\pi^2}{\beta_{nm}} \ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right) \ln\left(\frac{\Lambda}{\lambda}\right) \,. \tag{13.2.9}$$

Using this in the absolute value squared of Eq. (13.2.4) gives the effect of soft virtual photons on the rates for the process $\Gamma_{\beta\alpha}$ as

$$\Gamma^{\lambda}_{\beta\alpha} = \left(\frac{\lambda}{\Lambda}\right)^{A(\alpha \to \beta)} \Gamma^{\Lambda}_{\beta\alpha} \,, \tag{13.2.10}$$

where $\Gamma^{\lambda}_{\beta\alpha}$ and $\Gamma^{\Lambda}_{\beta\alpha}$ are the rates for the process $\alpha \to \beta$ including radiative corrections of soft photons only with momenta greater than λ or Λ, respectively, and A is the exponent

$$A(\alpha \to \beta) = -\frac{1}{8\pi^2} \sum_{nm} \frac{e_n e_m \eta_n \eta_m}{\beta_{nm}} \ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right) \,. \tag{13.2.11}$$

Note that this makes sense only because the correction factor $(\lambda/\Lambda)^A$ has turned out to be the ratio of a function of λ to the same function

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

of Λ, since the two rates in Eq. (13.2.10) can only depend on λ and Λ, respectively.

The exponent A is always positive. For instance, in the scattering of a single charged particle by a neutral particle or an external potential, we must add terms in Eq. (13.2.11) where both n and m are the initial or the final charged particle (in which case $\eta_n\eta_m = +1$ and $\beta_{nm} = 0$), or n is the initial or final charged particle and m is the other one (in which case $\eta_n\eta_m = -1$ and $\beta_{nm} = \beta$, where $1 > \beta > 0$.) This gives

$$A = -\frac{e^2}{8\pi^2}\left[4 - \frac{2}{\beta}\ln\left(\frac{1+\beta}{1-\beta}\right)\right],$$

which is positive for all $1 > \beta > 0$. Because A is positive the effect of the infrared divergences introduced by soft virtual photons when summed to all orders is to make the rate for any given charged particle process $\alpha \to \beta$ *vanish* in the limit $\lambda \to 0$.

* * *

Before we go on to consider how soft real photon emission cancels these infrared divergences, we should pause to note a technicality in the above calculation which as far as I know has always been ignored in the literature. In calculating these radiative corrections we have included diagrams in which the virtual photon is absorbed and emitted at the *same* external charged-particle line, as well as those in which it is emitted and absorbed at different lines. But as we learned in Chapter 10, in calculating the S-matrix we are not supposed to include radiative corrections arising from insertions of self-energy subgraphs in external lines. This might suggest that we should drop the terms $n = m$ in Eq. (13.2.11), but then the cancellation of infrared divergences we will find in the next section would not be complete.

The resolution of this problem can be found in the observation that soft virtual photons produce infrared divergences not only directly, but also through their effect on the renormalization constants Z_n of the charged particle fields. (The renormalization constant Z_n is the one usually called Z_2 in theories like quantum electrodynamics with a single charged field of spin $\frac{1}{2}$.) It is the counterterms proportional to $Z_n - 1$ that cancel the effect of radiative corrections in external lines. To be specific, the renormalized field of a charged particle of type n is a factor $Z_n^{-1/2}$ times the unrenormalized field, so when we calculate the S-matrix using renormalized fields (corresponding to the omission of radiative corrections in external lines) we are introducing an extra factor $\prod_n Z_n^{-1/2}$, the product running over all charged particles in the initial and final states. (Of course there are also factors $Z_n^{-1/2}$ for neutral particles, but these are not

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

infrared-divergent.) In a slightly different notation, this factor is

$$\prod_f Z_f^{-E_f/2},$$

where Z_f is the field renormalization constant for fields of type f, E_f is the number of external lines of type f, and the product now runs over *all* charged field types. However, these field renormalization constants also appear in the interiors of diagrams; expressing an interaction of type i that involves N_{if} charged-particle fields of type f in terms of renormalized fields introduces an infrared-divergent factor

$$\prod_f (Z_f)^{N_{if}/2}.$$

(For instance, the counterterm $-ie(Z_2-1)A_\mu\bar{\psi}\gamma^\mu\psi$ in Eq. (11.1.9) together with the ordinary electromagnetic interaction $-ieA_\mu\bar{\psi}\gamma^\mu\psi$ yields a total interaction $-iZ_2eA_\mu\bar{\psi}\gamma^\mu\psi$. It was the infrared divergence in this Z_2 factor that was responsible for the infrared divergence arising from the second term in the brackets in Eq. (11.3.23), and the last term in Eq. (11.4.14).) There is also an infrared divergence in the propagators of the renormalized fields; the propagator of a renormalized charged field of type f when expressed in terms of the propagator of the unrenormalized field introduces a factor Z_f^{-1}. Putting this all together, the total number of factors of Z_f for each charged field type f, introduced by the counterterms to interactions and to radiative corrections on internal as well as external lines, is

$$\tfrac{1}{2}\sum_i V_i N_{if} - I_f - \tfrac{1}{2}E_f,$$

where I_f and E_f are the numbers of internal and external lines of type f, and V_i is the number of vertices for interactions of type i. We have already noted in Section 6.3 that this quantity vanishes for each f. Thus the counterterms which cancelled the radiative corrections on external lines are themselves cancelled by the Z_f factors arising from internal lines and vertices. Eq. (13.2.11) is therefore correct as it stands, including the terms with $n = m$.

13.3 Real Soft Photons; Cancellation of Divergences

The resolution of the infrared divergence problem encountered in the previous section is found in the observation that it is not really possible to measure the rate $\Gamma_{\beta\alpha}$ for a reaction $\alpha \to \beta$ involving definite numbers of photons and charged particles, because photons of very low energy can always escape undetected. What can be measured is the rate $\Gamma_{\beta\alpha}(E, E_T)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

for such a reaction to take place with no unobserved photon having an energy greater than some small quantity E, and with not more than some small total energy E_T going into any number of unobserved photons. (Of course, $E \leq E_T$. In an experiment without soft photon detectors, one can rely on measurements of the energies of the 'hard' particles in α and β to put a limit E_T on the total energy going into soft photons, and in this case we just set $E = E_T$.) We now turn to a calculation of this rate.

The S-matrix for emitting N real soft photons in a process $\alpha \to \beta$ is obtained by contracting each of the N photon polarization indices $\mu_1, \mu_2, \cdots$ on the amplitude (13.1.8) with the appropriate coefficient function

$$\frac{\epsilon^*_\mu(\mathbf{q}, h)}{(2\pi)^{3/2}\sqrt{2|\mathbf{q}|}},$$

where $\mathbf{q}$ is the photon momentum, $h = \pm 1$ is its helicity, and ϵ^μ is the corresponding photon polarization 'vector'.* This gives a photon emission matrix element (the S-matrix element with delta function omitted) as

$$M^{\lambda}_{\beta\alpha}(\mathbf{q}_1, h_1, \mathbf{q}_2, h_2, \cdots) = M^{\lambda}_{\beta\alpha} \times \prod_{r=1}^{N} (2\pi)^{-3/2}(2|\mathbf{q}_r|)^{-1/2} \sum_n \frac{\eta_n e_n [p_n \cdot \epsilon^*(\mathbf{q}_r, h_r)]}{p_n \cdot q_r}. \tag{13.3.1}$$

(The superscript λ is to remind us that these amplitudes are to be calculated with an infrared cutoff λ on the momenta of virtual photons. Eventually we shall take $\lambda \to 0$. The presence of soft virtual photons does not interfere with the result (13.3.1) because of the factorization discussed in Section 13.1.) The differential rate for emitting N soft photons into a volume $\prod_r d^3q_r$ of momentum space is given by squaring this matrix element, summing over helicities, and multiplying with $\prod_r d^3q_r$. We recall from Eq. (8.5.7) that for $q^2 = 0$, the helicity sums take the form

$$\sum_{h=\pm 1} \epsilon_\mu(\mathbf{q}, h)\epsilon^*_\nu(\mathbf{q}, h) = \eta_{\mu\nu} + q_\mu c_\nu + q_\nu c_\mu, \tag{13.3.2}$$

where $\mathbf{c} \equiv -\mathbf{q}/2|\mathbf{q}|^2$ and $c^0 \equiv 1/2|\mathbf{q}|$. The charge conservation condition $\sum_n \eta_n e_n = 0$ allows us to drop the terms in Eq. (13.3.2) involving q_μ or q_ν,

* We are using ϵ^μ instead of e^μ for photon polarization vectors, to avoid confusion with our use here of e_n for electric charges.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

yielding a differential rate**

$$d\Gamma^{\lambda}_{\beta\alpha}(\mathbf{q}_1,\cdots\mathbf{q}_N) = \Gamma^{\lambda}_{\beta\alpha}\prod_{r=1}^{N}\frac{d^3q_r}{(2\pi)^3(2|\mathbf{q}_r|)}\sum_{nm}\frac{\eta_n\eta_m e_n e_m(p_n\cdot p_m)}{(p_n\cdot q_r)(p_m\cdot q_r)}\,. \tag{13.3.3}$$

To calculate the differential rate for the emission of N soft photons with definite energies $\omega_r \equiv |\mathbf{q}_r|$ we must integrate Eq. (13.3.3) over the directions of the photon momenta $\mathbf{q}_r$. These integrals are the same as those we encountered in the integrals (13.2.8),

$$-\pi(p_n\cdot p_m)\int\frac{d^2\hat{\mathbf{q}}}{(E_n-\hat{\mathbf{q}}\cdot\mathbf{p}_n)(E_m-\hat{\mathbf{q}}\cdot\mathbf{p}_m)} = \frac{2\pi^2}{\beta_{nm}}\ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right)\,. \tag{13.3.4}$$

Integrating Eq. (13.3.3) over photon directions thus gives the differential rate for photons of energy $\omega_1,\cdots\omega_N$:

$$d\Gamma^{\lambda}_{\beta\alpha}(\omega_1\cdots\omega_N) = \Gamma^{\lambda}_{\beta\alpha}A(\alpha\to\beta)^N\frac{d\omega_1}{\omega_1}\cdots\frac{d\omega_N}{\omega_N}\,, \tag{13.3.5}$$

where $A(\alpha\to\beta)$ is the same constant encountered in the previous section:

$$A(\alpha\to\beta) = -\frac{1}{8\pi^2}\sum_{nm}\frac{e_n e_m\eta_n\eta_m}{\beta_{nm}}\ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right)\,.$$

We see from Eq. (13.3.5) that an unrestricted integral over the energies of the emitted photons would introduce another infrared divergence. However, unitarity demands that if we use an infrared cutoff for the momenta of the virtual photons (as implied by the superscript λ) then we must use the same infrared cutoff for the real photons. To calculate the rate $\Gamma^{\lambda}_{\beta\alpha}(E,E_T)$ for the reaction $\alpha\to\beta$ with not more than energy E going into any one unobserved photon and not more than energy E_T going into any number of unobserved photons (with E and E_T chosen small enough to justify the approximations used in deriving Eq. (13.3.1)), we must integrate Eq. (13.3.5) over all photon energies, subject to the limits $E\geq\omega_r\geq\lambda$ and $\sum_r\omega_r\leq E_T$, then divide by $N!$ because this integral includes configurations that differ only by permutations of the N soft photons, and finally sum over N. This gives

$$\Gamma^{\lambda}_{\beta\alpha}(E,E_T) = \Gamma^{\lambda}_{\beta\alpha}\sum_{N=0}^{\infty}\frac{A(\alpha\to\beta)^N}{N!}\int_{E\geq\omega_r\geq\lambda,\sum_r\omega_r\leq E_T}\prod_{r=1}^{N}\frac{d\omega_r}{\omega_r}\,. \tag{13.3.6}$$

This integral would factor into the product of N integrals over the individual ω_r were it not for the restriction $\sum_r\omega_r\leq E$. This restriction may

** The result for $|\mathbf{q}|d\Gamma_{\beta\alpha}(\mathbf{q})/\Gamma_{\beta\alpha}$ in the case $N=1$ corresponds to the distribution of energy emitted classically by a discontinuously changing current density four-vector $J^\mu(x) = \sum_n{}^{(t)}\delta^3(\mathbf{x}-\mathbf{v}_n t)\,p_n^\mu e_n/E_n$, with the sum here running only over particles in the initial state for $t<0$ and in the final state for $t>0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

be implemented by including as a factor in the integrand a step function

$$\theta(E_T - \sum_r \omega_r) = \frac{1}{\pi}\int_{-\infty}^{\infty} du\, \frac{\sin E_T u}{u} \exp\left(iu\sum_r \omega_r\right) . \tag{13.3.7}$$

Eq. (13.3.6) then becomes

$$\Gamma^{\lambda}_{\beta\alpha}(E, E_T) = \frac{1}{\pi}\int_{-\infty}^{\infty} du\, \frac{\sin E_T u}{u} \exp\left(A(\alpha \to \beta)\int_{\lambda}^{E} \frac{d\omega}{\omega}\, e^{i\omega u}\right) \Gamma^{\lambda}_{\beta\alpha} . \tag{13.3.8}$$

The integral in the exponent can be done in the limit $\lambda \ll E$ by writing it as the sum of the integral of $(e^{i\omega u} - 1)/\omega$, in which we can set $\lambda = 0$, and the integral of $1/\omega$, which is trivial. Rescaling the u and ω variables, this gives for $\lambda \ll E$:

$$\Gamma^{\lambda}_{\beta\alpha}(E, E_T) \to \mathcal{F}\left(E/E_T; A(\alpha \to \beta)\right) \left(\frac{E}{\lambda}\right)^{A(\alpha\to\beta)} \Gamma^{\lambda}_{\beta\alpha} , \tag{13.3.9}$$

where

$$\begin{aligned}\mathcal{F}(x; A) &\equiv \frac{1}{\pi}\int_{-\infty}^{\infty} du\, \frac{\sin u}{u} \exp\left(A\int_0^x \frac{d\omega}{\omega}\,(e^{i\omega u} - 1)\right) \\ &= 1 - \frac{A^2\theta(x - \frac{1}{2})}{2}\int_{1-x}^{x} \frac{d\omega}{\omega} \ln\left(\frac{x}{1-\omega}\right) + \cdots . \end{aligned} \tag{13.3.10}$$

For E and E_T of the same order and $A \ll 1$ the factor $\mathcal{F}(E/E_T, A)$ in Eq. (13.3.9) is close to unity; for instance,

$$\mathcal{F}(1; A) \simeq 1 - \tfrac{1}{12}\pi^2 A^2 + \cdots .$$

Because $A(\alpha \to \beta) > 0$, the factor $(E/\lambda)^{A(\alpha\to\beta)}$ in Eq. (13.3.9) becomes infinite in the limit $\lambda \to 0$. However, Eq. (13.2.10) shows that the rate $\Gamma^{\lambda}_{\beta\alpha}$ vanishes in this limit:

$$\Gamma^{\lambda}_{\beta\alpha} = \left(\frac{\lambda}{\Lambda}\right)^{A(\alpha\to\beta)} \Gamma^{\Lambda}_{\beta\alpha} .$$

Using this in Eq. (13.3.9) shows that the infrared cutoff λ drops out in the limit $\lambda \ll E$:

$$\Gamma^{\lambda}_{\beta\alpha}(E, E_T) \to \mathcal{F}\left(E/E_T; A(\alpha \to \beta)\right) \left(\frac{E}{\Lambda}\right)^{A(\alpha\to\beta)} \Gamma^{\Lambda}_{\beta\alpha} . \tag{13.3.11}$$

We remind the reader that the energy Λ is just a convenient dividing point between 'soft' photons which are taken into account explicitly in Eq. (13.3.11) and 'hard' photons whose effects are buried in $\Gamma^{\Lambda}_{\beta\alpha}$. The right-hand side of Eq. (13.3.11) is independent of Λ because $\Gamma^{\Lambda}_{\beta\alpha} \propto \Lambda^A$. However, in theories with a small coupling constant like quantum electrodynamics it is frequently a good strategy to take Λ to be sufficiently

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

small compared with the typical energies W involved in the collision so that the approximations made here apply for photon energies less than Λ, but large enough so that $A(\alpha \to \beta) \ln(W/\Lambda) \ll 1$. Then it may be a good approximation to calculate $\Gamma^{\Lambda}_{\beta\alpha}$ in lowest-order perturbation theory, with the dominant radiative corrections for $E \ll \Lambda$ given by the factor $(E/\Lambda)^A$ in Eq. (13.3.11).

* * *

The same cancellation of infrared divergences occurs for soft gravitons.[3] The rate for any process $\alpha \to \beta$, with not more than an energy E going into soft gravitons, turns out to be proportional to E^B, where

$$B = \frac{G}{2\pi} \sum_{nm} \eta_n \eta_m m_n m_m \frac{1+\beta_{nm}^2}{\beta_{nm}\sqrt{1-\beta_{nm}^2}} \ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right) . \tag{13.3.12}$$

13.4 General Infrared Divergences

The infrared divergence due to soft photons that we have been considering up to now in this chapter is just one example of a variety of infrared divergences that are encountered in various physical theories. Another example is provided by quantum electrodynamics with massless charged particles. Here even after the cancellation of infrared divergences due to soft photons, we find a logarithmic divergence in the exponent A in Eq. (13.3.11). According to Eqs. (13.2.11) and (13.2.7), for a process in which all charged particles are electrons, in the limit $m_e \to 0$ the exponent goes as

$$A \to -\frac{1}{4\pi^2} \sum_n e_n^2 - \frac{1}{4\pi^2} \sum_{n \neq m} e_n e_m \eta_n \eta_m \ln\left(\frac{2|p_n \cdot p_m|}{m_e^2}\right) \to -\frac{\ln m_e}{2\pi^2} \sum_n e_n^2 .$$

(In the last step we have used the charge conservation condition $\sum_n e_n \eta_n = 0$.) The infrared divergence in this formula arises from soft photons that are emitted in a direction *parallel* to the momentum of one of the 'hard' electrons in the initial or final state, but it occurs also even if the photon like the electron is not soft, because the propagator denominator $(p_n \pm q)^2$ vanishes for $p_n^2 = q^2 = 0$ if $\mathbf{p}_n$ is parallel to $\mathbf{q}$. To be a little more specific, for $p_n^2 = q^2 = 0$ the integral of this factor* over photon directions takes

* This factor is not squared, because the divergence occurs only in the interference between this term in the S-matrix element and terms in which the photon is emitted from some other charged particle line $m \neq n$. For $m = n$ the integral (13.2.8) is proportional to m_n^2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

the form

$$\int d^2\hat{q}\,(p \pm q)^{-2} = \mp \frac{\pi}{\sqrt{\mathbf{p}^2\mathbf{q}^2}} \int_0^\pi \frac{\sin\theta\, d\theta}{1-\cos\theta}\,,$$

where θ is the angle between the momenta of the photon and the charged particle. This integral diverges logarithmically at $\theta = 0$.

Of course, in the real world there are no massless electrically charged particles, but in reactions in which the typical value E^2 of the scalar products $|p_n \cdot p_m|$ is much larger than m_e^2, it is of interest to identify the places where large $\ln(m_e/E)$ factors appear. The dominant radiative correction in this case is often given by the term $-\ln(m_e/E)\sum_n e_n^2/2\pi^2$ in A. More importantly, in quantum chromodynamics there are massless particles, the gluons, that carry a conserved quantum number known as color that is analogous to electric charge, so that infrared divergences arise from the emission of parallel hard gluons from hard gluons or other hard colored particles in the initial or final states.

These infrared divergences are not, in general, eliminated by summing over suitable sets of final states. However, Lee and Nauenberg[6] have pointed out that the infrared divergences can be made to cancel if we not only sum over suitable final states, but also assume a certain probabilistic distribution of *initial* states. What follows is a modified version of their argument, which will immediately make clear why in the case of electrodynamics with massive electrons it was sufficient to sum over final states.

For these purposes it is convenient to return to 'old-fashioned' perturbation theory, in which the S-matrix is given by Eq. (3.2.7) and Eq. (3.5.3) as

$$S_{ba} = \delta(b-a) - 2i\pi\delta(E_a - E_b)T_{ba}\,, \tag{13.4.1}$$

where

$$T_{ba} = V_{ba} + \sum_{\nu=1}^{\infty} \int dc_1 \cdots dc_\nu \, \frac{V_{bc_1}V_{c_1c_2}\cdots V_{c_\nu a}}{(E_a - E_{c_1} + i\epsilon)\cdots(E_a - E_{c_\nu} + i\epsilon)}. \tag{13.4.2}$$

(The integrals over $c_1 \cdots c_\nu$ should be understood to include sums over the spins and types of particles in these states as well as integrals over the three-momenta of these particles.) Infrared divergences arise from (and only from) the vanishing of one or more of the energy denominators in this expression.

However, not all vanishing energy denominators give rise to infrared divergences. A general intermediate state c may have $E_c = E_a$, but usually this is just one point in the interior of the range of integration, and the integral over this range is rendered convergent by the prescription implied

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

by the $i\epsilon$ in the denominator. In order for an intermediate state c to produce an infrared divergence, it is necessary that the energy $E_c = E_a$ be reached at the *endpoint* of the range of integration. This happens for instance if the first intermediate state c_1 in Eq. (13.4.2) consists of the particles in the initial state a, with any of the massless particles in this state replaced with *jets*, consisting of any number of nearly parallel massless particles with a total momentum equal to that of the particle the jet replaces. In this case, the endpoint at which $E_{c_1} = E_a$ is the point in momentum space at which all of the massless particles in each jet are parallel. More generally, we can have any number of the massless particles in a replaced with jets of nearly parallel massless particles, plus any number of additional soft massless particles. The set of all such states will be called $D(a)$. (To be precise, we need to introduce a small angle Θ as well as a small energy Λ to define what we mean by 'nearly parallel' and 'soft'. We will not bother to show the dependence of the set $D(a)$ on Θ and Λ.) The states in $D(a)$ are 'dangerous', in the sense that the vanishing of the energy denominator $E_a - E_{c_1}$ at the endpoint can introduce an infrared divergence; the endpoint at which $E_{c_1} = E_a$ is the point at which all massless particles in each jet are parallel, and all soft massless particles have zero energy.

Furthermore, if $c_1, \cdots c_n$ are each in the set $D(a)$, then an intermediate state c_{n+1} in $D(a)$ is also dangerous in the same sense. On the other hand, if some intermediate state c_m is *not* in $D(a)$, then a later state c_k with $k > m$ would not be dangerous even if it belonged to the set $D(a)$, because the configuration of hard particles or jets with three-momenta equal to those of particles in the state a would be just an ordinary point inside the range of integration. In exactly the same way, we may define a set of states $D(b)$ in which one or more of the massless particles in the state b are replaced with jets of nearly parallel massless particles, each having the same total three-momentum as the particle it replaces, and we add any number of soft massless particles. An intermediate state c_m is dangerous if it belongs to the set $D(b)$ and if the later states c_k with $k > m$ all belong to $D(b)$.

To isolate the effects of these dangerous states we rewrite Eq. (13.4.2) in the form

$$T_{ba} = V_{ba} + \sum_{\nu=1}^{\infty} \left(V \left[\frac{\mathscr{P}_a + \mathscr{P}_b + \mathscr{P}_{\notin a,b}}{E_a - H_0 + i\epsilon} V \right]^{\nu} \right)_{ba} , \tag{13.4.3}$$

where $\mathscr{P}_a$, $\mathscr{P}_b$, and $\mathscr{P}_{\notin a,b}$ are the projection operators respectively on $D(a)$, $D(b)$, and on all other states. (It is assumed here that none of the charged particles in b have momenta close to that of some charged particle in a, so that the sets $D(a)$ and $D(b)$ do not overlap.) Now, for $\Lambda \to 0$ and $\Theta \to 0$, the dangerous intermediate states occupy so little phase space that they may be neglected wherever they do not lead to infrared divergences. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

power series (13.4.3) therefore becomes

$$T_{ba} = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{v=0}^{\infty}\left(\left[V\frac{\mathscr{P}_b}{E_a - H_0 + i\epsilon}\right]^r V\left[\frac{\mathscr{P}_{\notin a,b}}{E_a - H_0 + i\epsilon}V\right]^v \times\left[\frac{\mathscr{P}_a}{E_a - H_0 + i\epsilon}V\right]^s\right)_{ba}. \tag{13.4.4}$$

This would be exact if all of the projection operators $\mathscr{P}_{\notin a,b}$ between the leftmost and the rightmost were replaced with $\mathscr{P}_a + \mathscr{P}_b + \mathscr{P}_{\notin a,b}$, and if $\mathscr{P}_b$ and $\mathscr{P}_a$ on the left and right were replaced with $\mathscr{P}_b + \mathscr{P}_a$, but as remarked above this would have a negligible effect on the final result when Λ and Θ are sufficiently small.

Eq. (13.4.4) may be written in a more compact form:

$$T_{ba} = \left(\Omega_b^{-\dagger} T_S \Omega_a^{+}\right)_{ba}, \tag{13.4.5}$$

where, for future use, we define Ω_α^+ and Ω_β^- for general states α and β as:

$$(\Omega_\alpha^+)_{ca} \equiv \sum_{r=0}^{\infty}\left(\left[\frac{\mathscr{P}_\alpha}{E_a - H_0 + i\epsilon}V\right]^r\right)_{ca}, \tag{13.4.6}$$

$$(\Omega_\beta^-)_{db} \equiv \sum_{r=0}^{\infty}\left(\left[\frac{\mathscr{P}_\beta}{E_b - H_0 - i\epsilon}V\right]^r\right)_{db}, \tag{13.4.7}$$

and T_S is the 'safe' operator**

$$(T_S)_{dc} \equiv \sum_{v=0}^{\infty}\left(V\left[\frac{\mathscr{P}_{\notin c,d}}{E_c - H_0 + i\epsilon}V\right]^v\right)_{dc}. \tag{13.4.8}$$

All of the infrared divergences have now been isolated in the two operator factors Ω_b^- and Ω_a^+.

To eliminate these infrared divergences it is now only necessary to note that if it were not for the projection operators on the dangerous states, the operators Ω_b^- and Ω_a^+ would be just the unitary operators that according to Eq. (3.1.16) convert free-particle states into 'out' or 'in' states, respectively. These operators are therefore unitary if confined to the subspaces $D(\beta)$ and $D(\alpha)$ of states that would be dangerous for some given final state β and some given initial state α. That is, for general α and β

$$\Omega_\beta^- \mathscr{P}_\beta \Omega_\beta^{-\dagger} = \mathscr{P}_\beta, \tag{13.4.9}$$

$$\Omega_\alpha^+ \mathscr{P}_\alpha \Omega_\alpha^{+\dagger} = \mathscr{P}_\alpha. \tag{13.4.10}$$

** In $(\Omega_b^-)_{db}$ we are using the fact that T_{ba} is calculated with $E_b = E_a$, and in $(T_S)_{dc}$ we are using the fact that the projection operators $\mathscr{P}_a$ make $(\Omega_a^+)_{ca}$ vanish unless E_c is very close to E_a. Also, the factors $\Omega_b^{-\dagger}$ and Ω_a^+ in Eq. (13.4.5) make $\mathscr{P}_{\notin c,d} = \mathscr{P}_{\notin a,b}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

The transition rate is therefore free of infrared divergences if summed over the subspaces of states that would be dangerous for any given final and initial states β and α:

$$\sum_{a\in D(\alpha)}\sum_{b\in D(\beta)}|T_{ba}|^2 = \mathrm{Tr}\left\{\Omega_\beta^{-}\mathscr{P}_\beta\Omega_\beta^{-\dagger}T_S\Omega_\alpha^{+}\mathscr{P}_\alpha\Omega_\alpha^{+\dagger}T_S^\dagger\right\}$$
$$= \mathrm{Tr}\left\{\mathscr{P}_\beta T_S\mathscr{P}_\alpha T_S^\dagger\right\} = \sum_{a\in D(\alpha)}\sum_{b\in D(\beta)}|(T_S)_{ba}|^2\,. \tag{13.4.11}$$

In order to be satisfied that this really does solve the general problem of infrared divergences, it is necessary to argue that it is only sums like that in Eq. (13.4.11) that are experimentally measurable. It is plausible that we should have to sum over dangerous final states in order to have a measurable transition rate, since it is not possible experimentally to distinguish an outgoing charged (or colored) massless particle from a jet of massless particles with nearly parallel momenta and the same total energy,[7] together with an arbitrary number of very soft quanta, all with the same total charge (or color). The sum over initial states is more problematic. Presumably one may argue that truly massless particles are always produced as jets accompanied by an ensemble of soft quanta that is uniform within some volume of momentum space. However, to the best of my knowledge no one has given a complete demonstration that the sums of transition rates that are free of infrared divergences are the only ones that are experimentally measurable.

This problem does not arise in quantum electrodynamics (with massive charged particles), where as we have seen it is only necessary to sum over final states in order to eliminate infrared divergences. The reason for this difference can be traced to the fact that in electrodynamics the states $a, b, c, \cdots$ are direct products of states (labelled with Greek letters) with fixed numbers of charged particles and hard photons, times states containing only soft photons having energy less than some small quantity Λ. Then for a reaction in which some set of soft photons f is produced in a reaction $\alpha \to \beta$ among charged particles and hard photons, Eq. (13.4.5) simplifies to

$$T_{\beta f,\alpha} = \left(\Omega^{-}(\beta)^\dagger\Omega^{+}(\alpha)\right)_{f0}(T_S)_{\beta\alpha}\,, \tag{13.4.12}$$

where 0 denotes the soft photon vacuum, and $\Omega^{\pm}$ are calculated as before, but in the reduced Hilbert space consisting only of soft photons, and with the interactions of these photons taken as the interaction Hamiltonian with all charged particles in the fixed states indicated by the arguments β or α. Just as before, these operators are unitary in the 'dangerous' Hilbert

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

space $\mathscr{D}$ of soft photons, so[†]

$$\sum_{f\in\mathscr{D}} |T_{\beta f,\alpha 0}|^2 = |(T_S)_{\beta\alpha}|^2 \left(\Omega^+(\alpha)^\dagger\Omega^-(\beta)\Omega^-(\beta)^\dagger\Omega^+(\alpha)\right)_{00}$$
$$= |(T_S)_{\beta\alpha}|^2 \left(\Omega^+(\alpha)^\dagger\Omega^+(\alpha)\right)_{00} = |(T_S)_{\beta\alpha}|^2 \qquad (13.4.13)$$

without having to sum over initial states.

13.5 Soft Photon Scattering*

In our treatment in this chapter of soft photon interactions, we have up to now considered only processes in which the soft photons are emitted or absorbed in a process $\alpha \to \beta$ which was going on anyway. It is also possible to make useful general statements about processes in which the process $\alpha \to \beta$ is trivial, and the soft photons play an essential part in producing an interesting reaction. We will consider here the simplest and most important example of this sort, the scattering of a soft photon from a massive particle of arbitrary type and spin, where α and β are single-particle states. The complication here is that the leading term in the soft photon scattering amplitude does not come from the pole terms, but from non-pole terms that are related to the pole terms by the condition of current conservation.

The S-matrix for photon scattering may be put in the form

$$S(q,\lambda;p,\sigma \to q',\lambda';p',\sigma') = i(2\pi)^4\delta^4(q+p-q'-p') \times \frac{\epsilon_\nu^*(\mathbf{q}',\lambda')\,\epsilon_\mu(\mathbf{q},\lambda)\,M^{\nu\mu}_{\sigma',\sigma}(q;\mathbf{p}',\mathbf{p})}{(2\pi)^6\sqrt{4q^0q'^0}}, \qquad (13.5.1)$$

where q and q' are the initial and final photon four-momenta, p and p' are the initial and final target four-momenta, λ and λ' are the initial and final photon helicities, $\epsilon_\nu(\mathbf{q}',\lambda')$ and $\epsilon_\mu(\mathbf{q},\lambda)$ are the corresponding photon polarization vectors, and σ and σ' are the initial and final target spin z-components. According to the theorem of Section 6.4, the amplitude $M^{\nu\mu}$ may be expressed as

$$(2\pi)^{-3}M^{\nu\mu}_{\sigma',\sigma}(q;\mathbf{p}',\mathbf{p}) = \int d^4x\, e^{iq\cdot x}\left(\Psi_{\mathbf{p}',\sigma'},\, T\{J^\nu(0),J^\mu(x)\}\Psi_{\mathbf{p},\sigma}\right) + \cdots \qquad (13.5.2)$$

[†] The reason that we are now not encountering any factor $(E/\Lambda)^4$ like that in Eq. (13.3.11) is that we are here identifying the maximum energy E of the real soft photon states over which we are summing with the maximum energy Λ of the 'dangerous' soft photon states over which we sum in calculating $\Omega^\pm$.

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

where $J^\mu(x)$ is the electromagnetic current, and the dots indicate possible 'seagull' terms such as those in the theory of charged scalar fields in which the two photons interact at a single vertex instead of with separate currents. We now repeat the standard polology arguments described in Chapter 10 and already used in Section 13.1. Inserting a complete set of intermediate states between the current operators in Eq. (13.5.2), integrating over x and isolating the one-particle intermediate states gives

$$M^{\nu\mu}(q;\mathbf{p}',\mathbf{p}) = \frac{G^\nu(\mathbf{p}',\mathbf{p}+\mathbf{q})G^\mu(\mathbf{p}+\mathbf{q},\mathbf{p})}{E(\mathbf{p}+\mathbf{q})-E(\mathbf{p})-q^0-i\epsilon} + \frac{G^\mu(\mathbf{p}',\mathbf{p}'-\mathbf{q})G^\nu(\mathbf{p}'-\mathbf{q},\mathbf{p})}{E(\mathbf{p}'-\mathbf{q})-E(\mathbf{p}')+q^0-i\epsilon} + N^{\nu\mu}(q;\mathbf{p}',\mathbf{p}), \quad (13.5.3)$$

where G^μ is the one-particle matrix element of the current

$$(2\pi)^{-3}G^\mu_{\sigma',\sigma}(\mathbf{p}',\mathbf{p}) \equiv \left(\Psi_{\mathbf{p}',\sigma'}, J^\mu(0)\,\Psi_{\mathbf{p},\sigma}\right) \quad (13.5.4)$$

and $N^{\nu\mu}$ represents the contribution of states other than the one-particle state itself, plus any direct two-photon interaction terms. (Eq. (13.5.3) is to be understood in the sense of matrix multiplication, with spin indices not shown explicitly.) About $N^{\nu\mu}$ we know very little, except that it does not have the singularity at $q^\mu \to 0$ exhibited by the first two terms, and therefore may be expanded in powers of q^μ.

We now use the current conservation (or gauge invariance) conditions:

$$q_\mu M^{\nu\mu}(q;\mathbf{p}',\mathbf{p}) = 0\,, \quad (13.5.5)$$

$$\mathbf{q}\cdot\mathbf{G}(\mathbf{p}+\mathbf{q},\mathbf{p}) = [E(\mathbf{p}+\mathbf{q})-E(\mathbf{p})]G^0(\mathbf{p}+\mathbf{q},\mathbf{p})\,, \quad (13.5.6)$$

$$\mathbf{q}\cdot\mathbf{G}(\mathbf{p}',\mathbf{p}'-\mathbf{q}) = [E(\mathbf{p}')-E(\mathbf{p}'-\mathbf{q})]G^0(\mathbf{p}',\mathbf{p}'-\mathbf{q})\,. \quad (13.5.7)$$

Applied to Eq. (13.5.3), these conditions yield the condition we need on $N^{\nu\mu}$:

$$q_\mu N^{\nu\mu}(q;\mathbf{p}',\mathbf{p}) = -G^\nu(\mathbf{p}',\mathbf{p}+\mathbf{q})\,G^0(\mathbf{p}+\mathbf{q},\mathbf{p}) + G^0(\mathbf{p}',\mathbf{p}'-\mathbf{q})\,G^\nu(\mathbf{p}'-\mathbf{q},\mathbf{p})\,. \quad (13.5.8)$$

We also note that $M^{\nu\mu}$ satisfies the 'crossing symmetry' condition

$$M^{\nu\mu}(q;\mathbf{p}',\mathbf{p}) = M^{\mu\nu}(p'-p-q;\mathbf{p}',\mathbf{p}) \quad (13.5.9)$$

and since the pole terms in Eq. (13.5.3) evidently satisfy this condition, so also does $N^{\nu\mu}$:

$$N^{\nu\mu}(q;\mathbf{p}',\mathbf{p}) = N^{\mu\nu}(p'-p-q;\mathbf{p}',\mathbf{p})\,. \quad (13.5.10)$$

We will use these conditions to determine the first terms in the expansion of $N^{\nu\mu}$ in powers of momenta.

First we need to say something about the expansion of the one-particle current matrix elements $G^\mu(\mathbf{p}',\mathbf{p})$ in powers of the momenta $\mathbf{p}'$ and $\mathbf{p}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

Space inversion invariance (to the extent that it is applicable) tells us that the expansion of G^0 and G^i (with $i = 1,2,3$) contain terms respectively of only even and odd order in the momenta. According to Eq. (10.6.3), the term in $G^0_{\sigma',\sigma}$ of zeroth order in momenta is $e\delta_{\sigma',\sigma}$, where e is the particle charge. The current conservation condition then tells us that to second order in momenta,

$$(\mathbf{p}' - \mathbf{p}) \cdot \mathbf{G}_{\sigma',\sigma}(\mathbf{p}', \mathbf{p}) = \left(\frac{\mathbf{p}'^2}{2m} - \frac{\mathbf{p}^2}{2m} \right) e\delta_{\sigma',\sigma} \, .$$

The terms in $\mathbf{G}$ of first order in momenta are thus given by $e(\mathbf{p}'+\mathbf{p})\delta_{\sigma',\sigma}/2m$, plus a possible first-order term orthogonal to $\mathbf{p}' - \mathbf{p}$, which rotational invariance tells us must be proportional to $(\mathbf{p}' - \mathbf{p}) \times \mathbf{J}_{\sigma',\sigma}$, where $\mathbf{J}$ is the familiar spin matrix of the charged particle. Summarizing these results, we have the expansions

$$G^0(\mathbf{p}', \mathbf{p}) = e1 + \text{quadratic} \, , \tag{13.5.11}$$

$$\mathbf{G}(\mathbf{p}', \mathbf{p}) = \frac{e1}{2m}(\mathbf{p}' + \mathbf{p}) + \frac{i\mu}{j}\mathbf{J} \times (\mathbf{p}' - \mathbf{p}) + \text{cubic} \, , \tag{13.5.12}$$

where '1' is the unit spin matrix, and 'quadratic' and 'cubic' refer to the order of the neglected terms in powers of the small momenta $\mathbf{p}$ and $\mathbf{p}'$. The coefficient μ/j in Eq. (13.5.12) is real because the current is Hermitian. With the coefficient written in this way (with j the spin of the charged particle), μ is the quantity known as the magnetic moment of the particle.

Now let us return to $N^{\nu\mu}$, and consider the expansion of Eq. (13.5.8) in powers of the small momenta q^μ, $\mathbf{p}$ and $\mathbf{p}'$. Taking $\nu = 0$ in Eq. (13.5.8) shows that $q_\mu N^{0\mu}$ is at least quadratic in these small quantities. There is no constant vector orthogonal to q^μ, so $N^{0\mu}$ must be at least of first order in small momenta. The crossing symmetry condition (13.5.10) then tells us that N^{i0} must also be at least of first order in small momenta. Taking $\nu = i$ in Eq. (13.5.8) and using Eq. (13.5.12) then tells us that

$$q_k N^{ik} = -\frac{e^2 q^i}{m} + \text{quadratic}$$

and hence

$$N^{ik} = -\frac{e^2}{m}\delta_{ik} + \text{linear} \, . \tag{13.5.13}$$

Since G^i is at least of first order in the small momenta, so are the pole terms in Eq. (13.5.3) for M^{ik}, leaving us in zero order with only the *non-pole* term N^{ik}:

$$M^{ik}(0;0,0) = N^{ik}(0;0,0) = -\frac{e^2}{m}\delta_{ik} \, . \tag{13.5.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

From this we can calculate the soft photon scattering cross-section. But there is no need for this calculation; now that we know that the photon scattering amplitude in the limit of zero momentum depends only on the target particle mass and charge, and is of second order in the charge, we can immediately use the results of *any* second-order calculation of the photon scattering cross-section for target particles of any given spin, such as our result (8.7.42) for the differential photon scattering cross-section in quantum electrodynamics:

$$\frac{d\sigma}{d\Omega} = \frac{e^4}{32\pi^2 m^2}(1 + \cos^2\theta)\,. \tag{13.5.15}$$

We now see that this is a universal formula, valid in the low-energy limit for target particles of mass m and charge e and of arbitrary type and spin, even if these particles are composite and strongly interacting, like atomic nuclei. Gell-Mann and Goldberger and Low[8] have shown that these results may be extended to give the next-to-leading term in the soft photon scattering amplitude in terms of the target particle's mass, charge, and magnetic moment.

13.6 The External Field Approximation*

It is intuitively obvious that a heavy charged particle like the nucleus of an atom acts approximately like the source of a classical external field. In this section we will see how to justify this approximation, and will gain some idea of its limitations.

Consider a Feynman diagram or a part of a Feynman diagram in which a heavy charged particle passing through the diagram from the initial to the final state emits N off-shell photons with four-momenta $q_1, q_2, \cdots q_N$ and polarization indices $\mu_1, \mu_2, \cdots \mu_N$. The sum of all such graphs or subgraphs (not including the N photon propagators) yields an amplitude

$$\begin{aligned}&\int d^4x_1\, d^4x_2 \cdots d^4x_N\, e^{-iq_1\cdot x_1} e^{-iq_2\cdot x_2} \cdots e^{-iq_N\cdot x_N} \\ &\qquad \times \langle \mathbf{p}', \sigma' | T\left\{ J^{\mu_1}(x_1),\, J^{\mu_2}(x_2),\, \cdots J^{\mu_N}(x_N) \right\} | \mathbf{p}, \sigma \rangle \\ &\equiv \mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1, q_2, \cdots q_N; p)\end{aligned} \tag{13.6.1}$$

with the matrix element calculated including all interactions in which the heavy particle may participate, including strong nuclear forces. This amplitude has a multiple pole at $q_1, q_2, \cdots q_N \to 0$, arising from terms in

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

the matrix elements of the product of currents in which the intermediate states consist of just the same heavy particle as in the initial and final states. This multiple pole dominates (13.6.1) when all components of $q_1, q_2, \cdots q_N$ are small compared with all energies and momenta associated with the dynamics of the (perhaps composite) heavy particle. In this case the methods of Section 10.2 give**

$$\mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1, q_2, \cdots q_N; p) \to \frac{(-i)^{N-1}}{2p^0(2\pi)^3}(2\pi)^4 \times \delta^4(p' + q_1 + q_2 + \cdots + q_N - p) \sum_{\sigma_1,\sigma_2,\cdots\sigma_{N-1}} \times \frac{\mathscr{G}^{\mu_1}_{\sigma',\sigma_1}(p)\,\mathscr{G}^{\mu_2}_{\sigma_1,\sigma_2}(p)\cdots\mathscr{G}^{\mu_N}_{\sigma_{N-1},\sigma}(p)}{[2p\cdot q_1 - i\epsilon][2p\cdot(q_1+q_2) - i\epsilon]\cdots[2p\cdot(q_1+\cdots+q_{N-1}) - i\epsilon]} + \text{permutations}, \tag{13.6.2}$$

where

$$\frac{\mathscr{G}^{\mu}_{\sigma',\sigma}(p)}{2p^0(2\pi)^3} \equiv \langle \mathbf{p}, \sigma' | J^{\mu}(0) | \mathbf{p}, \sigma \rangle \tag{13.6.3}$$

and '+ permutations' indicates that we are to sum over all permutations of the N photons. For applications to atomic systems it is important to recognize that (13.6.1) applies for particles of arbitrary spin that have strong as well as electromagnetic interactions, like atomic nuclei.

We also note that for particles of arbitrary spin and charge Ze, the matrix elements of the electric current between states of equal four-momenta are†

$$\langle p, \sigma' | J^{\mu}(0) | p, \sigma \rangle = \frac{Ze\,p^{\mu}\delta_{\sigma'\sigma}}{p^0(2\pi)^3}, \tag{13.6.4}$$

so that

$$\mathscr{G}^{\mu}_{\sigma'\sigma}(p) = 2Ze\,p^{\mu}\delta_{\sigma'\sigma}. \tag{13.6.5}$$

The important thing about Eq. (13.6.5) is that these matrices all commute,

** In perturbation theory, the denominators come from the denominators of the propagators:

$$(p' + q_1 + \cdots q_r)^2 + m^2 - i\epsilon \to 2p'\cdot(q_1 + \cdots q_r) - i\epsilon \to 2p\cdot(q_1 + \cdots q_r) - i\epsilon,$$

while the numerators of the propagators provide factors $\sum uu^{\dagger}$ that together with the photon emission vertex matrices yield the matrix elements (13.6.3). The matrix $\mathscr{G}^{\mu}$ differs from the matrix G^{μ} of the previous section by a factor $2p^0$.

† This is most easily proved by first noting that in the Lorentz frame in which the particle is at rest, rotational invariance requires that the matrix elements of the current have vanishing space components and a time component proportional to $\delta_{\sigma',\sigma}$, with no other dependence on σ or σ'. The constant of proportionality is supplied by Eq. (10.6.3), and a Lorentz transformation then gives Eq. (13.6.4).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

so their product can be factored out of the sum over permutations:

$$\mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1,q_2,\cdots q_N;p) \quad \rightarrow$$
$$\frac{(-i)^{N-1}(Ze)^N p^{\mu_1}p^{\mu_2}\cdots p^{\mu_N}}{p^0(2\pi)^3}(2\pi)^4\delta^4(p'+q_1+q_2+\cdots+q_N-p)\delta_{\sigma',\sigma}$$
$$\times\left[\frac{1}{[p\cdot q_1-i\epsilon][p\cdot(q_1+q_2)-i\epsilon]\cdots[p\cdot(q_1+\cdots+q_{N-1})-i\epsilon]}\right.$$
$$\left.+\text{ permutations}\right] . \tag{13.6.6}$$

To leading order in the qs the delta function here may be written

$$\delta^4(p'+q_1+\cdots+q_N-p)=p^0\delta^3(\mathbf{p}'+\mathbf{q}_1+\cdots+\mathbf{q}_N-\mathbf{p})\delta(p\cdot(q_1+\cdots+q_N)) . \tag{13.6.7}$$

Fortunately, it turns out that the result of summing over permutations here is much simpler than the individual terms. For $p\cdot(q_1+\cdots+q_N)=0$, we have

$$\left[\frac{1}{[p\cdot q_1-i\epsilon][p\cdot(q_1+q_2)-i\epsilon]\cdots[p\cdot(q_1+\cdots+q_{N-1})-i\epsilon]}\right.$$
$$\left.+\text{ permutations}\right]=(2i\pi)^{N-1}\delta(p\cdot q_1)\,\delta(p\cdot q_2)\cdots\delta(p\cdot q_{N-1}) . \tag{13.6.8}$$

For instance, for $N=2$ this reads:

$$\frac{1}{[p\cdot q_1-i\epsilon]}+\frac{1}{[p\cdot q_2-i\epsilon]}=\frac{1}{[p\cdot q_1-i\epsilon]}+\frac{1}{[-p\cdot q_1-i\epsilon]}=2i\pi\delta(p\cdot q_1) .$$

The general result (13.6.8) can be obtained most easily as the Fourier transform of the identity

$$\theta(\tau_1-\tau_2)\,\theta(\tau_2-\tau_3)\cdots\theta(\tau_{N-1}-\tau_N)+\text{permutations}=1 .$$

Inserting Eq. (13.6.8) in Eq. (13.6.6) gives our final result for the amplitude (13.6.1):

$$\mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1,q_2,\cdots q_N;p)\rightarrow(Ze)^N(2\pi)^N\,\delta_{\sigma',\sigma}\,p^{\mu_1}p^{\mu_2}\cdots p^{\mu_N}$$
$$\times\delta^3(\mathbf{p}'+\mathbf{q}_1+\mathbf{q}_2+\cdots+\mathbf{q}_N-\mathbf{p})\delta(p\cdot q_1)\,\delta(p\cdot q_2)\cdots\delta(p\cdot q_N) . \tag{13.6.9}$$

This result applies to relativistic as well as slowly moving heavy particles, and can be used to derive the 'Weizsäcker-Williams' approximation[9] for charged-particle scattering. In the special case of a non-relativistic heavy charged particle, with $|\mathbf{p}|\ll p^0$, Eq. (13.6.9) further simplifies to

$$\mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1,q_2,\cdots q_N;p)\rightarrow(Ze)^N(2\pi)^N n^{\mu_1}n^{\mu_2}\cdots n^{\mu_N}$$
$$\times\delta^3(\mathbf{p}'+\mathbf{q}_1+\mathbf{q}_2+\cdots+\mathbf{q}_N-\mathbf{p})\delta(q_1^0)\,\delta(q_2^0)\cdots\delta(q_N^0)\,\delta_{\sigma',\sigma}, \tag{13.6.10}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

where n is a unit time-like vector

$$n^0 = 1\,, \qquad \mathbf{n} = 0\,.$$

Now suppose that a single heavy non-relativistic particle of charge Ze with normalized momentum space wave function $\chi_\sigma(\mathbf{p})$ appears in both the initial and final states. Using the Fourier representation of the delta function in Eq. (13.6.10), the matrix element of $\mathscr{G}$ in this state is

$$\int d^3p\, d^3p'\, \chi^*_{\sigma'}(\mathbf{p}')\, \chi_\sigma(\mathbf{p})\, \mathscr{G}^{\mu_1\mu_2\cdots\mu_N}_{\sigma',\sigma}(q_1, q_2, \cdots q_N; p) \rightarrow$$
$$\int d^3X \sum_\sigma |\psi_\sigma(\mathbf{X})|^2 \prod_{r=1}^{N} 2\pi\, Ze\, n^{\mu_r} \delta(q_r^0) e^{-i\mathbf{q}_r\cdot\mathbf{X}} \tag{13.6.11}$$

where $\psi(\mathbf{X})$ is the coordinate space wave function:

$$\psi_\sigma(\mathbf{X}) \equiv (2\pi)^{-3/2} \int d^3p\, \chi_\sigma(\mathbf{p}) e^{i\mathbf{p}\cdot\mathbf{X}}\,. \tag{13.6.12}$$

Because of the factorization in Eq. (13.6.11), the effect of including a heavy charged particle in this state is then the same as that of adding any numbers of a new kind of vertex in the momentum space Feynman rules, in which light Dirac particles of charge $-e$ such as electrons interact with an external field, with each such vertex contributing to the overall amplitude a factor[†] (now including the photon propagator and electron–photon vertex)

$$i \int d^4q \left[\frac{-i}{(2\pi)^4} \frac{1}{q^2 - i\epsilon}\right] \left[2\pi\, Ze\, n_\mu \delta(q^0) e^{-i\mathbf{q}\cdot\mathbf{X}}\right] \left[(2\pi)^4\, e\, \gamma^\mu\, \delta^4(k - k' - q)\right] \tag{13.6.13}$$

where k and k' are the initial and final electron four-momenta. The complete scattering amplitude must then be averaged over the heavy particle position $\mathbf{X}$, with weight function $\sum_\sigma |\psi_\sigma(\mathbf{X})|^2$. The factor (13.6.13) is the same as would be produced by a new term in the interaction Lagrangian

$$\mathscr{L}_{\text{ext}}(x) = \mathscr{A}_\mu(x)\, J_e^\mu(x)\,, \tag{13.6.14}$$

where $J_e^\mu \equiv -ie\bar{\Psi}\gamma^\mu\Psi$ is the electric current of the electrons, and $\mathscr{A}^\mu$ is an external vector potential

$$\mathscr{A}^\mu(x) = \frac{1}{(2\pi)^4} \int d^4q\, e^{iq\cdot x} \left[\frac{2\pi\, Ze\, n^\mu \delta(q^0) e^{-i\mathbf{q}\cdot\mathbf{X}}}{q^2 - i\epsilon}\right]. \tag{13.6.15}$$

[†] The first factor here is the usual factor of i accompanying the constants in the interaction Lagrangian of the heavy charged particle in the Feynman rules.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

This, of course, is just the usual Coulomb potential:

$$\mathscr{A}^0(x) = \frac{Ze}{4\pi|\mathbf{x}-\mathbf{X}|}\,, \qquad \mathscr{A}(x) = 0. \tag{13.6.16}$$

If there is more than one heavy charged particle (as in a molecule) we must express $\mathscr{A}^\mu(x)$ as a sum of terms like (13.6.16), each with its own charge Ze and position $\mathbf{X}$.

It is useful to keep in mind what diagrams we are summing in using the external field approximation. Consider the interaction of a single electron (relativistic or non-relativistic) with a single heavy charged particle such as a proton or deuteron. If we ignore all other interactions, then the Feynman diagrams for the scattering of the electron due to its interaction with the external field are just those with any number of insertions of the electron-external field vertex (13.6.14) in the electron line. (See Figure 13.4.) But as shown by the sum over permutations in Eq. (13.6.2), these diagrams in the external field approximation come from diagrams in the underlying theory in which the photons attached to the electron line are attached to the heavy charged particle line in all possible orders. (See Figure 13.5.) The 'uncrossed ladder' diagrams (labelled L) of Figure 13.5 do *not* dominate this sum unless the electron as well as the heavy charged particle is non-relativistic. (These diagrams include contributions from terms in old-fashioned perturbation theory whose intermediate states contain the same particles as the initial and final states, leading to small energy denominators when the electron and heavy charged particle are both non-relativistic, while all other diagrams of Figure 13.5 correspond to intermediate states with either extra photons, electron–positron pairs, or heavy particle–antiparticle pairs, which are suppressed by large energy denominators.) The uncrossed ladders can be summed by solving an integral equation, known as the *Bethe–Salpeter equation*,[10] but there is no rationale for selecting out this subset of diagrams unless both particles are non-relativistic, in which case the Bethe–Salpeter equation just reduces to the ordinary non-relativistic Schrödinger equation, plus relativistic corrections associated with the spin–orbit coupling that can be treated as small perturbations. It must be said that the theory of relativistic effects and radiative corrections in bound states is not yet in entirely satisfactory shape.

In the derivation of the external field (13.6.16) we evaluated the interaction of the heavy charged particle with the electromagnetic field only to leading order in the photon momentum. There are corrections of higher order in the photon momentum arising from the heavy particle's magnetic dipole moment, electric quadrupole moment, etc. Also, of course, there are radiative corrections arising from Feynman diagrams beyond those of Figure 13.4, such as diagrams in which photons are emitted and absorbed

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

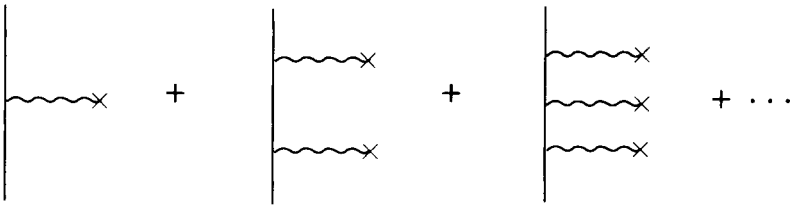

Figure 13.4. Diagrams for the scattering of an electron by an external electromagnetic field. Here straight lines represent the electron; wavy lines ending in crosses represent its interaction with an external field.

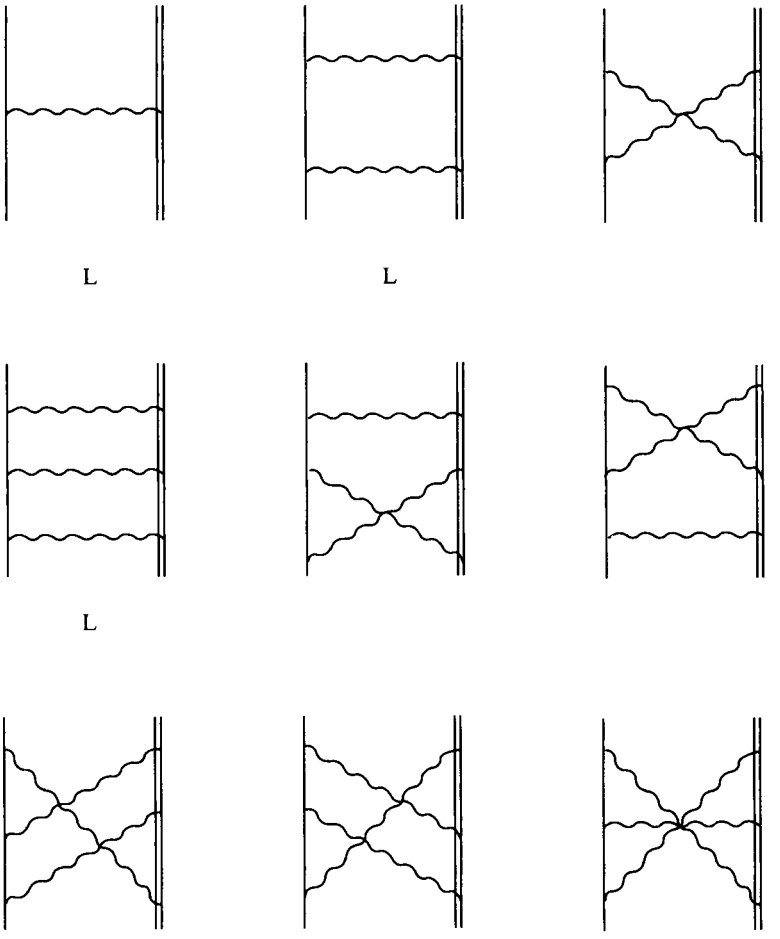

Figure 13.5. Diagrams for the scattering of an electron by a heavy, charged target particle, which in the limit of large target mass yield the same result as the diagrams of Figure 13.4. Here the single straight line is the electron; the double straight line is the heavy target particle; and wavy lines are virtual photons. Diagrams marked 'L' are called uncrossed ladder graphs; they dominate the sum when the electron as well as the target particle is non-relativistic.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

from the electron line or electron loops are inserted into photon lines. We shall see in the next chapter that in bound states the diagrams of Figure 13.4 must be included to all orders, but all other corrections of higher order in photon momenta or e may be included as perturbations to these diagrams.

Problems

1. Consider the process $e^+ + e^- \to \pi^+ + \pi^-$ in the center-of-mass frame at an energy of 1 GeV and scattering angle of 90°. Suppose that by measuring the energies of the final pions we determine that an energy of not more than $E_T \ll 1$ GeV is emitted in soft photons. How does the reaction rate depend on E_T?

2. Consider a massless spinless particle described by a scalar field ϕ, whose interaction Lagrangian density is of the form $\phi(x)J(x)$, where $J(x)$ involves only massive particle fields. Derive a formula for the rate of emission of arbitrary numbers of soft scalar particles in a process $\alpha \to \beta$, with these soft scalars having a total energy less than some small quantity E_T. Include radiative corrections due to soft scalars with energy less than some small quantity Λ.

3. Derive a formula for the next term beyond Eq. (13.5.14) for low-energy photon scattering on an arbitrary target.

4. Suppose that a spin one particle of very small mass m is described by a vector field $V^\mu(x)$, which couples only to a much heavier fermion described by a Dirac field $\psi(x)$, with an interaction Lagrangian density of the form $g V^\mu \bar{\psi}\gamma_\mu \psi$. Suppose that the heavy fermion normally decays into other particles that have no interaction with V^μ, releasing an energy W much greater than m. Consider such a decay process, but where an additional V^μ particle is emitted along with the other decay products, with the vector particle energy less than an upper limit E, in the range $W \gg E \gg m$. How does the rate of this process depend on E and m? Ignore radiative corrections.

5. Prove that Eq. (13.6.8) holds when $p \cdot (q_1 + \cdots q_N) = 0$.

References

1. F. Bloch and A. Nordsieck, *Phys. Rev.* **37**, 54 (1937); D. R. Yennie, S. C. Frautschi, and H. Suura, *Ann. Phys. (NY)* **13**, 379 (1961). Also

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

see K. T. Mahantappa, Ph.D. Thesis at Harvard University (1961), unpublished.

2. S. Weinberg, *Phys. Lett.* **9**, 357 (1964); *Phys. Rev.* **135**, B1049 (1964).
3. S. Weinberg, *Phys. Rev.* **140**, B515 (1965).
4. This phase was encountered in the perturbation series for non-relativistic Coulomb scattering, by R. H. Dalitz, *Proc. Roy. Soc. London* **206**, 509 (1951).
5. See, e.g., L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1949): Section 20.
6. T. D. Lee and M. Nauenberg, *Phys. Rev.* **133**, B1549 (1964). Also see T. Kinoshita, *J. Math. Phys.* **3**, 650 (1962); G. Sterman and S. Weinberg, *Phys. Rev. Lett.* **39**, 1416 (1977).
7. G. Sterman and S. Weinberg, Ref. 6.
8. F. E. Low, *Phys. Rev.* **96**, 1428 (1954); M. Gell-Mann and M. L. Goldberger, *Phys. Rev.* **96**, 1433 (1954). Also see S. Weinberg, in *Lectures on Elementary Particles and Quantum Field Theory — 1970 Brandeis Summer Institute in Theoretical Physics*, ed. by S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, MA, 1970).
9. E. J. Williams, *Kgl. Dan. Vid. Sel. Mat.-fys. Medd.* **XIII**, No. 4 (1935).
10. H. A. Bethe and E. E. Salpeter, *Phys. Rev.* **82**, 309 (1951); **84**, 1232 (1951).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.015

14

Bound States in External Fields

In our calculations of radiative corrections in Chapter 11 we went just one step beyond the lowest order in perturbation theory. However, there is a very important class of problems where even the simplest calculation requires that from the beginning we consider classes of Feynman diagrams of arbitrarily high order in coupling constants like e. These problems are those involving bound states – in electrodynamics, either ordinary atoms and molecules, or such exotic atoms as positronium or muonium.

It is easy to see that such problems necessarily involve a breakdown of ordinary perturbation theory. Consider for instance the amplitude for electron–proton scattering as a function of the center-of-mass energy E. As shown in Section 10.3, the existence of a bound state like the ground state of hydrogen implies the existence of a pole in this amplitude at $E = m_p + m_e - 13.6$ eV. However, no single term in the perturbation series for electron–proton scattering has such a pole. The pole therefore can only arise from a divergence of the sum over all diagrams at center-of-mass energies near $m_p + m_e$.

The reason for this divergence of the perturbation series is also easy to see, especially if for the moment we consider the time-ordered diagrams of old-fashioned perturbation theory instead of Feynman diagrams. Suppose that in the center-of-mass system the electron and proton both have momenta of magnitude $q \ll m_e$, and consider an intermediate state in which the electron and proton momenta are different but also of order q. The energy denominator factor contributed by this state will be of order $[q^2/m_e]^{-1}$. Each such state will also contribute a matrix element of the Coulomb interaction of order e^2/q^2 (the Fourier transform of e^2/r), and the corresponding momentum space integration will contribute a factor of order q^3. Putting this together, we see that each additional Coulomb interaction contributes an overall factor of order

$$[q^2/m_e]^{-1} \; [e^2/q^2] \; [q^3] = e^2 m_e/q \; .$$

Thus the perturbation theory should break down when q is less than or of the order of $e^2 m_e$, or in other words when the kinetic and potential

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

energies, which are of order q^2/m_e, are no larger than about $e^4 m_e$, which, of course, is of the order of the binding energy of hydrogen.

Our problem here is to learn how to use perturbation theory to evaluate radiative corrections in bound-state problems, summing to all orders those diagrams that need to be summed to all orders, and keeping only a finite number of those that don't.

14.1 The Dirac Equation

We shall limit ourselves in this chapter to problems in which bound states arise because of the Coulomb interaction of electrons (or muons) with heavy charged particles such as atomic nuclei. As shown in Section 13.6, this interaction may be taken into account by adding to the interaction Lagrangian a term* representing the effects of a c-number external vector potential $\mathscr{A}^\mu(x)$:

$$\begin{aligned}\mathscr{L}_{\mathscr{A}} = &-ie\bar{\Psi}\gamma^\mu\Psi\mathscr{A}_\mu - \tfrac{1}{2}(Z_3-1)(\partial^\mu\mathscr{A}^\nu - \partial^\nu\mathscr{A}^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu)\\ &-ie(Z_2-1)\mathscr{A}_\mu\bar{\Psi}\gamma^\mu\Psi\,,\end{aligned} \tag{14.1.1}$$

which is obtained by replacing the quantum vector potential A^μ with $A^\mu + \mathscr{A}^\mu$ in the interaction part of Eq. (11.1.6). For instance, for a single heavy particle of charge Ze at the origin,

$$\mathscr{A}^0(x) = \frac{Ze}{4\pi|\mathbf{x}|}\,, \qquad \mathscr{A}(x) = 0\,. \tag{14.1.2}$$

It is the interaction (14.1.1) that must be taken into account to all orders. In this section we will consider the theory with only this interaction, leaving radiative corrections to subsequent sections.

Physicists learn in kindergarten to approach this sort of problem by solving the wave equation of Dirac in the presence of the external field. It might seem unnecessary to derive this equation here, but as emphasized in Chapter 1, Dirac's original motivation for this equation as a sort of relativistic Schrödinger equation does not stand up to inspection. Also, in the course of our derivation we will discover the normalization conditions that have to be imposed on the solutions of the Dirac equation, which seemed somewhat *ad hoc* in Dirac's approach. The solutions of the Dirac equation discussed here will be important ingredients in our treatment of radiative corrections in the next section.

* In this chapter we return to the use of an upper case Ψ to denote the electron field in the Heisenberg picture, reserving a lower case ψ for the Dirac field with time-dependence governed solely by the c-number external field $\mathscr{A}^\mu(x)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

We will work here in a version of the Heisenberg picture, in which the time-dependence of operators is determined by a Hamiltonian including the external field interaction (14.1.1), but no other interactions. The electron field $\psi(x)$ in this picture satisfies the field equation

$$\left[\gamma^\lambda \frac{\partial}{\partial x^\lambda} + m + ie\gamma^\lambda \mathscr{A}_\lambda(x)\right] \psi(x) = 0 \,. \tag{14.1.3}$$

This is not the Dirac equation in the original sense of Dirac,[1] because $\psi(x)$ here is not a c-number wave function but a quantum operator. The c-number Dirac *wave functions* are defined by

$$u_N(x) \equiv (\Phi_0, \psi(x)\Phi_N) \,, \tag{14.1.4}$$

$$v_N(x) \equiv (\Phi_N, \psi(x)\Phi_0) \,, \tag{14.1.5}$$

where Φ_N are a complete orthonormal set of state-vectors, with Φ_0 the vacuum. It follows immediately from Eq. (14.1.3) that these functions satisfy the homogeneous Dirac equation

$$\left[\gamma^\lambda \frac{\partial}{\partial x^\lambda} + m + ie\gamma^\lambda \mathscr{A}_\lambda(x)\right] u_N(x) = \left[\gamma^\lambda \frac{\partial}{\partial x^\lambda} + m + ie\gamma^\lambda \mathscr{A}_\lambda(x)\right] v_N(x) = 0 \,. \tag{14.1.6}$$

We can also derive a normalization condition from the equal-time anti-commutation relations for the Dirac field. These are unaffected by the interaction (14.1.1), and therefore take the same form as for the free fields:

$$\{\psi(\mathbf{x}, t), \bar{\psi}(\mathbf{y}, t)\} = i\gamma^0 \delta^3(\mathbf{x} - \mathbf{y}) \,. \tag{14.1.7}$$

Taking the vacuum expectation value and inserting a sum over the states Φ_N, we find

$$\sum_N u_N(\mathbf{x}, t) u_N^\dagger(\mathbf{y}, t) + \sum_N v_N(\mathbf{x}, t) v_N^\dagger(\mathbf{y}, t) = \delta^3(\mathbf{x} - \mathbf{y}) \,, \tag{14.1.8}$$

it being understood that the sum over N includes an integral over continuum states as well as a sum over any discrete bound states.

We are chiefly interested in the case of a time-independent external field, like (14.1.2). In such cases, the states Φ_N may be taken as eigenstates of the Hamiltonian (including the interaction (14.1.1)) with energies E_N. Time-translation invariance then tells us that the $u_N(x)$ and $v_N(x)$ have the time-dependence:

$$u_N(\mathbf{x}, t) = e^{-iE_N t} u_N(\mathbf{x}), \qquad v_N(\mathbf{x}, t) = e^{+iE_N t} v_N(\mathbf{x}) \,. \tag{14.1.9}$$

The homogeneous Dirac equations (14.1.6) then become

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

$$i\gamma^0\left[\boldsymbol{\gamma}\cdot\boldsymbol{\nabla}+m+ie\gamma^\lambda\mathscr{A}_\lambda(\mathbf{x})\right]u_N(\mathbf{x})=E_N u_N(\mathbf{x})\,, \tag{14.1.10}$$

$$i\gamma^0\left[\boldsymbol{\gamma}\cdot\boldsymbol{\nabla}+m+ie\gamma^\lambda\mathscr{A}_\lambda(\mathbf{x})\right]v_N(\mathbf{x})=-E_N v_N(\mathbf{x})\,. \tag{14.1.11}$$

The minus sign on the right-hand side of Eq. (14.1.11) shows that the v_N are the famous 'negative-energy' solutions of Dirac. As shown by Eq. (14.1.8), these negative-energy solutions are needed to make up a complete set of wave functions. Of course, for moderate external fields there are no negative-energy *states* in the theory, so all E_N are positive, but there is still an important difference between the states with non-vanishing u_N or v_N: the definitions (14.1.4) and (14.1.5) show that a state can have $u_N \neq 0$ or $v_N \neq 0$ only if it has charge $-e$ or $+e$, respectively. It is in this sense that negative-energy solutions of the Dirac equation have something to do with the existence of antiparticles. However, this argument has nothing to do with the details of the Dirac equation, or even with the spin of the electron.

From the Dirac wave equations (14.1.10) and (14.1.11) we can easily see that wave functions of different energy are orthogonal. That is,

$$(E_M - E_N^*)\left(u_N^\dagger u_M\right)=\boldsymbol{\nabla}\cdot\left(u_N^\dagger\, i\gamma^0\boldsymbol{\gamma}u_M\right)\,,$$

so if $|\mathbf{x}|^2\left(u_N^\dagger\, i\gamma^0\boldsymbol{\gamma}u_M\right)$ remains bounded as $|\mathbf{x}|\to 0$ and $|\mathbf{x}|\to\infty$, then

$$\int d^3x\,\left(u_N^\dagger(\mathbf{x})u_M(\mathbf{x})\right)=0 \quad \text{if} \quad E_N\neq E_M^*\,. \tag{14.1.12}$$

With similar boundary conditions for the v_N, we find in the same way that

$$\int d^3x\,\left(v_N^\dagger(\mathbf{x})v_M(\mathbf{x})\right)=0 \quad \text{if} \quad E_N\neq E_M^*\,, \tag{14.1.13}$$

$$\int d^3x\,\left(u_N^\dagger(\mathbf{x})v_M(\mathbf{x})\right)=0 \quad \text{if} \quad E_N\neq -E_M^*\,. \tag{14.1.14}$$

Taking $N=M$, Eqs. (14.1.12) and (14.1.13) tell us that the energies are all real. Dropping the complex conjugation of E_M in Eqs. (14.1.12)–(14.1.14), we see that us of different energy are orthogonal, vs of different energy are orthogonal, and (as long as the potential is not strong enough to produce negative energy *states*) all us are orthogonal to all vs. By a suitable choice of the discrete quantum numbers that characterize the states along with the energy, we can then always arrange that

$$\int d^3x\,\left(u_N^\dagger(\mathbf{x})\,u_M(\mathbf{x})\right)=0 \quad \text{if} \quad N\neq M\,, \tag{14.1.15}$$

$$\int d^3x\,\left(v_N^\dagger(\mathbf{x})\,v_M(\mathbf{x})\right)=0 \quad \text{if} \quad N\neq M\,, \tag{14.1.16}$$

$$\int d^3x\,\left(u_N^\dagger(\mathbf{x})\,v_M(\mathbf{x})\right)=0\,. \tag{14.1.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Multiplying Eq. (14.1.8) on the right with $u_M(\mathbf{y})$ or $v_M(\mathbf{y})$, we find then that these wave functions must satisfy the normalization conditions

$$\int d^3y \left(u_N^\dagger(\mathbf{y})u_M(\mathbf{y})\right) = \int d^3y \left(v_N^\dagger(\mathbf{y})v_M(\mathbf{y})\right) = \delta_{NM} , \tag{14.1.18}$$

where δ_{NM} is a product of Kronecker deltas and momentum space delta functions, with normalization adapted to that used in defining $\sum_N$, in such a way that $\sum_N \delta_{NM} = 1$. These normalization conditions have nothing directly to do with any probabilistic interpretation of the Dirac wave functions, but arise instead from the anticommutation relations (14.1.7) for the fields.

Let us now specialize to the case of a pure electrostatic external field with $\mathscr{A} = 0$. In our standard representation of the Dirac matrices, we have

$$\gamma = i\begin{pmatrix} 0 & -\boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} , \qquad i\gamma^0 = \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

where $\boldsymbol{\sigma}$ is the usual three-vector of 2×2 Pauli matrices, and '1' and '0' here are the 2×2 unit and zero matrices. We introduce two-component wave functions f_N and g_N by setting

$$u_N = \frac{1}{\sqrt{2}}\begin{pmatrix} f_N + ig_N \\ f_N - ig_N \end{pmatrix} . \tag{14.1.19}$$

The energy eigenvalue condition (14.1.10) then takes the form:

$$(\boldsymbol{\sigma} \cdot \nabla) f_N = (E_N + e\mathscr{A}^0 + m) g_N , \tag{14.1.20}$$

$$(\boldsymbol{\sigma} \cdot \nabla) g_N = -(E_N + e\mathscr{A}^0 - m) f_N . \tag{14.1.21}$$

In the non-relativistic case where $e\mathscr{A}^0 r \approx Z\alpha \ll 1$ the binding energy $m - E_N$ is of order $Z^2\alpha^2 m$, while the gradient operator is of the order of $Z\alpha m$, so g_N is smaller than f_N by a factor of order $Z\alpha$. (To find the positron wave functions v_N we replace E_N everywhere by $-E_N$, so in this case f_N is smaller than g_N by the same factor.) We shall return to this non-relativistic case at the end of this section.

Physical states may be classified as even or odd under space inversion:

$$\mathsf{P}\Phi_N = \eta_N \Phi_N , \tag{14.1.22}$$

where η_N is a sign factor, ± 1. Recall that with the intrinsic parity of the electron defined to be $+1$, the Dirac field has the space-inversion property

$$\mathsf{P}\psi(\mathbf{x}, t)\mathsf{P}^{-1} = \beta\, \psi(-\mathbf{x}, t)$$

so Eqs. (14.1.4) and (14.1.22) show that the Dirac wave functions satisfy the parity condition

$$u_N(\mathbf{x}) = \eta_N \beta\, u_N(-\mathbf{x}) \tag{14.1.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

or equivalently

$$f_N(\mathbf{x}) = \eta_N f_N(-\mathbf{x}), \qquad g_N(\mathbf{x}) = -\eta_N g_N(-\mathbf{x}). \tag{14.1.24}$$

Note that the parity of the state is the same as the parity of $f_N(\mathbf{x})$, not $g_N(\mathbf{x})$.

Where the potential $\mathscr{A}^0$ is rotationally invariant, the solutions of the wave equations here may be classified according to their total angular momentum j and parity η. For a given j, the components f and g may be expanded in spherical harmonics with orbital angular momentum $\ell = j+\frac{1}{2}$ and $\ell = j-\frac{1}{2}$, but for a definite parity $\eta = (-1)^{j\mp\frac{1}{2}}$, Eqs. (14.1.24) show that we can only have $\ell = j \mp \frac{1}{2}$ in f and $\ell = j \pm \frac{1}{2}$ in g. The usual rules of angular-momentum addition then show that for a state of total angular momentum j, total angular-momentum z-component μ, and parity $(-1)^{j\mp\frac{1}{2}}$, the 'large' two-component wave function f has the form

$$f(\mathbf{x}) = \begin{pmatrix} C_{j\mp\frac{1}{2},\frac{1}{2}}(j\,\mu\,;\,\mu-\frac{1}{2}\quad\frac{1}{2})\,Y^{\mu-\frac{1}{2}}_{j\mp\frac{1}{2}}(\hat{\mathbf{x}}) \\ C_{j\mp\frac{1}{2},\frac{1}{2}}(j\,\mu\,;\,\mu+\frac{1}{2}\quad-\frac{1}{2})\,Y^{\mu+\frac{1}{2}}_{j\mp\frac{1}{2}}(\hat{\mathbf{x}}) \end{pmatrix} F(|\mathbf{x}|), \tag{14.1.25}$$

where C and Y are the usual Clebsch–Gordan coefficients and spherical harmonics.[2] Also, given any wave function of definite total angular momentum and parity we can construct another wave function with the same j and μ but opposite parity by applying the operator $\boldsymbol{\sigma}\cdot\hat{\mathbf{x}}$, so the 'small' components may be put in the form

$$g(\mathbf{x}) = (\boldsymbol{\sigma}\cdot\hat{\mathbf{x}}) \begin{pmatrix} C_{j\mp\frac{1}{2},\frac{1}{2}}(j\,\mu\,;\,\mu-\frac{1}{2}\quad\frac{1}{2})\,Y^{\mu-\frac{1}{2}}_{j\mp\frac{1}{2}}(\hat{\mathbf{x}}) \\ C_{j\mp\frac{1}{2},\frac{1}{2}}(j\,\mu\,;\,\mu+\frac{1}{2}\quad-\frac{1}{2})\,Y^{\mu+\frac{1}{2}}_{j\mp\frac{1}{2}}(\hat{\mathbf{x}}) \end{pmatrix} G(|\mathbf{x}|). \tag{14.1.26}$$

It is conventional to define the orbital angular momentum ℓ of the state as the orbital angular momentum of the 'large' components $f(\mathbf{x})$,

$$\ell = j \mp \tfrac{1}{2}, \tag{14.1.27}$$

so that the parity is always $(-1)^\ell$.

Inserting Eqs. (14.1.25) and (14.1.26) in Eqs. (14.1.20) and (14.1.21) yields the coupled differential equations

$$\frac{dG}{dr} + \frac{k+1}{r}G + (E + e\mathscr{A}^0 - m)F = 0, \tag{14.1.28}$$

$$\frac{dF}{dr} - \frac{k-1}{r}F - (E + e\mathscr{A}^0 + m)G = 0, \tag{14.1.29}$$

where for parity $\eta = (-1)^{j\mp\frac{1}{2}}$,

$$k \equiv \pm(j + \tfrac{1}{2}). \tag{14.1.30}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Let us now concentrate on the simple Coulomb field (14.1.2), for which $e\mathscr{A}^0 = Z\alpha/r$. The treatment of the Dirac equation in this case[3] is familiar, so it will be summarized briefly here just for completeness. It is easy to see that the solutions near the origin go as r^{s-1}, with $s^2 = k^2 - Z^2\alpha^2$. (Note that $k^2 \geq 1$, so the exponent s is real for $Z\alpha \leq 1$.) We must reject the solutions with $s < 0$ as being inconsistent with the normalization condition (14.1.18). The condition that the wave functions do not blow up for $r \to \infty$ then fixes the allowed values of the energy eigenvalues:

$$E_{n,j} = m\left[1 + \left(\frac{Z\alpha}{n - j - \frac{1}{2} + \sqrt{(j+\frac{1}{2})^2 - Z^2\alpha^2}}\right)^2\right]^{-1/2}, \qquad (14.1.31)$$

where n is a 'principal quantum number' with

$$j + \tfrac{1}{2} \leq n\,. \qquad (14.1.32)$$

It is noteworthy that these energies do not depend on the parity or ℓ, but only on n and j. For each n and j there are two solutions, corresponding to the two signs of k or the two possible parities, except that for $n = j + \frac{1}{2}$ we only have $k > 0$ and parity $(-1)^{j-\frac{1}{2}}$, so that $\ell = j - \frac{1}{2}$. With Eq. (14.1.32), this is the same as the familiar non-relativistic restriction that $\ell \leq n - 1$.

For light atoms with $Z\alpha \ll 1$, Eq. (14.1.31) yields the power series

$$E = m\left[1 - \frac{Z^2\alpha^2}{2n^2} + \frac{Z^4\alpha^4}{n^4}\left(\frac{3}{8} - \frac{n}{2j+1}\right) + \cdots\right]. \qquad (14.1.33)$$

The first two terms, of course, just represent the rest energy and the binding energy as given by the non-relativistic Schrödinger equation. The leading term that depends on j as well as n is the third term, the first relativistic correction. For $n = 1$ there is only one value of the total angular momentum, $j = \frac{1}{2}$, and since here $n = j + \frac{1}{2}$ there is also only one parity, $(-1)^{j-\frac{1}{2}} = +1$, corresponding to $\ell = 0$. It is therefore difficult to see the effects of the relativistic corrections in Eq. (14.1.33) in the $n = 1$ states of hydrogenic atoms, though as we shall see in Section 14.3, this has recently become possible. On the other hand, for $n = 2$ we have a $j = \frac{1}{2}$ state with both parities (i.e., $2s_{1/2}$ and $2p_{1/2}$) as well as a $2p_{3/2}$ state with $j = \frac{3}{2}$ and negative parity. Eq. (14.1.33) gives the splitting between the p states in hydrogen as

$$E(2p_{3/2}) - E(2p_{1/2}) = \frac{\alpha^4 m_e}{32} = 4.5283 \times 10^{-5}\,\text{eV}\,. \qquad (14.1.34)$$

Such relativistic line splitting is known as the *fine structure* of the atomic state. From the beginning it was known that this prediction is in good agreement with the observed fine structure. On the other hand, the Dirac

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

equation does not yield any energy difference in the $2s_{1/2}$ and $2p_{1/2}$ states, so this is a good place to look for the effects of further corrections, to be considered in Section 14.3.

Before closing this section we shall consider the approximate forms for the wave functions and matrix elements in the non-relativistic case for a general electrostatic potential $\mathscr{A}^0$. (For a Coulomb potential, this is the limit $Z\alpha \ll 1$.) Since here $E_N + m \simeq 2m \gg |e\mathscr{A}^0|$, the 'small' components of the electron wave function are given approximately in terms of the large components by

$$g_N \simeq (\boldsymbol{\sigma} \cdot \nabla) f_N / 2m \,. \tag{14.1.35}$$

Eq. (14.1.21) then becomes just the non-relativistic Schrödinger equation

$$\left[-\frac{\nabla^2}{2m} - e\mathscr{A}^0 \right] f_N \simeq (E_N - m) f_N \,. \tag{14.1.36}$$

Since there is no longer any coupling between spin and orbital degrees of freedom in the equation for f_N, we may find a complete set of solutions of this equation in the form

$$f_N = \chi_N \, \psi_N(\mathbf{x}) \,,$$

where χ_N is a two-component constant spinor, and $\psi_N(\mathbf{x})$ is an ordinary one-component solution of the Schrödinger equation. However, we often work with states that have definite values of the total angular momentum j, for which f_N is (for non-vanishing orbital angular momentum) a sum of such terms.

In the non-relativistic approximation, the four-component Dirac wave function takes the form

$$u_N \simeq \frac{1}{\sqrt{2}} \begin{bmatrix} (1 + i\boldsymbol{\sigma} \cdot \nabla / 2m) \, f_N \\ (1 - i\boldsymbol{\sigma} \cdot \nabla / 2m) \, f_N \end{bmatrix} \tag{14.1.37}$$

and Eq. (14.1.18) gives the normalization condition

$$\int d^3x (f_N^\dagger, f_M) \simeq \delta_{NM} - \tfrac{1}{4} (\mathbf{v}^2)_{NM} \,, \tag{14.1.38}$$

where

$$(\mathbf{v}^2)_{NM} \equiv -\frac{1}{m^2} \int d^3x f_N^\dagger(\mathbf{x}) \nabla^2 f_M(\mathbf{x}) \,.$$

In relating matrix elements in an external field to free-particle matrix elements, it is useful to note that the momentum space wave function in an energy eigenstate N may be written

$$u_N(\mathbf{p}) \equiv (2\pi)^{-3/2} \int d^3x \, e^{-i\mathbf{p}\cdot\mathbf{x}} u_N(\mathbf{x}) \simeq \sum_\sigma u(\mathbf{p}, \sigma) [f_N(\mathbf{p})]_\sigma \,, \tag{14.1.39}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

where $u(\mathbf{p},\sigma)$ is the free-particle Dirac spinor

$$u(\mathbf{p},\sigma) \simeq \frac{1}{\sqrt{2}} \begin{bmatrix} (1-\mathbf{p}\cdot\boldsymbol{\sigma}/2m)\,\chi_\sigma \\ (1+\mathbf{p}\cdot\boldsymbol{\sigma}/2m)\,\chi_\sigma \end{bmatrix},$$

$$\chi_{+\frac{1}{2}} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \chi_{-\frac{1}{2}} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and $f_N(\mathbf{p})$ is the Fourier transform of the two-component Schrödinger wave function

$$f_N(\mathbf{p}) \equiv (2\pi)^{-3/2} \int d^3x\, e^{-i\mathbf{p}\cdot\mathbf{x}} f_N(\mathbf{x}),$$

* * *

In closing, as an aid in calculating the effects of various perturbations, we note that the leading terms in the electron matrix elements of the sixteen independent 4×4 matrices are

$$(\bar{u}_M\, u_N) \simeq (f_M^\dagger f_N) - \frac{1}{4m^2}(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\,\boldsymbol{\sigma} \cdot \nabla f_N), \tag{14.1.40}$$

$$i(\bar{u}_M \gamma^0 u_N) \simeq (f_M^\dagger f_N) + \frac{1}{4m^2}(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\,\boldsymbol{\sigma} \cdot \nabla f_N), \tag{14.1.41}$$

$$(\bar{u}_M\, \boldsymbol{\gamma}\, u_N) \simeq \frac{1}{2m}[(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\,\boldsymbol{\sigma} f_N) - (f_M^\dagger \boldsymbol{\sigma}\,\boldsymbol{\sigma} \cdot \nabla f_N)], \tag{14.1.42}$$

$$(\bar{u}_M\, [\gamma^0, \boldsymbol{\gamma}]\, u_N) \simeq \frac{i}{m}[(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\,\boldsymbol{\sigma} f_N) + (f_M^\dagger \boldsymbol{\sigma}\,\boldsymbol{\sigma} \cdot \nabla f_N)], \tag{14.1.43}$$

$$(\bar{u}_M\, [\gamma^i, \gamma^j]\, u_N) \simeq 2i\epsilon_{ijk}\,(f_M^\dagger \sigma_k f_N), \tag{14.1.44}$$

$$(\bar{u}_M\, \gamma_5 \boldsymbol{\gamma}\, u_N) \simeq -i(f_M^\dagger \boldsymbol{\sigma} f_N), \tag{14.1.45}$$

$$(\bar{u}_M\, \gamma_5 \gamma^0\, u_N) \simeq \frac{1}{2m}[(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\, f_N) - (f_M^\dagger \boldsymbol{\sigma}\, \cdot \nabla f_N)], \tag{14.1.46}$$

$$(\bar{u}_M\, \gamma_5\, u_N) \simeq \frac{i}{2m}[(\nabla f_M^\dagger \cdot \boldsymbol{\sigma}\, f_N) + (f_M^\dagger \boldsymbol{\sigma}\, \cdot \nabla f_N)]. \tag{14.1.47}$$

14.2 Radiative Corrections in External Fields

We now consider radiative corrections to the results of the previous section, due to the interaction of electrons with the quantum electromagnetic field as well as the external field of the heavy charged particles. These radiative corrections can be calculated using Feynman diagrams of the usual sort, with the whole effect of the external field being to modify the propagator of the electron field in the presence of the external electromagnetic field (and to supply the external-field-dependent renormalization counterterms shown in Eq. (14.1.1)). To be specific, the effect of inserting any number

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

of vertices corresponding to the first term of the interaction (14.1.1) in an internal electron line of any graph is to replace the bare coordinate space propagator $-iS(x-y)$ with a corrected propagator

$$-iS_{\mathscr{A}}(x,y) \equiv -iS(x-y) + (-i)^2 \int d^4z_1\, S(x-z_1)e\gamma^\mu \mathscr{A}_\mu(z_1)S(z_1-y)$$
$$+(-i)^3 \int d^4z_1 \int d^4z_2\, S(x-z_1)e\gamma^\mu \mathscr{A}_\mu(z_1)S(z_1-z_2)e\gamma^\nu \mathscr{A}_\nu(z_2)S(z_2-y)$$
$$+\cdots, \tag{14.2.1}$$

where as usual

$$S(x-y) \equiv \frac{1}{(2\pi)^4}\int d^4p\, \frac{-i\gamma_\lambda p^\lambda + m}{p^2+m^2-i\epsilon} e^{ip\cdot(x-y)}.$$

(We must write $S_{\mathscr{A}}$ as a function of x and y rather than of $x-y$, because the external field invalidates translation invariance.) The theorem proved in Section 6.4 tells us that Eq. (14.2.1) is the same as

$$-iS_{\mathscr{A}}(x,y) = (\Phi_0, T\{\psi(x), \bar{\psi}(y)\}\Phi_0)_{\mathscr{A}} \tag{14.2.2}$$

with the subscript $\mathscr{A}$ on the right indicating that the vacuum state Φ_0 and electron field $\psi(x)$ are to be defined in a Heisenberg picture in which the only interaction taken into account is the interaction (14.1.1) with the external field. Inserting a complete set of intermediate states Φ_N in Eq. (14.2.2) yields an expression for the propagator in terms of the Dirac wave functions u_N and v_N introduced in the previous section

$$-iS_{\mathscr{A}}(x,y) = \theta(x^0-y^0)\sum_N u_N(x)\bar{u}_N(y) - \theta(y^0-x^0)\sum_M v_M(x)\bar{v}_M(y). \tag{14.2.3}$$

It is also possible to obtain the propagator (14.2.2) as the solution of the *inhomogeneous Dirac equation*:

$$\left[\gamma^\lambda \frac{\partial}{\partial x^\lambda} + m + ie\gamma^\lambda \mathscr{A}_\lambda(x)\right] S_{\mathscr{A}}(x,y) = \delta^4(x-y), \tag{14.2.4}$$

which follows from the field equations (14.1.3) and anticommutation relations (14.1.7), or formally from the perturbation series (14.2.1). Also Eq. (14.2.3) tells us that the propagator satisfies boundary conditions: its Fourier decomposition contains only 'positive frequency terms' proportional to $\exp(-iE(x^0-y^0))$ with $E>0$ for $x^0-y^0\to\infty$, and only 'negative frequency terms' proportional to $\exp(+iE(x^0-y^0))$ with $E>0$ for $x^0-y^0\to-\infty$. The inhomogeneous Dirac equation with these boundary conditions may be used to obtain a numerical solution[4] for this propagator even in cases where the external field is too strong to allow the use of the perturbation series (14.2.1). Once the propagator $S_{\mathscr{A}}(x,y)$ has

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

been calculated, the amplitudes for scattering in an external field can be calculated using ordinary Feynman diagrams, but with $S_{\mathscr{A}}(x,y)$ in place of $S(x-y)$ (and with $\mathscr{A}$-dependent renormalization counterterms inserted where appropriate).

Now let us see how to use the perturbation series with this corrected propagator to calculate the shifts of bound-state energy levels. Consider the full electron propagator $S'_{\mathscr{A}}(x,y)$, involving interactions of the electron with the quantum electromagnetic field as well as the external field:

$$-iS'_{\mathscr{A}}(x,y) \equiv (\Omega_0, T\{\Psi(x), \bar{\Psi}(y)\}\Omega_0)_{\mathscr{A}} \tag{14.2.5}$$

with $\Psi(x)$ the electron field in a Heisenberg picture including all interactions, and Ω_0 the vacuum eigenstate of the full Hamiltonian. For a time-independent external potential we can find a complete orthonormal set Ω_N of eigenstates of the full Hamiltonian with energies E'_N. Inserting a sum over these states in the operator product in Eq. (14.2.5), we find

$$\begin{aligned} -iS'_{\mathscr{A}}(x,y) = {} & \theta(x^0-y^0)e^{-iE'_N(x^0-y^0)}\sum_N U_N(\mathbf{x})\bar{U}_N(\mathbf{y}) \\ & -\theta(y^0-x^0)e^{-iE'_N(y^0-x^0)}\sum_N V_N(\mathbf{x})\bar{V}_N(\mathbf{y})\,, \end{aligned} \tag{14.2.6}$$

where

$$(\Omega_0, \Psi(\mathbf{x},t)\Omega_N) \equiv e^{-iE'_N t}U_N(\mathbf{x})\,, \tag{14.2.7}$$

$$(\Omega_N, \Psi(\mathbf{x},t)\Omega_0) \equiv e^{+iE'_N t}V_N(\mathbf{x})\,. \tag{14.2.8}$$

(The sum includes an integral over continuum states as well as a sum over discrete bound states. As before, U_N and V_N are non-zero only if the state Ω_N has charge $-e$ or $+e$, respectively.) We can redefine the propagator as a function of energy rather than time

$$S'_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) \equiv \int_{-\infty}^{\infty} dx^0\, e^{iE(x^0-y^0)}S'_{\mathscr{A}}(x,y)\,. \tag{14.2.9}$$

(Time-translation invariance dictates that $S'_{\mathscr{A}}(x,y)$ is a function of x^0-y^0 but not of x^0 and y^0 separately.) From Eq. (14.2.6) we see that

$$S'_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) = \sum_N \frac{U_N(\mathbf{x})\bar{U}_N(\mathbf{y})}{E'_N - E - i\epsilon} - \sum_N \frac{V_N(\mathbf{x})\bar{V}_N(\mathbf{y})}{E'_N + E - i\epsilon}\,. \tag{14.2.10}$$

In particular, $S'_{\mathscr{A}}(\mathbf{x},\mathbf{y};E)$ has a pole at any electron bound-state energy, and also at the negative of any positron bound-state energy. (Of course, positrons do not have bound states in the Coulomb field of an ordinary positively charged nucleus.)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Let us now consider the lowest-order radiative corrections to the complete propagator. The Feynman rules here give the complete propagator to this order as $S'_{\mathscr{A}} = S_{\mathscr{A}} + \delta S_{\mathscr{A}}$, with a correction term

$$\delta S_{\mathscr{A}}(x,y) = \int d^4z \int d^4w\, S_{\mathscr{A}}(x,z)\,\Sigma^*_{\mathscr{A}}(z,w)\,S_{\mathscr{A}}(w,y), \tag{14.2.11}$$

where $i\Sigma^*_{\mathscr{A}}$ is the sum of all one-loop diagrams with one incoming and one outgoing electron line (excluding final electron propagators) calculated using $S_{\mathscr{A}}(x,y)$ in place of $S(x-y)$ for internal electron lines, plus second-order renormalization counterterms. Using energy variables in place of time variables, this is

$$\delta S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) = \int d^3z \int d^3w\, S_{\mathscr{A}}(\mathbf{x},\mathbf{z};E)\,\Sigma^*_{\mathscr{A}}(\mathbf{z},\mathbf{w};E)\,S_{\mathscr{A}}(\mathbf{w},\mathbf{y};E)\,, \tag{14.2.12}$$

where

$$\Sigma^*_{\mathscr{A}}(\mathbf{z},\mathbf{w};E) \equiv \int dz^0\, e^{iE(z^0-w^0)}\Sigma^*_{\mathscr{A}}(z,w)\,. \tag{14.2.13}$$

The effect of these radiative corrections is to change the wave functions to $U_N = u_N + \delta u_N$ and $V_N = v_N + \delta v_N$ and the bound-state energies to $E'_N = E_N + \delta E_N$, so that the complete propagator is

$$\begin{aligned} S'_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) \simeq\; & S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) \\ & + \sum_N \frac{\delta u_N(\mathbf{x})\bar{u}_N(\mathbf{y}) + u_N(\mathbf{x})\delta\bar{u}_N(\mathbf{y})}{E_N - E} \\ & - \sum_N \frac{\delta v_N(\mathbf{x})\bar{v}_N(\mathbf{y}) + v_N(\mathbf{x})\delta\bar{v}_N(\mathbf{y})}{E_N + E} \\ & - \sum_N \frac{u_N(\mathbf{x})\bar{u}_N(\mathbf{y})\delta E_N}{(E_N - E)^2} + \sum_N \frac{v_N(\mathbf{x})\bar{v}_N(\mathbf{y})\delta E_N}{(E_N + E)^2}\,. \end{aligned} \tag{14.2.14}$$

(We are dropping the $i\epsilon$ terms because we are now not taking E to lie in the continuum of scattering states.) We see that the shift δE_N of an electron bound-state energy is given by the coefficient of $-u_N(\mathbf{x})\bar{u}_N(\mathbf{y})/(E_N - E)^2$ in the complete propagator. To calculate this, we note that Eq. (14.2.3) gives

$$S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) = \sum_N \frac{u_N(\mathbf{x})\bar{u}_N(\mathbf{y})}{E_N - E - i\epsilon} - \sum_N \frac{v_N(\mathbf{x})\bar{v}_N(\mathbf{y})}{E_N + E - i\epsilon}\,. \tag{14.2.15}$$

Inserting this in Eq. (14.2.12) gives

$$\begin{aligned} \delta S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E) = & \sum_{N,M} \frac{u_N(\mathbf{x})\bar{u}_M(\mathbf{y})}{(E_N - E)(E_M - E)} \\ & \times \int d^3z \int d^3w\, \bar{u}_N(\mathbf{z})\,\Sigma^*_{\mathscr{A}}(\mathbf{z},\mathbf{w};E)\,u_M(\mathbf{w}) + \cdots \end{aligned} \tag{14.2.16}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

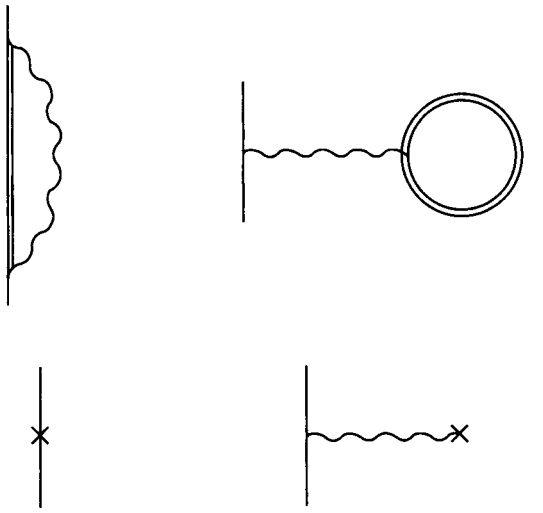

Figure 14.1. Lowest-order Feynman diagrams for the electron self-energy function $\Sigma^*_{\mathscr{A}}(x,y)$ in the presence of an external field. Here double straight lines represent electron propagators $S_{\mathscr{A}}$ that include effects of the external field; single straight lines are the incoming and outgoing electron lines; wavy lines are virtual photons; the cross represents renormalization counterterms.

where the dots denote additional terms involving at least one negative-energy pole. Comparing the coefficients of $(E_N - E)^{-2}$ here and in Eq. (14.2.14), we find

$$\delta E_N = -\int d^3x \int d^3y\, \bar{u}_N(\mathbf{x})\, \Sigma^*_{\mathscr{A}}(\mathbf{x},\mathbf{y};E_N)\, u_N(\mathbf{y})\,. \tag{14.2.17}$$

The u_N are solutions of the homogeneous Dirac equation satisfying the normalization condition (14.1.8), so this is very much like ordinary first-order perturbation theory, but with $-\Sigma^*$ in place of a perturbation to the Hamiltonian.

Generally δE_N turns out to be complex. This is simply a consequence of the instability of atomic energy levels to radiative decay to lower levels; we saw in Chapter 3 that an unstable state of energy E and decay rate Γ produces poles in various amplitudes at the complex energy $E - i\Gamma/2$. The imaginary part of Eq. (14.2.17) therefore equals $-\Gamma/2$, while its real part gives the energy shift.

The Feynman diagrams for Σ^* are shown in Figure 14.1. (Note that there are new photon tadpole diagrams here because the external field breaks the Lorentz invariance and charge-conjugation invariance that forbids such diagrams in the ordinary Feynman rules.) Application of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

position space Feynman rules to these diagrams gives

$$
\begin{aligned}
i\Sigma^*_{\mathscr{A}}(x,y) &= [e\gamma^\mu]\,[-iS_{\mathscr{A}}(x,y)]\,[e\gamma_\mu]\,[-iD(x-y)] \\
&- \left[e\gamma^\mu\delta^4(x-y)\right]\int d^4z\;\,[-iD(x-z)]\;\mathrm{Tr}\,\{[-iS_{\mathscr{A}}(z,z)]\,[e\gamma_\mu]\} \\
&-i(Z_2-1)(\gamma^\mu\partial_\mu+m)\delta^4(x-y)+i\delta m\,\delta^4(x-y) \\
&+e\gamma^\mu(Z_2-1)\delta^4(x-y)\mathscr{A}_\mu(x) \\
&+i(Z_3-1)[e\gamma_\mu]\delta^4(x-y)\int d^4z\;\,[-iD(x-z)]\,\partial_\nu(\partial^\nu\mathscr{A}^\mu(z)-\partial^\mu\mathscr{A}^\nu(z))
\end{aligned}
\tag{14.2.18}
$$

with the renormalization constants (Z_2-1), (Z_3-1), and δm calculated to second order in e. (The minus sign in the second term is the usual one that accompanies closed fermion loops.)

For strong external fields with $Z\alpha$ of order unity it is necessary to calculate the configuration space electron propagator $S_{\mathscr{A}}$ and the integrals in Eqs. (14.2.17) and (14.2.18) numerically.[4] However, for weak fields we can use the first few terms of the series (14.2.1) in Eq. (14.2.18) and calculate these integrals in closed form. For this purpose, it is more convenient to work in momentum space, defining:

$$
S_{\mathscr{A}}(x,y)=(2\pi)^{-4}\int d^4p'\,d^4p\;e^{ip'\cdot x}e^{-ip\cdot y}\;S_{\mathscr{A}}(p',p)\,, \tag{14.2.19}
$$

$$
\Sigma^*_{\mathscr{A}}(x,y)=(2\pi)^{-4}\int d^4p'\,d^4p\;e^{ip'\cdot x}e^{-ip\cdot y}\;\Sigma^*_{\mathscr{A}}(p',p)\,, \tag{14.2.20}
$$

$$
u_N(\mathbf{x})=(2\pi)^{-3/2}\int d^3p\;e^{i\mathbf{p}\cdot\mathbf{x}}u_N(\mathbf{p})\,, \tag{14.2.21}
$$

$$
\mathscr{A}^\mu(x)=\int d^4q\;e^{iq\cdot x}\mathscr{A}^\mu(q)\,. \tag{14.2.22}
$$

(We are here committing the impropriety of using the same symbol for a function and its Fourier transform, leaving it to the displayed arguments of the functions to indicate which is which.) Then Eqs. (14.2.1) and (14.2.18) become

$$
\begin{aligned}
S_{\mathscr{A}}(p',p) &= \frac{-i\not{p}+m}{p^2+m^2-i\epsilon}-ie\,\frac{-i\not{p}'+m}{p'^2+m^2-i\epsilon}\,\not{\mathscr{A}}(p'-p)\,\frac{-i\not{p}+m}{p^2+m^2-i\epsilon} \\
&\quad+\cdots\,,
\end{aligned}
\tag{14.2.23}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

and

$$\Sigma^*_{\mathscr{A}}(p',p) = \frac{ie^2}{(2\pi)^4}\int \frac{d^4k}{k^2-i\epsilon}\,\gamma^\mu S_{\mathscr{A}}(p'-k,p-k)\gamma_\mu$$
$$+[-(Z_2-1)(i\not{p}+m)+Z_2\delta m]\,\delta^4(p'-p) - ie(Z_2-1)\not{\mathscr{A}}(p'-p)$$
$$-\frac{ie^2\gamma^\mu}{(2\pi)^4}\frac{1}{(p-p')^2-i\epsilon}\int d^4q\ \mathrm{Tr}\{S_{\mathscr{A}}(q,q+p'-p)\gamma_\mu\}$$
$$+\frac{ie(Z_3-1)}{(p-p')^2-i\epsilon}\left[(p-p')^2\not{\mathscr{A}}(p'-p)-(\not{p}-\not{p}')(p-p')\cdot\mathscr{A}(p'-p)\right]. \tag{14.2.24}$$

Because the external field is time-independent, $S_{\mathscr{A}}(x,y)$ and $\Sigma^*_{\mathscr{A}}(x,y)$ can depend on x^0 and y^0 only through the difference x^0-y^0, so $S_{\mathscr{A}}(p',p)$ and $\Sigma^*_{\mathscr{A}}(p',p)$ as well as $\mathscr{A}^\mu(p'-p)$ must be proportional to $\delta(p'^0-p^0)$:

$$\mathscr{A}^\mu(p'-p) = \delta(p'^0-p^0)\mathscr{A}^\mu(\mathbf{p}'-\mathbf{p})\,, \tag{14.2.25}$$

$$S_{\mathscr{A}}(p',p) = \delta(p'^0-p^0)S_{\mathscr{A}}(\mathbf{p}',\mathbf{p};p^0)\,, \tag{14.2.26}$$

$$\Sigma^*_{\mathscr{A}}(p',p) = \delta(p'^0-p^0)\Sigma^*_{\mathscr{A}}(\mathbf{p}',\mathbf{p};p^0)\,. \tag{14.2.27}$$

The energy shift is then given by Eqs. (14.2.17) and (14.2.13) as

$$\delta E_N = -\int d^3p'\int d^3p\ \bar{u}_N(\mathbf{p}')\,\Sigma^*_{\mathscr{A}}(\mathbf{p}',\mathbf{p};E_N)\,u_N(\mathbf{p})\,, \tag{14.2.28}$$

with $\Sigma^*_{\mathscr{A}}(\mathbf{p}',\mathbf{p};E_N)$ given by Eqs. (14.2.23), (14.2.24), and (14.2.27). This is the master formula we shall use in the next section to calculate energy shifts in weak external fields.

14.3 The Lamb Shift in Light Atoms

Let us now consider radiative corrections to the energy levels of a non-relativistic electron in a general electrostatic field, such as an electron in the Coulomb field of a light nucleus with $Z\alpha \ll 1$. It is natural in this limit to treat the Coulomb field as a weak perturbation, but we will see that this would lead to an infrared divergence, related to that found in Section 11.3. The infrared divergence is really fictitious, because the four-momenta $\mathbf{p}, E_N$ and $\mathbf{p}', E_N$ are not on the electron mass shell, but it does force us to proceed with some care.

Usually this problem is dealt with by dividing the integral over virtual photon energies into a low-energy range, within which we can treat the electrons non-relativistically but must include effects to all orders in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

external field, and a high-energy range, within which we have to include relativistic effects but may include only effects of lowest order in the external field. Instead, we shall here introduce a fictitious photon mass μ, chosen to be much larger than typical electron kinetic energies, but much less than typical electron momenta. For a Coulomb field, this amounts to the requirement that

$$(Z\alpha)^2 m_e \ll \mu \ll Z\alpha m_e \,. \tag{14.3.1}$$

We write the photon propagator in the first two terms of Eq. (14.2.24) (including the formulas for the counterterms $Z_2 - 1$ and $Z_2\delta m$) in the form

$$\frac{1}{k^2 - i\epsilon} = \left[\frac{1}{k^2 + \mu^2 - i\epsilon}\right] + \left[\frac{1}{k^2 - i\epsilon} - \frac{1}{k^2 + \mu^2 - i\epsilon}\right] . \tag{14.3.2}$$

The energy shift is correspondingly a sum of two terms, a 'high-energy' and a 'low-energy' term: The high-energy term is calculated by using the first term in the photon propagator (14.3.2) in the first three terms in Eq. (14.2.24), and adding the result to the last two terms (the vacuum polarization terms) in Eq. (14.2.24) which are not infrared divergent anyway; the low-energy term is calculated using the second term in Eq. (14.3.2) in the first three terms of Eq. (14.2.24). One advantage of this procedure is that we shall be able to use the results of the relativistic calculations of Sections 11.3 and 11.4 directly, without the rather tricky conversion from a photon mass to an infrared energy cutoff. Of course, in the end we shall have to check that the dependence on the photon mass μ in the high-energy and low-energy contributions to the energy shift cancel, leaving the total energy shift μ-independent.

A High-Energy Term

Because μ is taken much larger than the atomic binding energies we can here keep only terms of lowest order in the external field. The one-loop radiative correction to atomic energy levels in a general time-independent external vector potential $\mathscr{A}^\mu(\mathbf{x})$ is given in momentum space by Eq. (14.2.28), with the self-energy insertion $\Sigma_{\mathscr{A}}(p', p)$ given by Eqs. (14.2.24) and (14.2.23). The terms of zeroth order in the external field simply cancel: the δm term cancels the first term with $\mathscr{A} = 0$; the $Z_3 - 1$ term cancels the third term with $\mathscr{A} = 0$; and the $Z_2 - 1$ terms vanish because $u(p)$ satisfies the Dirac equation. The term in $\Sigma_{\mathscr{A}}(p', p)$ that is of first order in $\mathscr{A}^\mu$ may be put in the form

$$\Sigma_{\mathscr{A}1}(p', p) = -ie\mathscr{A}_\mu(p' - p)\,\Gamma_1^\mu(p', p) \tag{14.3.3}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

with

$$\Gamma_1^\mu(p',p) = \frac{ie^2}{(2\pi)^4}\int \frac{d^4k}{k^2+\mu^2-i\epsilon} \times \gamma^\nu\left[\frac{-i(\not p'-\not k)+m_e}{(p'-k)^2+m_e^2-i\epsilon}\right]\gamma^\mu\left[\frac{-i(\not p-\not k)+m_e}{(p-k)^2+m_e^2-i\epsilon}\right]\gamma_\nu + (Z_2-1)\gamma^\mu - \frac{ie^2\gamma_\nu}{(2\pi)^4}\frac{1}{(p-p')^2-i\epsilon} \times \int d^4l\ \mathrm{Tr}\left\{\left[\frac{-i\not l+m_e}{l^2+m_e^2}\right]\gamma^\mu\left[\frac{-i\not l-i\not p'+i\not p+m_e}{(l+p'-p)^2+m_e^2}\right]\gamma^\nu\right\} - \frac{Z_3-1}{(p'-p)^2-i\epsilon}\left[(p-p')^2\eta^{\mu\nu}-(p'-p)^\mu(p'-p)^\nu\right]\gamma_\nu. \quad (14.3.4)$$

Comparison of the first two terms with Eqs. (11.3.1) and (11.3.8) and the second two terms with Eqs. (11.3.9), (11.2.3), and (11.2.15) reveals that $\Gamma_1^\mu(p',p)$ is the complete one-loop vertex function, including vacuum polarization and all counterterms, whose mass-shell matrix elements we have already calculated in Section 11.3. Using Eqs. (14.2.26) and (14.2.25), this contribution to the energy shift (14.2.28) is given by

$$[\delta E_N]_{\text{high energy}} = ie\int d^3p'\int d^3p\ (\bar u_N(\mathbf{p}')\,\Gamma_1^\mu(\mathbf{p}',E_N,\mathbf{p},E_N)\,u_N(\mathbf{p})) \times \mathscr{A}_\mu(\mathbf{p}'-\mathbf{p})\,. \quad (14.3.5)$$

(This could have been guessed at, by simply replacing the γ^μ in the interaction of the electron with the external field with Γ_1^μ.) As discussed in Section 14.1, because $Z\alpha \ll 1$ we can approximate the Dirac wave function u_N in Eq. (14.3.5) as

$$[u_N(\mathbf{p})]_\alpha = \sum_\sigma u_\alpha(\mathbf{p},\sigma)\,[f_N(\mathbf{p})]_\sigma\,, \quad (14.3.6)$$

where f_N is the non-relativistic two-component wave function of an electron in the external Coulomb field, and $u(\mathbf{p},\sigma)$ is the four-component normalized solution of the momentum space Dirac equation

$$[i\gamma_\mu p^\mu + m_e]\,u(\mathbf{p},\sigma) = 0 \quad (14.3.7)$$

for spin z-component σ. Since $u_N(\mathbf{p})$ approximately satisfies the free-particle Dirac equation, Eq. (10.6.15) gives the general form of the matrix element of Γ_1^μ as

$$\bar u_M(\mathbf{p}')[\gamma^\mu+\Gamma_1^\mu(p',p)]u_N(\mathbf{p}) = \bar u_M(\mathbf{p}')\left[\gamma^\mu F_1(q^2)+\tfrac{1}{2}i\,[\gamma^\mu,\gamma^\nu]\,q_\nu F_2(q^2)\right]u_N(\mathbf{p})\,, \quad (14.3.8)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

where $q \equiv p' - p$. The wave functions $u_N(\mathbf{p})$ fall off very rapidly for $|\mathbf{p}| \gg Z\alpha m_e$, so we only need $F_1(\mathbf{q}^2)$ and $F_2(\mathbf{q}^2)$ in the limit $|q^2| \ll m_e^2$. In this limit, Eqs. (11.3.31), (10.6.18), and (11.3.16) give

$$F_1(q^2) \simeq 1 + \frac{e^2}{24\pi^2}\left(\frac{q^2}{m_e^2}\right)\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5} + \frac{3}{4}\right], \tag{14.3.9}$$

$$F_2(q^2) \simeq \frac{e^2}{16 m_e \pi^2}. \tag{14.3.10}$$

Let us first consider the contribution of the F_1 term in Eq. (14.3.8), which makes by far the largest contribution to the energy shift, and raises the most interesting problems in its calculation. For a pure electrostatic field with $\mathscr{A} = 0$, Eqs. (14.3.5), (14.3.8), and (14.3.9) give

$$[\delta E_N]_{F_1} = -\frac{e^2}{24\pi^2 m_e^2}\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5} + \frac{3}{4}\right] \times \int d^3p' \int d^3p\, \bar{u}_N(\mathbf{p}')\Big(-ie\mathscr{A}^0(\mathbf{p}' - \mathbf{p})\Big)\gamma^0(\mathbf{p}' - \mathbf{p})^2 u_N(\mathbf{p}). \tag{14.3.11}$$

To calculate this contribution, we may use the leading term in the non-relativistic matrix element (14.1.41), and find

$$[\delta E_N]_{F_1} = -\frac{e^2}{24\pi^2 m_e^2}\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5} + \frac{3}{4}\right] \times \int d^3p' \int d^3p\, f_N^\dagger(\mathbf{p}')\, e\mathscr{A}^0(\mathbf{p}' - \mathbf{p})[\mathbf{p}' - \mathbf{p}]^2 f_N(\mathbf{p}) \tag{14.3.12}$$

or in position space

$$[\delta E_N]_{F_1} = \frac{e^2}{24\pi^2 m_e^2}\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5} + \frac{3}{4}\right] \int d^3x\, f_N^\dagger(\mathbf{x})\,[e\nabla^2\mathscr{A}^0(\mathbf{x})] f_N(\mathbf{x}). \tag{14.3.13}$$

In particular, for the Coulomb potential (14.1.2), we have $e\nabla^2\mathscr{A}^0(\mathbf{x}) = -Ze^2\delta^3(\mathbf{x})$, and the label N consists of a principal quantum number n and angular-momentum quantum numbers j, m, ℓ, while (11.2.41) gives $[f_{njm\ell}(0)]_\sigma = 2(Z\alpha m_e/n)^{3/2}\delta_{\ell,0}\delta_{\sigma,m}/\sqrt{4\pi}$. The energy shift (14.3.11) is then

$$[\delta E_{jn\ell}]_{F_1} = -\frac{2Z^4\alpha^5 m_e}{3\pi n^3}\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5} + \frac{3}{4}\right]\delta_{\ell,0}. \tag{14.3.14}$$

(The lack of dependence of δE on the total angular momentum z-component m is guaranteed by rotational invariance.) The term $\frac{2}{5}$ in the brackets in Eqs. (14.3.12) and (14.3.13) arises from vacuum polarization, and yields just the energy shift calculated somewhat heuristically in Section 11.2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Before going on to calculate the magnetic and low-energy contributions to the energy shift, it is worth noting that the result we have obtained so far yields a fair order-of- magnitude estimate of the Lamb shift without further work. We can anticipate that the low-energy terms will contain a term proportional to $\ln(\mu/B)$, with a coefficient such as to cancel the μ-dependence in Eq. (14.3.12). The constant B here is an energy which must be included to make the argument of the logarithm dimensionless; since it is the binding of the electron in the atom that will eventually provide our infrared cutoff, we may guess that B is a typical atomic binding energy, of order $B \simeq (Z\alpha)^2 m_e$. The total energy shift in a state N with principal quantum number n and orbital angular momentum ℓ is therefore of the form

$$\delta E_N = -\frac{2Z^4\alpha^5 m_e}{3\pi n^3}\left[\ln\left(Z^4\alpha^4\right)\delta_{\ell,0} + O(1)\right] . \tag{14.3.15}$$

For the $2s$ state of hydrogen, the logarithmic term alone gives

$$\delta E_{2s} \simeq -\frac{\alpha^5 m_e}{12\pi}\ln\left(\alpha^4\right) = 5.5 \times 10^{-6}\ \text{eV} = 1300\,\text{MHz} \times 2\pi\hbar .$$

As we shall see, the 'O(1)' terms in Eq. (14.3.15) will lower the total energy shift by about 25%.

Now let us consider the contribution of the F_2 term in the matrix element of Γ_1^μ, which as we saw in Section 10.6 may be interpreted as a radiative correction to the magnetic moment of the electron. Using Eqs. (14.3.10), (14.3.8), and (14.3.6) in Eq. (14.3.5), we find that this term gives an energy shift

$$[\delta E_N]_{F_2} = -\frac{e^2}{32\pi^2 m_e}\int d^3p' \int d^3p\ \left(\bar{u}_N(\mathbf{p}')\,[\gamma^\mu,\gamma^\nu]u_N(\mathbf{p})\right) \times e\mathscr{A}_\mu(\mathbf{p}'-\mathbf{p})\,(p'-p)_\nu \tag{14.3.16}$$

or in position space

$$[\delta E_N]_{F_2} = \frac{ie^2}{64\pi^2 m_e}\int d^3x\ \left(\bar{u}_N(\mathbf{x})\,[\gamma^\mu,\gamma^\nu]u_N(\mathbf{x})\right) e\mathscr{F}_{\mu\nu}(\mathbf{x}) , \tag{14.3.17}$$

where

$$\mathscr{F}_{\mu\nu}(\mathbf{x}) \equiv \partial_\mu \mathscr{A}_\nu(\mathbf{x}) - \partial_\nu \mathscr{A}_\mu(\mathbf{x}) . \tag{14.3.18}$$

For a pure electrostatic field with $\mathscr{A} = 0$, this is

$$[\delta E_N]_{F_2} = \frac{-ie^2}{32\pi^2 m_e}\int d^3x\ \left(\bar{u}_N(\mathbf{x})\,[\boldsymbol{\gamma},\gamma^0]u_N(\mathbf{x})\right)\cdot\nabla[e\mathscr{A}^0(\mathbf{x})] . \tag{14.3.19}$$

In the non-relativistic limit $Z\alpha \ll 1$, we can use the approximate result

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

(14.1.43), which here reads

$$(\bar{u}_N[\gamma^0,\gamma]u_N) \simeq \frac{i}{m_e}[(\nabla f_N^\dagger \cdot \boldsymbol{\sigma}\,\boldsymbol{\sigma} f_N) + (f_N^\dagger \boldsymbol{\sigma}\,\boldsymbol{\sigma} \cdot \nabla f_N)]$$
$$= \frac{i}{m_e}[\nabla(f_N^\dagger f_N) - i(\nabla f_N^\dagger \times \boldsymbol{\sigma}) f_N - i f_N^\dagger(\boldsymbol{\sigma} \times \nabla f_N)]\,. \quad (14.3.20)$$

Using this in Eq. (14.3.8) and integrating by parts, this part of the energy shift is

$$[\delta E_N]_{F_2} = \frac{e^2}{32\pi^2 m_e^2}\int d^3x \Big[-|f_N(\mathbf{x})|^2 \nabla^2(e\mathscr{A}^0(\mathbf{x}))$$
$$+2i f_N^\dagger(\mathbf{x})\,\boldsymbol{\sigma}\cdot(\nabla(e\mathscr{A}^0(\mathbf{x})) \times \nabla f_N(\mathbf{x}))\Big]\,. \quad (14.3.21)$$

Combining Eqs. (14.3.12) and (14.3.21) gives the total high-energy contribution to the energy shift in an arbitrary electrostatic potential $\mathscr{A}^0$:

$$[\delta E_N]_{\text{high energy}} =$$
$$\frac{e^2}{24\pi^2 m_e^2}\left[\ln\left(\frac{\mu^2}{m_e^2}\right) + \frac{2}{5}\right]\int d^3x\, f_N^\dagger(\mathbf{x})\,[e\nabla^2\mathscr{A}^0(\mathbf{x})] f_N(\mathbf{x})$$
$$+\frac{ie^2}{16\pi^2 m_e^2}\int d^3x\; f_N^\dagger(\mathbf{x})\,\boldsymbol{\sigma}\cdot(\nabla(e\mathscr{A}^0(\mathbf{x})) \times \nabla f_N(\mathbf{x}))\,. \quad (14.3.22)$$

B Low-Energy Term

The low-energy contribution to the energy shift is obtained from the first three terms in Eq. (14.2.24), making the replacement in the photon propagator

$$\frac{1}{k^2 - i\epsilon} \;\to\; \frac{1}{k^2 - i\epsilon} - \frac{1}{k^2 + \mu^2 - i\epsilon}\,. \quad (14.3.23)$$

This substitution will eventually serve to cut off the integral over components of the photon four-momentum k at values of order μ, but it is not possible to see this until we carefully take mass renormalization into account, so we shall defer making any non-relativistic approximations until then. Also, we are now including photon momenta as small or smaller than the binding energies of the atomic states, so we must treat the electrostatic forces responsible for this binding to all orders.

Instead of working with the momentum space formula (14.2.24), it will be convenient to return to the configuration space formula (14.2.18). This gives the low- energy contribution to the electron self-energy function as

$$[\Sigma^*_{\mathscr{A}}(x,y)]_{\text{low energy}} = ie^2\gamma^\rho\, S_{\mathscr{A}}(x,y)\,\gamma_\rho\, D(x-y;\mu) + \delta m_e(\mu)\,\delta^4(x-y)$$
$$-(Z_2(\mu)-1)(\gamma^\mu[\partial_\mu + ie\mathscr{A}_\mu] + m_e)\delta^4(x-y)\,, \quad (14.3.24)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

where $D(x-y;\mu)$ is the modified photon propagator

$$D(x-y;\mu)=\frac{1}{(2\pi)^4}\int d^4k\, e^{ik\cdot(x-y)}\left[\frac{1}{k^2-i\epsilon}-\frac{1}{k^2+\mu^2-i\epsilon}\right], \quad (14.3.25)$$

and the counterterms $Z_2(\mu)-1$ and $\delta m(\mu)$ are calculated using this modified propagator. Converting from time to energy variables, the low-energy contribution to the function (14.2.13) is then

$$\begin{aligned}[\Sigma^*_{\mathscr{A}}(\mathbf{x},\mathbf{y};E)]_{\text{low energy}} &= \frac{ie^2}{(2\pi)^4}\int d^4k\,\gamma^\rho\, S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E-k^0)\,\gamma_\rho \\ &\quad\times\left[\frac{1}{k^2-i\epsilon}-\frac{1}{k^2+\mu^2-i\epsilon}\right]e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \\ &\quad-(Z_2(\mu)-1)\left(\boldsymbol{\gamma}\cdot\nabla+i\gamma^0E+ie\gamma^\nu\mathscr{A}_\nu+m_e\right)\delta^3(\mathbf{x}-\mathbf{y}) \\ &\quad+\delta m_e(\mu)\,\delta^3(\mathbf{x}-\mathbf{y})\,. \end{aligned} \quad (14.3.26)$$

The low-energy contribution to the energy shift is then given by Eq. (14.2.17) as

$$\begin{aligned}[\delta E_N]_{\text{low energy}} &= -\int d^3x\int d^3y\,\bar u_N(\mathbf{x})\,[\Sigma^*_{\mathscr{A}}(\mathbf{x},\mathbf{y};E_N)]_{\text{low energy}}\,u_N(\mathbf{y}) \\ &= \frac{-ie^2}{(2\pi)^4}\int d^4k\int d^3x\int d^3y\,\bar u_N(\mathbf{x})\,\gamma^\rho\,S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E_N-k^0)\,\gamma_\rho\,u_N(\mathbf{y}) \\ &\quad\times\left[\frac{1}{k^2-i\epsilon}-\frac{1}{k^2+\mu^2-i\epsilon}\right]e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \\ &\quad-\delta m_e(\mu)\int d^3x\,\bar u_N(\mathbf{x})u_N(\mathbf{x})\quad. \end{aligned} \quad (14.3.27)$$

Note that the terms proportional to $Z_2(\mu)-1$ have dropped out because the Dirac wave function $u_N(\mathbf{x})$ satisfies the Dirac equation Eq. (14.1.10). For the electron propagator in the presence of the Coulomb field, we use Eq. (14.2.15):

$$S_{\mathscr{A}}(\mathbf{x},\mathbf{y};E)=\sum_M\frac{u_M(\mathbf{x})\bar u_M(\mathbf{y})}{E_M-E-i\epsilon}-\sum_M\frac{v_M(\mathbf{x})\bar v_M(\mathbf{y})}{E_M+E-i\epsilon},$$

with the sums in the first and second terms running over all one-electron and one-positron states, respectively. The k^0 integrals can be done most easily by closing the contour of integration with a large semi-circle in the lower half-plane in the first term and in the upper half-plane in the second term:

$$\begin{aligned}\int dk^0\left(\frac{1}{k^2+\mu^2-i\epsilon}\right)\left(\frac{1}{E_M\mp E_N\pm k^0-i\epsilon}\right) &= \\ =\frac{i\pi}{\sqrt{\mathbf{k}^2+\mu^2}}&\left(\frac{1}{E_M\mp E_N+\sqrt{\mathbf{k}^2+\mu^2}-i\epsilon}\right)\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

and likewise if μ is replaced with zero. The energy shift (14.3.27) now becomes

$$[\delta E_N]_{\text{low energy}} = -\frac{e^2}{2(2\pi)^3}\int d^3k \sum_M \times \left[\Gamma^\rho_{MN}(\mathbf{k})^* \Gamma_{\rho MN}(\mathbf{k})\left(\frac{1}{|\mathbf{k}|(E_M - E_N + |\mathbf{k}| - i\epsilon)} - \frac{1}{\sqrt{\mathbf{k}^2+\mu^2}(E_M - E_N + \sqrt{\mathbf{k}^2+\mu^2} - i\epsilon)}\right) - \tilde{\Gamma}^\rho_{MN}(\mathbf{k})^* \tilde{\Gamma}_{\rho MN}(\mathbf{k})\left(\frac{1}{|\mathbf{k}|(E_M + E_N + |\mathbf{k}| - i\epsilon)} - \frac{1}{\sqrt{\mathbf{k}^2+\mu^2}(E_M + E_N + \sqrt{\mathbf{k}^2+\mu^2} - i\epsilon)}\right)\right] - \delta m_e(\mu)\int d^3x\, \bar{u}_N(\mathbf{x})u_N(\mathbf{x})\,, \tag{14.3.28}$$

where

$$\Gamma^\rho_{MN}(\mathbf{k}) \equiv \int d^3y\, e^{-i\mathbf{k}\cdot\mathbf{y}}\, \bar{u}_M(\mathbf{y})\gamma^\rho u_N(\mathbf{y})\,, \tag{14.3.29}$$

$$\tilde{\Gamma}^\rho_{MN}(\mathbf{k}) \equiv \int d^3y\, e^{-i\mathbf{k}\cdot\mathbf{y}}\, \bar{v}_M(\mathbf{y})\gamma^\rho u_N(\mathbf{y})\,. \tag{14.3.30}$$

(Of course, the 'sum' over M in Eq. (14.3.28) receives contributions only from electron states in the first term and positron states in the second term.) Eq. (14.3.28) could have been derived more directly from old-fashioned perturbation theory; the energy denominators $E_M - E_N + \omega$ and $E_M + E_N + \omega$ are the result of subtracting the energy E_N of the initial state from the energy of an intermediate state consisting of either an electron of energy E_M and a photon of energy ω, or else a positron of energy E_M, a photon of energy ω, and both the final and the initial electron. (See Figure 14.2.)

Before making any approximations to Eq. (14.3.28), it will be convenient to express the time components of the matrix elements Γ^ρ_{MN} and $\tilde{\Gamma}^\rho_{MN}$ in terms of the corresponding space components, using relations* derived

* To derive Eq. (14.3.31), note that

$$k_i\Gamma^i_{MN}(\mathbf{k}) = -i\int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, \nabla\cdot(\bar{u}_M(\mathbf{x})\gamma u_N(\mathbf{x}))$$

$$= i\int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, \partial_0\left[\left(\bar{u}_M(\mathbf{x})\gamma^0 u_N(\mathbf{x})\right)\, e^{-i(E_N - E_M)x^0}\right]_{x^0=0} = (E_N - E_M)\Gamma^0_{MN}(\mathbf{k})\,.$$

Eq. (14.3.32) is derived in the same way.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

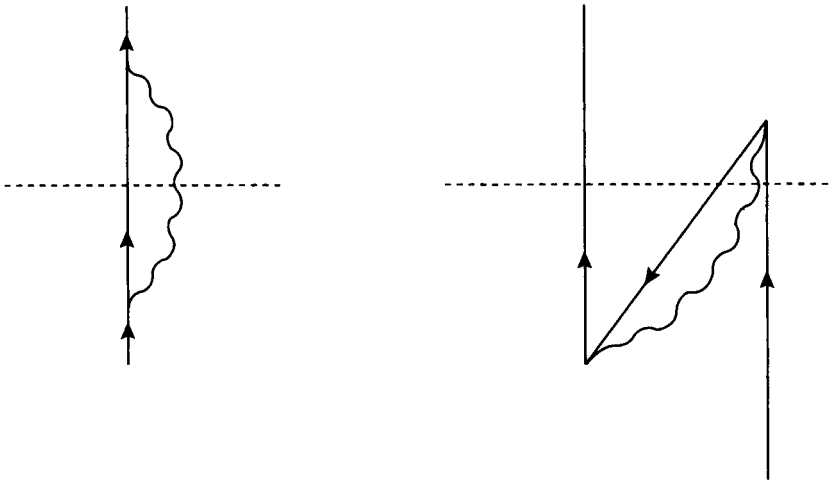

Figure 14.2. Old-fashioned perturbation theory diagrams for the low-energy part of the electron energy shift. Here solid lines are electrons; wavy lines are photons; the dashed line cuts through particle lines corresponding to the immediate states appearing in the first two terms in (14.3.28)

from the conservation of the electric current:

$$k_i \Gamma^i_{MN}(\mathbf{k}) = (E_N - E_M)\Gamma^0_{MN}(\mathbf{k}) , \tag{14.3.31}$$

$$k_i \tilde{\Gamma}^i_{MN}(\mathbf{k}) = (E_N + E_M)\tilde{\Gamma}^0_{MN}(\mathbf{k}) . \tag{14.3.32}$$

Furthermore, by using the completeness relation (14.1.8) it is straightforward to show that:

$$\sum_M \left[|\Gamma^0_{MN}(\mathbf{k})|^2 + |\tilde{\Gamma}^0_{MN}(\mathbf{k})|^2 \right] = 1 \tag{14.3.33}$$

and

$$\begin{aligned} &\sum_M \left[|\Gamma^0_{MN}(\mathbf{k})|^2 (E_M - E_N) - |\tilde{\Gamma}^0_{MN}(\mathbf{k})|^2 (E_M + E_N) \right] \\ &= \sum_M \left[-\Gamma^{0*}_{MN}(\mathbf{k})\, \mathbf{k} \cdot \boldsymbol{\Gamma}_{MN}(\mathbf{k}) - \tilde{\Gamma}^{0*}_{MN}(\mathbf{k})\, \mathbf{k} \cdot \tilde{\boldsymbol{\Gamma}}_{MN}(\mathbf{k}) \right] \\ &= -i\mathbf{k} \cdot \int d^3x\, \bar{u}_N(\mathbf{x}) \boldsymbol{\gamma} u_N(\mathbf{x}) = 0 , \end{aligned} \tag{14.3.34}$$

the last step following from the parity condition (14.1.23). In this way,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Eq. (14.3.28) can be rewritten

$$
\begin{aligned}
&[\delta E_N]_{\text{low energy}} = \\
&-\frac{e^2}{2(2\pi)^3}\int d^3k \sum_M \left[\frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/\mathbf{k}^2\right)}{|\mathbf{k}|\,(E_M - E_N + |\mathbf{k}| - i\epsilon)} \right. \\
&\qquad \left. - \frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/(\mathbf{k}^2+\mu^2)\right)}{\sqrt{\mathbf{k}^2+\mu^2}\,(E_M - E_N + \sqrt{\mathbf{k}^2+\mu^2} - i\epsilon)} \right] \\
&+\frac{e^2}{2(2\pi)^3}\int d^3k \sum_M \left[\frac{\left(|\tilde{\mathbf{\Gamma}}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\tilde{\mathbf{\Gamma}}_{MN}(\mathbf{k})|^2/\mathbf{k}^2\right)}{|\mathbf{k}|\,(E_M + E_N + |\mathbf{k}|)} \right. \\
&\qquad \left. - \frac{\left(|\tilde{\mathbf{\Gamma}}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\tilde{\mathbf{\Gamma}}_{MN}(\mathbf{k})|^2/(\mathbf{k}^2+\mu^2)\right)}{\sqrt{\mathbf{k}^2+\mu^2}\,(E_M + E_N + \sqrt{\mathbf{k}^2+\mu^2})} \right] \\
&-\frac{e^2}{(2\pi)^3}\int d^3k \sum_M |\tilde{\Gamma}^0_{MN}(\mathbf{k})|^2 \left(\frac{1}{\mathbf{k}^2} - \frac{1}{\mathbf{k}^2+\mu^2}\right) \\
&+ \tfrac{1}{2}\alpha\mu - \delta m_e(\mu)\int d^3x\, \bar{u}_N(\mathbf{x}) u_N(\mathbf{x})\,, \qquad (14.3.35)
\end{aligned}
$$

In the next-to-last term, we have used the elementary integral

$$\int d^3k \left(\frac{1}{\mathbf{k}^2} - \frac{1}{\mathbf{k}^2+\mu^2}\right) = 2\mu\pi^2\,.$$

So far, this has been an exact rewriting of Eq. (14.3.28). We must now invoke several approximations. First, consider mass renormalization. We have already calculated $\delta m_e(\mu)$ to order α in Section 11.4; it is

$$\delta m_e(\mu) = \frac{2m_e\pi^2e^2}{(2\pi)^4}\int_0^1 dx\,[1+x]\,\ln\left(\frac{m_e^2x^2 + \mu^2(1-x)}{m_e^2x^2}\right)\,. \qquad (14.3.36)$$

Although in Section 11.4 we regarded μ as a regulator mass, to be taken much larger than m_e, we can just as well use Eq. (14.3.36) to provide a value for $\delta m_e(\mu)$ in the case that interests us here, $\mu \ll m_e$. In this limit Eq. (14.3.36) gives

$$\delta m_e(\mu) \to \frac{\alpha\mu}{2}\left[1 - \frac{3\mu}{2\pi m_e} + \ldots\right]\,. \qquad (14.3.37)$$

We also recall that for $Z\alpha \ll 1$, $u_N(\mathbf{x})$ is given by Eq. (14.1.37) as

$$u_N(\mathbf{x}) = \frac{1}{\sqrt{2}}\begin{bmatrix} (1 - \boldsymbol{\sigma}\cdot\mathbf{v}/2 + \cdots)f_N(\mathbf{x}) \\ (1 + \boldsymbol{\sigma}\cdot\mathbf{v}/2 + \cdots)f_N(\mathbf{x}) \end{bmatrix}\,, \qquad (14.3.38)$$

where dots indicate terms of higher order in $Z\alpha$; $\mathbf{v}$ is the non-relativistic velocity operator $-i\nabla/m_e$; and $f_N(\mathbf{x})$ is a two-component spinor solution

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

of the non-relativistic Schrödinger equation, normalized according to Eq. (14.1.38) so that

$$\int |f_N(\mathbf{x})|^2 = 1 - \tfrac{1}{4}(\mathbf{v}^2)_{NN} + \cdots \quad . \tag{14.3.39}$$

This gives the coefficient of $\delta m_e(\mu)$ in Eq. (14.3.35) as

$$\int d^3x\, \bar{u}_N(\mathbf{x}) u_N(\mathbf{x}) = 1 - \tfrac{1}{2}(\mathbf{v}^2)_{NN} + \cdots \quad . \tag{14.3.40}$$

We note immediately that the leading term in $-\delta m_e(\mu) \int d^3x\, \bar{u}_N u_N$ cancels the term $\alpha\mu/2$ in Eq. (14.3.35). Indeed, we could have anticipated this cancellation, because the term $\alpha\mu/2$ in Eq. (14.3.35) survives in the limit $Z\alpha \to 0$, and the definition of $m_e(\mu)$ as the renormalized electron mass implies that there must be no energy shift in this limit. By the same argument we can anticipate that the term of order $\alpha\mu^2/m_e$ in $\delta m_e(\mu)$ (which is larger than of order $\alpha(Z\alpha)^4 m_e$, and therefore cannot simply be neglected) cancels the second and third terms in Eq. (14.3.35), which are also of this order.** On the other hand, the product of the $\alpha\mu^2/m_e$

** The cancellation can be shown as follows. We anticipate that the second and third terms in Eq. (14.3.35) are small enough so they can be evaluated using the extreme nonrelativistic approximation $\beta u_N(\mathbf{x}) = u_N(\mathbf{x})$ for the Dirac equation satisfied by $u_N(\mathbf{x})$. On the other hand, although the Coulomb force may be neglected in the positron wave functions $v_M(\mathbf{x})$, the sum over M in the third term receives important contributions from relativistic positrons, so we use the approximation $v_{\mathbf{p},\sigma}(\mathbf{x}) \simeq (2\pi)^{-3/2} v(\mathbf{p},\sigma) e^{i\mathbf{p}\cdot\mathbf{x}}$, where $v(\mathbf{p},\sigma)$ is the positron spinor introduced in Section 5.5, normalized so that $\bar{v}(\mathbf{p},\sigma')v(\mathbf{p},\sigma) = \delta_{\sigma',\sigma}$. Thus the sums over M in the second and third terms of Eq. (14.3.35) are approximately given by

$$\tfrac{1}{2}\left[\sum_M \tilde{\Gamma}^{i*}_{MN}(\mathbf{k})\tilde{\Gamma}^{j}_{MN}(\mathbf{k}) + (i \leftrightarrow j)\right] \simeq \delta_{ij}\left(\frac{\sqrt{\mathbf{k}^2+m_e^2}+m_e}{2\sqrt{\mathbf{k}^2+m_e^2}}\right),$$

$$\sum_M |\tilde{\Gamma}^{0}_{MN}(\mathbf{k})|^2 \simeq \left(\frac{\sqrt{\mathbf{k}^2+m_e^2}-m_e}{2\sqrt{\mathbf{k}^2+m_e^2}}\right).$$

To leading order in μ/m_e, the second and third terms in Eq. (14.3.35) are then, respectively,

$$\frac{e^2}{4m_e(2\pi)^3}\int d^3k \left[\frac{2}{k} - \frac{(3-k^2/(k^2+\mu^2))}{\sqrt{k^2+\mu^2}}\right]\left(\frac{\sqrt{\mathbf{k}^2+m_e^2}+m_e}{2\sqrt{\mathbf{k}^2+m_e^2}}\right) \simeq \frac{\alpha\mu^2}{4\pi m_e}$$

and

$$-\frac{e^2}{2(2\pi)^3}\int d^3k \left(\frac{1}{\mathbf{k}^2} - \frac{1}{\mathbf{k}^2+\mu^2}\right)\left(\frac{\sqrt{\mathbf{k}^2+m_e^2}-m_e}{2\sqrt{\mathbf{k}^2+m_e^2}}\right) \simeq -\frac{\alpha\mu^2}{\pi m_e}.$$

(Eq. (14.3.32) rules out the possibility that a relativistic correction to the latter expression might not be suppressed by the factor $\mathbf{k}^2/m_e^2$ that appears in this expression for $|\mathbf{k}|^2 \ll m_e^2$.) These two terms are cancelled by the term $+3\alpha\mu^2/4\pi m_e$ in $-\delta m_e(\mu)\int d^3x \bar{u}_N(\mathbf{x}) u_N(\mathbf{x})$. Finally, we note that relativistic corrections to the above estimates for the sums over positron states would involve additional factors of $v^2/c^2 \approx (Z\alpha)^2$, yielding contributions of order $\alpha(Z\alpha)^2\mu^2/m_e \ll \alpha(Z\alpha)^4 m_e$, which justifies the non-relativistic approximations used here.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

term in δm_e with the second term in the matrix element (14.3.40) is of order $(Z\alpha)^2\alpha\mu^2/m_e \ll \alpha(Z\alpha)^4 m_e$, and may therefore be neglected. To order $\alpha(Z\alpha)^4 m_e$, the only remaining effect of mass renormalization is to leave us with the product of the leading term in $\delta m_e(\mu)$ with the term of order $(Z\alpha)^2$ in $\int d^3x\, \bar{u}_N u_N$:

$$-\left[\frac{e^2\mu}{8\pi}\right]\left[-\frac{1}{2}(\mathbf{v}^2)_{NN}\right] = \frac{e^2\mu}{16\pi}(\mathbf{v}^2)_{NN} \,.$$

(This is the effect of mass renormalization on the electron kinetic energy, mentioned in Section 1.3.) It turns out that this is just the negative of what the first term in Eq. (14.3.35) would be if we neglected the difference between energy levels. To see this, note that the integral in this term is effectively cut off at $|\mathbf{k}| \sim \mu \ll Z\alpha m_e$, so we can evaluate the matrix element $\mathbf{\Gamma}_{MN}(\mathbf{k})$ in the limit $\mathbf{k} \to 0$. To lowest order in $Z\alpha$, Eq. (14.1.42) gives

$$\mathbf{\Gamma}_{MN}(0) = (\mathbf{v})_{MN} \tag{14.3.41}$$

and using the completeness of the solutions f_N of the non-relativistic Schrödinger equation, we then have

$$\sum_M \Gamma^{i*}_{MN}(\mathbf{k})\Gamma^j_{MN}(\mathbf{k}) \simeq (v^i v^j)_{NN} \,, \tag{14.3.42}$$

so to this order

$$\begin{aligned}
&-\frac{e^2}{2(2\pi)^3}\int d^3k \sum_M \left[\frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/\mathbf{k}^2\right)}{\mathbf{k}^2}\right. \\
&\qquad \left. - \frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/(\mathbf{k}^2+\mu^2)\right)}{\mathbf{k}^2+\mu^2}\right] \\
&\simeq -\frac{e^2}{2(2\pi)^3}(\mathbf{v}^2)_{NN}\int d^3k \left[\frac{2}{3\mathbf{k}^2} - \frac{\left(1-\mathbf{k}^2/3(\mathbf{k}^2+\mu^2)\right)}{\mathbf{k}^2+\mu^2}\right] \\
&= -\frac{e^2\mu}{16\pi}(\mathbf{v}^2)_{NN} \qquad .
\end{aligned}$$

Thus after mass renormalization we are left with just the first term in Eq. (14.3.35), less the same with energy differences $E_N - E_M$ dropped:

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

$$[\delta E_N]_{\text{low energy}} = \frac{e^2}{2(2\pi)^3}\int d^3k \sum_M (E_M - E_N) \times \left[\frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/\mathbf{k}^2\right)}{\mathbf{k}^2(E_M - E_N + |\mathbf{k}| - i\epsilon)} - \frac{\left(|\mathbf{\Gamma}_{MN}(\mathbf{k})|^2 - |\mathbf{k}\cdot\mathbf{\Gamma}_{MN}(\mathbf{k})|^2/(\mathbf{k}^2+\mu^2)\right)}{(\mathbf{k}^2+\mu^2)(E_M - E_N + \sqrt{\mathbf{k}^2+\mu^2} - i\epsilon)}\right] . \quad (14.3.43)$$

Again using Eq. (14.3.41), this is

$$[\delta E_N]_{\text{low energy}} = \frac{e^2}{2(2\pi)^3}\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 \times \int d^3k \left[\frac{2}{3\mathbf{k}^2(E_M - E_N + |\mathbf{k}| - i\epsilon)} - \frac{1 - \mathbf{k}^2/3(\mathbf{k}^2+\mu^2)}{(\mathbf{k}^2+\mu^2)(E_M - E_N + \sqrt{\mathbf{k}^2+\mu^2} - i\epsilon)}\right] . \quad (14.3.44)$$

Even though typical values of the electron momentum are much larger than typical atomic energy differences, this is not true of typical values of $|\mathbf{k}|$ in this integral, because the integral would be infrared divergent if we did not keep the $E_M - E_N$ terms in the denominators. The integral in Eq. (14.3.44) may be evaluated in the limit $\mu \gg |E_M - E_N| \sim (Z\alpha)^2 m_e$ by dividing the range of integration of $|\mathbf{k}|$ into two segments, from zero to λ and from λ to infinity, with λ chosen so that $|E_M - E_N| \ll \lambda \ll \mu$ but otherwise arbitrary. In this way we find

$$\int_0^\infty k^2 dk \left[\frac{2}{3k^2(E_M - E_N + k) - i\epsilon} - \frac{1 - k^2/3(k^2+\mu^2)}{(k^2+\mu^2)(E_M - E_N + \sqrt{k^2+\mu^2} - i\epsilon)}\right] \simeq \frac{2}{3}\left[\ln\left(\frac{\mu}{2|E_M - E_N|}\right) + \frac{5}{6} + i\pi\theta(E_N - E_M)\right] .$$

The imaginary term here reflects the possibility of decay of the atom in state N to states M of lower energy. This term contributes to the decay rate, given by the imaginary part of the energy shift. We are interested here in the real part of the energy shift, and so will drop this imaginary term in what follows. Eq. (14.3.44) now gives

$$[\delta E_N]_{\text{low energy}} = \frac{e^2}{6\pi^2}\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 \left[\ln\left(\frac{\mu}{2|E_N - E_M|}\right) + \frac{5}{6}\right] . \quad (14.3.45)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

C Total Energy Shift

We need to make a connection between the sum in Eq. (14.3.45) and the matrix element in the high-energy term (14.3.22). For this purpose, let's first see what value the sum in Eq. (14.3.45) would have if we could ignore the logarithm. We note that $(E_M - E_N)\mathbf{v}_{NM} = [\mathbf{v}, H]_{NM}$, so

$$\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 = \tfrac{1}{2}\sum_M \Big([v^i, H]_{NM}\, v^i_{MN} + v^i_{NM}\, [H, v^i]_{MN}\Big)$$
$$= -\frac{1}{2m_e^2}\Big([p^i, [p^i, H]]\Big)_{NN} .$$

The only term in the non-relativistic Hamiltonian H that does not commute with the momentum operator $\mathbf{p}$ is the potential term, $-e\mathscr{A}^0(\mathbf{x})$, so this gives

$$\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 = -\frac{e}{2m_e^2}\Big(\nabla^2 \mathscr{A}^0(\mathbf{x})\Big)_{NN} . \tag{14.3.46}$$

Inspection of Eqs. (14.3.45) and (14.3.22) now shows that the term proportional to $\ln\mu$ in the high-energy term is cancelled by the term proportional to $\ln\mu$ in the low-energy term:

$$\delta E_N = [\delta E_N]_{\text{high energy}} + [\delta E_N]_{\text{low energy}}$$
$$= \frac{e^2}{6\pi^2}\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 \left[\ln\left(\frac{m_e}{2|E_N - E_M|}\right) + \frac{5}{6} - \frac{1}{5}\right]$$
$$-\frac{e^2}{16\pi^2 m_e^2}\Big(\boldsymbol{\sigma}\cdot\nabla(e\mathscr{A}^0(\mathbf{x}))\times\mathbf{p}\Big)_{NN} . \tag{14.3.47}$$

So far, this has been for a general electrostatic field $\mathscr{A}^0(\mathbf{x})$. Let us now specialize to a pure Coulomb field, with

$$\mathscr{A}^0(\mathbf{x}) = Ze/|\mathbf{x}| . \tag{14.3.48}$$

In this case, Eq. (14.3.46) reads

$$\sum_M (E_M - E_N)|\mathbf{v}_{MN}|^2 = \frac{Ze^2}{2m_e^2}\Big(\delta^3(\mathbf{x})\Big)_{NN} = \frac{Ze^2}{2m_e^2}\Big(f_N^\dagger(0) f_N(0)\Big) . \tag{14.3.49}$$

This is non-vanishing only for $\ell = 0$. Also, the matrix element in the last term in Eq. (14.3.47) has the value

$$\Big(\boldsymbol{\sigma}\cdot\nabla(e\mathscr{A}^0(\mathbf{x}))\times\mathbf{p}\Big)_{NN} = -Ze\Big(\frac{1}{r^3}\boldsymbol{\sigma}\cdot\mathbf{L}\Big)_{NN} , \tag{14.3.50}$$

which is non-vanishing only for $\ell \neq 0$. It is therefore useful at this point to divide our consideration between the two cases, $\ell = 0$ and $\ell \neq 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

i. $\ell = 0$

It is convenient here to define a mean excitation energy ΔE_N:

$$\sum_M |\mathbf{v}_{MN}|^2 (E_M - E_N) \ln |E_N - E_M| \equiv \ln \Delta E_N \sum_M |\mathbf{v}_{MN}|^2 (E_M - E_N)$$
$$= \frac{Ze^2}{2m_e^2} \ln \Delta E_N \left(f_N^\dagger(0) f_N(0) \right) . \tag{14.3.51}$$

For s-wave hydrogenic states, the label N consists of a principal quantum number n and spin z-component m, and $[f_{nm}(0)]_\sigma = 2(Z\alpha m_e/n)^{3/2}\delta_{\sigma,m}/\sqrt{4\pi}$, so

$$\left(f_N^\dagger(0) f_N(0) \right) = \frac{1}{\pi} \left(\frac{Z\alpha m_e}{n} \right)^3 . \tag{14.3.52}$$

Using Eqs. (14.3.51) and (14.3.52) in Eq. (14.3.47) gives the energy shift in these states as

$$[\delta E]_{n,\ell=0} = \frac{4\alpha(Z\alpha)^4 m_e}{3\pi n^3} \left[\ln \left(\frac{m_e}{2\Delta E_{n,\ell=0}} \right) + \frac{19}{30} \right] . \tag{14.3.53}$$

ii. $\ell \neq 0$

For non-vanishing orbital angular momentum the sum (14.3.49) vanishes, so the definition (14.3.51) is inappropriate. Instead, it is conventional here to define a mean excitation energy ΔE_N by

$$\sum_M (E_M - E_N) |\mathbf{v}_{MN}|^2 \ln |E_N - E_M| \equiv \frac{2(Z\alpha)^4 m_e}{n^3} \ln \left(\frac{2\Delta E_N}{Z^2\alpha^2 m_e} \right) . \tag{14.3.54}$$

(Because Eq. (14.3.49) vanishes, it makes no difference what units are used to measure $E_N - E_M$ in Eq. (14.3.54).) Also, in a state of total angular momentum j and orbital angular momentum ℓ, the scalar product $\boldsymbol{\sigma} \cdot \mathbf{L}$ has the familiar value $j(j+1) - \ell(\ell+1) - \frac{3}{4}$, and for principal quantum number n the operator $1/r^3$ has the expectation value

$$\int d^3r |f_N|^2 / r^3 = \frac{2Z^3\alpha^3 m_e^3}{n^3 \ell(\ell+1)(2\ell+1)} . \tag{14.3.55}$$

Putting this all together in Eq. (14.3.47), we have for $\ell \neq 0$:

$$[\delta E]_{jn\ell} = -\frac{4\alpha(Z\alpha)^4 m_e}{3\pi n^3} \ln \left(\frac{2\Delta E_{jn\ell}}{Z^2\alpha^2 m_e} \right)$$
$$+ \frac{\alpha(Z\alpha)^4 m_e}{2\pi n^3} \left[\frac{j(j+1) - \ell(\ell+1) - \frac{3}{4}}{\ell(\ell+1)(2\ell+1)} \right] . \tag{14.3.56}$$

It only remains to use these results to give numerical value for the energy shifts. The mean excitation energies here must be calculated

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

numerically; using non-relativistic hydrogen wave functions, they have the values[5]:

$$\Delta E_{1s} = 19.769266917(6)\ \text{Ry}\,,$$

$$\Delta E_{2s} = 16.63934203(1)\ \text{Ry}\,,$$

$$\Delta E_{2p} = 0.9704293186(3)\ \text{Ry}\,,$$

where 1 Ry $\equiv m_e\alpha^2/2 = 13.6057$ eV. Eq. (14.3.53) then gives

$$[\delta E]_{1s} = \frac{4\alpha^5 m_e}{3\pi}\left[\ln\left(\frac{m_e}{2\Delta E_{1s}}\right) + \frac{19}{30}\right] = 3.3612 \times 10^{-6}\ \text{eV}$$
$$= 2\pi\hbar \times 8127.4\ \text{MHz}\,, \tag{14.3.57}$$

$$[\delta E]_{2s} = \frac{\alpha^5 m_e}{6\pi}\left[\ln\left(\frac{m_e}{2\Delta E_{2s}}\right) + \frac{19}{30}\right] = 4.2982 \times 10^{-5}\ \text{eV}$$
$$= 2\pi\hbar \times 1039.31\ \text{MHz}\,, \tag{14.3.58}$$

while Eq. (14.3.56) gives

$$[\delta E]_{2p_{1/2}} = \frac{\alpha^5 m_e}{6\pi}\left[\ln\left(\frac{\alpha^2 m_e}{2\Delta E_{2p}}\right) - \frac{1}{8}\right] = -5.3267 \times 10^{-8}\ \text{eV}$$
$$= 2\pi\hbar \times -12.88\ \text{MHz}\,. \tag{14.3.59}$$

The classic Lamb shift is the energy difference between the $2s$ and $2p_{\frac{1}{2}}$ states of the hydrogen atom, states that would be degenerate in the absence of radiative corrections. Our calculation has given

$$[\delta E]_{2s} - [\delta E]_{2p_{1/2}} = 4.35152 \times 10^{-6}\ \text{eV} = 2\pi\hbar \times 1052.19\ \text{MHz}\,.$$

This is just the same as the old result of Kroll and Lamb[6] and French and Weisskopf,[7] which they obtained using the techniques of old-fashioned perturbation theory. Earlier in this section we made a crude estimate of 1300 MHz by considering only the high-energy contribution to the $2s$ energy shift, with an infrared cutoff guessed to be of order $\alpha^2 m_e = 2$ Ry. We can now see that this was an overestimate, arising mostly from the fact that the true value of the effective infrared cutoff $\Delta E_{2s} = 16.64$ Ry is considerably larger than we had guessed. On the other hand, as described in Section 1.3, in 1947 Hans Bethe[8] was able to make a rather good estimate of the Lamb shift, 1040 MHz, by considering only the *low-energy* contribution to the $2s$ energy shift, with an *ultraviolet* cutoff guessed to be m_e. (Bethe made the first estimate of the excitation energy, $\Delta E_{2s} \simeq 17.8$ Ry.)

The calculation of the Lamb shift described here has been improved by the inclusion of higher-order radiative corrections and nuclear size

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

and recoil effects. At present the greatest uncertainty is due to a doubt about the correct value of the rms charge radius r_p of the proton. For $r_p = 0.862 \times 10^{-13}$ cm or $r_p = 0.805 \times 10^{-13}$ cm, one calculation[9] gives a Lamb shift of either 1057.87 MHz or 1057.85 MHz, while another[10] gives either 1057.883 MHz or 1057.865 MHz. Given the uncertainty in proton radius, the agreement is excellent with the present experimental value,[11] 1057.845(9) MHz. The accuracy of this experimental value is limited chiefly by the $\sim$ 100 MHz natural linewidth of the $2p$ state in hydrogen, so further improvements here will be very difficult.

In the last few years there has been an important improvement in measurements of the energy shift of the $1s$ ground state itself, by direct comparison of the frequency of the $1s$–$2s$ resonance with four times the frequencies of the $2s$–$4s$ and $2s$–$4d$ two-photon resonances. These s and d states are much narrower than the $2p$ state, so these frequency differences can be measured more accurately than the classic Lamb shift. For a brief time it seemed that there was a discrepancy here between theory and experiment. Calculations[12,13] showed that for a proton radius $r_p = 0.862(11) \times 10^{-13}$ cm or $0.805(11) \times 10^{-13}$ cm, the inclusion of proton size and other corrections increases the theoretical $1s$ energy shift from the above result of 8127.4 MHz to 8173.12(6) MHz or 8172.94(9) MHz, respectively. This result for a proton radius $r_p = 0.862(11) \times 10^{-13}$ cm, which is believed to be more reliable, was not quite in agreement with the measured value[13] of 8172.86(5) MHz. But a later calculation[14] that adopts this proton radius and includes terms of order $\alpha^2(Z\alpha)^5$ yields energy shifts for the $1s$, $2s$, and $4s$ states in agreement with experiment. So apparently quantum electrodynamics wins again.

Problems

1. Consider a charged scalar particle of mass $m \neq 0$, described by a field $\phi(x)$ whose only interaction is with an external time-independent electromagnetic field $\mathscr{A}^\mu(\mathbf{x})$. Let Φ_N be a complete set of normalized one-boson or one-antiboson states with energies E_N, and define $u_N(\mathbf{x})e^{-iE_Nt} \equiv (\Phi_0, \phi(\mathbf{x}, t)\Phi_N)$ and $v_N(\mathbf{x})e^{iE_Nt} \equiv (\Phi_N, \phi(\mathbf{x}, t)\Phi_0)$, where Φ_0 is the vacuum. Show that the u_N and v_N together form a complete set, and give formulas for the coefficients of u_N and v_N in the expansion of a general function $f(\mathbf{x})$.

2. Suppose we include radiative corrections in the theory of Problem 1. Let $i\Pi^*(x, y)$ be the sum of all diagrams with one incoming and one outgoing charged scalar line (excluding final scalar propagators) to order α. Derive a formula for the shift in the energies E_N of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

one-boson states due to these radiative corrections, in terms of $u_N(\mathbf{x})$ and $\Pi^*(x, y)$.

3. Use the results of Section 14.3 to calculate the radiative decay rate of the $2p$ states of hydrogen.

4. Suppose that the electron has an interaction with a light scalar field ϕ, of the form $g\phi\bar{\psi}_e\psi_e$. Suppose that the scalar mass m_ϕ is in the range $(Z\alpha)^2 m_e \ll m_\phi \ll Z\alpha m_e$. Calculate the change in the energy of the $1s$ state of hydrogenic atoms due to this interaction.

5. Carry out the calculation of Problem 4 for $m_\phi = 0$.

References

1. P. A. M. Dirac, *Proc. Roy. Soc. (London)* **A117**, 610 (1928).

2. See, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957); M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957).

3. See, e.g., L. I. Schiff, *Quantum Mechanics*, (McGraw-Hill, New York, 1949): Section 43. The original references are C. G. Darwin, *Proc. Roy. Soc. (London)* **A118**, 654 (1928); *ibid.*, **A120**, 621 (1928); W. Gordon, *Zeit. f. Phys.* **48**, 11 (1928).

4. G. E. Brown, J. S. Langer, and G. W. Schaefer, *Proc. Roy. Soc. (London)* **A 251**, 92 (1959); G. E. Brown and D. F. Mayers, *Proc. Roy. Soc. (London)* **A 251**, 105 (1959); A. M. Desiderio and W. R. Johnson, *Phys. Rev.* **A3**, 1267 (1971).

5. R. W. Huff, *Phys. Rev.* **186**, 1367 (1969).

6. N. M. Kroll and W. E. Lamb, *Phys. Rev.* **75**, 388 (1949)

7. J. B. French and V. F. Weisskopf, *Phys. Rev.* **75**, 1240 (1949)

8. H. A. Bethe, *Phys. Rev.* **72**, 339 (1947).

9. J. R. Sapirstein and D. R. Yennie, in *Quantum Electrodynamics*, ed. by T. Kinoshita (World Scientific, Singapore, 1990): p. 575, and references quoted therein.

10. H. Grotch, *Foundations of Physics* **24**, 249 (1994).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

11. S. R. Lundeen and F. M. Pipkin, *Phys. Rev. Lett.* **46**, 232 (1981); S. R. Lundeen and F. M. Pipkin, *Metrologia* **22**, 9 (1986). For a review, see F. M. Pipkin, in *Quantum Electrodynamics*, ed. by T. Kinoshita (World Scientific, Singapore, 1990): p. 697.

12. M. Weitz, A. Huber, F. Schmidt-Kaler, D. Leibfried, and T. W. Hänsch, *Phys. Rev. Lett.* **72**, 328 (1994).

13. M. Weitz, F. Schmidt-Kaler, and T. W. Hänsch, *Phys. Rev. Lett.* **68**, 1120 (1992), and Ref. 12.

14. K. Pachucki, *Phys. Rev. Lett.* **72**, 3154 (1994).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167.016

Author Index

Where page numbers are given in italics, they refer to publications cited in bibliographies and lists of references.

Aaron, R. *470*
Aharony, A. *168*
Aitken, A. C. *106*
Amado, R. D. *470*
Anderson, C. D. 12, 30, *43, 45*
Aramaki, S. *39*
Argyres, P. C. *257*
Artin, E. *424*

Bagger, J. *338, 532*
Bailey, J. *498*
Bakamijian, B. *189*
Bargmann, V. *106*
Belinfante, F. J. *44*, 316, *338*
Berezin, F. A. 403, 404, *424*
Bergmann, P. G. *338*
Bethe, H. 29, 33, 36, *45, 47*, 560, *563*, 593, *596*
Beyer, R. T. *39, 43*
Bhabha, H. J. 29, *45*
Bloch, F. 33, *46, 562*
Bjorken, J. D. *532*
Blackett, P. M. 13
Bogoliubov, N. N. 512, *532*
Bohr, N. 3, 11, 19, 32, *42, 44, 168*, 198, *256*
Born, M. 3, 15, 16, 17, 18, 19, 23, 25, *40, 42, 43, 46*, 115, *166*, 292
Breit, G. 36, *48*, 162, *167, 168*
de Broglie, L. 3, *40*
Brown, G. E. *596*
Brown, L. *39, 40, 45, 46, 167, 532*
Burgoyne, N. *257*
Butler, C. C. 30, *45*, 123, *167*

Cao, T. Y. *39*
Carlson, J. F. 29, 30, 32, *45*
Cartan, E. *256*
Casimir, H. B. G. *338*
Cassen, B. *167*
Chadwick, J. *45*
Chew, G. *47, 471*
Chinowsky, W. *167*
Christenson, J. H. *106, 167*
Cohen, R. S. *44, 256*
Collins, P. D. B. *471, 532*
Compton, A. H. *41*, 362, 364, 504
Condon, E. U. *167*
Conversi, M. 30, *45*
Crichton, J. H. *189*
Cronin, J. W. *106*, 132, *167*
Curie–Joliot, I. *45*

Dalitz, R. H. 141, *168, 563*
Dancoff, S. M. 33, 34, *46*
Darwin, C. G. 10, *41, 42, 596*
Davisson, C. J. 3, *40*
De Witt, C. M. 377, 418, *424, 471*
Deans, W. M. *41*
Dedijer, S. *257*
Deser, S. *563*
Desiderio, A. M. *596*
DeWitt, B. S. *39, 424*
Dirac, P. A. M. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 19, 22, 23, 24, 27, 28, 29, 32, 33, 34, *39, 40, 41, 42, 43, 44, 46, 47*, 49, *105*, 200, 213, 218, *256*, 292, 325, 328, 329, 330, *338*, 345, 376, *424*, 457, *470*, 489, 565, 566, 567, *596*
Donoghue, J. F. *533*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Drell, S. D. *532*
Dresden, M. *167*
Drinkwater, J. W. *47*
Drühl, K. *424*
Dyson, F. J. 37, *48*, *106*, 144, *168*, *258*, 259, 287, *291*, 376, 499, *532*

Eckart, 153, 156, 162, 165
Edmonds, A. R. *106*, *167*, *168*, *257*, *596*
Ehrenfest, P. 15, 29, *43*, *45*
Einstein, A. 12, 13, 18, 19, *43*, 55, 518
Elsasser, W. 3, *40*
Epstein, S. *46*
Erickson, G. W. *498*
Euler, H. 32, *46*, 523, 524, 526, *533*

Fabri, E. *168*
Faddeev, L. D. *190*, 376, 378, *424*
Feenberg, E. *167*, *168*
Feinberg, G. *167*, 530, *533*
Feinberg, J. *338*
Fermi, E. 11, 19, 23, *39*, *42*, *43*, *44*, 292
Ferretti, B. *167*
Feynman, R. P. 33, 36, 37, 38, *47*, *48*, 259, 276, 280, 286, 353, 354, 355, 360, *375*, 376, 377, 400, 411, 413, 417, *423*, *424*, 426, 430, 459, 472, 474, 486, *498*, 559, 572
Fierz, M. 20, *40*, *44*, *46*, *47*, *257*
Fitch, V. L. *106*, 133, *167*
Flanders, H. *106*, *375*
Fock, V. 22, 23, *41*, *44*, *375*
Foley, H. M. *48*
Frautschi, S. C. *471*, *562*
Fredenhagen, K. *424*
French, J. B. 35, 38, *47*, 593, *596*
Friedman, J. I. *106*, *167*
Fröhlich, H. *48*
Froissart, M. *168*
Fukuda, H. *47*
Furry, W. H. 23, 27, 28, 32, *44*, *46*, 428, *470*

Gaberdiel, R. *424*
Gamow, G. 7, *39*, *42*
Garwin, R. *106*, *167*
Gasser, J. *533*
Gell-Mann, M. 123, 132, *167*, *291*, 463, *471*, 556, *563*
Georgi, H. *257*
Germer, L. H. 3, *40*
Gerstein, I. S. *424*
Glimm, J. *423*
Goldberger, M. L. *166*, 463, *471*, 556, *563*
Goldstone, J. 178, *189*
Gordon, W. 4, 7, 10, 13, 25, 27, *41*, *42*, 200, 211, 239, 277, *596*
Goudsmit. S. 5, 10, *41*
Green, M. B. *375*
Grisaru, M. *563*
Grotch, H. *596*
Gudehus, T. *257*
Gürsey, F. *106*

Haag, R. *424*
Hafstad, L. D. *167*
Hahn, Y. *532*
Halter, J. *533*
Hänsch, T. W. *596*
Harvey, J. *533*
Hawking, S. W. *533*
Heisenberg, W. 3, 4, 10, 15, 16, 17, 20, 21, 23, 24, 25, 29, 31, 32, 33, *40*, *41*, *42*, *43*, *44*, *45*, *46*, 109, *166*, 292, 519, *532*, *533*
Heitler, W. 29, *45*, *48*
Hepp, K. 512, *532*
Herzberg, G. *45*
Heydenberg, N. *167*
Hibbs, A. R. *423*
Hocking, J. G. *106*
Hoddeson, L. *39*, *40*, *46*, *167*
Howard, J. C. *470*
Hubbard, J. 461, *470*
Huber, A. *596*
Huff, R. W. *596*
Hugenholtz, N. M. 178, *190*

Infeld, L. *46*
Inönü, E. 62, 105
Inoue, T. *45*
Israel, W. *533*
Ito, D. *48*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Jackiw, R. *40, 424*
Jackson, J. D. *471*
Jaffe, A. *423*
Jammer, M. *39*
Jauch, J. M. *257*
Johnson, W. R. *596*
Joliot, F. *45*
Jordan, P. 3, 10, 15, 16, 17, 18, 19, 23, 25, *40, 41, 43, 44*, 292
Joos, H. *257*
Jost, R. *258*
Jouvet, B. *470*
Julian, R. S. *48*

Kabir, P. 530, *532*
Kahn, B. *48*
Kaiser, G. *257*
Källen, G. 457, *470*
Kamefuchi, S. *424, 532*
Kanesawa, S. *48*
Kemmer, N. 31, *46*
Kinoshita, T. *563, 596*
Klein, O. 4, 7, 13, 25, 27, 29, *41, 44*, 200, 211, 239, 277, 368, *375*
Koba, Z. *46, 48*
Kockel, B. 31, *46, 533*
Kollath, R. *168*
Kramers, H. A. 35, *47*, 81, *106, 168*, 469, *471*
Kroll, N. M. 38, *47*, 593, *596*
de Kronig, R. 469, *471*
Kubo, R. *190*
Kusch, P. *48*

Laidlaw, M. G. G. 418, *424*
Lamb, W. E. 35, 36, 38, *47, 48*, 593, *596*
Landau, L. D. *106, 168*, 418, 452
Landé, A. *41*
Lang, S. *424*
Langer, J. S. *596*
Lattes, C. 30
Lederman, L. *106, 167*
Lee, T. D. *106*, 127, 132, *167*, 178, *189*, 228, *424*, 549, *563*
Lehmann, H. 438, 457, 463, *470*
Liebfried, D. *596*
Leinaas, J. M. *424*
Leutwyler, H. *533*
Lewis, H. W. *46*
Lifshitz, E. M. *106, 168*
Lippmann, B. 111, *166*
Liu, H. H. T. *498*
London, F. *375*
Low, F. E. *291*, 556, *563*
Lüders, G. 245, *257*
Ludwig, G. *40*
Lundeen, S. R. *596*
Lyubarski, G. Ya. *257*

Mackey, G. W. *105*
Mandelstam, S. *424*
Marshak, R. E. 38, *45*
Maskawa, T. 329, 330, *338*
Matthews, P. T. *532*
Mayers, D. F. *596*
Mehra, J. *39, 42*
Michelson, A. A. *41*
Miller, A. I. *39, 44, 498*
Miyamoto, Y. *47*
Møller, C. 29, *45, 46, 166*
Myrheim, J. *424*

Nafe, J. E. *48*
Nagel, D. E. *48*
Nakajima, H. 329, 330, *338*
Nappi, C. R. *257*
Nash, C. *106*
Nauenberg, M. 549, *563*
Neddermeyer, S. H. 30, *45*
Ne'eman, Y. 123, *167*
Nelson, E. B. *48*
Newton, R. G. *166*
Nishina, Y. 29, *44*, 368, *375*
Nordheim, L. 30, *45*
Nordsieck, A. 33, *46, 562*
Novikov, V. *533*

Occhialini, G. P. S. 13, 30, *45*
Ohnuki, Y. *424*
Omnes, R. *471*
Oppenheimer, J. R. 12, 23, 27, 28, 29, 30, 31, 35, 37, 38, *43, 44, 45, 46*
Osterwalder, K. *424*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Pachucki, K. *596*
Pais, A. *39*, 132, *167*
Pancini, E. 30, *45*
Parasiuk, O. 512, *532*
Paschen, F. *41*
Pasternack, S. *47*
Pauli, W. 3, 11, 14, 19, 20, 21, 22, 23, 24, 26, 28, 31, *40, 42, 43, 46, 47*, 245, *257*, 292, 494, *498*, 517
Peierls, R. E. *44, 46, 47, 168*
Pendleton, H. *563*
Perlmutter, A.. *257*
Petermann, A. *498*
Piccioni, O. 30, *45, 167*
Pipkin, F. M. *596*
Placzek, G. *168*
Podolsky, B. 22, *44*
Polchinski, J. *424*, 526, *533*
Pomeranchuk, I. Ia. 468, 469, *471*
Popov, V. N. 377, *424*
Powell, C. F. 30, *45*
Present, R. S. *167*

Rabi, I. I. 36, *48*
Racah, G. *45*
Ramsauer, 165
Rarita, W. 232, *257*
Rechenberg, H. *45*
Regge, T. 468, 469, *471*
Retherford, R. C. *47*
Richardson, O. *47*
Rindani, S. D. *257*
Roberts, J. E. *424*
Robertson, H. P. *42*
Rochester, G. D. 30, *45*, 123, *167*
Rohrlich, F. *257*
Rose, M. E. *46, 106, 167, 168, 257, 596*
Rosenbluth, M. N. 457, *470*
Rosenfeld, L. 19, *44*, 198, *256, 338*
Rüger, S. M. *424*

Sakata, S. *45, 532*
Salam, A, *39, 42*, 512, *532*
Salpeter, E. E. 560, *563*
Sapirstein, J. R. *596*
Schaefer, G. W. *596*
Schearer, J. F. *41*
Schiff, L. I. *41, 43, 168, 563, 596*
Schmidt–Kaler, F. *596*
Schrader, R. *424*
Schrödinger, E. 3, 4, 6, 13, 27, *41*, 109
Schroer, B. *257*
Schubert, K. R. *167*
Schur, I. *257*
Schwarz, J. H. *375*
Schweber, S. S. *39, 40*
Schwinger, J. 14, 33, 35, 36, 37, 38, *40, 43, 46, 47, 48*, 111, *166*, 232, *257*, 259, *291, 375*, 376, *424, 470*, 489, *498*
Screaton, R. *471*
Seiler, R. *257*
Sen, S. *106*
Serber, R. *46, 48*
Shifman, M. A. *533*
Sivakumar, M. *257*
Slater, J. C. *42*
Sommerfeld, A. 3, 5, 6, 11, *41*
Spaarnay, M. J. *338*
Stachel, J. *44, 256*
Steinberger, J. *167*
Sterman, G. *563*
Stevenson, E. C. 30
Stoner, E. C. *42*
Stora, R. *39*
Stratonovich, R. L. 461, *470*
Streater, R. F. *257, 258*
Street, J. C. 30
Suura, H. *498, 562*
Swieca, J. A. *257*
Symanzik, K. 438, 464, *470*

't Hooft, G. 377, 477, *424, 498*
Takahashi, Y. 447, *470*
Tamm, I. 12, *43, 44, 375*
Tate, J. E. *424*
Tati, T. *48*
Telegdi, V. L. *106, 167*
Thirring, W. 463, *471*
Thomas, L. H. *42, 189*
Thomson, J. J. 369
Titchmarsh, E. C. *190*
Toll, J. S. *471*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Tomonaga, S.-I. 36, 37, 38, *40, 46, 47, 48*, 259, 376
Townsend 165
Tung, W.-K. *106*
Turlay, R. *106, 167*
Turnbull, H. W. *106*
Tuve, M. A. *167*

Uehling, E. A. 34, *46, 47*, 484, *498*
Uhlenbeck, G. E. 5, *41*
Umezawa, H. *532*

Vainshtein, A. I. *533*
Vaughn, M. J. *470*
Velo, G. *257*
Veltman, M. 477, *498*
Villars, F. 494, *498*

van der Waerden, B. L. *40, 43, 257*
Waller, I., 31, *45*
Ward, J. C. 447, *470*
Watson, K. M. 130, *166, 167*
Webb, N. *45*
Weinberg, S. *40, 42, 47, 105, 167, 168, 190, 256, 257, 338, 375, 424, 470, 471, 532, 533, 563*
Weinrich, M. *106, 167*
Weisskopf, V. F. 24, 26, 28, 31, 32, 34, 35, 37, *40, 43, 44, 45, 46, 47*, 593, *596*
Weitz, M. *596*
Wentzel, G. *40, 42, 46*
Wess, J. *338, 532*
West, T. *375*
Weyl, H. 12, *43, 375*
Wheeler, J. A. 4, 33, *46, 47, 166*
Whittaker, E. *40*
Wichmann, E. H. *189, 256, 498*
Wick, G. C. *105, 291*, 475, 486, 497, *498*
Wightman, A. S. *105, 257, 258*
Wigner, E. P. 19, *39, 42, 43, 44*, 51, 62, 68, 91, 93, 100, 103, 104, *105, 106*, 108, 153, 156, 162, *168*, 236
Wilczek, F. *424*
Williams, E. J. *563*
Williams, R. C. *47*
Williams, W. E. *47*
Wilson, W. *41*, 525, 528, *533*
Witten, E. *40, 257, 375*
Wu, C. S. *106*, 127, *167*

Yang, C. N. *106*, 127, 132, *167, 168*, 178, *189*, 228, *257, 375*
Yang. C. P. *168*
Yennie, D. R. *562, 596*
Yukawa, H. 30, *45*, 159, 436, *470*

Zacharias, J. R. *48*
Zakharov, V. I. *533*
Zimmerman, W. 438, 463, *470*, 512, *532*
Zumino, B. *257*
Zwanziger, D. *257*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:33, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Subject Index

abelian groups, defined, 55
absorption of photons, 18
accidental symmetries, 529–31
action, 299, 307
alpha decay, 160
anomalous Zeeman effect, 5
annihilation and creation operators, 16, 19–20, 23–4, 26–8, 169, 173–7
antiparticles, 13–14, 23–8, 104, 149–50, 199, 567
antiunitary and antilinear operators, defined, 51
anyons, 420
auxiliary fields, 302–3, 314
axial gauge, 346

baryon number, defined, 122
Belinfante tensor, 316
Berezin integration, defined, 403
beta decay, 23, 29, 127, 146, 228, 519
Bhabha (electron–positron) scattering, 29
Boltzmann H-theorem, 151
boosts, defined, 61
Born approximation, 115, 156
Bose–Einstein statistics, 11, 170–2, 418–20
BPHZ prescription, 512–13
braid group, 420
Bethe–Salpeter equation, 560
bound states, *see* composite particles
Breit–Wigner formula, 162–3
bremstrahlung, 29
broken symmetry, 443, 451

canonical commutation and anticommutation relations, 16, 19–22, 293–8, 529
canonical transformations, 329
Casimir effect, 297
causality, 145, 198, 463
center-of-mass frame, 139
central charges, 83
charge, *see* electric charge
charge conjugation (C)
  accidental symmetry, 521, 530–1
  defined, 121, 131–2
  for photons, 427–8
  intrinsic charge conjugation phases, 131
  non-conservation, 132
  transformation of creation operators, 177
  transformation of Dirac fields, 226–7
  transformation of fermion bilinears, 229
  transformation of general irreducible fields, 241–2
  transformation of scalar fields, 206
  transformation of vector fields, 213
  also see specific particle types
Chew–Frautschi plot, 469, 471
chiral transformation, 520
circular polarization, 359
Clebsch–Gordan coefficients, 124, 152, 154, 156, 233–4, 242, 569
Clifford algebra, 214
closed p-forms, defined, 369
cluster decomposition principle, 169, 177–89, 197, 259

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Co^{60} decay, 127, 130–1
coherent states, 189
color, 549
compact and non-compact groups, 231
composite particles, 110, 461–2
Compton (electron–photon) scattering, 29, 362–9
connected amplitudes, 178–82, 270, 282, 286, 389, 413
conservation laws
  for angular momentum, 118
  for charge, 122, 199, 427
  for current, 212, 307, 478, 586
  for energy and momentum, 117–18, 425–7
  limitations, 253, 537–8
  also see specific symmetries and conserved quantities
constraints, 325–31
  in electrodynamics, 344, 346–7
cosmic rays, 29, 123
Coulomb gauge, 251, 346–50, 365
Coulomb energy, 350, 353, 355–6, 560
CP-invariance
  for degenerate multiplets, 104
  non-conservation in K^0–$\bar{K}^0$ decay, 132–3
counterterms in quantum electrodynamics, 472–3
CPT-invariance, 104, 133, 244–6, 459
creation operators, *see* annihilation and creation operators
cross sections
  defined, 137–9
  high energy limit, 158–9
  partial wave expansions, 155–6
crossing symmetry, 269, 467, 554
cumulants, 178

Dalitz plot, 141
dangerous states, 550–2
decay rates, general formula, 136–7
Δ baryon, 165
Δ_+ function, 202
Δ_F function, 276
de Rham cohomology, 370
density matrix, 360
differential forms, *see* p-forms
diffraction scattering, 148–9, 158
dimensional regularization, 449, 477–80, 497
dimensionality, of fields and couplings, 502, 519, 525–7
Dirac brackets, 328–31, 332–7
  in electrodynamics, 347–9
Dirac equation, 1, 6–14, 225, 565–72
Dirac matrices, 8–9, 214–19
  slash notation, defined, 358
  traces, 361, 372–4
Dirac representation of homogeneous Lorentz group, 213–19
dispersion relations, 460, 462–9
distorted wave Born approximation, 146–7
dotted and undotted indices, 230
duality, 232, 371
Dyson series, 144, 259–60

eclipsing binaries, 368
effective field theories, 499, 523–5
electric charge, 341
  conservation, 122, 537
  renormalization, 342, 442–8, 473, 480–3
electric charge radius, defined, 493
electric dipole moments, 81, 521
electron
  charge radius, 493
  classical theory, 31, 369, 496
  magnetic moment, 6, 10, 14, 36, 457, 468, 520
  spin, 6–9
elliptic polarization, 360
energy shifts of atomic states, 31–2, 574–8
  1s energy shifts, 594
  also see Lamb shift, Uehling effect, Muonic atoms
energy–momentum tensor, 310–12
entropy, 151
equivalence principle, 537
exact p-forms, 369

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

excitation energies, in hydrogen, 592–3
exclusion principle, 11
exterior derivatives, 369
external fields, 266, 287–90, 412–13, 556–62, 572–8
Euclidean path integrals, 384
Euler constant, 479, 497
Euler–Lagrange equations, 300

Faddeev equations, 188
Fermi–Dirac statistics, 12, 171–2, 267–70, 418–20
Feynman diagrams, 36–7, 259–91
  for electrodynamics, 355–8
Feynman gauge, 355, 417
Feynman parameters, 474, 486, 497
fields, *see* quantum fields
field equations, 200, 211-12, 239; *also see* quantum fields
field renormalization, 331–2, 436–42, 452, 461, 473, 479, 484, 543–4
field-translation-invariant scalar theory, 521–3
fine structure, 4–6, 570
fine structure constant, 2, 5
first class constraints, *see* constraints
flavors (of leptons), 529
floating cutoff, 525–8
forests, 512–13
form factors, 452–7, 485–93, 580–2
Froissart bound, 159
functionals, notation, 299
Furry's theorem, 428–9, 509

Galilean invariance, 62, 145
γ_5, 217–18
Gaussian integrals, 420–3
Grassmann variables, defined, 401
gauge transformations, 251–2, 339–43, 345, 370, 448–52
  also see Lorentz gauge, Coulomb gauge, temporal gauge, axial gauge, unitarity gauge, Feynman gauge
general relativity, 255, 312, 316, 369, 518–19, 521
generators of symmetries, 307–14
global symmetries, defined, 307
gluons, 549
graviton, 73–4, 253, 521, 537, 548
groups, defined, 52; *also see* abelian groups, homotopy groups, Lie groups, little groups, representations, semi-simple groups

Hamiltonian,
  for complex scalar field, 22
  for Dirac equation, 8
  for electrodynamics, 349–50
  for free particles, 176
  for interacting Dirac field, 323
  for interacting scalar fields, 199, 302
  for interacting vector field, 321
  for one-dimensional scalar field, 15–17
Heisenberg picture, 109, 288, 297, 425
helicity
  defined, 72
  limitations for massless particle fields, 253–4
  limited to integers and half-integers, 90
Hilbert space, defined, 49
hole theory, 12, 29, 31
homotopy groups and classes, 89–90, 96–100, 419–20
  defined, 100
Hubbard–Stratonovich transformation, 461
hypercharge, 123
hyperons, 123
  discovered, 30

'in' and 'out' states, 107–12, 116
induced emission of photons, 18
induced representations, 65
infinities, *see* ultraviolet divergences, infrared divergences
infrared divergences, 32–3, 491, 496, 539–53
Inönü–Wigner contraction, 62
interaction picture, 143, 287
  for derivative coupling, 318–20
  for Dirac field, 323–5

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

for electrodynamics, 353
for interacting scalar field, 304–5
for vector field, 320–3
internal symmetries, 121–2
invariant subgroups and subalgebras, 70, 88
irreducible representations, defined, 64
irrelevant couplings, 503
isospin (isotopic spin), 123

J/ψ meson, 225
Jacobi identity, 83
jets, 550, 552

K mesons, 123, 132–3
discovered, 30
Källen–Lehmann representation, 457–62
kinetic theory, 150–1
Klein–Gordon–Schrödinger equation, 1, 13, 21
Kramers degeneracy, 81
Kramers–Kronig relation, 469

Lagrangians, 20, 297–306
for complex scalar field, 21
for interacting scalar field, 302
integration by parts, 305–6
irregular, 326
Lamb shift, 34–6, 484, 578–94
Lee–Nauenberg theorem, 549–53
Legendre transformations, 297, 301
lepton number, 122, 530
leptons, 472
Lie groups, 53–5
linear polarization, 359
Lippmann–Schwinger equation, 111, 142
little groups, 64–6
$m \neq 0$: 68–9
$m = 0$: 69–73
local symmetries, 342
defined, 64
for different momenta, 65–6
for massless particles, 248
also see gauge transformations
loops, 186–7, 270, 282–3, 358, 413
Lorentz (or Landau) gauge, 212, 346, 418
Lorentz transformations
action on creation operators, 177
action on general states, 57
action on one-particle states, 62–73
defined, 55–7
homogeneous Lorentz group, defined, 57; general representations, 229–33
in canonical formalism, 314–18
in off-shell matrix elements, 427
in perturbation theory, 144–5, 277–9, 259
of cross-sections and decay rates, 138
of S-matrix, 116–21
Poincaré algebra, 58–61
Poincaré group, defined, 57
proper orthochronous Lorentz group, 58
LSZ theorem, *see* reduction formula

M-matrix, defined, 117
magnetic moments, of spin 1/2 particles, 456–7, 485–90, 520, 555–6
also see electron, muon
Majorana fermions, 226, 242
Mandelstam variables, 515
many time formalism, 22
marginal couplings, 503
mass renormalization, 439–42, 473, 495–6, 587–9
matrix mechanics, 3–4,
Maxwell equations, 1, 252, 339, 342, 370
Møller (electron–electron) scattering, 29
momentum operator, 61, 310–11
muon, 38
discovered, 30
magnetic moment, 489–90, 520
muonic atoms, 483–4

N^{14} molecular spectrum, 29
negative energy 'states', 10–14, 23–7, 567
neutron, 29
neutron–proton scattering, 158

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

neutron scattering on complex nuclei, 157, 160
Noether's theorem, 307
non-compact groups, *see* compact and non-compact groups
non-linear σ-model, 337, 377, 392–3
normal ordering, 175, 200, 262
nuclear forces, 29–30, 434–6

old-fashioned perturbation theory, *see* perturbation theory
ω meson, 227
one-particle-irreducible, defined, 439
optical theorem, 148
Osterwalder–Schrader axioms, 384, 424
overlapping divergences, 511–15

p-forms, 369–72
pair production, 29
parastatistics, 420
parity and space inversion (P)
accidental symmetry, 521, 530–1
defined, 58
intrinsic parities, 125–7
non-conservation in weak interactions, 127
of bound electron states, 568–9, 586
transformation of creation operators, 177
transformation of Dirac fields, 221, 224–5
transformation of general irreducible fields, 239–40
transformation of $J_{\mu\nu}$ and P_μ, 74–6
transformation of one-particle states, 76–9, 103
transformation of scalar fields, 205–6
transformation of vector fields, 213
also see specific particle types
partial wave expansions, 151–9
path integrals, 259, 376–7, 524
derivation, 378–84
for derivative coupled scalars, 391–2
for fermions, 399–413
for massive vector fields, 393–5
for non-linear σ-model, 392–3
for quantum electrodynamics, 413–18
for S-matrix, 385–9
Lagrangian version, 389–95
used to derive Feynman rules, 395–8
Pauli matrices, defined, 217
Pauli term, 14, 517, 520
Pauli–Villars regulator, 486, 494
perturbation theory
old-fashioned, 142, 259, 549, 564, 585–6, 593
time-dependent, 142–5
also see Dyson series, Feynman diagrams, path integrals
phase shifts, 129–30, 155
phase space factors, 139–41
photon, 3
charge conjugation phase, 229
helicity and polarization, 73–4, 250–1, 359–61, 368
masslessness, 343, 452, 477
photon–photon scattering, 32, 509, 523–5
propagator, 353–5
also see soft photons, quantum electrodynamics
pion–nucleon coupling constant, 435
pions, 13, 38, 435–6, 521
$\pi^\pm$ decay, 127
π^0 decay, 131
intrinsic charge-conjugation phase of π^0, 131, 229
intrinsic parity, 127
prediction and discovery, 30
scattering on nucleons, 463, 469
Poincaré group and algebra, *see* Lorentz transformations
Poincaré's theorem, 99, 370
Poisson brackets, defined, 327
polar decomposition theorem, 88, 530
polarization
massive spin one particle, 209–10, 212
masslessness, 343
photon, 73–4
Polchinski's theorem, 526–8

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

poles in scattering amplitudes, 428–36, 554, 564
Pomeranchuk's theorem, 468
positivity of energy, 75–6, 302
positron, 12–13, 567–8, 585
positronium, 227
power counting theorem, 505
primary constraints, *see* constraints
principal value function, 113
probabilities in quantum mechanics, 7, 27–8, 33, 50
projective representations, 53, 81–91, 96–100
propagators, 262, 274–80, 397–8, 506, 574–8
proton
  charge, 445
  charge radius, 38, 594
  form factors, 457
  identified as holes, 12
  in nuclei, 29
PT-non-conservation, 130–1

quantum chromodynamics, 123, 531, 549
quantum electrodynamics, 29, 31–8, 339–62, 413–18, 472–97, 503, 507–15, 529–31, 564–94
  also see quantum fields, electron, photon, gauge transformations, ultra-violet divergences, infrared divergences
quantum fields,
  Dirac, 23–4, 219–29
  early theory, 15–31
  free fields, 191–200
  general quantum fields, 233–44
  massless particles, 246–55
  redefinitions, 331–2
  scalar, 21–2, 24–8, 201–6
  uniqueness of irreducible fields, 238
  vector, 207–12
quantum mechanics, 49–50
quarks, 217, 531

radiative corrections, *see* renormalization, quantum electrodynamics, ultraviolet divergences, infrared divergences, self-energy functions, vertex function, vacuum polarization, soft photons
Ramsauer–Townsend effect, 165
Rarita–Schwinger field, 232, 423
rates, general formula, 134–6
rays, defined, 49–50
reduction formula, 436–8
redundant couplings, 331–2, 522–3
Regge trajectories, 468–9
relativistic wave equations, 1–14
relevant couplings, 530
renormalizability, 37, 499, 502–3, 516–25
renormalization, 34–8, 506–16
  also see electric charge, field renormalization, mass renormalization
representations
  of groups, defined, 53
  of homogeneous Lorentz group, 229-33
renormalization group, 490, 516, 525
resonances, 159–65
ρ-mesons, 165, 225, 227, 469
Riemann–Lebesgue theorem, 181
Rosenbluth formula, 457
rotation matrices ($D^j_{\sigma'\sigma}(R)$), 68
Rydberg unit (Ry), defined, 593

S-matrix, 4, 33
  C, CP and CPT, 131–4
  defined, 113
  internal symmetries, 121–4
  Lorentz invariance, 116–21
  parity, 124–7
  PT, 130–1, 133
  S-operator, 114
  time reversal, 127–30
  unitarity, 113–14, 147–51
  also see T-matrix, M-matrix
scattering amplitude (f), 148, 153–4, 466–7
scattering lengths, defined, 157
Schrödinger picture, 109
Schwinger action principle, 288

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

Schwinger terms, 449
second class constraints, *see* constraints
secondary constraints, *see* constraints
self-energy functions (Π^*, Σ^*), 439, 473–80, 493–6, 508–9, 512–15, 575–9, 583–4
semi-simple Lie groups and algebras, defined, 70, 86
separable interactions, 165
Shelter Island Conference (1947), 34–8
simply connected spaces, defined, 84
$SL(2,C)$ group, 87, 90
soft photons, 534–48, 553–6
space inversion, *see* parity and space inversion
spectral functions, *see* Källen–Lehmann representation
spherical harmonics, 153, 569
spin matrices, defined, 230
spin–statistics connection, 238
spin sums, 210, 224, 236–7, 252, 360–1, 365, 545
$Spin(d)$ group, 90
spontaneous emission of photons, 17–19, 590
standing wave states, 166
strangeness and strange particles, 123
  see K mesons, hyperons
string theory, 1, 15, 244, 371, 525
strong interactions, 30
structure constants, defined, 54
$SU(2)$ group, 88, 123, 230
$SU(3)$ symmetry, 123
subtractions in dispersion relations, 460, 465
superficial divergences, 500–2, 510
superrenormalizable interactions, 503, 507
superselection rules, 53, 90–1
supersymmetry, 325, 506, 526
symmetries, 50–5, 91–6, 306–14, 425
  also see conservation laws, groups, Lorentz invariance, parity, charge conjugation, time reversal, internal symmetries, isospin, $SU(3)$
T-matrix, 111, 116, 141–2, 152, 549
τ–θ problem, 127
temporal gauge, 346
Thomson scattering, 369, 556
threshold behavior, 157–8
time-ordered products, defined, 143, 280
time reversal (T)
  consequences for S-matrix, 127–30
  defined, 58
  non-conservation, 134
  transformation of creation operators, 177
  transformation of degenerate multiplets, 100–4
  transformation of Dirac fields, 227–8
  transformation of general irreducible fields, 242–3
  transformation of $J_{\mu\nu}$ and P_μ, 74–6
  transformation of one-particle states, 77–81
  transformation of scalar fields, 206
  transformation of vector fields, 213
topology
  of Lorentz and rotation groups, 86–90
  also see homotopy groups, simply connected spaces, de Rham cohomology
traces, *see* Dirac matrices
tree graphs, defined, 283
two-cocycles, defined, 82

Uehling effect, 34, 484, 581
ultraviolet divergences, 31–8, 476–7, 482–3, 485, 491, 494–7, 500–5
unitarity gauge, 346
unitarity of S-matrix, 113–16, 129, 147–51, 155, 161, 521
universal covering group, 90

vacuum energy, 24, 26
vacuum polarization, 32, 34, 473–85, 581
vacuum state, 24, 27, 176

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

vertex function (Γ_μ), 446, 485–8, 507–8, 579–80
  wave function, 386–8

$W^{\pm}$ particles, 207
Ward and Ward–Takahashi identities, 445–8, 476, 508, 511
Watson's theorem, 130
wave packets, 109–10
Wick rotation, 475–6
Wick's theorem, 261
Wigner–Eckart theorem, 153
Wigner rotation, 68–73
Wigner three-j symbols, 237
Wilson's renormalization method, 525–8

Z^0 particle, 159–60, 207
Z_2 group, 88

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:19:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644167

In *The Quantum Theory of Fields*, Nobel Laureate Steven Weinberg combines his exceptional physical insight with his gift for clear exposition to provide a self-contained, comprehensive, and up-to-date introduction to quantum field theory. Volume II gives a current and self-contained account of the methods of quantum field theory, and how they have led to an understanding of the weak, strong, and electromagnetic interactions of the elementary particles. The presentation of modern mathematical methods is throughout interwoven with accounts of the problems of elemenatary particle physics and condensed matter physics to which they have been applied. Many topics are includded that are not usually found in books on quantum field theory. The book contains much original material, and is peppered with examples and insights from the author's experience as a leader of elementary particle physics. Problems are included at the end of each chapter.

From reviews of Volume I

'...an impressively lucid and thorough presentation of the subject... Weinberg manages to present difficult topics with richness of meaning and marvellous clarity. Full of valuable insights, his treatise is sure to become a classic, doing for quantum field theory what Dirac's *Quantum Mechanics* did for quantum mechanics. I eagerly await the publication of the second volume.'

S. S. Schweber, *Nature*

'For over twenty years there has been no good modern textbook on the subject. For all that time, Steven Weinberg has been promising to write one. That he has finally done it is cause for celebration among those who try to teach and try to learn the subject. Weinberg's book is for serious students of field theory... it is the first textbook to treat quantum field theory the way it is used by physicists today.'

Howard Georgi, *Science*

'Steven Weinberg, who contributed to the development of quantum chromodynamics and shared the Nobel Prize in Physics for his contributions to the electroweak theory, has written a definitive text on the physical foundations of quantum field theory. His book differs significantly from the long line of previous books on quantum field theory... To summarize, *Foundations* builds the structure of quantum field theory on the sure footing of physical insight. It is beautifully produced and meticulously edited... and it is a real bargain in price. If you want to learn quantum field theory, or have already learned it and want to have a definitive reference at hand, purchase this book.'

O. W. Greenberg, *Physics Today*

'In addition to a superb treatment of all the conventional topics there are numerous sections covering areas that are not normally emphasized, such as the subject of field redefinitions, higher-rank tensor fields and an unusually clear and thorough treatment of infrared effects... this latest book reinforces his high scholarly standards. It provides a unique exposition that will prove invaluable both to new research students as well as to experienced research workers. Together

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

with Volume II, this will become a classic text on a subject of central importance to a wide area of theoretical physics.'

M. B. Green, *CERN Courier*

'I believe that what readers will find particularly helpful in this volume is the consistency of the whole approach, and the emphasis on quantities and properties that are directly useful to particle physicists. This is particularly true for those who are interested in the more phenomenological aspects. The reader only needs limited background knowledge, and a clear line is followed throughout the book, making it easy to follow. The author presents extremely thorough but elementary discusssions of important physical questions, some of which seem to be an original way of addressing the subject.'

J. Zinn-Justin, *Physics World*

'This is a well-written book by one of the masters of the subject... it is certainly destined to become a standard text book and should find its way to the shelves of every physics library.'

J. Madore, *Classical and Quantum Gravity*

'The book starts out with an excellent historical introduction, not found anywhere else, giving citations to many by now classic papers... a valuable reference work as well as a textbook for graduate students.'

G. Roepstorff, *Zentralblatt für Mathematik*

From reviews of Volume II

'It is a majestic exposition. The two volumes are structured in a logical way. Everything is explained with incisive clarity. Weinberg always goes to the heart of any argument, and includes many things that cannot be found elsewhere in the literature. Often I find myself thinking: "Ah! Now I understand that properly."... I find it hard to imagine a better treatment of quantum field theory than Weinberg's. All serious students and researchers will want to have these volumes on their shelves.'

John C. Taylor, *Nature*

'Weinberg's *Modern Applications* goes to the boundaries of our present understanding of field theory. It is unmatched by any other book on quantum field theory for its depth, generality and definitive character, and it will be an essential reference for serious students and researchers in elementary particle physics.'

O. W. Greenberg, *Physics Today*

'... Steven Weinberg is one of our most gifted makers of theoretical tools as well as a virtuoso in their use. His new book conveys both the satisfaction of understanding nature and the feel of the atelier, for the "modern applications" of its subtitle include both the derivation of physical consequences and the development of new tools for understanding and applying field theory itself... *Modern Applications* is a splendid book, with abundant useful references to the original literature. It is a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

very interesting read from cover to cover, for the wholeness Weinberg's personal perspective gives to quantum field theory and particle physics.'

Chris Quigg, *Science*

'Experienced researchers and beginning graduate students will delight in the gems of wisdom to be found in these pages. This book combines exposition of technical detail with physical insight in a unique manner that confirms the promise of Volume I and I have no doubt that these two volumes will rapidly constitute the classic treatment of this important subject.'

M. B. Green, *CERN Courier*

'... a valued reference and a mine of useful information for professional field theorists.'

Tom Kibble, *New Scientist*

'... a clear presentation of the subject, explaining the underlying concepts in much depth and in an accessible style. I expect that these volumes will become the first source we turn to when trying to answer the challenging questions asked by bright postgraduates when they first encounter quantum field theory... I have no doubt that *The Quantum Theory of Fields* will soon be found on the bookshelves of most particle theorists, and that it will be one of the main sources used in the preparation of lectures on the subject for postgraduate students.'

C. T. C. Sachrajda, *The Times Higher Education Supplement*

'... Weinberg has produced a masterpiece that will be a standard reference on the field for a long time to come.'

B. E. Y. Svensson, *Elementa*

From reviews of Volume III

'... has produced a treatise that many of us had long awaited, perhaps without fully realizing it... with the publication of *The Quantum Theory of Fields*, Vol. III, has performed an analogous service for supersymmetry... Although this volume is the third in a trilogy, it is quite different from its two predecessors, and it stands on its own... May a new generation of students imbibe its content and spirit.'

Physics Today

'The third volume of *The Quantum Theory of Fields* is a self-contained introduction to the world of supersymmetry and supergravity. It will be useful both for experienced researchers in the field and for students who want to take the first steps towards learning about supersymmetry. Unlike other books in this field, it covers the wide spectrum of possible applications of supersymmetry in physics.'

Hans Peter Nilles, *Nature*

'Weinberg is of course one of the creators of modern quantum field theory, as well as of its physical culmination, the standard model of all (nongravitational) interactions. It is... very timely that this latest part of his monograph, devoted to supersymmetry and supergravity, has just appeared. As a text, it has been

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

pretested by Weinberg for a freestanding one-year graduate course; as a clear organizing reference to this extremely vast field, it will help the experts as well... Weinberg's style of presentation is as clear and meticulous as in his previous works.'

Stanley Deser, *Journal of General Relativity and Gravitation*

'Weinberg tries to be as elementary and clear as possible and steers clear of more sophisticated mathematical tools. Together with the previous volumes, this volume will serve as an invaluable reference to researchers and a textbook for graduate students.'

G. Roepstorff, *Zentralblatt MATH*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

The Quantum Theory of Fields

Volume II
Modern Applications

Steven Weinberg
University of Texas at Austin

CAMBRIDGE
UNIVERSITY PRESS

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo,
Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge, CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521670548

© S.Weinberg 1995, 2005

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1995
Eigth printing 2004
Paperback edition first published 2005
11th printing 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-55001-7 Volume 1 hardback
ISBN 978-0-521-67053-1 Volume 1 paperback
ISBN 978-0-521-55002-4 Volume 2 hardback
ISBN 978-0-521-67054-8 Volume 2 paperback
ISBN 978-0-521-66000-6 Volume 3 hardback
ISBN 978-0-521-67055-5 Volume 3 paperback
ISBN 978-0-521-67056-2 three-volume set

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Contents

Sections marked with an asterisk are somewhat out of the book's main line of development and may be omitted in a first reading.

PREFACE TO VOLUME II xvii

NOTATION xx

15 NON-ABELIAN GAUGE THEORIES 1

15.1 Gauge Invariance 2

Gauge transformations □ Structure constants □ Jacobi identity □ Adjoint representation □ Yang–Mills theory □ Covariant derivatives □ Field strength tensor □ Finite gauge transformations □ Analogy with general relativity

15.2 Gauge Theory Lagrangians and Simple Lie Groups 7

Gauge field Lagrangian □ Metric □ Antisymmetric structure constants □ Simple, semisimple, and $U(1)$ Lie algebras □ Structure of gauge algebra □ Compact algebras □ Coupling constants

15.3 Field Equations and Conservation Laws 12

Conserved currents □ Covariantly conserved currents □ Inhomogeneous field equations □ Homogeneous field equations □ Analogy with energy-momentum tensor □ Symmetry generators

15.4 Quantization 14

Primary and secondary first-class constraints □ Axial gauge □ Gribov ambiguity □ Canonical variables □ Hamiltonian □ Reintroduction of A_α^0 □ Covariant action □ Gauge invariance of the measure

15.5 The De Witt–Faddeev–Popov Method 19

Generalization of axial gauge results □ Independence of gauge fixing functionals □ Generalized Feynman gauge □ Form of vertices

15.6 Ghosts 24

Determinant as path integral □ Ghost and antighost fields □ Feynman rules for ghosts □ Modified action □ Power counting and renormalizability

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

15.7 BRST Symmetry 27

Auxiliary field h_α □ BRST transformation □ Nilpotence □ Invariance of new action □ BRST-cohomology □ Independence of gauge fixing □ Application to electrodynamics □ BRST-quantization □ Geometric interpretation

15.8 Generalizations of BRST Symmetry* 36

De Witt notation □ General Faddeev–Popov–De Witt theorem □ BRST transformations □ New action □ Slavnov operator □ Field-dependent structure constants □ Generalized Jacobi identity □ Invariance of new action □ Independence of gauge fixing □ Beyond quadratic ghost actions □ BRST quantization □ BRST cohomology □ Anti-BRST symmetry

15.9 The Batalin–Vilkovisky Formalism* 42

Open gauge algebras □ Antifields □ Master equation □ Minimal fields and trivial pairs □ BRST-transformations with antifields □ Antibrackets □ Anticanonical transformations □ Gauge fixing □ Quantum master equation

Appendix A A Theorem Regarding Lie Algebras 50

Appendix B The Cartan Catalog 54

Problems 58

References 59

16 EXTERNAL FIELD METHODS 63

16.1 The Quantum Effective Action 63

Currents □ Generating functional for all graphs □ Generating functional for connected graphs □ Legendre transformation □ Generating functional for one-particle-irreducible graphs □ Quantum-corrected field equations □ Summing tree graphs

16.2 Calculation of the Effective Potential 68

Effective potential for constant fields □ One loop calculation □ Divergences □ Renormalization □ Fermion loops

16.3 Energy Interpretation 72

Adiabatic perturbation □ Effective potential as minimum energy □ Convexity □ Instability between local minima □ Linear interpolation

16.4 Symmetries of the Effective Action 75

Symmetry and renormalization □ Slavnov–Taylor identities □ Linearly realized symmetries □ Fermionic fields and currents

Problems 78

References 78

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

17 RENORMALIZATION OF GAUGE THEORIES 80

17.1 The Zinn-Justin Equation 80

Slavnov–Taylor identities for BRST symmetry □ External fields $K_n(x)$ □ Antibrackets

17.2 Renormalization: Direct Analysis 82

Recursive argument □ BRST-symmetry condition on infinities □ Linearity in $K_n(x)$ □ New BRST symmetry □ Cancellation of infinities □ Renormalization constants □ Nonlinear gauge conditions

17.3 Renormalization: General Gauge Theories* 91

Are 'non-renormalizable' gauge theories renormalizable? □ Structural constraints □ Anticanonical change of variables □ Recursive argument □ Cohomology theorems

17.4 Background Field Gauge 95

New gauge fixing functions □ True and formal gauge invariance □ Renormalization constants

17.5 A One-Loop Calculation in Background Field Gauge 100

One-loop effective action □ Determinants □ Algebraic calculation for constant background fields □ Renormalization of gauge fields and couplings □ Interpretation of infinities

Problems 109

References 110

18 RENORMALIZATION GROUP METHODS 111

18.1 Where do the Large Logarithms Come From? 112

Singularities at zero mass □ 'Infrared safe' amplitudes and rates □ Jets □ Zero mass singularities from renormalization □ Renormalized operators

18.2 The Sliding Scale 119

Gell-Mann–Low renormalization □ Renormalization group equation □ One-loop calculations □ Application to ϕ^4 theory □ Field renormalization factors □ Application to quantum electrodynamics □ Effective fine structure constant □ Field-dependent renormalized couplings □ Vacuum instability

18.3 Varieties of Asymptotic Behavior 130

Singularities at finite energy □ Continued growth □ Fixed point at finite coupling □ Asymptotic freedom □ Lattice quantization □ Triviality □ Universal coefficients in the beta function

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

18.4 Multiple Couplings and Mass Effects 139

Behavior near a fixed point □ Invariant eigenvalues □ Nonrenormalizable theories □ Finite dimensional critical surfaces □ Mass renormalization at zero mass □ Renormalization group equations for masses

18.5 Critical Phenomena* 145

Low wave numbers □ Relevant, irrelevant, and marginal couplings □ Phase transitions and critical surfaces □ Critical temperature □ Behavior of correlation length □ Critical exponent □ $4-\epsilon$ dimensions □ Wilson–Fisher fixed point □ Comparison with experiment □ Universality classes

18.6 Minimal Subtraction 148

Definition of renormalized coupling □ Calculation of beta function □ Application to electrodynamics □ Modified minimal subtraction □ Non-renormalizable interactions

18.7 Quantum Chromodynamics 152

Quark colors and flavors □ Calculation of beta function □ Asymptotic freedom □ Quark and gluon trapping □ Jets □ e^+–e^- annihilation into hadrons □ Accidental symmetries □ Non-renormalizable corrections □ Behavior of gauge coupling □ Experimental results for g_s and Λ

18.8 Improved Perturbation Theory* 157

Leading logarithms □ Coefficients of logarithms

Problems 158

References 159

19 SPONTANEOUSLY BROKEN GLOBAL SYMMETRIES 163

19.1 Degenerate Vacua 163

Degenerate minima of effective potential □ Broken symmetry or symmetric superpositions? □ Large systems □ Factorization at large distances □ Diagonalization of vacuum expectation values □ Cluster decomposition

19.2 Goldstone Bosons 167

Broken global symmetries imply massless bosons □ Proof using effective potential □ Proof using current algebra □ F factors and vacuum expectation values □ Interactions of soft Goldstone bosons

19.3 Spontaneously Broken Approximate Symmetries 177

Pseudo-Goldstone bosons □ Tadpoles □ Vacuum alignment □ Mass matrix □ Positivity

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

19.4 Pions as Goldstone Bosons 182

$SU(2) \times SU(2)$ chiral symmetry of quantum chromodynamics □ Breakdown to isospin □ Vector and axial-vector weak currents □ Pion decay amplitude □ Axial form factors of nucleon □ Goldberger-Treiman relation □ Vacuum alignment □ Quark and pion masses □ Soft pion interactions □ Historical note

19.5 Effective Field Theories: Pions and Nucleons 192

Current algebra for two soft pions □ Current algebra justification for effective Lagrangian □ σ-model □ Transformation to derivative coupling □ Nonlinear realization of $SU(2) \times SU(2)$ □ Effective Lagrangian for soft pions □ Direct justification of effective Lagrangian □ General effective Lagrangian for pions □ Power counting □ Pion–pion scattering for massless pions □ Identification of F-factor □ Pion mass terms in effective Lagrangian □ Pion–pion scattering for real pions □ Pion–pion scattering lengths □ Pion–nucleon effective Lagrangian □ Covariant derivatives □ $g_A \neq 1$ □ Power counting with nucleons □ Pion–nucleon scattering lengths □ σ-terms □ Isospin violation □ Adler–Weisberger sum rule

19.6 Effective Field Theories: General Broken Symmetries 211

Transformation to derivative coupling □ Goldstone bosons and right cosets □ Symmetric spaces □ Cartan decomposition □ Nonlinear transformation rules □ Uniqueness □ Covariant derivatives □ Symmetry breaking terms □ Application to quark mass terms □ Power counting □ Order parameters

19.7 Effective Field Theories: $SU(3) \times SU(3)$ 225

$SU(3)$ multiplets and matrices □ Goldstone bosons of broken $SU(3) \times SU(3)$ □ Quark mass terms □ Pseudoscalar meson masses □ Electromagnetic corrections □ Quark mass ratios □ Higher terms in Lagrangian □ Nucleon mass shifts

19.8 Anomalous Terms in Effective Field Theories* 234

Wess–Zumino–Witten term □ Five-dimensional form □ Integer coupling □ Uniqueness and de Rham cohomology

19.9 Unbroken Symmetries 238

Persistent mass conjecture □ Vafa–Witten proof □ Small non-degenerate quark masses

19.10 The $U(1)$ Problem 243

Chiral $U(1)$ symmetry □ Implications for pseudoscalar masses

Problems 246

References 247

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

20 OPERATOR PRODUCT EXPANSIONS 252

20.1 The Expansion: Description and Derivation 253

Statement of expansion □ Dominance of simple operators □ Path-integral derivation

20.2 Momentum Flow* 255

ϕ^2 contribution for two large momenta □ Renormalized operators □ Integral equation for coefficient function □ ϕ^2 contribution for many large momenta

20.3 Renormalization Group Equations for Coefficient Functions 263

Derivation and solution □ Behavior for fixed points □ Behavior for asymptotic freedom

20.4 Symmetry Properties of Coefficient Functions 265

Invariance under spontaneously broken symmetries

20.5 Spectral Function Sum Rules 266

Spectral functions defined □ First, second, and third sum rules □ Application to chiral $SU(N) \times SU(N)$ □ Comparison with experiment

20.6 Deep Inelastic Scattering 272

Form factors W_1 and W_2 □ Deep inelastic differential cross section □ Bjorken scaling □ Parton model □ Callan–Gross relation □ Sum rules □ Form factors T_1 and T_2 □ Relation between T_r and W_r □ Symmetric tensor operators □ Twist □ Operators of minimum twist □ Calculation of coefficient functions □ Sum rules for parton distribution functions □ Altarelli–Parisi differential equations □ Logarithmic corrections to Bjorken scaling

20.7 Renormalons* 283

Borel summation of perturbation theory □ Instanton and renormalon obstructions □ Instantons in massless ϕ^4 theory □ Renormalons in quantum chromodynamics

Appendix Momentum Flow: The General Case 288

Problems 292

References 293

21 SPONTANEOUSLY BROKEN GAUGE SYMMETRIES 295

21.1 Unitarity Gauge 295

Elimination of Goldstone bosons □ Vector boson masses □ Unbroken symmetries and massless vector bosons □ Complex representations □ Vector field propagator □ Continuity for vanishing gauge couplings

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

21.2 Renormalizable ξ-Gauges 300

Gauge fixing function □ Gauge-fixed Lagrangian □ Propagators

21.3 The Electroweak Theory 305

Lepton-number preserving symmetries □ $SU(2) \times U(1)$ □ $W^{\pm}$, Z^0, and photons □ Mixing angle □ Lepton-vector boson couplings □ $W^{\pm}$ and Z^0 masses □ Muon decay □ Effective fine structure constant □ Discovery of neutral currents □ Quark currents □ Cabibbo angle □ c quark □ Third generation □ Kobayashi–Maskawa matrix □ Discovery of $W^{\pm}$ and Z^0 □ Precise experimental tests □ Accidental symmetries □ Nonrenormalizable corrections □ Lepton nonconservation and neutrino masses □ Baryon nonconservation and proton decay

21.4 Dynamically Broken Local Symmetries* 318

Fictitious gauge fields □ Construction of Lagrangian □ Power counting □ General mass formula □ Example: $SU(2) \times SU(2)$ □ Custodial $SU(2) \times SU(2)$ □ Technicolor

21.5 Electroweak–Strong Unification 327

Simple gauge groups □ Relations among gauge couplings □ Renormalization group flow □ Mixing angle and unification mass □ Baryon and lepton nonconservation

21.6 Superconductivity* 332

$U(1)$ broken to Z_2 □ Goldstone mode □ Effective Lagrangian □ Conservation of charge □ Meissner effect □ Penetration depth □ Critical field □ Flux quantization □ Zero resistance □ ac Josephson effect □ Landau–Ginzburg theory □ Correlation length □ Vortex lines □ $U(1)$ restoration □ Stability □ Type I and II superconductors □ Critical fields for vortices □ Behavior near vortex center □ Effective theory for electrons near Fermi surface □ Power counting □ Introduction of pair field □ Effective action □ Gap equation □ Renormalization group equations □ Conditions for superconductivity

Appendix General Unitarity Gauge 352

Problems 353

References 354

22 ANOMALIES 359

22.1 The π^0 Decay Problem 359

Rate for $\pi^0 \to 2\gamma$ □ Naive estimate □ Suppression by chiral symmetry □ Comparison with experiment

22.2 Transformation of the Measure: The Abelian Anomaly 362

Chiral and non-chiral transformations □ Anomaly function □ Chern–Pontryagin density □ Nonconservation of current □ Conservation of gauge-non-invariant

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

current □ Calculation of $\pi^0 \to 2\gamma$ □ Euclidean calculation □ Atiyah–Singer index theorem

22.3 Direct Calculation of Anomalies: The General Case 370

Fermion non-conserving currents □ Triangle graph calculation □ Shift vectors □ Symmetric anomaly □ Bardeen form □ Adler–Bardeen theorem □ Massive fermions □ Another approach □ Global anomalies

22.4 Anomaly-Free Gauge Theories 383

Gauge anomalies must vanish □ Real and pseudoreal representations □ Safe groups □ Anomaly cancellation in standard model □ Gravitational anomalies □ Hypercharge assignments □ Another $U(1)$?

22.5 Massless Bound States* 389

Composite quarks and leptons? □ Unbroken chiral symmetries □ 't Hooft anomaly matching conditions □ Anomaly matching for unbroken chiral $SU(n) \times SU(n)$ with $SU(N)$ gauge group □ The case $N = 3$ □ □ Chiral $SU(3) \times SU(3)$ must be broken □ 't Hooft decoupling condition □ Persistent mass condition

22.6 Consistency Conditions 396

Wess–Zumino conditions □ BRST cohomology □ Derivation of symmetric anomaly □ Descent equations □ Solution of equations □ Schwinger terms □ Anomalies in Zinn-Justin equation □ Antibracket cohomology □ Algebraic proof of anomaly absence for safe groups

22.7 Anomalies and Goldstone Bosons 408

Anomaly matching □ Solution of anomalous Slavnov–Taylor identities □ Uniqueness □ Anomalous Goldstone boson interactions □ The case $SU(3) \times SU(3)$ □ Derivation of Wess–Zumino–Witten interaction □ Evaluation of integer coefficient □ Generalization

Problems 416

References 417

23 EXTENDED FIELD CONFIGURATIONS 421

23.1 The Uses of Topology 422

Topological classifications □ Homotopy □ Skyrmions □ Derrick's theorem □ Domain boundaries □ Bogomol'nyi inequality □ Cosmological problems □ Instantons □ Monopoles and vortex lines □ Symmetry restoration

23.2 Homotopy Groups 430

Multiplication rule for $\pi_1(\mathcal{M})$ □ Associativity □ Inverses □ $\pi_1(S_1)$ □ Topological conservation laws □ Multiplication rule for $\pi_k(\mathcal{M})$ □ Winding number

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

23.3 Monopoles 436

$SU(2)/U(1)$ model □ Winding number □ Electromagnetic field □ Magnetic monopole moment □ Kronecker index □ 't Hooft–Polyakov monopole □ Another Bogomol'nyi inequality □ BPS monopole □ Dirac gauge □ Charge quantization □ $G/(H' \times U(1))$ monopoles □ Cosmological problems □ Monopole–particle interactions □ G/H monopoles with G not simply connected □ Irrelevance of field content

23.4 The Cartan–Maurer Integral Invariant 445

Definition of the invariant □ Independence of coordinate system □ Topological invariance □ Additivity □ Integral invariant for $S_1 \mapsto U(1)$ □ Bott's theorem □ Integral invariant for $S_3 \mapsto SU(2)$

23.5 Instantons 450

Evaluation of Cartan–Maurer invariant □ Chern–Pontryagin density □ One more Bogomol'nyi inequality □ $\nu = 1$ solution □ General winding number □ Solution of $U(1)$ problem □ Baryon and lepton non-conservation by electroweak instantons □ Minkowskian approach □ Barrier penetration □ Thermal fluctuations

23.6 The Theta Angle 455

Cluster decomposition □ Superposition of winding numbers □ P and CP non-conservation □ Complex fermion masses □ Suppression of P and CP non-conservation by small quark masses □ Neutron electric dipole moment □ Peccei–Quinn symmetry □ Axions □ Axion mass □ Axion interactions

23.7 Quantum Fluctuations around Extended Field Configurations 462

Fluctuations in general □ Collective parameters □ Determinental factor □ Coupling constant dependence □ Counting collective parameters

23.8 Vacuum Decay 464

False and true vacua □ Bounce solutions □ Four dimensional rotational invariance □ Sign of action □ Decay rate per volume □ Thin wall approximation

Appendix A Euclidean Path Integrals 468

Appendix B A List of Homotopy Groups 472

Problems 473

References 474

AUTHOR INDEX 478

SUBJECT INDEX 484

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

OUTLINE OF VOLUME I

1 HISTORICAL INTRODUCTION

2 RELATIVISTIC QUANTUM MECHANICS

3 SCATTERING THEORY

4 THE CLUSTER DECOMPOSITION PRINCIPLE

5 QUANTUM FIELDS AND ANTIPARTICLES

6 THE FEYNMAN RULES

7 THE CANONICAL FORMALISM

8 ELECTRODYNAMICS

9 PATH-INTEGRAL METHODS

10 NON-PERTURBATIVE METHODS

11 ONE-LOOP RADIATIVE CORRECTIONS IN QUANTUM ELECTRODYNAMICS

12 GENERAL RENORMALIZATION THEORY

13 INFRARED EFFECTS

14 BOUND STATES IN EXTERNAL FIELDS

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Preface To Volume II

This volume describes the advances in the quantum theory of fields that have led to an understanding of the electroweak and strong interactions of the elementary particles. These interactions have all turned out to be governed by principles of gauge invariance, so we start here in Chapters 15–17 with gauge theories, generalizing the familiar gauge invariance of electrodynamics to non-Abelian Lie groups.

Some of the most dramatic aspects of gauge theories appear at high energy, and are best studied by the methods of the renormalization group. These methods are introduced in Chapter 18, and applied to quantum chromodynamics, the modern non-Abelian gauge theory of strong interactions, and also to critical phenomena in condensed matter physics. Chapter 19 deals with general spontaneously broken global symmetries, and their application to the broken approximate $SU(2) \times SU(2)$ and $SU(3) \times SU(3)$ symmetries of quantum chromodynamics. Both the renormalization group method and broken symmetries find some of their most interesting applications in the context of operator product expansions, discussed in Chapter 20.

The key to the understanding of the electroweak interactions is the spontaneous breaking of gauge symmetries, which are explored in Chapter 21 and applied to superconductivity as well as to the electroweak interactions. Quite apart from spontaneous symmetry breaking is the possibility of symmetry breaking by quantum-mechanical effects known as anomalies. Anomalies and various of their physical implications are presented in Chapter 22. This volume concludes with a discussion in Chapter 23 of extended field configurations, which can arise either as new ingredients in physical states, such as skyrmions, monopoles, or vortex lines, or as non-perturbative quantum corrections to path integrals, where anomalies play a crucial role.

It would not be possible to provide a coherent account of these developments if they were presented in a historical order. I have chosen instead to describe the material of this book in an order that seems to me to work best pedagogically — I introduce each topic at a point where

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.001

the motivation as well as the mathematics can be understood with the least possible reference to material in subsequent chapters, even where logic might suggest a somewhat different order. For instance, instead of having one long chapter to introduce non-Abelian gauge theories, this material is split between Chapters 15 and 17, because Chapter 15 provides a motivation for the external field formalism introduced in Chapter 16, and this formalism is necessary for the work of Chapter 17.

In the course of this presentation, the reader will be introduced to various formal devices, including BRST invariance, the quantum effective action, and homotopy theory. The Batalin–Vilkovisky formalism is presented as an optional side track. It is introduced in Chapter 15 as a compact way of formulating gauge theories, whether based on open or closed symmetry algebras, and then used in Chapter 17 to study the cancellation of infinities in 'non-renormalizable' gauge theories, including general relativity, and in Chapter 22 to show that certain gauge theories are anomaly-free to all orders of perturbation theory. The effective field theory approach is extensively used in this volume, especially in applications to theories with broken symmetry, including the theory of superconductivity. I have struggled throughout for the greatest possible clarity of presentation, taking time to show detailed calculations where I thought it might help the reader, and dropping topics that could not be clearly explained in the space available.

The guiding aim of both Volumes I and II of this book is to explain to the reader why quantum field theory takes the form it does, and why in this form it does such a good job of describing the real world. Volume I outlined the foundations of the quantum theory of fields, emphasizing the reasons why nature is described at accessible energies by effective quantum field theories, and in particular by gauge theories. (A list of chapters of Volume I is given at the end of the table of contents of this volume.) The present volume takes quantum field theory and gauge invariance as its starting points, and concentrates on their implications.

This volume should be accessible to readers who have some familiarity with the fundamentals of quantum field theory. It is not assumed that the reader is familiar with Volume I (though it wouldn't hurt). Aspects of group theory and topology are explained where they are introduced.

Some of the formal methods described in this volume (such as BRST invariance and the renormalization group) have important applications in speculative theories that involve supersymmetry or superstrings. I am enthusiastic about the future prospects of these theories, but I have not included them in this book, because it seems to me that they require a whole book to themselves. (Perhaps supersymmetry and supergravity will be the subjects of a Volume III.) I have excluded some other interesting topics here, such as finite temperature field theory, lattice gauge calcula-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.001

tions and the large N_c approximation, because they were not needed to provide either motivation or mathematical techniques for the rest of the book, and the book was long enough.

The great volume of the literature on quantum field theory and its applications makes it impossible for me to read or quote all relevant articles. I have tried to supply citations to the classic papers on each topic, as well as to papers that describe further developments of material covered here, and to references that present detailed calculations, data, or proofs referred to in the text. As before, the mere absence of a citation should not be interpreted as a claim that the material presented is original, but some of it is.

In my experience this volume provides enough material for a one-year course for graduate students on advanced topics in quantum field theory, or on elementary particle physics. Selected parts of Volumes I and II would be suitable as the basis of a compressed one-year course on both the foundations and the modern applications of quantum field theory. I have supplied problems for each chapter. Some of these problems aim simply at providing exercise in the use of techniques described in the chapter; others are intended to suggest extensions of the results of the chapter to a wider class of theories.

* * *

I must acknowledge my special intellectual debt to colleagues at the University of Texas, notably Luis Boya, Phil Candelas, Bryce and Cecile De Witt, Willy Fischler, Joaquim Gomis, and Vadim Kaplunovsky, and especially Jacques Distler. Also, Luis Alvarez-Gaumè, Sidney Coleman, John Dixon, Tony Duncan, Jürg Fröhlich, Arthur Jaffe, Marc Henneaux, Roman Jackiw, Joe Polchinski, Michael Tinkham, Cumrun Vafa, Don Weingarten, Edward Witten and Bruno Zumino gave valuable help with special topics. Jonathan Evans read through the manuscript of this volume, and made many valuable suggestions. For corrections to the first printing of this volume I am indebted to several students and colleagues, including Stephen Adler, Mark Byrd, Vincent Liu, Herbert Neuberger, Chun-yen Wang, and especially Michio Masujima. Thanks are due to Alyce Wilson, who prepared the illustrations and typed the LaTeX input files until I learned how to do it, to Terry Riley for finding countless books and articles, and to Jan Duffy for many helps. I am grateful to Maureen Storey and Alison Woollatt of Cambridge University Press for working to ready this book for publication, and especially to my editor, Rufus Neal, for his continued friendly good advice.

STEVEN WEINBERG

Austin, Texas
December, 1995

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.001

Notation

Latin indices i, j, k, and so on generally run over the three spatial coordinate labels, usually taken as 1, 2, 3. Where specifically indicated, they run over values 1, 2, 3, 4, with $x^4 = it$.

Greek indices μ, ν, etc. from the middle of the Greek alphabet generally run over the four spacetime coordinate labels 1, 2, 3, 0, with x^0 the time coordinate.

Greek indices α, β, etc. from the beginning of the Greek alphabet generally run over the generators of a symmetry algebra.

Repeated indices are generally summed, unless otherwise indicated.

The spacetime metric $\eta_{\mu\nu}$ is diagonal, with elements $\eta_{11} = \eta_{22} = \eta_{33} = 1$, $\eta_{00} = -1$.

The d'Alembertian is defined as $\Box \equiv \eta^{\mu\nu}\partial^2/\partial x^\mu \partial x^\nu = \nabla^2 - \partial^2/\partial t^2$, where ∇^2 is the Laplacian $\partial^2/\partial x^i \partial x^i$.

The 'Levi–Civita tensor' $\epsilon^{\mu\nu\rho\sigma}$ is defined as the totally antisymmetric quantity with $\epsilon^{0123} = +1$.

Spatial three-vectors are indicated by letters in boldface.

Three-vectors in isospin space are indicated by arrows.

A hat over any vector indicates the corresponding unit vector: Thus, $\hat{\mathbf{v}} \equiv \mathbf{v}/|\mathbf{v}|$.

A dot over any quantity denotes the time-derivative of that quantity.

Dirac matrices γ_μ are defined so that $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\eta_{\mu\nu}$. Also, $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$, and $\beta = i\gamma^0 = \gamma_4$.

The step function $\theta(s)$ has the value $+1$ for $s > 0$ and 0 for $s < 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.002

The complex conjugate, transpose, and Hermitian adjoint of a matrix or vector A are denoted A^*, A^{T}, and $A^\dagger = A^{*\mathrm{T}}$, respectively. The Hermitian adjoint of an operator O is denoted $O^\dagger$, except where an asterisk is used to emphasize that a vector or matrix of operators is not transposed. +H.c. or +c.c. at the end of an expression indicates the addition of the Hermitian adjoint or complex conjugate of the foregoing terms. A bar on a Dirac spinor u is defined by $\bar{u} = u^\dagger\beta$. The antifield of a field χ in the Batalin–Vilkovisky formalism is denoted $\chi^\ddagger$ rather than χ^* to distinguish it from the ordinary complex conjugate or the antiparticle field.

Units are usually used with $\hbar$ and the speed of light taken to be unity. Throughout $-e$ is the rationalized charge of the electron, so that the fine structure constant is $\alpha = e^2/4\pi \simeq 1/137$.

Numbers in parenthesis at the end of quoted numerical data give the uncertainty in the last digits of the quoted figure. Where not otherwise indicated, experimental data are taken from 'Review of Particle Properties,' *Phys. Rev.* **D50**, 1173 (1994).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.002

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.002

15

Non-Abelian Gauge Theories

The quantum field theories that have proved successful in describing the real world are all non-Abelian gauge theories, theories based on principles of gauge invariance more general than the simple $U(1)$ gauge invariance of quantum electrodynamics. These theories share with electrodynamics the attractive feature, outlined at the end of Section 8.1, that the existence and some of the properties of the gauge fields follow from a principle of invariance under local gauge transformations. In electrodynamics, fields $\psi_n(x)$ of charge e_n undergo the gauge transformation $\psi_n(x) \rightarrow \exp(ie_n\Lambda(x))\psi_n(x)$ with arbitrary $\Lambda(x)$. Since $\partial_\mu\psi_n(x)$ does not transform like $\psi_n(x)$, we must introduce a field $A_\mu(x)$ with the gauge transformation property $A_\mu(x) \rightarrow A_\mu(x)+\partial_\mu\Lambda(x)$, and use it to construct a gauge-covariant derivative $\partial_\mu\psi_n(x)-ie_nA_\mu(x)\psi_n(x)$, which transforms just like $\psi_n(x)$ and can therefore be used with $\psi_n(x)$ to construct a gauge-invariant Lagrangian. In a similar way, the existence and some of the properties of the gravitational field $g_{\mu\nu}(x)$ in general relativity follow from a symmetry principle, under general coordinate transformations.* Given these distinguished precedents, it was natural that local gauge invariance should be extended to invariance under local non-Abelian gauge transformations.

In the original 1954 work of Yang and Mills,[1] the non-Abelian gauge group was taken to be the $SU(2)$ group of isotopic spin rotations, and the vector fields analogous to the photon field were interpreted as the fields of strongly-interacting vector mesons of isotopic spin unity. This proposal immediately encountered the obstacle that these vector mesons would have to have zero mass, like photons, and it seemed that any such particles would already have been detected. Another problem was that, like all strong-interaction theories at that time, there was nothing that

* Of course, both local gauge invariance and general covariance can be realized in a trivial way, by taking $A_\mu(x)$ and $g_{\mu\nu}(x)$ to be non-dynamical c-number functions that simply characterize a choice of phase or coordinate system, respectively. These symmetries become physically significant when we treat $A_\mu(x)$ and $g_{\mu\nu}(x)$ as dynamical fields, over which we integrate in calculating S-matrix elements.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

could be done with it; it seemed that the large coupling constant of the theory would preclude any use of perturbation theory.

Gauge theories were soon generalized to arbitrary non-Abelian gauge groups,[2] and their quantization continued to be studied mathematically, notably by Feynman,[3] Faddeev and Popov,[4] and De Witt,[5] in part as a warming up exercise for the harder problem of quantizing general relativity. They showed that the naive Feynman rules obtained by simply inspecting the Lagrangian need to be supplemented by additional 'ghost' loops. However, the physical relevance of these theories did not begin to be understood until the late 1960s. It eventually turned out that *all* of the observed interactions of elementary particles are generated by vector fields associated with local gauge symmetries; the corresponding spin 1 particles are either very heavy, as a result of a spontaneous breakdown of the gauge symmetry, or are 'trapped', as a result of the rise of the coupling constant at long distances. These matters will be the subjects respectively of Chapters 21 and 18. In this chapter we shall explore the formulation of the non-Abelian gauge theories, and the derivation of their Feynman rules.

15.1 Gauge Invariance

We assume that the Lagrangian of our theory is invariant under a set of infinitesimal transformations on the matter fields $\psi_\ell(x)$

$$\delta\psi_\ell(x) = i\,\epsilon^\alpha(x)(t_\alpha)_\ell{}^m\psi_m(x)\,, \tag{15.1.1}$$

with some set of independent constant matrices** t_α, and with real infinitesimal parameters $\epsilon^\alpha(x)$ which (as for gauge transformations in electrodynamics) are allowed to depend on position in spacetime. We assume that these symmetry transformations are the infinitesimal part of a Lie group; as shown in Section 2.2, this requires that the t_α obey commutation relations

$$[t_\alpha, t_\beta] = i\,C^\gamma{}_{\alpha\beta}t_\gamma\,, \tag{15.1.2}$$

where $C^\gamma{}_{\alpha\beta}$ are a set of real constants, known as the structure constants of the group. The antisymmetry of the commutator immediately tells us

** In this book we shall generally label symmetry generators with letters α, β, etc. from the beginning of the Greek alphabet, in order to keep these labels distinct from the indices μ, ν, etc. from the middle of the Greek alphabet that are used to label spacetime coordinates. Later, in dealing with broken symmetries, we will often use letters a, b, etc. from the beginning of the Latin alphabet to label generators of spontaneously broken symmetries, and letters i, j, etc. from the middle of the Latin alphabet to label generators of unbroken symmetries.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

that the structure constants are similarly antisymmetric:

$$C^{\gamma}{}_{\alpha\beta} = -C^{\gamma}{}_{\beta\alpha} \,. \tag{15.1.3}$$

Also, from the Jacobi identity

$$0 = \Big[[t_\alpha, t_\beta], t_\gamma\Big] + \Big[[t_\gamma, t_\alpha], t_\beta\Big] + \Big[[t_\beta, t_\gamma], t_\alpha\Big] \tag{15.1.4}$$

we see that the Cs satisfy the further constraint

$$0 = C^{\delta}{}_{\alpha\beta} C^{\epsilon}{}_{\delta\gamma} + C^{\delta}{}_{\gamma\alpha} C^{\epsilon}{}_{\delta\beta} + C^{\delta}{}_{\beta\gamma} C^{\epsilon}{}_{\delta\alpha} \,. \tag{15.1.5}$$

Any set of constants $C^{\gamma}{}_{\alpha\beta}$ that satisfy Eqs. (15.1.3) and (15.1.5) define at least one set of matrices $t^{\mathrm{A}}{}_{\alpha}$:

$$(t^{\mathrm{A}}{}_{\alpha})^{\beta}{}_{\gamma} \equiv -i\, C^{\beta}{}_{\gamma\alpha} \,, \tag{15.1.6}$$

that satisfy the commutation relations (15.1.2) with structure constants $C^{\gamma}{}_{\alpha\beta}$:

$$[t^{\mathrm{A}}{}_{\alpha}, t^{\mathrm{A}}{}_{\beta}] = i\, C^{\gamma}{}_{\alpha\beta} t^{\mathrm{A}}{}_{\gamma} \,. \tag{15.1.7}$$

This is known as the 'adjoint' (or 'regular') representation of the Lie algebra with structure constants $C^{\alpha}{}_{\beta\gamma}$.

For example, in the original Yang–Mills theory, the matter fields were the doublet consisting of proton and neutron fields ψ_p and ψ_n:

$$\psi = \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix}$$

and the t_α with $\alpha = 1, 2, 3$ were the isospin matrices

$$t_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad t_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad t_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

These satisfy the commutation relations (15.1.2) with

$$C^{\gamma}{}_{\alpha\beta} = \epsilon_{\gamma\alpha\beta} \,,$$

where as usual $\epsilon_{\gamma\alpha\beta}$ is $+1$ or -1 if γ, α, β is an even or an odd permutation of 1, 2, 3, respectively, and vanishes otherwise. We recognize this as the same as the Lie algebra (2.4.18) of the three-dimensional rotation group; the matrices t_α here furnish what we recognize as the spin 1/2 representation of this Lie algebra. The matrices (15.1.6) of the adjoint representation are here (in a basis with rows and columns labelled 1,2,3):

$$t_1^A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix} , \qquad t_2^A = \begin{bmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{bmatrix} , \qquad t_3^A = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

This is the spin 1 representation of the Lie algebra of the rotation group.

Now consider what is needed to make the Lagrangian invariant under the transformations (15.1.1). If there were no derivatives acting on the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

fields, the task would be easy — any function of the matter fields that was invariant under the transformation (15.1.1) with ϵ^α constant would also be invariant with ϵ^α arbitrary real functions of the spacetime coordinates. This is not the case if the Lagrangian involves derivatives of the fields (as it must), because with position-dependent functions $\epsilon^\alpha(x)$, the derivatives of the matter fields do not transform like the fields themselves. Differentiating Eq. (15.1.1) gives

$$\delta\Big(\partial_\mu\psi_\ell(x)\Big) = i\,\epsilon^\alpha(x)(t_\alpha)_\ell{}^m\Big(\partial_\mu\psi_m(x)\Big) + i\Big(\partial_\mu\epsilon^\alpha(x)\Big)(t_\alpha)_\ell{}^m\psi_m(x)\,. \quad (15.1.8)$$

To make the Lagrangian invariant, we need a field $A^\alpha{}_\mu$, whose transformation rule involves a term $\partial_\mu\epsilon^\alpha$, which can be used to cancel the second term in Eq. (15.1.8). Since this field carries an α-index, we would expect it also to undergo a matrix transformation like Eq. (15.1.1), but with t_α replaced with the adjoint representation matrices (15.1.6). Let us therefore tentatively take the transformation relation of these new 'gauge' fields as

$$\delta A^\beta{}_\mu = \partial_\mu\epsilon^\beta \;+\; i\,\epsilon^\alpha(t^{\rm A}{}_\alpha)^\beta{}_\gamma A^\gamma{}_\mu$$

or, using Eq. (15.1.6),

$$\delta A^\beta{}_\mu = \partial_\mu\,\epsilon^\beta \;+\; C^\beta{}_{\gamma\alpha}\epsilon^\alpha A^\gamma{}_\mu\,. \quad (15.1.9)$$

This allows us to construct a 'covariant derivative':†

$$(D_\mu\psi(x))_\ell = \partial_\mu\psi_\ell(x) - i\,A^\beta{}_\mu(x)(t_\beta)_\ell{}^m\psi_m(x)\,. \quad (15.1.10)$$

As planned, the term $\partial_\mu\epsilon^\beta$ in the transformation of $A^\beta{}_\mu$ in the second term of Eq. (15.1.10) cancels the term proportional to $\partial_\mu\epsilon^\beta$ in the transformation of the first term, leaving us with

$$\begin{aligned}\delta\,(D_\mu\,\psi)_\ell = {} & i\,\epsilon^\alpha(t_\alpha)_\ell{}^m\partial_\mu\psi_m - i\,C^\beta{}_{\gamma\alpha}\,\epsilon^\alpha\,A^\gamma{}_\mu(t_\beta)_\ell{}^m\psi_m \\ & + A^\gamma{}_\mu\,(t_\gamma)_\ell{}^m(t_\alpha)_m{}^n\psi_n\,\epsilon^\alpha\end{aligned}$$

or, using Eq. (15.1.2),

$$\delta(D_\mu\psi)_\ell \;=\; i\,\epsilon^\alpha\,(t_\alpha)_\ell{}^m\,(D_\mu\psi)_m\,, \quad (15.1.11)$$

so that $D_\mu\psi$ transforms just like ψ itself.

We also need to worry about derivatives of the gauge field. In order to eliminate the term $\partial_\nu\partial_\mu\epsilon^\beta$ in the transformation of $\partial_\nu A^\beta{}_\mu$, we antisymmetrize with respect to μ and ν, just as in electrodynamics. However, we still have terms in the transformation of $\partial_\nu A^\beta{}_\mu - \partial_\mu A^\beta{}_\nu$ proportional to first derivatives of $\epsilon(x)$, arising from the second term in Eq. (15.1.9). The easiest way to construct a 'covariant curl', $F^\gamma{}_{\nu\mu}$ in whose transformation

† As discussed in the next section, in writing Eq. (15.1.10) we are tacitly supposing that any coupling-constant factors like the electric charge are included in the t_β, and hence also in the structure constants.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

rule all such derivatives of $\epsilon(x)$ cancel, is to consider the commutator of two covariant derivatives acting on a matter field ψ:

$$\Big([D_\nu, D_\mu]\psi\Big)_\ell = -i\,(t_\gamma)_\ell{}^m F^\gamma{}_{\nu\mu}\psi_m\,, \tag{15.1.12}$$

where

$$F^\gamma{}_{\nu\mu} \equiv \partial_\nu A^\gamma{}_\mu - \partial_\mu A^\gamma{}_\nu + C^\gamma{}_{\alpha\beta}A^\alpha{}_\nu A^\beta{}_\mu\,. \tag{15.1.13}$$

Eq. (15.1.12) makes it obvious that $F^\gamma{}_{\nu\mu}$ must transform just like a matter field that happens to belong to the adjoint representation:

$$\delta F^\beta{}_{\nu\mu} \equiv i\,\epsilon^\alpha (t^{\rm A}{}_\alpha)^\beta{}_\gamma F^\gamma{}_{\nu\mu} = \epsilon^\alpha C^\beta{}_{\gamma\alpha}F^\gamma{}_{\nu\mu}\,. \tag{15.1.14}$$

The reader may check by direct calculation (using the relation (15.1.5)) that the quantity $F^\alpha{}_{\nu\mu}$ defined in (15.1.13) actually has the simple transformation rule (15.1.14).

For some purposes, it is useful to know that these infinitesimal gauge transformations can be upgraded to finite transformations. A group element can be parameterized by a set of real functions $\Lambda^\alpha(x)$ so that it acts on a general matter field $\psi_\ell(x)$ through the matrix transformation

$$\psi_\ell(x) \to \psi_{\ell\Lambda}(x) = \Big[\exp\Big(i\,t_\alpha\Lambda^\alpha(x)\Big)\Big]_\ell{}^m \psi_m(x)\,. \tag{15.1.15}$$

We want the covariant derivative to transform in the same way:

$$(\partial_\mu - i\,t_\alpha A^\alpha{}_{\mu\Lambda})\psi_\Lambda = \exp(i\,t_\alpha\Lambda^\alpha)\,(\partial_\mu - i\,t_\alpha A^\alpha_\mu)\psi\,, \tag{15.1.16}$$

so we must impose on $A^\alpha{}_\mu$ the transformation rule $A^\alpha_\mu \to A^\alpha_{\mu\Lambda}$, with

$$\partial_\mu \exp(i\,t_\beta\Lambda^\beta) - i\,t_\beta \exp(i\,t_\alpha\Lambda^\alpha)\,A^\beta_{\mu\Lambda} = -i\exp(i\,t_\alpha\Lambda^\alpha)t_\beta A^\beta$$

or in other words

$$t_\alpha A^\alpha{}_{\mu\Lambda} = \exp(i\,t_\beta\Lambda^\beta)\,t_\alpha A^\alpha{}_\mu \exp(-it_\beta\Lambda^\beta) - i\,\Big[\partial_\mu \exp(it_\beta\Lambda^\beta)\Big]\exp(-it_\beta\Lambda^\beta)\,. \tag{15.1.17}$$

Eqs. (15.1.15) and (15.1.17) reduce to the previous transformation rules (15.1.1) and (15.1.9) in the limit where $\Lambda^\alpha(x)$ is an infinitesimal $\epsilon^\alpha(x)$.

From Eq. (15.1.17), we can see that by a suitable choice of $\Lambda^\beta(x)$, it is always possible to make $A^\alpha{}_{\mu\Lambda}(x)$ vanish at any *one* point, say $x = z$. (Simply take $\Lambda^\alpha(z)$ to vanish, and $\partial\Lambda^\alpha(x)/\partial x^\mu = -A^\alpha{}_\mu(x)$ at $x = z$.) Also, it is always possible to choose $\Lambda^\beta(x)$ so that any *one* spacetime component of $A^\alpha{}_{\mu\Lambda}(x)$ vanishes for all α everywhere in at least a finite domain around any given point. For instance, to make $A^\alpha{}_{3\Lambda}(x)$ vanish, we must solve the set of *ordinary* first-order differential equations for the parameters $\Lambda^\beta(x)$:

$$\partial_3 \exp(it_\beta\Lambda^\beta) = -i\exp(i\,t_\beta\Lambda^\beta)\,t_\alpha A^\alpha{}_3\,, \tag{15.1.18}$$

which always have a solution in at least a finite domain around any ordinary point.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

However, in general it is not possible to choose $\Lambda^\alpha(x)$ to make all four components $A^\alpha{}_{\mu\Lambda}(x)$ vanish in a finite region. For this purpose, we would have to be able to satisfy the *partial* differential equations

$$\partial_\mu \exp(it_\beta \Lambda^\beta) = -i \exp(i t_\beta \Lambda^\beta)\, t_\alpha A^\alpha{}_\mu \,, \tag{15.1.19}$$

which cannot be solved unless certain integrability conditions are satisfied. In particular, if $A^\alpha{}_{\mu\Lambda}(x)$ vanishes everywhere then so does $F^\alpha{}_{\mu\nu\Lambda}(x)$, but since the field strength transforms homogeneously, $F^\alpha{}_{\mu\nu\Lambda}(x)$ can vanish only if $F^\alpha{}_{\mu\nu}(x)$ does. A gauge field $A^\alpha{}_\mu(x)$ is called a 'pure gauge' field if there exists a gauge transformation which makes it vanish everywhere. It is not difficult to show that the condition that $F^\alpha{}_{\mu\nu}$ should vanish everywhere is not only necessary but also sufficient for $A^\alpha{}_\mu(x)$ to be expressible in any simply connected region as a pure gauge field.[6]

* * *

There is a deep analogy between the construction here of objects that transform simply under gauge transformations and the construction in general relativity of objects that transform covariantly under general coordinate transformations. Just as we use the gauge field to construct covariant derivatives $D_\mu \psi_\ell$ of matter fields with the same gauge transformation properties as the matter fields themselves, so we use the affine connection $\Gamma^\mu{}_{\nu\lambda}(x)$ to construct covariant derivatives of tensors $T^{\rho\sigma\cdots}{}_{\kappa\lambda\cdots}$:

$$T^{\rho\cdots}{}_{\kappa\cdots;\nu} \equiv \partial_\nu T^{\rho\cdots}{}_{\kappa\cdots} + \Gamma^\rho{}_{\nu\lambda} T^{\lambda\cdots}{}_{\kappa\cdots} + \cdots - \Gamma^\mu{}_{\nu\kappa} T^{\rho\cdots}{}_{\mu\cdots} - \cdots \,,$$

which are themselves tensors. Also, from the derivatives of the gauge field we constructed a field strength $F^\alpha{}_{\mu\nu}$ with the gauge transformation property of a matter field belonging to the adjoint representation of the gauge group; correspondingly, from the derivatives of the affine connection we may construct a quantity:

$$R^\lambda{}_{\mu\nu\kappa} = \frac{\partial \Gamma^\lambda{}_{\mu\nu}}{\partial x^\kappa} - \frac{\partial \Gamma^\lambda{}_{\mu\kappa}}{\partial x^\nu} + \Gamma^\eta{}_{\mu\nu}\Gamma^\lambda{}_{\kappa\eta} - \Gamma^\eta{}_{\mu\kappa}\Gamma^\lambda{}_{\nu\eta} \,,$$

which transforms as a tensor, the Riemann–Christoffel curvature tensor. The commutator of two gauge-covariant derivatives D_μ and D_ν may be expressed in terms of the field-strength tensor $F^\alpha{}_{\mu\nu}$; similarly, the commutator of two covariant derivatives with respect to x^ν and x^κ may be expressed in terms of the curvature:

$$T^{\lambda\cdots}{}_{\mu\cdots;\nu;\kappa} - T^{\lambda\cdots}{}_{\mu\cdots;\kappa;\nu} = R^\lambda{}_{\sigma\nu\kappa}\, T^{\sigma\cdots}{}_{\mu\cdots} + \cdots - R^\sigma{}_{\mu\nu\kappa}\, T^{\lambda\cdots}{}_{\sigma\cdots} - \cdots \,.$$

The necessary and sufficient condition for the existence of a gauge in which the gauge field vanishes in a finite simply connected region is the vanishing of the field-strength tensor, and the necessary and sufficient condition for the existence of a coordinate system in which the affine connection vanishes

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

in a finite simply connected region is the vanishing of the Riemann–Christoffel curvature tensor. The analogy breaks down in one important respect: in general relativity the affine connection is itself constructed from first derivatives of the metric tensor, while in gauge theories the gauge fields are not expressed in terms of any more fundamental fields.

15.2 Gauge Theory Lagrangians and Simple Lie Groups

The transformation rules of the gauge-field tensor $F^{\alpha}{}_{\mu\nu}$ and the matter fields ψ and their gauge-covariant derivatives do not involve the derivatives of the transformation parameters $\epsilon^{\alpha}(x)$, so if the Lagrangian is constructed solely from these ingredients, and if it is invariant under global transformations with ϵ^{α} constant, then it is invariant under gauge transformations with general position-dependent $\epsilon^{\alpha}(x)$. We therefore assume that the Lagrangian satisfies these conditions: that is,

$$\mathscr{L} = \mathscr{L}(\psi, D_\mu\psi, D_\nu D_\mu\psi, \cdots, F^{\alpha}{}_{\mu\nu}, D_\rho F^{\alpha}{}_{\mu\nu} \cdots) \tag{15.2.1}$$

with the invariance condition:

$$\begin{aligned}&\frac{\partial\mathscr{L}}{\partial\psi_\ell} i(t_\alpha)_\ell{}^m\psi_m + \frac{\partial\mathscr{L}}{\partial(D_\mu\psi_\ell)} i(t_\alpha)_\ell{}^m(D_\mu\psi)_m\\ &+ \frac{\partial\mathscr{L}}{\partial(D_\nu D_\mu\psi)_\ell} i(t_\alpha)_\ell{}^m(D_\nu D_\mu\psi)_m + \cdots + \frac{\partial\mathscr{L}}{\partial F^{\beta}{}_{\mu\nu}} C^{\beta}{}_{\gamma\alpha}F^{\gamma}{}_{\mu\nu}\\ &+ \frac{\partial\mathscr{L}}{\partial D_\rho F^{\beta}{}_{\mu\nu}} C^{\beta}{}_{\gamma\alpha}D_\rho F^{\gamma}{}_{\mu\nu} + \cdots = 0\,.\end{aligned} \tag{15.2.2}$$

On the other hand, the Lagrangian may not depend on the gauge field itself, except insofar as it appears in $F^{\alpha}{}_{\mu\nu}$ and in gauge-covariant derivatives D_μ. In particular, a mass term $-\frac{1}{2} m^2{}_{\alpha\beta} A_{\alpha\mu}A_\beta{}^\mu$ is ruled out.

We shall concentrate now on the terms in the Lagrangian that depend only on $F^{\alpha}{}_{\mu\nu}$. Just as in electrodynamics, for any massless particle of unit spin the Lagrangian must contain a free-particle term quadratic in $\partial_\mu A^{\alpha}{}_{\nu} - \partial_\nu A^{\alpha}{}_{\mu}$, and gauge invariance then dictates that this free-particle term should appear as part of a term quadratic in the field-strength tensor $F^{\alpha}{}_{\mu\nu}$. Lorentz invariance and parity conservation dictate its form as

$$\mathscr{L}_A = -\tfrac{1}{4}\, g_{\alpha\beta}F^{\alpha}{}_{\mu\nu}F^{\beta\mu\nu} \tag{15.2.3}$$

with a constant matrix $g_{\alpha\beta}$. If we do not assume parity (or CP or T) conservation, then we may also include in the Lagrangian a term

$$\mathscr{L}'_A = -\tfrac{1}{2}\,\theta_{\alpha\beta}\,\epsilon^{\mu\nu\rho\sigma}F^{\alpha}{}_{\mu\nu}F^{\beta}{}_{\rho\sigma}$$

with another constant matrix $\theta_{\alpha\beta}$. This term is actually a derivative, and therefore does not affect the field equations or the Feynman rules. Such a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

term would, however, have non-perturbative quantum mechanical effects, to be discussed in Section 23.6.

Before going on to consider the properties of the matrix $g_{\alpha\beta}$, it is worth drawing attention to the fact that it is not possible to introduce a kinematic term for the gauge field $A^{\alpha}{}_{\mu}(x)$ without also including interactions, the terms in Eq. (15.2.3) arising from the quadratic part of the field strength $F^{\alpha}{}_{\mu\nu}$ defined by Eq. (15.1.13). This is one more respect in which non-Abelian gauge theories resemble general relativity, where the kinematic part of the Lagrangian for the gravitational field is contained in the Einstein–Hilbert Lagrangian density $-\sqrt{g}R/8\pi G$, which also contains self-interactions of the field. The reasons in both cases are similar: the gravitational field interacts with itself because it interacts with anything that carries energy and momentum, and the gauge field interacts with itself because it interacts with anything that transforms according to a non-trivial representation (in this case the adjoint representation) of the gauge group. This is in contrast to the case of electrodynamics, where the photon does not carry electric charge, the quantum number with which it interacts, and it is consequently possible to introduce a kinematic term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ for the electromagnetic field that does not entail interactions.

The numerical matrix $g_{\alpha\beta}$ may be taken symmetric, and must be taken real to give a real Lagrangian. In order for this term to satisfy the gauge-invariance requirement (15.2.2), we must have for all δ:

$$g_{\alpha\beta}F^{\alpha}{}_{\mu\nu}C^{\beta}{}_{\gamma\delta}F^{\gamma\mu\nu} = 0\,.$$

In order for this to be true without having to impose any functional relations among the Fs, the matrix $g_{\alpha\beta}$ must satisfy the condition:

$$g_{\alpha\beta}C^{\beta}{}_{\gamma\delta} = -\,g_{\gamma\beta}C^{\beta}{}_{\alpha\delta}\,. \tag{15.2.4}$$

There is one more important condition on the matrix $g_{\gamma\beta}$. Just as in quantum electrodynamics, the rules of canonical quantization and the positivity properties of the quantum mechanical scalar product require that the matrix $g_{\alpha\beta}$ in the Lagrangian (15.2.3) must be positive-definite. (That is, $g_{\alpha\beta}u^{\alpha}u^{\beta}$ is positive for all real u, and vanishes for some real u only if $u^{\alpha} = 0$ for all α.) This is analogous to the requirement that in the kinematic Lagrangian $-\frac{1}{2}Z\,\partial_{\mu}\phi\partial^{\mu}\phi - \frac{1}{2}m^2\phi^2$ for a real scalar field ϕ, the constant Z must be positive-definite.

These requirements on the matrix $g_{\alpha\beta}$ have far reaching implications. They form one of a set of three equivalent conditions:

a: There exists a real symmetric positive-definite matrix $g_{\alpha\beta}$ that satisfies the invariance condition (15.2.4).

b: There is a basis for the Lie algebra (that is, a set of generators $\tilde{t}_{\alpha} = \mathscr{S}_{\alpha\beta}t_{\beta}$, with $\mathscr{S}$ a real non-singular matrix) for which the structure

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

constants $\tilde{C}^{\alpha}{}_{\beta\gamma}$ are antisymmetric not only in the lower indices β and γ but in all three indices α, β and γ. (In this basis it is convenient to drop the distinction between upper and lower indices α, β, etc., and write $\tilde{C}_{\alpha\beta\gamma}$ in place of $\tilde{C}^{\alpha}{}_{\beta\gamma}$.)

c: The Lie algebra is the direct sum of commuting compact simple and $U(1)$ subalgebras.*

Appendix A of this chapter presents a proof of the equivalence of the conditions **a**, **b**, and **c**.[7]

Before going on to discuss the physical implications of this result, it will be useful to say a bit more about the condition of compactness. We will not use this here, but a compact Lie algebra consists of the generators of a compact Lie group: one for which the invariant volume of the group is finite. For instance, the rotation group is compact; the Lorentz group is not. As a simple example of a simple Lie algebra that is not compact, consider the commutation relations

$$[t_1, t_2] = -i t_3, \qquad [t_2, t_3] = i t_1, \qquad [t_3, t_1] = i t_2 .$$

The structure constant here is real, but not completely antisymmetric; its non-vanishing components are

$$C^3{}_{12} = -C^3{}_{21} = -1, \quad C^1{}_{23} = -C^1{}_{32} = 1, \quad C^2{}_{31} = -C^2{}_{13} = 1 .$$

The metric given by Eq. (15.A.10) is here diagonal, with elements:

$$g_{11} = g_{22} = -g_{33} = -2 .$$

This is not a positive matrix, so the Lie algebra is not compact. It is in fact the Lie algebra of the non-compact group $O(2,1)$, the Lorentz group in two space and one time dimensions.

---

* Some definitions: A subalgebra $\mathscr{H}$ of a Lie algebra $\mathscr{G}$ is a linear space, spanned by certain real linear combinations $t_i = \mathscr{S}_{i\alpha} t_\alpha$ of the generators t_α of $\mathscr{G}$, such that $\mathscr{H}$ is itself a Lie algebra, in the sense that the commutators of the t_i with each other are of the form $[t_i, t_j] = i C^k{}_{ij} t_k$. A subalgebra $\mathscr{H}$ is called *invariant* if the commutator of any element of the whole algebra $\mathscr{G}$ with any element of the subalgebra $\mathscr{H}$ is in the subalgebra $\mathscr{H}$. A *simple* Lie algebra is one without invariant subalgebras. A $U(1)$ subalgebra of $\mathscr{G}$ is one with just a single generator that commutes with all generators of the whole algebra $\mathscr{G}$. A *semi-simple* Lie algebra is one that has no invariant Abelian subalgebras, i.e., invariant subalgebras whose generators all commute with each other. Semi-simple Lie algebras are direct sums of simple (but not $U(1)$) Lie algebras. A simple or semi-simple Lie algebra is said to be *compact* if the matrix $\mathrm{Tr}\{t^A{}_\alpha t^A{}_\beta\} = -C^\gamma{}_{\alpha\delta} C^\delta{}_{\beta\gamma}$ is positive-definite. The meaning and importance of the properties of simplicity and compactness will be discussed further below. In saying that a Lie algebra $\mathscr{G}$ is a direct sum of subgroups $\mathscr{H}_n$, it is meant that it is possible to find a basis for $\mathscr{G}$ with generators t_{na} for which the structure constants take the form

$$C^{\ell c}{}_{na\,mb} = \delta_{\ell m} \delta_{mn} C^{(n)c}{}_{ab} ,$$

where $C^{(n)c}{}_{ab}$ is the structure constant of the subalgebra $\mathscr{H}_n$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Two sets of generators that differ by a *real* non-singular linear transformation are considered to span the same Lie algebra, and generate the same group. This is not true for complex linear transformations of generators. In particular, any simple Lie algebra can be put into a compact form by a change of phase of the generators in a suitable basis. For instance, for the Lie algebra of the above example, it is only necessary to define new generators $t'_1 = it_1$, $t'_2 = it_2$, $t'_3 = t_3$, for which the commutation relations are

$$[t'_1, t'_2] = i\,t'_3, \quad [t'_2, t'_3] = i\,t'_1, \quad [t'_3, t'_1] = i\,t'_2\,.$$

The structure constant is now real and totally antisymmetric: $C^a{}_{bc} = \epsilon_{abc}$. Here $g_{ab} = 2\delta_{ab}$, and the algebra is compact. We recognize this, of course, as the familiar algebra of the compact group $O(3)$ of rotations in three dimensions. To see that this is always possible for any simple Lie algebra, note that the matrix g_{ab} defined by Eq. (15.A.10) is real, symmetric, and non-singular, so that by a real orthogonal transformation it may be put in a diagonal form with non-zero elements along the main diagonal. It is then only necessary to multiply all the generators that correspond in this basis to the negative diagonal elements of g_{ab} by factors i.

We note without proof that the finite-dimensional representations of compact Lie groups are all unitary, and the finite-dimensional representations of compact Lie algebras are correspondingly all Hermitian. Furthermore, it is easy to see that the only Lie algebras that can have any non-trivial representation by independent finite-dimensional Hermitian matrices t_α are direct sums of $U(1)$ and compact simple Lie algebras. To show this, we may simply define

$$g_{\alpha\beta} \equiv \mathrm{Tr}\{t_\alpha t_\beta\}\,.$$

This matrix is obviously positive-definite, because $g_{\alpha\beta}u^\alpha u^\beta = \mathrm{Tr}\{(u^\alpha t_\alpha)^2\}$ is positive for any real u^α and vanishes only if $u^\alpha t_\alpha = 0$, which is not possible unless all u^α vanish because the t^α are assumed independent. Furthermore this $g_{\alpha\beta}$ satisfies the invariance condition (15.2.4), as can be seen by multiplying the commutation relation (15.1.2) with t_δ and taking the trace; this gives

$$i\,C^\gamma{}_{\alpha\beta}\mathrm{Tr}\{t_\gamma t_\delta\} = \mathrm{Tr}\{[t_\alpha, t_\beta]\,t_\delta\} = \mathrm{Tr}\{t_\delta t_\alpha t_\beta - t_\beta t_\alpha t_\delta\}\,,$$

which is obviously antisymmetric in β and δ. Having verified **a**, we can rely on the above theorem to infer condition **c**, so that the Lie algebra must be a direct sum of compact simple and $U(1)$ subalgebras.

Let's now return to the physics of gauge theories. In this section we have inferred the existence of a positive symmetric real matrix $g_{\alpha\beta}$, that satisfies the invariance condition (15.2.4), from the necessity of constructing a suitable kinematic term in the Lagrangian for the gauge field, and in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Appendix A of this chapter we have shown that this result is equivalent to a condition on the Lie algebra, that it is a direct sum of compact simple and $U(1)$ subalgebras. For our purposes, the important thing about this result is that the simple Lie algebras are all of certain limited types and dimensionalities. For instance, it is easy to see that there is no simple Lie algebra with less than three generators, because in one or two dimensions there can be no non-zero totally antisymmetric structure constants with three indices. With three generators, an invariant subalgebra can be avoided by taking $C^3{}_{12}$, $C^2{}_{31}$, and $C^1{}_{23}$ all non-zero. In the basis in which the structure constant is real and totally antisymmetric, there is obviously only one possibility:

$$C_{\alpha\beta\gamma} = c\,\epsilon_{\alpha\beta\gamma} .$$

Here c is an arbitrary non-zero real constant, which can be eliminated by a change of scale of the generators, $t_\alpha \to t_\alpha/c$, so the Lie algebra is

$$[t_\alpha, t_\beta] = i\,\epsilon_{\alpha\beta\gamma} t_\gamma .$$

This may be recognized as the Lie algebra of the three-dimensional rotation group $O(3)$, and also of the group $SU(2)$ of unitary unimodular matrices in two dimensions, and was used as a basis for the original non-Abelian gauge theory of Yang and Mills. Continuing in the same way, it can be shown that there are no simple Lie algebras with 4, 5, 6, or 7 generators, one with 8 generators, and so on. Mathematicians (notably Killing and E. Cartan) have been able to catalog all simple Lie algebras. The compact forms of the simple Lie algebras form several infinite classes of algebras of the 'classical' Lie groups — the unitary unimodular, unitary orthogonal, and unitary symplectic groups — plus just five exceptional Lie algebras. This catalog is presented in Appendix B of this chapter.

It is also shown in Appendix A that under the equivalent conditions **a**, **b**, or **c**, the metric takes the form

$$g_{ma,nb} = g_m^{-2}\delta_{mn}\delta_{ab} \tag{15.2.5}$$

with real g_m, where m and n label the simple or $U(1)$ subalgebras, and a and b label the individual generators of these subalgebras. We can eliminate the constants g_m^{-2} by a rescaling of the gauge fields

$$A^\mu_{ma} \to \tilde{A}^\mu_{ma} \equiv g_m^{-1} A^\mu_{ma} , \tag{15.2.6}$$

but then in order to keep the same formulas (15.1.10) and (15.1.13) for $D_\mu\psi$ and $F_\alpha{}^{\mu\nu}$, we must also redefine the matrices t_α and the structure constants

$$t_{ma} \to \tilde{t}_{ma} = g_m t_{ma} , \tag{15.2.7}$$

$$C^{(m)}_{cab} \to \tilde{C}^{(m)}_{cab} = g_m C^{(m)}_{cab} . \tag{15.2.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

That is, we can always *define* the scale of the gauge fields (now dropping the tildes) so that g_m in Eq. (15.2.5) is unity:

$$g_{\alpha\beta} = \delta_{\alpha\beta} \,, \tag{15.2.9}$$

but then the transformation matrices t_α and the structure constants $C_{\alpha\beta\gamma}$ contain an unknown multiplicative factor g_m for each simple or $U(1)$ subalgebra. These factors are the *coupling constants* of the gauge theory. Alternatively, it is sometimes more convenient to adopt some fixed though arbitrary normalization for the t_α and structure constants within each simple or $U(1)$ subalgebra, in which case the coupling constants appear in the gauge-field Lagrangian (15.2.3) as the factors g_m^{-2} in Eq. (15.2.5).

15.3 Field Equations and Conservation Laws

Using Eq. (15.2.9) for the matrix $g_{\alpha\beta}$ in Eq. (15.2.3), the full Lagrangian density is

$$\mathscr{L} = -\tfrac{1}{4} F_{\alpha\mu\nu}F_\alpha{}^{\mu\nu} + \mathscr{L}_M(\psi, D_\mu\psi) \,, \tag{15.3.1}$$

where in the absence of gauge fields $\mathscr{L}_M(\psi, \partial_\mu\psi)$ would be the 'matter' Lagrangian density. We could, in principle, include a dependence of $\mathscr{L}_M$ on $F_{\alpha\mu\nu}$ as well as higher covariant derivatives $D_\nu D_\mu\psi$, $D_\lambda F_{\alpha\mu\nu}$, etc., but we exclude these non-renormalizable terms here for the same reason as in electrodynamics: as discussed in Section 12.3, such terms would be highly suppressed at ordinary energies by negative powers of some very large mass. For this reason the standard model of the weak, electromagnetic and strong interactions has a Lagrangian of the general form (15.3.1).

The equations of motion of the gauge field are here

$$\begin{aligned}\partial_\mu \frac{\partial\mathscr{L}}{\partial(\partial_\mu A_{\alpha\nu})} &= -\partial_\mu F_\alpha{}^{\mu\nu} = \frac{\partial\mathscr{L}}{\partial A_{\alpha\nu}} \\ &= -F_\gamma{}^{\nu\mu}C_{\gamma\alpha\beta}A_{\beta\mu} - i\frac{\partial\mathscr{L}_M}{\partial D_\nu\psi}\, t_\alpha\psi\end{aligned}$$

and so

$$\partial_\mu F_\alpha{}^{\mu\nu} = -\mathscr{J}_\alpha{}^\nu \,, \tag{15.3.2}$$

where $\mathscr{J}_\alpha{}^\nu$ is the current:

$$\mathscr{J}_\alpha{}^\nu \equiv -F_\gamma{}^{\nu\mu}C_{\gamma\alpha\beta}A_{\beta\mu} - i\frac{\partial\mathscr{L}_M}{\partial D_\nu\psi}t_\alpha\psi \,. \tag{15.3.3}$$

The current $\mathscr{J}_\alpha{}^\nu$ is conserved in the ordinary sense

$$\partial_\nu\mathscr{J}_\alpha{}^\nu = 0 \,, \tag{15.3.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

as can be seen either from the Euler–Lagrange equations for ψ and the invariance condition (15.2.2) or, more easily, directly from the field equations (15.3.2).

The derivatives in Eqs. (15.3.2) and (15.3.4) are ordinary derivatives, not the gauge-covariant derivatives D_ν, so the gauge invariance of these equations is somewhat obscure. It can be made manifest by rewriting Eq. (15.3.2) in terms of the gauge-covariant derivative of the field strength

$$\begin{aligned} D_\lambda F_\alpha{}^{\mu\nu} &\equiv \partial_\lambda F_\alpha{}^{\mu\nu} - i(t^A_\beta)_{\alpha\gamma} A_{\beta\lambda} F_\gamma{}^{\mu\nu} \\ &= \partial_\lambda F_\alpha{}^{\mu\nu} - C_{\alpha\gamma\beta} A_{\beta\lambda} F_\gamma{}^{\mu\nu} . \end{aligned} \tag{15.3.5}$$

Then Eq. (15.3.2) reads

$$D_\mu F_\alpha{}^{\mu\nu} = -J_\alpha{}^\nu , \tag{15.3.6}$$

where $J_\alpha{}^\nu$ is the current of the matter fields alone

$$J_\alpha{}^\nu \equiv -i \frac{\partial \mathscr{L}_M}{\partial D_\nu \psi} t_\alpha \psi . \tag{15.3.7}$$

This is gauge-covariant, if $\mathscr{L}_M$ is gauge-invariant. Also, by operating on Eq. (15.3.6) with D_ν, using the commutation relation

$$[D_\nu, D_\mu] F_\alpha^{\rho\sigma} = -i(t^A{}_\gamma)_{\alpha\beta} F_{\gamma\nu\mu} F_\beta{}^{\rho\sigma} = -C_{\gamma\alpha\beta} F_{\gamma\nu\mu} F_\beta{}^{\rho\sigma} ,$$

we see that $J_\alpha{}^\nu$ satisfies a gauge-covariant conservation law

$$D_\nu J_\alpha{}^\nu = 0 , \tag{15.3.8}$$

rather than the ordinary conservation law (15.3.4) obeyed by the full current $\mathscr{J}_\alpha{}^\nu$. Also, it is straightforward (using Eq. (15.1.5)) to derive the identities:

$$D_\mu F_{\alpha\nu\lambda} + D_\nu F_{\alpha\lambda\mu} + D_\lambda F_{\alpha\mu\nu} = 0 , \tag{15.3.9}$$

which hold whether or not the gauge fields satisfy the field equations.

These results serve to underscore the profound analogy mentioned in Section 15.1 between non-Abelian gauge theories and general relativity. In general relativity there is a matter energy-momentum tensor $T^\nu{}_\mu$, analogous to J^μ, which satisfies a generally covariant conservation law $T^\nu{}_{\mu;\nu} = 0$, and stands on the right-hand side of the Einstein field equations in their generally covariant form, $R^\nu{}_\mu - \frac{1}{2}\delta^\nu{}_\mu R = -8\pi G T^\nu{}_\mu$. However, $T^\nu{}_\mu$ is not conserved in the ordinary sense: $\partial_\nu T^\nu{}_\mu$ does not vanish. On the other hand, moving the non-linear terms on the left-hand side of the Einstein equation to the right-hand side gives a field equation[8]

$$\left(R^\nu{}_\mu - \frac{1}{2} \delta^\nu{}_\mu R \right)_{\text{LINEAR}} = -8\pi G \tau^\nu{}_\mu ,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

where $\tau^{\nu}{}_{\mu}$ is the non-tensor

$$\tau^{\nu}{}_{\mu} \equiv T^{\nu}{}_{\mu} + \frac{1}{8\pi G}\left(R^{\nu}{}_{\mu} - \frac{1}{2}\,\delta^{\nu}{}_{\mu}R\right)_{\text{NONLINEAR}},$$

analogous to $\mathscr{J}_{\alpha}{}^{\nu}$. Like $\mathscr{J}_{\alpha}{}^{\nu}$, $\tau^{\nu}{}_{\mu}$ is conserved in the ordinary sense

$$\partial_{\nu}\tau^{\nu}{}_{\mu} = 0$$

and may be regarded as the current of energy and momentum:

$$P_{\mu} = \int \tau^{0}{}_{\mu}\, d^{3}x\,.$$

It contains a purely gravitational term, because gravitational fields carry energy and momentum; without this term, $\tau^{\nu}{}_{\mu}$ could not be conserved. Similarly, $\mathscr{J}^{\nu}_{\alpha}$ contains a gauge-field term (the first term on the right in Eq. (15.3.3)) because for non-Abelian groups (those with $C^{\gamma}_{\alpha\beta} \neq 0$) the gauge fields carry the quantum numbers with which they interact. Because $\mathscr{J}_{\alpha}{}^{\nu}$ is conserved in the ordinary sense, it can be regarded as the current of these quantum numbers, with the symmetry generators given by the time-independent quantities

$$T_{\alpha} = \int \mathscr{J}_{\alpha}{}^{0} d^{3}x\,. \tag{15.3.10}$$

(Also, the homogeneous equations (15.3.9) involve covariant derivatives, just as do the Bianchi identities of general relativity.) In contrast, none of these complications arises in quantum electrodynamics, because photons do not carry the quantum number, electric charge, with which they interact.

15.4 Quantization

We now proceed to quantize the gauge theories described in the previous two sections. The Lagrangian density is taken in the form (15.3.1):

$$\mathscr{L} = -\tfrac{1}{4}\,F_{\alpha\mu\nu}F_{\alpha}{}^{\mu\nu} + \mathscr{L}_{M}(\psi, D_{\mu}\psi)\,, \tag{15.4.1}$$

with

$$F_{\alpha\mu\nu} \equiv \partial_{\mu}A_{\alpha\nu} - \partial_{\nu}A_{\alpha\mu} + C_{\alpha\beta\gamma}A_{\beta\mu}A_{\gamma\nu}\,,$$
$$D_{\mu}\psi \equiv \partial_{\mu}\psi - it_{\alpha}A_{\alpha\mu}\psi\,.$$

We cannot immediately quantize this theory by setting commutators equal to i times the corresponding Poisson brackets. The problem is one of constraints. In the terminology of Dirac, described in Section 7.6, there is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

a primary constraint that

$$\Pi_{\alpha 0} \equiv \frac{\partial \mathscr{L}}{\partial(\partial_0 A^0_\alpha)} = 0 \tag{15.4.2}$$

and a secondary constraint provided by the field equation for A^0_α:

$$\begin{aligned} -\partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu A_{\alpha 0})} + \frac{\partial \mathscr{L}}{\partial A_{\alpha 0}} &= \partial_\mu F_\alpha{}^{\mu 0} + F_\gamma{}^{\mu 0} C_{\gamma\alpha\beta} A_{\beta\mu} + J_\alpha{}^0 \\ &= \partial_k \Pi_\alpha{}^k + \Pi_\gamma{}^k C_{\gamma\alpha\beta} A_{\beta k} + J_\alpha{}^0 = 0 , \end{aligned} \tag{15.4.3}$$

where $\Pi^k_\alpha \equiv \partial\mathscr{L}/\partial(\partial_0 A_{\alpha k}) = F^{k0}_\alpha$ is the 'momentum' conjugate to $A_{\alpha k}$, with k running over the values 1, 2, 3. The Poisson brackets of $\Pi_{\alpha 0}$ and $\partial_k \Pi_\alpha{}^k + \Pi_\gamma{}^k C_{\gamma\alpha\beta} A_{\beta k} + J_\alpha{}^0$ vanish (because the latter quantity is independent of $A_\alpha{}^0$), so these are first class constraints, which cannot be dealt with by replacing Poisson brackets with Dirac brackets.

As in the case of electrodynamics, we deal with these constraints by choosing a gauge. The Coulomb gauge adopted for electrodynamics would lead to painful complications here,* so instead we will work in what is known as *axial gauge*, based on the condition

$$A_{\alpha 3} = 0 . \tag{15.4.4}$$

The canonical variables of the gauge field are then $A_{\alpha i}$, with i now running over the values 1 and 2, together with their canonical conjugates

$$\Pi_{\alpha i} \equiv \frac{\partial \mathscr{L}}{\partial(\partial_0 A_{\alpha i})} = -F_\alpha{}^{0i} = \partial_0 A_{\alpha i} - \partial_i A_{\alpha 0} + C_{\alpha\beta\gamma} A_{\beta 0} A_{\gamma i} . \tag{15.4.5}$$

The field $A_{\alpha 0}$ is not an independent canonical variable, but rather is defined in terms of the other variables by the constraint (15.4.3). To see this, note that the 'electric' field strengths $F_\alpha{}^{\mu 0}$ are

$$F_\alpha{}^{i0} = \Pi_{\alpha i} , \quad F_\alpha{}^{30} = \partial_3 A_\alpha{}^0 , \tag{15.4.6}$$

so the constraint (15.4.3) reads

$$-(\partial_3)^2 A^0_\alpha = \partial_i \Pi_{\alpha i} + \Pi_{\gamma i} C_{\gamma\alpha\beta} A_{\beta i} + J_\alpha{}^0 , \tag{15.4.7}$$

which can easily be solved (with reasonable boundary conditions) to give $A_\alpha{}^0$ as a functional of $\Pi_{\gamma i}$, $A_{\beta i}$, and $J_\alpha{}^0$. (We are using a summation

* In addition to purely algebraic complications, Coulomb gauge (like many other gauges) has a problem known as the Gribov ambiguity:[9] even with the condition that $\mathbf{A}_\alpha$ vanishes at spatial infinity, for each solution of the Coulomb gauge condition $\nabla \cdot \mathbf{A}_\alpha = 0$ there are other solutions that differ by finite gauge transformations. The Gribov ambiguity will not bother us here, because we quantize in axial gauge where it is absent, and we shall use other gauges like Lorentz gauge only to generate a perturbation series.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

convention, with indices i, j, etc. summed over the values 1 and 2.) It should be noted that the canonical conjugate to the matter field ψ_ℓ is

$$\pi_\ell = \frac{\partial \mathscr{L}}{\partial(\partial_0 \psi_\ell)} = \frac{\partial \mathscr{L}_M}{\partial(D_0 \psi_\ell)}\,, \tag{15.4.8}$$

so the time component of the matter current can be expressed in terms of the canonical variables of the matter fields *alone*

$$J_\alpha{}^0 = -i \frac{\partial \mathscr{L}_m}{\partial(D_0\psi)_\ell} (t_\alpha)_\ell{}^m \psi_m = -i\,\pi_\ell (t_\alpha)_\ell{}^m \psi_m\,. \tag{15.4.9}$$

Hence Eq. (15.4.7) defines $A_\alpha{}^0$ at a given time as a functional of the canonical variables $\Pi_{\gamma i}$, $A_{\beta i}$, π_ℓ, and ψ_m at the same time.

Now that we have identified the canonical variables in this gauge, we can proceed to the construction of a Hamiltonian. The Hamiltonian density is

$$\begin{aligned}\mathscr{H} &= \Pi_{\alpha i}\partial_0 A_{\alpha i} + \pi_\ell \partial_0 \psi_\ell - \mathscr{L} \\ &= \Pi_{\alpha i}\left(F_{\alpha 0 i} + \partial_i A_{\alpha 0} - C_{\alpha\beta\gamma} A_{\beta 0} A_{\gamma i}\right) + \pi_\ell \partial_0 \psi_\ell \\ &\quad - \tfrac{1}{2} F_{\alpha 0 i} F_{\alpha 0 i} + \tfrac{1}{2} F_{\alpha i j} F_{\alpha i j} + \tfrac{1}{2} F_{\alpha i 3} F_{\alpha i 3} \\ &\quad - \tfrac{1}{2} F_{\alpha 0 3} F_{\alpha 0 3} - \mathscr{L}_M\,.\end{aligned} \tag{15.4.10}$$

Using Eqs. (15.4.4) and (15.4.6), this is

$$\begin{aligned}\mathscr{H} = \mathscr{H}_M &+ \Pi_{\alpha i}(\partial_i A_{\alpha 0} - C_{\alpha\beta\gamma} A_{\beta 0} A_{\gamma i}) + \tfrac{1}{2}\,\Pi_{\alpha i}\Pi_{\alpha i} \\ &+ \tfrac{1}{4} F_{\alpha i j} F_{\alpha i j} + \tfrac{1}{2}\,\partial_3 A_{\alpha i}\partial_3 A_{\alpha i} - \tfrac{1}{2}\partial_3 A_{\alpha 0}\partial_3 A_{\alpha 0}\,,\end{aligned} \tag{15.4.11}$$

where $\mathscr{H}_M$ is the matter Hamiltonian density:

$$\mathscr{H}_M \equiv \pi_\ell\,\partial_0 \psi_\ell - \mathscr{L}_M\,. \tag{15.4.12}$$

Following the general rules derived in Section 9.2, we can now use this Hamiltonian density to calculate matrix elements as path integrals over $A_{\alpha i}$, $\Pi_{\alpha i}$, ψ_ℓ, and π_ℓ, with weighting factor $\exp(iI)$, where

$$I = \int d^4x \left[\Pi_{\alpha i}\partial_0 A_{\alpha i} + \pi_\ell \partial_0 \psi_\ell - \mathscr{H} + \epsilon \text{ terms}\right], \tag{15.4.13}$$

in which the 'ϵ terms' serve only to supply the correct imaginary infinitesimal terms in propagator denominators. (See Section 9.2.) We note that Eqs. (15.4.7) and (15.4.9) give A_α^0 as a functional of the canonical variables, *linear* in $\Pi_{\alpha i}$ and π_ℓ. Inspection of Eq. (15.4.11) shows then (assuming $\mathscr{L}_M$ to be no more than quadratic in $D_\mu\psi$) that the integrand of the complete action (15.4.13) is no more than quadratic in $\Pi_{\alpha i}$ and π_ℓ. We could therefore carry out the path integral over these canonical 'momenta' by the usual rules of Gaussian integration. The trouble with this procedure is that the coefficients of the terms in Eq. (15.4.13) of second order

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

in $\Pi_{\alpha i}$ are functions of the $A_{\alpha i}$, so the Gaussian integral would yield an awkward field-dependent determinant factor. Also, the whole formalism at this point looks hopelessly non-Lorentz-invariant.

Instead of proceeding in this way, we will apply a trick like that used in the path integral formulation of electrodynamics in Section 9.6. Note that if for a moment we think of $A_{\alpha 0}$ as an independent variable, then the action (15.4.13) is evidently quadratic in $A_{\alpha 0}$, with the coefficient of the second-order term $A_{\alpha 0}(x)A_{\beta 0}(y)$ equal to the field-independent kernel $(\partial_3)^2\delta^4(x-y)$. As we saw in the appendix to Chapter 9, the integral of such a Gaussian over $A_{\alpha 0}(x)$ is, up to a constant factor, equal to the value of the integrand at the stationary 'point' of the argument of the exponential. But the variational derivative of the action here is

$$\frac{\delta I}{\delta A_{\alpha 0}} = -\frac{\partial \mathscr{H}}{\partial A_{\alpha 0}} = J_\alpha{}^0 + \partial_i \Pi_{\alpha i} + C_{\beta\alpha\gamma}\Pi_{\beta i}A_{\gamma i} - \partial_3^2 A_{\alpha 0}\,,$$

so the stationary 'point' of the action is the solution of the constraint equation (15.4.7). Hence, instead of using for $A_{\alpha 0}$ the solution of Eq. (15.4.7), we can just as well treat it as an independent variable of integration.

With $A_{\alpha 0}$ now regarded as an independent variable, the Hamiltonian $\int d^3x \mathscr{H}$ is evidently quadratic in $\Pi_{\alpha i}$, with the coefficient of the second-order term $\Pi_{\alpha i}(x)\Pi_{\beta j}(y)$ given by the field-independent kernel $\frac{1}{2}\delta^4(x-y)\delta_{ij}$. Assuming that the same is true for the matter variable π_ℓ, we can evaluate path integrals over π_ℓ and $\Pi_{\alpha i}$ up to a constant factor by simply setting π_ℓ and $\Pi_{\alpha i}$ at the stationary 'points' of the action corresponding to Eq. (15.4.1):

$$0 = \frac{\delta I}{\delta \pi_\ell} = \partial_0 \psi_\ell - \frac{\partial \mathscr{H}_M}{\partial \pi_\ell}\,,$$

$$0 = \frac{\delta I}{\delta \Pi_{\alpha i}} = \partial_0 A_{\alpha i} - \Pi_{\alpha i} - \partial_i A_{\alpha 0} + C_{\alpha\beta\gamma}A_{\beta 0}A_{\gamma i} = F_{\alpha 0 i} - \Pi_{\alpha i}\,.$$

Inserting these back into Eq. (15.4.13) gives

$$\begin{aligned} I &= \int d^4x \left[\mathscr{L}_M + \tfrac{1}{2} F_{\alpha 0 i} F_{\alpha 0 i}\right. \\ &\qquad \left. - \tfrac{1}{4} F_{\alpha i j}F_{\alpha i j} - \tfrac{1}{2}\partial_3 A_{\alpha i}\partial_3 A_{\alpha i} + \tfrac{1}{2}(\partial_3 A_{\alpha 0})^2\right] \\ &= \int d^4x\, \mathscr{L}\,, \end{aligned} \tag{15.4.14}$$

where $\mathscr{L}$ is the Lagrangian (15.3.1) with which we started! In other words, *we are to do path integrals over $\psi_\ell(x)$ and all four components of $A_{\alpha\mu}(x)$, with a manifestly covariant weighting factor* $\exp(iI)$ *given by Eqs. (15.4.14) and (15.3.1)*, but with the axial-gauge condition enforced by inserting a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

factor

$$\prod_{x,\alpha} \delta\Big(A_{\alpha 3}(x)\Big) \,. \tag{15.4.15}$$

As long as $\mathcal{O}_A, \mathcal{O}_B \cdots$ are gauge-invariant, we have

$$\begin{aligned} \langle T\{\mathcal{O}_A \mathcal{O}_B \cdots\}\rangle_{\text{VACUUM}} \;\propto\; & \int \left[\prod_{\ell,x} d\psi_\ell(x)\right] \left[\prod_{\alpha,\mu,x} dA_{\alpha\mu}(x)\right] \\ & \times \mathcal{O}_A \mathcal{O}_B \cdots \exp\{iI + \epsilon \text{ terms}\} \prod_{x,\alpha} \delta\left(A_{\alpha 3}(x)\right) \,, \end{aligned} \tag{15.4.16}$$

with Lorentz- and gauge-invariant action I given by Eq. (15.4.14).

* * *

For future reference, we note that the volume element $\prod_{\alpha,\mu,x} dA_{\alpha\mu}(x)$ for the integration over gauge fields in (15.4.16) is gauge-invariant, in the sense that

$$\prod_{\alpha,\mu,x} dA_{\Lambda\,\alpha\mu}(x) = \prod_{\alpha,\mu,x} dA_{\alpha\mu}(x) \,, \tag{15.4.17}$$

where $A_{\Lambda\,\alpha\mu}(x)$ is the result of acting on $A_{\alpha\mu}(x)$ with a gauge transformation having transformation parameters $\Lambda_\alpha(x)$. It will be enough to show that this is true for transformations near the identity, say with infinitesimal transformation parameters $\lambda_\alpha(x)$. In this case,

$$A^\mu_{\lambda\,\alpha} = A^\mu_\alpha + \partial^\mu \lambda_\alpha + C_{\alpha\beta\gamma} A^\mu_\beta \lambda_\gamma \,,$$

so the volume elements are related by

$$\prod_{\alpha,\mu,x} dA_{\lambda\,\alpha\mu}(x) = \text{Det}(\mathcal{N}) \prod_{\alpha,\mu,x} dA_{\alpha\mu}(x) \,,$$

where $\mathcal{N}$ is the 'matrix':

$$\mathcal{N}_{\alpha\mu x,\beta\nu y} = \frac{\delta A_{\lambda\,\alpha\mu}(x)}{\delta A_{\beta\nu}(y)} = \delta^4(x-y)\,\delta^\nu_\mu \Big[\delta_{\alpha\beta} + C_{\alpha\beta\gamma}\lambda_\gamma(x)\Big] \,.$$

The determinant of $\mathcal{N}$ is unity to first order in λ_γ because the trace $C_{\alpha\alpha\gamma}$ vanishes.

In this chapter we shall assume that the volume element $\prod_{n,x} d\psi_n(x)$ for the integration over matter fields is also gauge-invariant. There are important subtleties here, to which we shall return in Chapter 22, but as shown there this assumption turns out to be valid in our present non-Abelian gauge theories of strong and electroweak interactions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

15.5 The De Witt–Faddeev–Popov Method

Our formula (15.4.16) for the path integral was derived in a gauge that is convenient for canonical quantization, but the Feynman rules that would be derived from this formula would hide the underlying rotational and Lorentz invariance of the theory. In order to derive manifestly Lorentz-invariant Feynman rules, we need to change the gauge.

We first note that Eq. (15.4.16) is (up to an unimportant constant factor) a special case of a general class of functional integrals, of the form:

$$\mathcal{I} = \int \left[\prod_{n,x} d\phi_n(x)\right] \mathcal{G}[\phi]\, B\Big[f[\phi]\Big] \operatorname{Det} \mathcal{F}[\phi]\,, \tag{15.5.1}$$

where $\phi_n(x)$ are a set of gauge and matter fields; $\prod_{n,x} d\phi_n(x)$ is a volume element; and $\mathcal{G}[\phi]$ is a functional of the $\phi_n(x)$, satisfying the gauge-invariance condition:

$$\mathcal{G}[\phi_\lambda] \prod_{n,x} d\phi_{\lambda n}(x) = \mathcal{G}[\phi] \prod_{n,x} d\phi_n(x)\,, \tag{15.5.2}$$

where $\phi_{\lambda n}(x)$ is the result of operating on ϕ with a gauge transformation having parameters $\lambda_\alpha(x)$. (Usually when this is satisfied both the functional $\mathcal{G}$ and the volume element are separately invariant, but Eq. (15.5.2) is all we need here.) Also, $f_\alpha[\phi; x]$ is a non-gauge-invariant 'gauge-fixing functional' of these fields that also depends on x and α; $B[f]$ is some numerical functional defined for general functions $f_\alpha(x)$ of x and α; and $\mathcal{F}$ is the 'matrix':

$$\mathcal{F}_{\alpha x,\beta y}[\phi] \equiv \left.\frac{\delta f_\alpha[\phi_\lambda; x]}{\delta \lambda_\beta(y)}\right|_{\lambda=0}. \tag{15.5.3}$$

(In accordance with our usual notation for functionals of functions or of functionals, $B\Big[f[\phi]\Big]$ is understood to depend on the values taken by $f_\alpha[\phi; x]$ for all values of the undisplayed variables α and x, with the displayed variable, the function $\phi_n(x)$, held fixed.) Eq. (15.5.1) does not represent the widest possible generalization of Eq. (15.4.16); we will see in Section 15.7 that there is a further generalization that is needed for some purposes. We start here with Eq. (15.5.1) because it will help to motivate the formalism of Section 15.7, and it is adequate for dealing with non-Abelian gauge theories in the most convenient gauges.

We now must check that the path integral (15.4.16) is in fact a special case of Eq. (15.5.1). In Eq. (15.4.16) the fields $\phi_n(x)$ consist of both $A_{\alpha\mu}(x)$ and matter fields $\psi_\ell(x)$, and

$$f_\alpha[A, \psi; x] = A_{\alpha 3}(x)\,, \tag{15.5.4}$$

$$B[f] = \prod_{x,\alpha} \delta\Big(f_\alpha(x)\Big)\,, \tag{15.5.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

$$\mathscr{G}[A,\psi] = \exp\{iI + \epsilon \text{ terms}\}\ \mathscr{O}_A\mathscr{O}_B\cdots\,, \tag{15.5.6}$$

$$\prod_{n,x} d\phi_n(x) = \left[\prod_{\ell,x} d\psi_\ell(x)\right]\ \left[\prod_{\alpha,\mu,x} dA^\mu_\alpha(x)\right]\,. \tag{15.5.7}$$

(We are now dropping the distinction between upper and lower indices $\alpha,\beta,\cdots$.) Comparison of Eq. (15.4.16) with Eqs. (15.5.1)–(15.5.3) shows that these path integrals are indeed the same, aside from the factor $\text{Det}\,\mathscr{F}[\phi]$. For the particular gauge-fixing functional (15.5.4), this factor is field-independent: if $A^3_\alpha(x) = 0$, then the change in $A^3_\alpha(x)$ under a gauge transformation with parameters $\lambda_\alpha(x)$ is

$$A^3_{\lambda\alpha}(x) = \partial_3\,\lambda_\alpha(x) = \int d^4y\ \lambda_\alpha(y)\,\partial_3\delta^4(x-y)\,,$$

so that here Eq. (15.5.3) is the field-independent 'matrix'

$$\mathscr{F}_{\alpha x,\beta y}[\phi] = \delta_{\alpha\beta}\,\partial_3\delta^4(x-y)\,.$$

The determinant in Eq. (15.5.1) is therefore also field-independent in this gauge. As discussed in Chapter 9, field-independent factors in the functional integral affect only the vacuum-fluctuation part of expectation values and S-matrix elements, and so are irrelevant to the calculation of the connected parts of the S-matrix.

The point of recognizing the functional integral (15.4.16) for non-Abelian gauge theories as a special case of the general path integral (15.5.1) is that in this form we may freely change the gauge. Specifically, we have a theorem, that *the integral (15.5.1) is actually independent (within broad limits) of the gauge-fixing functional $f_\alpha[\phi;x]$, and depends on the choice of the functional $B[f]$ only through an irrelevant constant factor.*

Proof: Replace the integration variable ϕ everywhere in Eq. (15.5.1) with a new integration variable ϕ_Λ, with $\Lambda^\alpha(x)$ any arbitrary (but fixed) set of gauge transformation parameters:

$$\mathscr{I} = \int \left[\prod_{n,x} d\phi_{\Lambda n}(x)\right]\ \mathscr{G}[\phi_\Lambda]B\Big[f[\phi_\Lambda]\Big]\,\text{Det}\,\mathscr{F}[\phi_\Lambda]\,. \tag{15.5.8}$$

(This step is a mathematical triviality, like changing an integral $\int_{-\infty}^{\infty} f(x)dx$ to read $\int_{-\infty}^{\infty} f(y)dy$, and does not yet make use of our assumptions regarding gauge invariance.) Now use the assumed gauge invariance (15.5.2) of the measure $\Pi d\phi$ times the functional $\mathscr{G}[\phi]$ to rewrite this as

$$\mathscr{I} = \int \left[\prod_{n,x} d\phi_n(x)\right]\ \mathscr{G}[\phi]B\Big[f[\phi_\Lambda]\Big]\ \text{Det}\,\mathscr{F}[\phi_\Lambda]\,. \tag{15.5.9}$$

Since $\Lambda^\alpha(x)$ was arbitrary, the left-hand side here cannot depend on it. Integrating over $\Lambda^\alpha(x)$ with some suitable weight-functional $\rho[\Lambda]$ (to be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

chosen below) thus gives

$$\mathscr{I}\int\left[\prod_{\alpha,x}d\Lambda^{\alpha}(x)\right]\rho[\Lambda]=\int\left[\prod_{n,x}d\phi_n(x)\right]\mathscr{G}[\phi]C[\phi]\,, \tag{15.5.10}$$

where

$$C[\phi]\equiv\int\left[\prod_{\alpha,x}d\Lambda^{\alpha}(x)\right]\rho[\Lambda]\,B\Big[f[\phi_\Lambda]\Big]\,\mathrm{Det}\,\mathscr{F}[\phi_\Lambda]\,. \tag{15.5.11}$$

Now, Eq. (15.5.3) gives

$$\mathscr{F}_{\alpha x,\beta y}[\phi_\Lambda]=\frac{\delta f_\alpha[(\phi_\Lambda)_\lambda;x]}{\delta\lambda^\beta(y)}\bigg|_{\lambda=0}\,. \tag{15.5.12}$$

We are assuming that these transformations form a group; that is, we may write the result of performing the gauge transformation with parameters $\Lambda^\alpha(x)$ followed by the gauge transformation with parameters $\lambda^\alpha(x)$ as the action of a single 'product' gauge transformation with parameters $\tilde{\Lambda}^\alpha(x;\Lambda,\lambda)$,

$$(\phi_\Lambda)_\lambda=\phi_{\tilde{\Lambda}(\Lambda,\lambda)}\,. \tag{15.5.13}$$

Using the chain rule of partial (functional) differentiation, we have then

$$\mathscr{F}_{\alpha x,\beta y}[\phi_\Lambda]=\int\mathscr{J}_{\alpha x,\gamma z}[\phi,\Lambda]\mathscr{R}^{\gamma z}{}_{\beta y}[\Lambda]d^4z\,, \tag{15.5.14}$$

where

$$\mathscr{J}_{\alpha x,\gamma z}[\phi,\Lambda]\equiv\frac{\delta f_\alpha[\phi_{\tilde{\Lambda}};x]}{\delta\tilde{\Lambda}^\gamma(z)}\bigg|_{\tilde{\Lambda}=\Lambda}=\frac{\delta f_\alpha[\phi_\Lambda;x]}{\delta\Lambda^\gamma(z)} \tag{15.5.15}$$

and

$$\mathscr{R}^{\gamma z}{}_{\beta y}[\Lambda]=\frac{\delta\tilde{\Lambda}^\gamma(z;\Lambda,\lambda)}{\delta\lambda^\beta(y)}\bigg|_{\lambda=0}\,. \tag{15.5.16}$$

It follows that

$$\mathrm{Det}\,\mathscr{F}[\phi_\Lambda]=\mathrm{Det}\,\mathscr{J}[\phi,\Lambda]\;\;\mathrm{Det}\,\mathscr{R}[\Lambda]\,. \tag{15.5.17}$$

We note that Det $\mathscr{J}[\phi,\Lambda]$ is nothing but the Jacobian of the transformation of integration variables from the $\Lambda^\alpha(x)$ to (for a fixed ϕ) the $f_\alpha[\phi_\Lambda;x]$. Hence, if we choose the weight-function $\rho(\Lambda)$ as

$$\rho(\Lambda)=1\Big/\mathrm{Det}\,\mathscr{R}[\Lambda] \tag{15.5.18}$$

then

$$\begin{aligned}C[\phi]&=\int\left[\prod_{\alpha,x}d\Lambda^{\alpha}(x)\right]\;\mathrm{Det}\,\mathscr{J}[\phi,\Lambda]\,B\Big[f[\phi_\Lambda]\Big]\\&=\int\left[\prod_{\alpha,x}df_\alpha(x)\right]\;B\;[f]\equiv C\,,\end{aligned} \tag{15.5.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

which is clearly independent of ϕ. (Eq. (15.5.18) may be recognized by the reader as giving the invariant (Haar) measure on the space of group parameters.) We have then at last

$$\mathcal{I} = \frac{C\int\left[\prod_{n,x} d\phi_n(x)\right]\mathcal{G}[\phi]}{\int\left[\prod_{\alpha,x} d\Lambda^\alpha(x)\right]\rho[\Lambda]} . \tag{15.5.20}$$

This is clearly independent of our choice of $f_\alpha[\phi;x]$, which has been reduced to a mere variable of integration, and it depends on $B[f]$ only through the constant C, as was to be proved.

Before proceeding with the applications of this theorem, we should pause to note a tricky point in the derivation. The integrals in the numerator and denominator of Eq. (15.5.20) are both ill-defined for the same reason. Since $\mathcal{G}[\phi]$ is assumed to be gauge-invariant, its integral over ϕ cannot possibly converge; the integrand is constant along all 'orbits,' obtained by sending ϕ into ϕ_λ with all possible $\lambda^\alpha(x)$. Likewise, the integrand in the denominator is divergent, because $\rho(\Lambda)\Pi d\Lambda$ is nothing but the usual invariant volume element for integrating over the group, and this is also constant along 'orbits' $\Lambda \to \tilde{\Lambda}(\Lambda,\lambda)$. This divergence can be eliminated in both the numerator and denominator of Eq. (15.5.20) by formulating the theory on a finite spacetime lattice, in which case the volume of the gauge group is just the volume of the global Lie group itself times the number of lattice sites. Because the gauge-fixing factor $B[f]$ eliminates this divergence in the original definition (15.5.1) of the left-hand side of Eq. (15.5.20), we may presume that, as the number of lattice sites goes to infinity, it cancels between the numerator and denominator of the right-hand side of Eq. (15.5.20).

Now to the point. We have seen that the vacuum expectation value (15.4.16) in axial gauge is given by a functional integral of the general form (15.5.1). Armed with the above theorem, we conclude then that

$$\langle T\{\mathcal{O}_A\mathcal{O}_B\cdots\}\rangle_V \propto \int\left[\prod_{\ell,x} d\psi_\ell(x)\right]\left[\prod_{\alpha,\mu,x} dA^\mu{}_\alpha(x)\right]$$
$$\times\, \mathcal{O}_A\mathcal{O}_B\cdots\exp\{iI+\epsilon\text{ terms}\}\, B\Big[f[A,\psi]\Big]\operatorname{Det}\mathcal{F}[A,\psi] \tag{15.5.21}$$

for (almost) any choice of $f_\alpha[A,\psi;x]$ and $B[f]$. We are now therefore free to use Eq. (15.5.21) to derive the Feynman rules in a more convenient gauge.

The path integrals that we understand how to calculate are of Gaussians times polynomials, so we will generally take

$$B[f] = \exp\left(-\frac{i}{2\xi}\int d^4x\, f_\alpha(x)f_\alpha(x)\right) \tag{15.5.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

with arbitrary real parameter ξ. With this choice, the effect of the factor B in Eq. (15.5.21) is just to add a term to the effective Lagrangian

$$\mathscr{L}_{\text{EFF}} = \mathscr{L} - \frac{1}{2\xi} f_\alpha f_\alpha \,. \tag{15.5.23}$$

The simplest Lorentz-invariant choice of the gauge-fixing function f_α is the same as in electrodynamics:

$$f_\alpha = \partial_\mu A^\mu_\alpha \,. \tag{15.5.24}$$

The bare gauge-field propagator can then be calculated just as in quantum electrodynamics. The free-vector-boson part of the effective action can be written

$$\begin{aligned} I_{0A} &= -\int d^4x \left[\frac{1}{4}(\partial_\mu A_{\alpha\nu} - \partial_\nu A_{\alpha\mu})(\partial^\mu A_\alpha{}^\nu - \partial^\nu A_\alpha{}^\mu) \right. \\ &\qquad \left. + \frac{1}{2\xi}(\partial_\mu A_\alpha{}^\mu)(\partial_\nu A_\alpha{}^\nu) \;+\; \epsilon \text{ terms} \right] \\ &= -\frac{1}{2}\int d^4x \, \mathscr{D}_{\alpha\mu x,\beta\nu y} \, A_\alpha{}^\mu(x) A_\beta{}^\nu(y) \,, \end{aligned}$$

where

$$\begin{aligned} \mathscr{D}_{\alpha\mu x,\beta\nu y} &= \eta_{\mu\nu} \frac{\partial^2}{\partial x^\lambda \partial y_\lambda} \delta^4(x-y)\delta_{\alpha\beta} \\ &\quad - \left(1 - \frac{1}{\xi}\right) \frac{\partial^2}{\partial x^\mu \partial y^\nu} \delta^4(x-y)\delta_{\alpha\beta} \;+\; \epsilon \text{ terms} \\ &= (2\pi)^{-4}\delta_{\alpha\beta} \int d^4p \left[\eta_{\mu\nu}(p^2 - i\epsilon) - \left(1 - \frac{1}{\xi}\right) p_\mu p_\nu \right] e^{ip\cdot(x-y)} \,. \end{aligned}$$

Taking the reciprocal of the matrix in square brackets, we find the propagator:

$$\begin{aligned} \Delta_{\alpha\mu,\beta\nu}(x,y) &= (\mathscr{D}^{-1})_{\alpha\mu x,\beta\nu y} \\ &= (2\pi)^{-4} \int d^4p \left[\eta_{\mu\nu} + (\xi - 1)\frac{p_\mu p_\nu}{p^2} \right] \frac{e^{ip\cdot(x-y)}}{p^2 - i\epsilon} \,. \end{aligned} \tag{15.5.25}$$

This is a generalization of both Landau and Feynman gauges, which are recovered by taking $\xi = 0$ and $\xi = 1$, respectively. For $\xi \to 0$, the functional (15.5.22) oscillates very rapidly except near $f_\alpha = 0$, so this functional acts like a delta-function imposing the Landau gauge condition $\partial_\mu A^\mu = 0$, leading naturally to a propagator satisfying the corresponding condition $\partial^\mu \Delta_{\alpha\mu,\beta\nu} = 0$. For non-zero values of ξ the functional $B[f]$ does not pick out gauge fields satisfying any specific gauge condition on the field $A_{\alpha\mu}$, but it is common to refer to the propagator (15.5.25) as being in a 'generalized Feynman gauge' or 'generalized ξ-gauge'. It is often a good

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

strategy to calculate physical amplitudes with ξ left arbitrary, and then at the end of the calculation check that the results are ξ-independent.

With one qualification, the Feynman rules are now obvious: the contributions of vertices are to be read off from the interaction terms in the original Lagrangian $\mathscr{L}$, with gauge-field propagators given by Eq. (15.5.25), and matter-field propagators calculated as before. To be specific, the trilinear interaction term in $\mathscr{L}$

$$-\tfrac{1}{2}\, C_{\alpha\beta\gamma}(\partial_\mu A_{\alpha\nu} - \partial_\nu A_{\alpha\mu})A_\beta{}^\mu A_\gamma{}^\nu$$

corresponds to a vertex to which are attached three vector boson lines. If these lines carry (incoming) momenta p, q, k and Lorentz and gauge-field indices $\mu\alpha, \nu\beta, \lambda\gamma$, then according to the momentum-space Feynman rules, the contribution of such a vertex to the integrand is

$$i(2\pi)^4\delta^4(p+q+k)\,[-i\,C_{\alpha\beta\gamma}]\left[p_\nu\eta_{\mu\lambda} - p_\lambda\eta_{\mu\nu} + q_\lambda\eta_{\nu\mu} - q_\mu\eta_{\nu\lambda} + k_\mu\eta_{\lambda\nu} - k_\nu\eta_{\lambda\mu}\right]. \tag{15.5.26}$$

Also, the A^4 interaction term in $\mathscr{L}$,

$$-\tfrac{1}{4}\, C_{\epsilon\alpha\beta}C_{\epsilon\gamma\delta}\, A_{\alpha\mu}A_{\beta\nu}A_\gamma{}^\mu A_\delta{}^\nu\ ,$$

corresponds to a vertex to which are attached four vector boson lines. If these lines carry (incoming) momenta p, q, k, ℓ, and Lorentz and gauge indices $\mu\alpha$, $\nu\beta$, $\rho\gamma$, and $\sigma\delta$, then the contribution of such a vertex to the integrand is

$$\begin{aligned} &i(2\pi)^4\delta^4(p+q+k+\ell)\ \times\ \Big[-C_{\epsilon\alpha\beta}\,C_{\epsilon\gamma\delta}(\eta_{\mu\rho}\eta_{\nu\sigma} - \eta_{\mu\sigma}\eta_{\nu\rho}) \\ &-C_{\epsilon\alpha\gamma}\,C_{\epsilon\delta\beta}(\eta_{\mu\sigma}\eta_{\rho\nu} - \eta_{\mu\nu}\eta_{\sigma\rho}) - C_{\epsilon\alpha\delta}\,C_{\epsilon\beta\gamma}(\eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\rho}\eta_{\sigma\nu})\Big]. \end{aligned} \tag{15.5.27}$$

(Recall that the structure constants $C_{\alpha\beta\gamma}$ contain coupling constant factors, so the factors (15.5.26) and (15.5.27) are respectively of first and second order in coupling constants.)

The one complication in the Feynman rules with which we have not yet dealt is the presence in Eq. (15.5.21) of the factor Det $\mathscr{F}$, which for general gauges is *not* a constant. We now turn to a consideration of this factor.

15.6 Ghosts

We now consider the effect of the factor Det $\mathscr{F}$ in Eq. (15.5.21) on the Feynman rules for a non-Abelian gauge theory. In order to be able to treat this effect as a modification of the Feynman rules, recall that as shown in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Section 9.5, the determinant of any matrix $\mathscr{F}_{\alpha x,\beta y}$ may be expressed as a path integral

$$\text{Det}\,\mathscr{F} \propto \int \left[\prod_{\alpha,x} d\omega^*_\alpha(x)\right] \left[\prod_{\alpha,x} d\omega_\alpha(x)\right] \exp(iI_{GH})\,, \tag{15.6.1}$$

where

$$I_{GH} \equiv \int d^4x\, d^4y\, \omega^*_\alpha(x)\, \omega_\beta(y)\, \mathscr{F}_{\alpha x,\beta y}\,. \tag{15.6.2}$$

Here ω^*_α and ω_α are a set of independent anticommuting classical variables, and the constant of proportionality is field-independent. (We have to choose the ω_α and ω^*_α field variables to be fermionic in order to reproduce the factor Det $\mathscr{F}$; had we chosen these field variables to be bosonic, the path integral (15.6.1) would have been proportional to (Det $\mathscr{F}$)$^{-1}$.) The fields ω^*_α and ω_α are not necessarily related by complex conjugation; indeed, in Section 15.7 we shall see that for some purposes we need to assume that ω^*_α and ω_α are independent *real* variables. The whole effect of the factor Det $\mathscr{F}$ is the same as that of including $I_{GH}[\omega,\omega^*]$ in the full effective action, and integrating over 'fields' ω and ω^*. That is, for arbitrary gauge-fixing functionals $f_\alpha(x)$,

$$\langle T\{\mathscr{O}_A\cdots\}\rangle_V \propto \int \left[\prod_{n,x} d\psi_n(x)\right] \left[\prod_{\alpha,\mu,x} dA_{\alpha\mu}(x)\right]$$
$$\times \left[\prod_{\alpha,x} d\omega_\alpha(x)\, d\omega^*_\alpha(x)\right] \exp\left(i I_{\text{MOD}}[\psi,A,\omega,\omega^*]\right) \mathscr{O}_A\cdots\,, \tag{15.6.3}$$

where I_{MOD} is a modified action

$$I_{\text{MOD}} = \int d^4x \left[\mathscr{L} - \frac{1}{2\xi} f_\alpha f_\alpha\right] + I_{GH}\,. \tag{15.6.4}$$

The fields ω_α and ω^*_α are Lorentz scalars (at least in covariant gauges) but satisfy Fermi statistics. The connection between spin and statistics is not really violated here, because there are no particles described by these fields that can appear in initial or final states. For that reason, ω_α and ω^*_α are called the fields of 'ghost' and 'antighost' particles. Inspection of Eq. (15.6.2) shows that the action respects the conservation of a quantity known as 'ghost number,' equal to +1 for ω_α, −1 for ω^*_α, and zero for all other fields.

The Feynman rules for the ghosts are simplest in the case in which the 'matrix' $\mathscr{F}$ may be expressed as

$$\mathscr{F} = \mathscr{F}_0 + \mathscr{F}_1\,, \tag{15.6.5}$$

where $\mathscr{F}_0$ is field-independent and of zeroth order in coupling constants, while $\mathscr{F}_1$ is field-dependent and proportional to one or more coupling

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

constant factors. In this case, the ghost propagator is just

$$\Delta_{\alpha\beta}(x,y) = -(\mathscr{F}_0^{-1})_{\alpha x,\beta y} \tag{15.6.6}$$

and the ghost vertices are to be read off from the interaction term

$$I'_{GH} = \int d^4x\, d^4y\, \omega^*_\alpha(x)\, \omega_\beta(y)(\mathscr{F}_1)_{\alpha x,\beta y}\ . \tag{15.6.7}$$

For instance, in the generalized ξ-gauge discussed in the previous section, we have

$$f_\alpha = \partial_\mu A^\mu_\alpha \tag{15.6.8}$$

and for infinitesimal gauge parameters λ_α, Eq. (15.1.9) gives:

$$A^\mu_{\alpha\lambda} = A^\mu_\alpha + \partial^\mu \lambda_\alpha + C_{\alpha\gamma\beta}\lambda_\beta A^\mu_\gamma$$

so that

$$\begin{aligned}\mathscr{F}_{\alpha x,\beta y} &= \frac{\delta \partial_\mu A^\mu_{\alpha\lambda}(x)}{\delta \lambda_\beta(y)}\bigg|_{\lambda=0} \\ &= \Box\, \delta^4(x-y) + C_{\alpha\gamma\beta}\frac{\partial}{\partial x^\mu}\left[A^\mu_\gamma(x)\, \delta^4(x-y)\right]\ .\end{aligned} \tag{15.6.9}$$

This is of the form (15.6.5), with

$$(\mathscr{F}_0)_{\alpha x,\beta y} = \Box\delta^4(x-y)\, \delta_{\alpha\beta}\ , \tag{15.6.10}$$

$$(\mathscr{F}_1)_{\alpha x,\beta y} = -C_{\alpha\beta\gamma}\frac{\partial}{\partial x^\mu}\left[A^\mu_\gamma(x)\delta^4(x-y)\right]\ . \tag{15.6.11}$$

From Eqs. (15.6.6) and (15.6.10), we see that the ghost propagator is

$$\Delta_{\alpha\beta}(x,y) = \delta_{\alpha\beta}(2\pi)^{-4}\int d^4p\, (p^2 - i\epsilon)^{-1}\, e^{ip\cdot(x-y)}\ , \tag{15.6.12}$$

so in this gauge the ghosts behave like spinless fermions of zero mass, transforming according to the adjoint representation of the gauge group. Using Eqs. (15.6.7) and (15.6.11) and integrating by parts, we find that the ghost interaction term in the action is now

$$I'_{\mathrm{GH}} = \int d^4x\, C_{\alpha\beta\gamma}\frac{\partial \omega^*_\alpha}{\partial x^\mu} A^\mu_\gamma\, \omega_\beta\ . \tag{15.6.13}$$

This interaction corresponds to vertices to which are attached one outgoing ghost line, one incoming ghost line, and one vector boson line. If these lines carry (incoming) momenta p, q, k respectively and gauge group indices α, β, γ respectively, and the gauge field carries a vector index μ, then the contribution of such a vertex to the integrand is given by the momentum-space Feynman rules as

$$i(2\pi)^4\delta^4(p+q+k)\ \times\ ip_\mu C_{\alpha\beta\gamma}\ . \tag{15.6.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

The ghosts propagate around loops, with single vector boson lines attached at each vertex along the loops, and with an extra minus sign supplied for each loop as is usual for fermionic field variables.

The extra minus sign for ghost loops suggests that each ghost field ω_α together with the associated antighost field ω_α^* represents something like a negative degree of freedom. These negative degrees of freedom are necessary because in using covariant gauge field propagators we are really over-counting; the physical degrees of freedom are the components of $A_\alpha^\mu(x)$, *less* the parameters $\Lambda_\alpha(x)$ needed to describe a gauge transformation.

In summary, the modified action (15.6.4) may be written in generalized ξ-gauge as

$$I_{\text{MOD}} = \int d^4x \, \mathscr{L}_{\text{MOD}} \tag{15.6.15}$$

with a modified Lagrangian density:

$$\begin{aligned}\mathscr{L}_{\text{MOD}} = \mathscr{L}_{\text{M}} &- \frac{1}{4}F_\alpha^{\mu\nu}F_{\alpha\mu\nu} - \frac{1}{2\xi}(\partial_\mu A_\alpha^\mu)(\partial_\nu A_\alpha^\nu) \\ &-\partial_\mu\omega_\alpha^*\,\partial^\mu\omega_\alpha + C_{\alpha\beta\gamma}(\partial_\mu\omega_\alpha^*)\,A_\gamma^\mu\,\omega_\beta\ .\end{aligned} \tag{15.6.16}$$

It is important that this Lagrangian is *renormalizable* (if the matter Lagrangian $\mathscr{L}_{\text{M}}$ is), in the elementary sense that its terms involve products of fields and their derivatives of total dimensionality (in powers of mass) four or less. (The kinematic term $-\partial_\mu\omega_\alpha^*\,\partial^\mu\omega_\alpha$ in Eq. (15.6.16) fixes the dimensionality of the fields ω and ω^* to be mass to the power unity, just like ordinary scalar and gauge fields.) However, there is more to renormalizability than power counting; it is necessary also that there be a counterterm to absorb every divergence. In the next section we shall consider a remarkable symmetry that will be used in Section 17.2 to show that non-Abelian gauge theories are indeed renormalizable in this sense, and that can even take the place of the Faddeev–Popov–De Witt approach that we have been following.

15.7 BRST Symmetry

Although the Faddeev–Popov–De Witt method described in the previous two sections makes the Lorentz invariance of the theory manifest, it still rests on a choice of gauge, and hence naturally it hides the underlying gauge invariance of the theory. This is a serious problem in trying to prove the renormalizability of the theory — gauge invariance restricts the form of the terms in the Lagrangian that are available as counterterms to absorb ultraviolet divergences, but once we choose a gauge, how do we

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

know that gauge invariance still restricts the ways that the infinities can appear?

Remarkably, however, even after we choose a gauge, the path integral still does have a symmetry related to gauge invariance. This symmetry was discovered by Becchi, Rouet, and Stora,[10] (and independently by Tyutin,[11]) in 1975, several years after the work of Faddeev and Popov and De Witt, and is known in honor of its discoverers as BRST symmetry. This symmetry will be presented more-or-less as it was originally discovered, as a by-product of the method of Faddeev, Popov, and De Witt, but as we shall see it can also be regarded as a replacement for the Faddeev–Popov–De Witt approach.

We have seen in Eqs. (15.6.3) and (15.6.4) that the Feynman rules for a non-Abelian gauge theory may be obtained from a path integral over matter, gauge, and ghost fields, with a modified action, which we may write

$$I_{\rm MOD} = I_{\rm EFF} + I_{GH} = \int d^4x\, \mathscr{L}_{\rm MOD}\,, \tag{15.7.1}$$

$$\mathscr{L}_{\rm MOD} \equiv \mathscr{L} - \frac{1}{2\xi} f_\alpha f_\alpha + \omega^*_\alpha \Delta_\alpha\,, \tag{15.7.2}$$

where we have now introduced the quantity

$$\Delta_\alpha(x) \equiv \int d^4y\, \mathscr{F}_{\alpha x,\beta y}[A,\psi]\, \omega_\beta(y)\,. \tag{15.7.3}$$

This is for the choice

$$B[f] \propto \exp\left(-\frac{i}{2\xi}\int d^4x\, f_\alpha f_\alpha\right) \tag{15.7.4}$$

of the gauge-fixing functional in Eq. (15.5.21). For our present purposes, it will be helpful to rewrite $B[f]$ as a Fourier integral:

$$B[f] = \int \left[\prod_{\alpha,x} dh_\alpha(x)\right] \exp\left[\frac{i\xi}{2}\int d^4x\, h_\alpha h_\alpha\right] \exp\left[i\int d^4x\, f_\alpha h_\alpha\right]\,. \tag{15.7.5}$$

We must now do our path integrals over the field h_α (often known as a 'Nakanishi–Lautrup' field[11a]) as well as over matter, gauge, ghost and antighost fields, with a new modified action

$$I_{\rm NEW} = \int d^4x \left(\mathscr{L} + \omega^*_\alpha \Delta_\alpha + h_\alpha f_\alpha + \tfrac{1}{2}\xi h_\alpha h_\alpha\right)\,. \tag{15.7.6}$$

This modified action is not gauge-invariant — indeed, it had better not be, if we are to be able to use it in path integrals. However, it *is* invariant under a 'BRST' symmetry transformation, parameterized by an

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

infinitesimal constant θ that *anticommutes* with ω_α, ω_α^*, and all fermionic matter fields. For a given θ, the BRST transformation is

$$\delta_\theta \psi = i t_\alpha \theta \omega_\alpha \psi \,, \tag{15.7.7}$$

$$\delta_\theta A_{\alpha\mu} = \theta D_\mu \omega_\alpha = \theta[\partial_\mu \omega_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu} \omega_\gamma] \,, \tag{15.7.8}$$

$$\delta_\theta \omega_\alpha^* = -\theta h_\alpha \,, \tag{15.7.9}$$

$$\delta_\theta \omega_\alpha = -\tfrac{1}{2}\theta\, C_{\alpha\beta\gamma} \omega_\beta \omega_\gamma \,, \tag{15.7.10}$$

$$\delta_\theta h_\alpha = 0 \,. \tag{15.7.11}$$

(Recall that in fermionic path integrals, there is no connection between ω_α and ω_α^*, so that Eq. (15.7.9) does not need to be the adjoint of Eq. (15.7.10).) Because h_α is BRST-invariant, we could if we like replace the Gaussian factor $\exp(\frac{1}{2} i\xi \int h_\alpha h_\alpha)$ in Eq. (15.7.5) with an arbitrary smooth functional of h_α, yielding an arbitrary functional $B[f]$, without affecting the BRST invariance of the action. However, for the purposes of diagrammatic calculation and renormalization it will help to leave $B[f]$ as a Gaussian.

In checking the invariance of the action (15.7.1), it will be very useful first to note that the transformation (15.7.7)–(15.7.11) is *nilpotent*; that is, if F is any functional of $\psi, A, \omega, \omega^*$, and h, and we define sF by

$$\delta_\theta F \equiv \theta s F \tag{15.7.12}$$

then*

$$\delta_\theta (sF) = 0 \tag{15.7.13}$$

or equivalently

$$s(sF) = 0 \,. \tag{15.7.14}$$

It is straightforward to verify this nilpotence when δ_θ acts on a single field. First, acting on a matter field,

$$\begin{aligned} \delta_\theta s\psi = i t_\alpha\, \delta_\theta(\omega_\alpha \psi) &= -\tfrac{1}{2} i\, C_{\alpha\beta\gamma} t_\alpha \theta \omega_\beta \omega_\gamma \psi - t_\alpha t_\beta \omega_\alpha \theta \omega_\beta \psi \\ &= -\tfrac{1}{2} i\, C_{\alpha\beta\gamma} t_\alpha \theta \omega_\beta \omega_\gamma \psi + t_\alpha t_\beta \theta \omega_\alpha \omega_\beta \psi \,. \end{aligned}$$

The product $\omega_\alpha \omega_\beta$ in the second term on the right is antisymmetric in α and β, so we can replace $t_\alpha t_\beta$ in this term with $\frac{1}{2}[t_\alpha, t_\beta]$, and this term thus cancels the first term:

$$ss\psi = 0 \,. \tag{15.7.15}$$

* In the original work on BRST symmetry the functional $B[f]$ was left in the form (15.7.4), so that h_α was replaced in Eq. (15.7.9) with $-f_\alpha/\xi$, and the BRST transformation was only nilpotent when acting on functions of ω_α and the gauge and matter fields, but not of ω_α^*.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Next, acting on a gauge field, we have

$$\begin{aligned}\delta_\theta s A_{\alpha\mu} &= \delta_\theta D_\mu \omega_\alpha \\ &= \partial_\mu \delta_\theta \omega_\alpha + C_{\alpha\beta\gamma}\delta_\theta A_{\beta\mu}\omega_\gamma + C_{\alpha\beta\gamma}A_{\beta\mu}\delta_\theta\omega_\gamma \\ &= \theta\Big(-\tfrac{1}{2}C_{\alpha\beta\gamma}\partial_\mu(\omega_\beta\omega_\gamma) + C_{\alpha\beta\gamma}(\partial_\mu\omega_\beta)\omega_\gamma \\ &\quad + C_{\alpha\beta\gamma}C_{\beta\delta\epsilon}A_{\delta\mu}\omega_\epsilon\omega_\gamma - \tfrac{1}{2}C_{\alpha\beta\gamma}C_{\gamma\delta\epsilon}A_{\beta\mu}\omega_\delta\omega_\epsilon\Big) \\ &= \theta\Big(\tfrac{1}{2}C_{\alpha\beta\gamma}(\partial_\mu\omega_\beta)\omega_\gamma + \tfrac{1}{2}C_{\alpha\beta\gamma}(\partial_\mu\omega_\gamma)\omega_\beta \\ &\quad - C_{\alpha\beta\gamma}C_{\gamma\delta\epsilon}A_{\delta\mu}\omega_\epsilon\omega_\beta - \tfrac{1}{2}C_{\alpha\beta\gamma}C_{\gamma\delta\epsilon}A_{\beta\mu}\omega_\delta\omega_\epsilon\Big) .\end{aligned}$$

The first two terms of the final expression cancel because $C_{\alpha\beta\gamma}$ is antisymmetric in β and γ, and the third and fourth terms cancel because of the Jacobi identity (15.1.5), so

$$ssA_{\alpha\mu} = 0 . \tag{15.7.16}$$

Eqs. (15.7.9) and (15.7.11) show immediately that

$$ss\omega^*_\alpha = 0 \tag{15.7.17}$$

and

$$ssh_\alpha = 0 . \tag{15.7.18}$$

Finally,

$$\begin{aligned}\delta_\theta s\omega_\alpha &= -\tfrac{1}{2}C_{\alpha\beta\gamma}\delta_\theta(\omega_\beta\omega_\gamma) \\ &= \tfrac{1}{4}\theta\left(C_{\alpha\beta\gamma}C_{\beta\delta\epsilon}\,\omega_\delta\omega_\epsilon\omega_\gamma - C_{\alpha\beta\gamma}C_{\gamma\delta\epsilon}\,\omega_\beta\omega_\delta\omega_\epsilon\right) \\ &= \tfrac{1}{2}\theta\, C_{\alpha\beta\gamma}C_{\gamma\delta\epsilon}\omega_\delta\omega_\epsilon\omega_\beta .\end{aligned}$$

But the product $\omega_\beta\omega_\delta\omega_\epsilon$ is antisymmetric, so the Jacobi identity (15.1.5) shows that this too vanishes

$$ss\omega_\alpha = 0 . \tag{15.7.19}$$

Now consider a product of two fields ϕ_1 and ϕ_2, either or both of which may be ψ, A, ω, ω^*, or h, not necessarily at the same point in spacetime. Then

$$\delta_\theta(\phi_1\phi_2) = \theta(s\phi_1)\phi_2 + \phi_1\theta(s\phi_2) = \theta\Big[(s\phi_1)\phi_2 \pm \phi_1 s\phi_2\Big] ,$$

where the sign $\pm$ is plus if ϕ_1 is bosonic, minus if ϕ_1 is fermionic. That is,

$$s(\phi_1\phi_2) = (s\phi_1)\phi_2 \pm \phi_1 s\phi_2 .$$

Since as we have seen $\delta_\theta(s\phi_1) = \delta_\theta(s\phi_2) = 0$, the effect of a BRST transformation on $s(\phi_1\phi_2)$ is

$$\delta_\theta s(\phi_1\phi_2) = (s\phi_1)\theta(s\phi_2) \pm \theta(s\phi_1)(s\phi_2) .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

But $s\phi$ always has statistics opposite to ϕ, so moving θ to the left in the first term on the right-hand side introduces a sign factor $\mp$:

$$\delta_\theta s(\phi_1\phi_2) = \theta\Big[\mp(s\phi_1)(s\phi_2) \pm (s\phi_1)(s\phi_2)\Big] = 0\,.$$

Continuing in this way, we see that BRST transformations are nilpotent acting on any product of fields at arbitrary spacetime points:

$$\delta_\theta\, s(\phi_1\phi_2\phi_3\cdots) = 0\,.$$

Any functional $F[\phi]$ can be written as a sum of multiple integrals of such products with c-number coefficients, so likewise

$$\delta_\theta\, sF[\phi] = \theta ssF[\phi] = 0\,. \tag{15.7.20}$$

This completes the proof of the nilpotency of the BRST transformation.

Now let us return to the verification of the BRST invariance of the action (15.7.6). First note that for any functional of matter and gauge fields alone, the BRST transformation is just a gauge transformation with infinitesimal gauge parameter

$$\lambda_\alpha(x) = \theta\omega_\alpha(x)\,. \tag{15.7.21}$$

Therefore the first term in Eq. (15.7.6) is automatically BRST-invariant:

$$\delta_\theta \int d^4x\, \mathscr{L} = 0\,. \tag{15.7.22}$$

To calculate the effect of a BRST transformation on the rest of the action (15.7.6), note that its effect on the gauge-fixing function is just the gauge transformation (15.7.21), so

$$\begin{aligned}\delta_\theta f_\alpha[x;A,\psi] &= \int \frac{\delta f_\alpha[x;A_\lambda,\psi_\lambda]}{\delta\lambda^\beta(y)}\bigg|_{\lambda=0} \theta\,\omega_\beta(y)\,d^4y \\ &= \theta\int \mathscr{F}_{\alpha x,\beta y}[A,\psi]\,\omega_\beta(y)\,d^4y\end{aligned}$$

or in terms of the quantity (15.7.3)

$$\delta_\theta f_\alpha[x;A,\psi] = \theta\Delta_\alpha(x;A,\psi,\omega)\,. \tag{15.7.23}$$

(Note that $\mathscr{F}$ is a bosonic quantity, so there is no sign change in moving θ to the left here.) Also recall that $\delta_\theta\omega^*_\alpha = -\theta h_\alpha$ and $\delta_\theta h_\alpha = 0$. Therefore the terms in the integrand of the 'new' action (15.7.6) other than $\mathscr{L}$ may be written

$$\omega^*_\alpha\Delta_\alpha + h_\alpha f_\alpha + \tfrac{1}{2}\xi h_\alpha h_\alpha = -s(\omega^*_\alpha f_\alpha + \tfrac{1}{2}\xi\omega^*_\alpha h_\alpha) \tag{15.7.24}$$

or in other words

$$I_{\rm NEW} = \int d^4x\, \mathscr{L} + s\Psi\,, \tag{15.7.25}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

where

$$\Psi \equiv -\int d^4x \, (\omega_\alpha^* f_\alpha + \tfrac{1}{2}\xi\omega_\alpha^* h_\alpha) \ . \tag{15.7.26}$$

The nilpotence of the BRST transformation tells us immediately that the term $s\Psi$ as well as $\int d^4x \, \mathscr{L}$ is BRST-invariant.

In a sense the converse of this result also applies: we shall see in Section 17.2 that a renormalizable Lagrangian that obeys BRST invariance and the other symmetries of the Lagrangian (15.7.25) must take the form of Eq. (15.7.25), aside from changes in the values of various constant coefficients. But this is not enough to establish the renormalizability of these theories. BRST symmetry transformations act non-linearly on the fields, and in this case there is no simple connection between the symmetries of the Lagrangian and the symmetries of matrix elements and Greens functions. Using the external field methods developed in the next chapter, it will be shown in Section 17.2 that the ultraviolet divergent terms in Feynman amplitudes (though not the finite parts) do obey a sort of renormalized BRST invariance, which allows the proof of renormalizability to be completed.

Eq. (15.7.25) shows that the physical content of any gauge theory is contained in the *kernel* of the BRST operator (that is, in a general BRST-invariant term $\int d^4x \, \mathscr{L} + s\Psi$), modulo terms in the *image* of the BRST transformation (that is, terms of the form $s\Psi$). The kernel modulo the image of any nilpotent transformation is said to form the *cohomology* of the transformation. There is another sense in which the physical content of a gauge theory may be identified with the cohomology of the BRST operator.[12] It is a fundamental physical requirement that matrix elements between physical states should be independent of our choice of the gauge-fixing function f_α, or in other words, of the functional Ψ in Eq. (15.7.25). The change in any matrix element $\langle\alpha|\beta\rangle$ due to a change $\tilde{\delta}\Psi$ in Ψ is

$$\tilde{\delta}\langle\alpha|\beta\rangle = i\langle\alpha|\tilde{\delta}I_{\rm NEW}|\beta\rangle = i\langle\alpha|s\tilde{\delta}\Psi|\beta\rangle \ . \tag{15.7.27}$$

(We use a tilde here to distinguish this arbitrary change in the gauge-fixing function from a BRST transformation or a gauge transformation.) We can introduce a fermionic BRST 'charge' Q, defined so that for any field operator Φ,

$$\delta_\theta\Phi = i[\theta\, Q, \Phi] = i\theta\, [Q, \Phi]_\mp \ ,$$

or in other words,

$$[Q, \Phi]_\mp = is\Phi \ , \tag{15.7.28}$$

the sign being $-$ or $+$ according as Φ is bosonic or fermionic. The nilpotence of the BRST transformation then gives

$$0 = -ss\Phi = [Q, [Q, \Phi]_\mp]_\pm = [Q^2, \Phi]_- \ .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

For this to be satisfied for all operators Φ, it is necessary for Q^2 either to vanish or be proportional to the unit operator. But Q^2 cannot be proportional to the unit operator, since it has a non-vanishing ghost quantum number**, so it must vanish:

$$Q^2 = 0 \,. \tag{15.7.29}$$

From Eqs. (15.7.27) and (15.7.28), we have

$$\tilde{\delta}\langle\alpha|\beta\rangle = \langle\alpha|[Q, \tilde{\delta}\Psi]|\beta\rangle \,. \tag{15.7.30}$$

In order for this to vanish for all changes $\tilde{\delta}\Psi$ in Ψ, it is necessary that

$$\langle\alpha|Q = Q|\beta\rangle = 0 \,. \tag{15.7.31}$$

Thus physical states are in the kernel of the nilpotent operator Q. Two physical states that differ only by a state vector in the image of Q, that is, of form $Q|\cdots\rangle$, evidently have the same matrix element with all other physical states, and are therefore physically equivalent. Hence *independent* physical states correspond to states in the kernel of Q, modulo the image of Q — that is, they correspond to the cohomology of Q.

To see how this works in practice, let us consider the simple example of pure electrodynamics.† Taking the gauge-fixing function as $f = \partial_\mu A^\mu$ and integrating over the auxiliary field h, the BRST transformation (15.7.8)–(15.7.10) is here

$$s\,A_\mu = \partial_\mu \omega \,, \qquad s\,\omega^* = \partial_\mu A^\mu/\xi \,, \qquad s\,\omega = 0 \,. \tag{15.7.32}$$

We expand the fields in normal modes††

$$\begin{aligned} A^\mu(x) &= (2\pi)^{-3/2} \int \frac{d^3p}{\sqrt{2p^0}} \left[a^\mu(\mathbf{p})\, e^{ip\cdot x} + a^{\mu *}(\mathbf{p})\, e^{-ip\cdot x}\right] \,, \\ \omega(x) &= (2\pi)^{-3/2} \int \frac{d^3p}{\sqrt{2p^0}} \left[c(\mathbf{p})\, e^{ip\cdot x} + c^*(\mathbf{p})\, e^{-ip\cdot x}\right] \,, \\ \omega^*(x) &= (2\pi)^{-3/2} \int \frac{d^3p}{\sqrt{2p^0}} \left[b(\mathbf{p})\, e^{ip\cdot x} + b^*(\mathbf{p})\, e^{-ip\cdot x}\right] \,. \end{aligned} \tag{15.7.33}$$

** Recall that the ghost quantum number is defined as $+1$ for ω_α, -1 for ω_α^*, and 0 for all gauge and matter fields.

† Eqs. (15.6.11) and (15.6.7) show that because the structure constants vanish in electrodynamics, the ghosts here are not coupled to other fields. Nevertheless, electrodynamics provides a good example of the use of BRST symmetry in identifying physical states. Indeed, in analyzing the physicality conditions on 'in' and 'out' states we ignore interactions, so for this purpose a non-Abelian gauge theory is treated like several copies of quantum electrodynamics.

†† Just as $\omega^*(x)$ is not to be thought of as the Hermitian adjoint of $\omega(x)$, b^* and c^* are not the adjoints of c and b. But since $A^\mu(x)$ is Hermitian, $\omega(x)$ is Hermitian if Q is.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Matching coefficients of $e^{\pm ip\cdot x}$ on both sides of Eq. (15.7.28) yields

$$\begin{aligned}
&[Q, a^\mu(\mathbf{p})]_- = -p^\mu c(\mathbf{p}) , \qquad [Q, a^{\mu *}(\mathbf{p})]_- = p^\mu c^*(\mathbf{p}) , \\
&[Q, b(\mathbf{p})]_+ = p^\mu a_\mu(\mathbf{p})/\xi , \qquad [Q, b^*(\mathbf{p})]_+ = p^\mu a^*_\mu(\mathbf{p})/\xi , \\
&[Q, c(\mathbf{p})]_+ = [Q, c^*(\mathbf{p})]_+ = 0 .
\end{aligned} \tag{15.7.34}$$

Consider any state $|\psi\rangle$ satisfying the physicality condition (15.7.31):

$$Q|\psi\rangle = 0 . \tag{15.7.35}$$

The states $|e, \psi\rangle = e_\mu a^{\mu *}(\mathbf{p})|\psi\rangle$ with one additional photon then satisfy the physicality condition $Q|e, \psi\rangle = 0$ if $e_\mu p^\mu = 0$. Also, the state $|\psi\rangle' \equiv b^*(\mathbf{p})|\psi\rangle$ satisfies

$$Q|\psi\rangle' = p^\mu a^*_\mu(\mathbf{p})|\psi\rangle/\xi , \tag{15.7.36}$$

so $|e + \alpha p, \psi\rangle = |e, \psi\rangle + \xi\alpha Q|\psi\rangle'$, and is therefore physically equivalent to $|e, \psi\rangle$. From this we conclude that e^μ is physically equivalent to $e^\mu + \alpha p^\mu$, which is the usual 'gauge-invariance' condition on photon polarization vectors. On the other hand,

$$Qb^*(\mathbf{p})|\psi\rangle = p^\mu a^*_\mu(\mathbf{p})|\psi\rangle \neq 0 ,$$

so $b^*|\psi\rangle$ does not satisfy the physicality condition (15.7.31). Also, for any e_μ with $e \cdot p \neq 0$,

$$c^*(\mathbf{p})|\psi\rangle = Qe_\mu a^{*\mu}(\mathbf{p})|\psi\rangle/e \cdot p$$

so $c^*|\psi\rangle$ is BRST-exact, and hence equivalent to zero. *Thus the physical Hilbert space is free of ghosts and antighosts.*

To maintain Lorentz invariance, we must interpret all four components of $a^\mu(\mathbf{p})$ as annihilation operators, in the sense that

$$0 = a_\mu(\mathbf{p})|0\rangle , \tag{15.7.37}$$

where $|0\rangle$ is the BRST-invariant vacuum state. But the canonical commutation relations derived from the BRST-invariant action (say, with $\xi = 1$) give

$$[a_\mu(\mathbf{p}), a^*_\nu(\mathbf{p}')]_- = \eta_{\mu\nu}\delta^3(\mathbf{p} - \mathbf{p}') , \tag{15.7.38}$$

corresponding to the propagator in Feynman gauge. This violates the usual positivity rules of quantum mechanics, because Eqs. (15.7.37) and (15.7.38) yield[13]

$$\langle 0|a_0(\mathbf{p})\, a^*_0(\mathbf{p}')|0\rangle = -\langle 0|0\rangle . \tag{15.7.39}$$

Nevertheless we can rest assured that all amplitudes among physical states satisfy the usual positivity conditions, because these states satisfy Eq. (15.7.31), and for such states the transition amplitudes are the same

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

as they would be in a more physical gauge like Coulomb or axial gauge, where there is no problem of positivity or unitarity.

The Faddeev–Popov–De Witt formalism described so far necessarily yields an action that is bilinear in the ghost fields ω_α^* and ω_α. This is adequate for renormalizable Yang–Mills theories with the gauge-fixing function $f_\alpha = \partial_\mu A_\alpha^\mu$, but not in more general cases. For instance, as we shall see in Section 17.2, in other gauges renormalizable Yang–Mills theories need $\omega^*\omega^*\omega\omega$ terms in the Lagrangian density to serve as counterterms for the ultraviolet divergences in loop graphs with four external ghost lines.

Fortunately the Faddeev–Popov–De Witt formalism represents only one way of generating a class of equivalent Lagrangians that yield the same unitary S-matrix. The BRST formalism provides a more general approach, that dispenses altogether with the Faddeev–Popov–De Witt formalism. In this approach, one takes the action to be the most general local functional of matter, gauge, ω^A, ω^{*A} and h^A fields with ghost number zero that is invariant under the BRST transformation (15.7.7)–(15.7.11) and under any other global symmetries of the theory. (For renormalizable theories one would also limit the Lagrangian density to operators of dimensionality four or less, but this restriction plays no role in the following discussion.) In the next section we shall prove, in a context more general than Yang–Mills theories, that the most general action of this sort is the sum of a functional of the matter and gauge fields (collectively called ϕ) alone, plus a term given by the action of the BRST operator s on an arbitrary functional Ψ of ghost number -1:

$$I_{\rm NEW}[\phi,\omega,\omega^*,h] = I_0[\phi] + s\,\Psi[\phi,\omega,\omega^*,h]\,, \tag{15.7.40}$$

as for instance in the Faddeev–Popov–De Witt action (15.7.25), but with $s\Psi$ now not necessarily bilinear in ghost and antighost fields.

By the same argument as before, the S-matrix elements for states that are annihilated by the BRST generator Q are independent of the choice of Ψ in Eq. (15.7.40), so if there is any choice of Ψ for which the ghosts decouple, then the ghosts decouple in general. In Yang–Mills theories, such a Ψ is provided by quantization of the theory in axial gauge, so in such theories ghosts decouple for arbitrary choices of the functional $\Psi[\phi,\omega,\omega^*,h]$, not just those choices like (15.7.26) that are generated by the Faddeev–Popov–De Witt formalism.

We can go further, and free ourselves of all dependence on canonical quantization in Lorentz-non-invariant gauges like axial gauge. Again, take the action to be the most general functional of gauge, matter, ω^A, ω^{*A} and h^A fields with ghost number zero, that is invariant under the BRST transformation (15.7.7)–(15.7.11) and under any other global symmetries of the theory, including Lorentz invariance. From the BRST invariance

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

of the action we can infer the existence of a conserved nilpotent BRST generator Q. With the ghost and antighost fields treated as Hermitian, Q is also Hermitian. The space of physical states is defined as above as consisting of states annihilated by Q, with two states treated as equivalent if their difference is Q acting on another state. It has been shown that for Yang–Mills theories this space is free of ghosts and antighosts and has a positive-definite norm, and that the S-matrix in this space is unitary.[13a]

This procedure is known as *BRST quantization.* It has been extended to theories with other local symmetries, such as general relativity and string theories. Unfortunately, it seems so far to be necessary to give separate proofs in each case that the BRST-cohomology is ghost-free and that the S-matrix acting in this space is unitary. The key point in these proofs is that, for each negative-norm degree of freedom, such as the time components of the gauge fields in Yang–Mills theories, there is one local symmetry that allows this degree of freedom to be transformed away.

* * *

Although we shall not use it here, there is a beautiful geometric interpretation[14] of the ghosts and the BRST symmetry that should be mentioned. The gauge fields A_α^μ may be written as one-forms $A_\alpha \equiv A_{\alpha\mu} dx^\mu$, where dx^μ are a set of anticommuting c-numbers. (See Section 5.8.) This can be combined with the ghost to compose a one-form $\mathscr{A}_\alpha \equiv A_\alpha + \omega_\alpha$ in an extended space. Also, the ordinary exterior derivative $d \equiv dx^\mu \partial/\partial x^\mu$ may be combined with the BRST operator s to form an exterior derivative $\mathscr{D} \equiv d + s$ in this space, which is nilpotent because $s^2 = d^2 = sd + ds = 0$.

The next chapter will introduce external field methods, which will be used along with the BRST symmetry in Chapter 17 to complete the proof of the renormalizability of non-Abelian gauge theories.

15.8 Generalizations of BRST Symmetry*

The BRST symmetry described in the previous section has a useful generalization to the quantization of a wide class of theories, including general relativity and string theories. In all these cases, we deal with an action $I[\phi]$ and measure $[d\phi] \equiv \prod_r d\phi^r$ that are invariant under a set of infinitesimal transformations

$$\phi^r \to \phi^r + \epsilon^A \delta_A \phi^r \,. \tag{15.8.1}$$

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

This is an abbreviated 'De Witt' notation, with r and A including spacetime coordinates as well as discrete labels, and sums including integrals over these coordinates. For instance, for the gauge transformation (15.1.9), the index A consists of a group index α and a spacetime coordinate x, with $\epsilon^{\alpha x} \equiv \epsilon^{\alpha}(x)$, while the index r consists of a vector index μ as well as a group index α and a spacetime coordinate x, with $\phi^{\mu\alpha x} \equiv A^{\alpha}{}_{\mu}(x)$; in the notation of Eq. (15.8.1), the variation $\delta_A \phi^r$ in the transformation (15.1.9) reads

$$\delta_{\beta y}\phi^{\mu\alpha x} = \delta^{\beta}_{\alpha}\frac{\partial}{\partial x^{\mu}}\delta^4(x-y) + C^{\beta}{}_{\gamma\alpha}\phi^{\mu\gamma x}\delta^4(x-y)\,.$$

As in the special case of Yang–Mills theories discussed in the previous section, BRST invariance can be used as a substitute for the Faddeev–Popov–De Witt formulation of these theories, one that is applicable even where the Faddeev–Popov–De Witt approach fails. Nevertheless, in order to motivate the introduction of BRST invariance, we shall begin here with the Faddeev–Popov–De Witt formulation of theories with general local symmetries, and then go on to consider further generalizations.

By following the same arguments used to derive Eq. (15.5.21), we obtain the general Faddeev–Popov–De Witt theorem:

$$\frac{C}{\Omega}\int [d\phi]\; e^{iI[\phi]}\, V[\phi] = \int [d\phi]\; e^{iI[\phi]}\, B[f[\phi]]\; \mathrm{Det}(\delta_A f_B[\phi])\, V[\phi]\,, \quad (15.8.2)$$

where $V[\phi]$ is an arbitrary functional of ϕ^r that is invariant under the gauge transformations (15.8.1); $f_A[\phi]$ are a set of gauge-fixing functionals** of the ϕ^r, chosen so that the 'matrix' $\delta_A f_B[\phi]$ has a non-vanishing determinant, and $B[f]$ is a more-or-less arbitrary functional of the f_A (as for instance $\prod_A \delta(f_A)$). The constant Ω is the volume of the gauge group, and the constant C is defined (as in Eq. (15.5.19)) by

$$C \equiv \int [df]\; B[f]\,. \quad (15.8.3)$$

As we have seen, in gauge theories the importance of Eq. (15.8.2) is that it tells us that the integral on the right-hand side is independent of the choice of the gauge-fixing functionals f_A, and depends on $B[f]$ only through the constant C. Where some meaning can be given to the usually infinite group volume Ω, as in gauge theories on a finite spacetime

** We use the same letters A, B, etc. to label the f_A as the gauge variations δ_A, in order to emphasize that there must be as many gauge-fixing functionals as there are independent gauge transformations. However, in some cases like string theory it is natural to use gauge-fixing functionals f^a for which, although the index a runs over 'as many' values as the index A on the gauge variations δ_A, the values taken by these indices are quite different. No change is needed in the present formalism as long as we can define $f_A = c_{Aa}f^a$, with c_{Aa} field-independent and non-singular.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

lattice, Eq. (15.8.2) can also have value as a formula for the integral on the left-hand side.

To define a nilpotent BRST transformation, we must first express the functional $B[f]$ as a Fourier transform

$$B[f] = \int [dh]\ \exp(ih^A f_A)\, \mathscr{B}[h]\ , \tag{15.8.4}$$

where $[dh] \equiv \prod_A dh^A$. Also, the determinant can be expressed as an integral over fermionic c-number fields[†] ω^{*A} and ω^A:

$$\mathrm{Det}(\delta_A f_B[\phi]) \propto \int [d\omega^*]\,[d\omega]\ \exp\left(i\omega^{*B}\omega^A \delta_A f_B\right)\ , \tag{15.8.5}$$

where $[d\omega^*] \equiv \prod_A d\omega^{*A}$ and $[d\omega] \equiv \prod_A d\omega^A$, and as usual '$\propto$' means proportional with field-independent factors. Inserting these in Eq. (15.8.2) gives a general formula for the gauge-fixed path integral

$$\int [d\phi]\ \exp\left(iI[\phi]\right) B[f[\phi]]\ \mathrm{Det}(\delta_A f_B[\phi])\, V[\phi]$$
$$\propto \int [d\phi]\,[dh]\,[d\omega^*]\,[d\omega]\ \exp\left(iI_{\mathrm{NEW}}[\phi, h, \omega, \omega^*]\right) \mathscr{B}[h]\ V[\phi]\ , \tag{15.8.6}$$

where I_{NEW} is the new total action:

$$I_{\mathrm{NEW}}[\phi, h, \omega, \omega^*] = I[\phi] + h^A f_A[\phi] + \omega^{*B}\omega^A \delta_A f_B[\phi]\ . \tag{15.8.7}$$

As mentioned in Section 15.6, we can think of the ghost fields as compensation for the fact that we are integrating over all ϕ^r, including those ϕ^r that differ only by gauge transformations (15.8.1). Because ghosts are fermions, loops of ghost lines carry extra minus signs that allow these loops to compensate for the integration over gauge-equivalent ϕs. But for this to work, there must be just as many ghost fields ω^A as there are independent gauge transformations. That is, since the ω^A are independent, the gauge transformations (15.8.1) must all be independent. This is the case for gauge transformations in Yang–Mills theory and coordinate transformations in general relativity, but not always. The classic example of a theory with non-independent gauge transformations is the theory of p-form gauge fields, described in Section 8.8. A p-form A (an antisymmetric tensor of rank p) undergoes a gauge transformation $A \to A + d\phi$, where ϕ is a $(p-1)$-form, and $d\phi$ is its exterior derivative (the antisymmetrized derivative). Because d is nilpotent, for $p \geq 2$ we can shift ϕ by an amount $d\psi$ without changing the gauge transformation, so there is a sort of invariance under gauge transformations of gauge

[†] It is common in string theories and elsewhere to find the ghost fields ω^{*A} and ω^A written b^A (or b_A) and c^A, respectively.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

transformations, in which the transformation parameters are the $(p-2)$-forms ψ. In such cases we must compensate for introducing too many ghosts by also introducing 'ghosts of ghosts.'[15] For $p \geq 3$ we need to compensate further by introducing 'ghosts of ghosts of ghosts,' and so on. In what follows we shall assume that the gauge transformations (15.8.1) are all independent, so that the ghost fields ω^A (and the antighosts ω^{*A}) are all we need.

Although the original symmetry (15.8.1) has been eliminated by the insertion of the non-gauge-invariant functional $B[f]$, the new total action has an exact symmetry under the infinitesimal BRST transformations

$$\chi \to \chi + \theta s\chi\,, \tag{15.8.8}$$

where χ is any of the ϕ^r, ω^A, ω^{A*}, or h^A; θ is an infinitesimal anticommuting c-number; and s is the *Slavnov operator*

$$s = \omega^A \delta_A \phi^r \frac{\delta_L}{\delta \phi^r} - \frac{1}{2}\omega^B \omega^C f^A{}_{BC} \frac{\delta_L}{\delta \omega^A} - h^A \frac{\delta_L}{\delta \omega^{*A}}\,. \tag{15.8.9}$$

In Eq. (15.8.9) the subscript L denotes left differentiation, defined so that if $\delta F = \delta\chi\, G$, then $\delta_L F/\delta\chi = G$, and $f^A{}_{BC}$ is the structure constant[††] appearing in the commutation relation

$$[\delta_B, \delta_C] = f^A{}_{BC}\delta_A\,. \tag{15.8.10}$$

The $f^A{}_{BC}$ are field-independent in non-Abelian gauge theories and in string theories, though not always, but the BRST formalism is not limited to this case. A straightforward calculation gives

$$\begin{aligned} s^2 = {} & \frac{1}{2}\omega^A \omega^B \left[\delta_A \phi^s \frac{\delta_L(\delta_B \phi^r)}{\delta \phi^s} - \delta_B \phi^s \frac{\delta_L(\delta_A \phi^r)}{\delta \phi^s} - f^C{}_{AB}\delta_C \phi^r\right] \frac{\delta_L}{\delta \phi^r} \\ & - \frac{1}{2}\omega^B \omega^C \omega^D \left[f^E{}_{BC} f^A{}_{DE} + \delta_D \phi^r \frac{\delta_L f^A{}_{BC}}{\delta \phi^r}\right] \frac{\delta_L}{\delta \omega^A}\,. \end{aligned} \tag{15.8.11}$$

Hence the condition that the BRST transformation be nilpotent is equivalent to the commutation relation (15.8.10), together with a consistency condition

$$f^E{}_{[BC} f^A{}_{D]E} + \delta_{[D}\phi^r \left(\delta_L f^A{}_{BC]}/\delta \phi^r\right) = 0\,, \tag{15.8.12}$$

where the brackets in subscripts indicate antisymmetrization with respect to the enclosed indices B, C, and D. Eq. (15.8.12) may be derived from the commutation relation (15.8.10) in the same way as the usual Jacobi

[††] For instance, for a gauge transformation acting on a matter field $\psi(x)$ we have $\delta_{\beta y}\psi(x) = it_\beta \psi(x)\delta^4(x-y)$, and so $\delta_{\beta y}\delta_{\gamma z}\psi(x) = -t_\gamma t_\beta \psi(x)\delta^4(x-y)\delta^4(x-z)$. Hence in this case we have $f^{\alpha x}{}_{\beta y\, \gamma z} = C^\alpha{}_{\beta\gamma}\delta^4(x-y)\delta^4(x-z)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

identity, and takes the place of the Jacobi identity for symmetries with field-dependent structure constants.

To show that the transformation (15.8.8) is a symmetry of $I_{\rm NEW}$, we note (recalling that θ anticommutes with ω^{*A}) that Eq. (15.8.7) may be rewritten

$$I_{\rm NEW}[\phi, h, \omega, \omega^*] = I[\phi] - s(\omega^{*A} f_A) \,. \tag{15.8.13}$$

The term $I[\phi]$ is BRST-invariant, because on the fields ϕ^r a BRST transformation is just a gauge transformation (15.8.1) with ϵ^A replaced with $\theta\omega^A$, which commutes with all ϕ^r. The term $s(\omega^{*A} f_A)$ is BRST-invariant because BRST transformations are nilpotent.

For several reasons we may need to consider a wider class of actions than those that can be constructed by the Faddeev–Popov–De Witt approach, by simply requiring that the action is invariant under the BRST transformation (15.8.8). As a step toward showing that such an action yields physically sensible results, we shall now prove the general result (already used in the previous section) that the most general BRST-invariant functional of ghost number zero is the sum of a functional of the ϕ alone, plus a term given by the action of the BRST operator s on an arbitrary functional Ψ of ghost number -1:

$$I_{\rm NEW}[\phi, \omega, \omega^*, h] = I_0[\phi] + s\,\Psi[\phi, \omega, \omega^*, h] \tag{15.8.14}$$

as for instance in the Faddeev–Popov–De Witt action (15.8.13). In brief, the BRST cohomology consists of gauge-invariant functionals $I[\phi]$ of the fields ϕ^r alone.

To prove Eq. (15.8.14), we note that the BRST transformation (15.8.8)–(15.8.9) does not change the total number of h^A and ω^{*A} fields, so if we expand I in a series of terms I_N that contain definite total numbers N of h^A and ω^{*A} fields, then there can be no cancellations in sI between terms with different N, so each term must be separately BRST-invariant:

$$sI_N = 0 \,. \tag{15.8.15}$$

We next introduce what is called a Hodge operator:

$$t \equiv \omega^{*A} \frac{\delta}{\delta h^A} \,. \tag{15.8.16}$$

It is straightforward to check the anticommutation relation

$$\{s, t\} = -\omega^{*A} \frac{\delta_L}{\delta \omega^{*A}} - h^A \frac{\delta}{\delta h^A} \,. \tag{15.8.17}$$

Applying the operator $\{s, t\}$ to I_N and using Eq. (15.8.15) then gives

$$stI_N = -NI_N \,, \tag{15.8.18}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

so each I_N except for I_0 is BRST-exact, in the sense that it may be written as the operator s acting on some other functional. The complete functional I may therefore be written in the form $I_0 + s\Psi$, with

$$\Psi = -\sum_{N=1}^{\infty} \frac{tI_N}{N} . \tag{15.8.19}$$

The term I_0 is by definition independent of ω^{*A} and h^A, and since we assume it has zero ghost number it must also be independent of ω^A, as was to be proved.

To show the invariance of physical matrix elements under changes in the definition of the gauge-fixing functional Ψ, we define a fermionic 'charge' Q, such that the change under a BRST transformation of any operator Φ is

$$\delta_\theta \Phi = i[\theta Q, \Phi] = i\theta [Q, \Phi]_\mp \tag{15.8.20}$$

with the top or bottom sign in $[x, y]_\mp \equiv xy \mp yx$ according as Φ is bosonic or fermionic. Just as in the previous section, the nilpotence of the BRST transformation then tells us that $Q^2 = 0$. Matrix elements of gauge-invariant operators between physical states will be independent of the choice of Ψ if and only if the physical states $|\alpha\rangle$ and $\langle\beta|$ satisfy

$$Q|\alpha\rangle = \langle\beta|Q = 0 , \tag{15.8.21}$$

so that physically distinguishable physical states are again in one-to-one correspondence with elements of the cohomology of Q. The general BRST-invariant action (15.8.14) will therefore yield physically sensible results for any gauge-fixing functional Ψ if we can find *some* Ψ, like axial gauge in Yang–Mills theories, in which the ghosts do not interact with other fields. If this ghost-free choice of Ψ is inconvenient for actual calculation, as for instance axial gauge is inconvenient because it violates Lorentz invariance, we can adopt any gauge-fixing Ψ we like, and still be confident that there is a unitary S-matrix with no ghosts in initial or final states.

This approach works well in string theories, where so-called light-cone quantization takes the place of axial gauge. But in other theories like general relativity there is no way of choosing a coordinate system in which the ghosts decouple. Such theories may be dealt with by the BRST-quantization method described at the end of the previous section, using BRST invariance to prove that the S-matrix in a physical ghost-free Hilbert space is unitary.

The discovery[17] of invariance under an 'anti-BRST' symmetry[18] showed that, despite appearances, there is a similarity between the roles of ω^A and ω^{*A}, which remains somewhat mysterious.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

15.9 The Batalin–Vilkovisky Formalism*

This section will describe a powerful formalism, widely known as the Batalin–Vilkovisky[19] method. It is developed in the Lagrangian framework, but has its roots in the earlier Batalin–Fradkin–Vilkovisky formalism,[20] which had been derived in the Hamiltonian framework. (The two schemes have been proved perturbatively equivalent.[21]) As we shall see in Section 17.1, the same formal machinery had been developed even earlier by Zinn-Justin[22] in order to deal with the renormalization of gauge theories. There are at least three areas where this formalism has proved invaluable:

(i) Up to this point, we have considered only irreducible symmetries with an algebra that closes in the sense of Eq. (15.8.10). In some theories, such as supergravity (without auxiliary fields)[23] the algebra is open: it closes only when the field equations are satisfied, so that terms appear in Eq. (15.8.10) proportional to $\delta I/\delta\chi^n$. Similar terms will also then appear in the consistency conditions (15.8.12). Eq. (15.8.11) then shows that in such theories s^2 will not vanish, but rather will equal a linear combination of the derivatives $\delta I/\delta\chi^n$. As we will see in this section, the Batalin–Vilkovisky method allows us to deal with very general gauge theories, including those with open or reducible gauge symmetry algebras.

(ii) As mentioned above, essential aspects of the Batalin–Vilkovisky formalism were originally developed by Zinn-Justin in order to prove the renormalizability of gauge theories. The crucial point, to be explained in Section 17.1, is that although the sum of all one-particle-irreducible diagrams in a background field does not obey the BRST symmetries of the original action, it does share one of the key properties of the action, known as the master equation.

(iii) The Batalin–Vilkovisky method provides a convenient way of analyzing the possible violations of symmetries of the action by quantum effects. It is used for this purpose in Section 22.6.

The starting point of the Batalin–Vilkovisky formalism is the introduction of what are called 'antifields,' one for each field in the theory. We let χ^n run over all the fields ϕ^r, ω^A, ω^{*A}, and h^A, and for each χ^n we introduce an external antifield** $\chi_n^\ddagger$, with the same Bose or Fermi statistics

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** The symbol ‡ is used here in place of the more usual * in order to emphasize that it has nothing whatever to do with complex conjugation or charge conjugation. In particular, the antighost field ω^{*A} is *not* the same as the antifield $\omega_A^\ddagger$ of ω^A.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

and opposite ghost number as that of the BRST-transformed field $s\chi^n$. That is, $\chi^\ddagger_n$ has the opposite statistics to χ^n, and a ghost number equal to $-\text{gh}(\chi^n) - 1$, where $\text{gh}(\chi^n)$ is the ghost number of χ^n. In the simplest cases, including Yang–Mills theories and quantum gravity, the original gauge-invariant action $I[\phi]$ is supplemented with a term coupling the antifields $\chi^\ddagger_n$ to the $s\chi^n$, giving an action

$$S[\chi, \chi^\ddagger] \equiv I[\phi] + (s\chi^n)\,\chi^\ddagger_n \,. \tag{15.9.1}$$

This satisfies what is called the *master equation*

$$0 = \frac{\delta_R S}{\delta \chi^\ddagger_n}\,\frac{\delta_L S}{\delta \chi^n} \,, \tag{15.9.2}$$

with 'R' and 'L' here denoting right- and left-differentiation. To check this, note that the terms in Eq. (15.9.2) of zeroth order in antifields just yield the condition of gauge invariance

$$0 = \left(s\phi^r\right)\frac{\delta_L I}{\delta \phi^r} = \omega^A\,\delta_A I[\phi] \,, \tag{15.9.3}$$

while the terms linear in the antifields $\phi^\ddagger_n$ provide the condition of nilpotence:

$$0 = \left(s\chi^m\right)\frac{\delta_L(s\chi^n)}{\delta \chi^m} = s^2\chi^n \,. \tag{15.9.4}$$

The $\chi^\ddagger_n$ are external fields, and must be given suitable values before we use $S[\chi, \chi^\ddagger]$ to calculate the S-matrix. For this purpose, we introduce an arbitrary fermionic functional $\Psi[\chi]$ with ghost number -1, and set[†]

$$\chi^\ddagger_n = \frac{\delta \Psi[\chi]}{\delta \chi^n} \,. \tag{15.9.5}$$

Then Eq. (15.9.1) becomes

$$S[\phi, \delta\Psi/\delta\chi] = I[\phi] + (s\chi^n)\,\delta\Psi[\chi]/\delta\chi^n = I[\phi] + s\Psi[\chi] \,. \tag{15.9.6}$$

Comparison with Eq. (15.8.14) shows that this is the same as the gauge-fixed action $I_{\text{NEW}}[\chi]$. Thus, using the same arguments as in the previous section, physical matrix elements are not affected by small changes in Ψ. The action (15.8.7) constructed by the Faddeev–Popov–De Witt method corresponds to the choice $\Psi = -\omega^{*A} f_A$, for which $\phi^\ddagger_r = \omega^{*A}\,\delta f_A/\delta\phi^r$, $\omega^\ddagger_C = 0$, and $\omega^{*\ddagger}_A = -f_A$.

So far, nothing new has been accomplished. The first new point is that the master equation (15.9.2) can be used for more general theories

[†] It is not necessary to distinguish between left- and right-differentiation here, because either χ^n or $\delta\Psi/\delta\chi^n$ must be bosonic.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

by letting $S[\chi, \chi^\ddagger]$ be a non-linear functional of the antifields $\chi_n^\ddagger$. (For reducible theories we must also include ghosts for ghosts among the χ^n, as discussed in the previous section, along with their antifields.) As above, we take the statistics of $\chi_n^\ddagger$ to be opposite to that of χ^n, and its ghost number equal to $-\mathrm{gh}(\chi^n) - 1$, and we require $S[\chi, \chi^\ddagger]$ to be a bosonic operator of ghost number zero. Because ω^{*A} and h_A have *linear* BRST transformations they are unaffected by the complications affecting the other χ^n (in this connection, see Section 16.4), so they and their antifields enter in the action $S[\chi, \chi^\ddagger]$ in the same way as in Eq. (15.9.1). That is,

$$S = S_{\min}[\phi, \omega, \phi^\ddagger, \omega^\ddagger] - h^A \omega_A^{*\ddagger} , \tag{15.9.7}$$

where $\phi_r^\ddagger$, $\omega_A^\ddagger$, and $\omega_A^{*\ddagger}$ are the antifields of ϕ^r, ω^A, and ω^{*A}, with ghost numbers -1, -2, and 0, respectively, and $S_{\min}[\phi, \omega, \phi^\ddagger, \omega^\ddagger]$ is some bosonic functional of ghost number zero. The last term in (15.9.7) has no effect on the master equation, so $S_{\min}$ satisfies the master equation by itself.††

Because $S_{\min}$ has ghost number zero, its expansion in powers of antifields must take the form

$$\begin{aligned} S_{\min} = {} & I[\phi] + \omega^A f_A^r[\phi]\, \phi_r^\ddagger + \tfrac{1}{2}\omega^A \omega^B f^C{}_{AB}[\phi]\, \omega_C^\ddagger \\ & + \tfrac{1}{2}\omega^A \omega^B f^{rs}{}_{AB}[\phi]\, \phi_r^\ddagger \phi_s^\ddagger + \omega^A \omega^B \omega^C f^{rD}{}_{ABC}[\phi]\, \phi_r^\ddagger \omega_D^\ddagger \\ & + \tfrac{1}{2}\omega^A \omega^B \omega^C \omega^D f^{EF}{}_{ABCD}[\phi]\, \omega_E^\ddagger \omega_F^\ddagger + \ldots \end{aligned} \tag{15.9.8}$$

The term in the master equation (15.9.2) of zeroth order in antifields (and hence first order in ω^A) yields

$$0 = f_A^r[\phi] \frac{\delta I[\phi]}{\delta \phi^r} , \tag{15.9.9}$$

which is just the statement that $I[\phi]$ is invariant under the transformation

$$\phi^r \to \phi^r + \epsilon^A f_A^r[\phi] \tag{15.9.10}$$

with ϵ^A arbitrary infinitesimals. The term in the master equation proportional to $\phi_s^\ddagger$ on the right and $\omega^A \omega^B$ on the left yields

$$\begin{aligned} 0 = {} & f_A^r[\phi] \frac{\delta f_B^s[\phi]}{\delta \phi^r} - f_B^r[\phi] \frac{\delta f_A^s[\phi]}{\delta \phi^r} + f^C{}_{AB}[\phi]\, f_C^s[\phi] \\ & + \frac{\delta I[\phi]}{\delta \phi^r} f^{rs}_{AB}[\phi] , \end{aligned} \tag{15.9.11}$$

which, when the field equations $\delta I/\delta\phi = 0$ are satisfied, becomes the commutation relation (with structure constant $f^C{}_{AB}[\phi]$) for the transformation (15.9.10). The other term in the master equation linear in antifields

†† The fields ϕ^r, ω^A, $\phi_r^\ddagger$, $\omega_A^\ddagger$ are sometimes called *minimal variables*, while fields like ω^{*A} and h_A which together with their antifields enter bilinearly as in Eq. (15.9.7), are called *trivial pairs*.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

is proportional to $\omega^A\omega^B\omega^C$ on the left and $\omega^\ddagger_D$ on the right; it yields

$$0 = f^r_{[A}[\phi]\frac{\delta f^D{}_{BC]}[\phi]}{\delta\phi^r} - f^E{}_{[AB}[\phi]f^D{}_{C]E}[\phi] + f^{rD}_{ABC}[\phi]\frac{\delta I[\phi]}{\delta\phi^r}\,, \quad (15.9.12)$$

where square brackets in the subscripts indicate antisymmetrization with respect to the enclosed indices A, B, and C. When the field equations are satisfied this becomes the generalized Jacobi identity (15.8.12). Eq. (15.9.11) is necessary for the consistency of the symmetry condition (15.9.9) (assuming that the f^r_A furnish a complete set of gauge symmetries), while Eq. (15.9.12) is necessary for the consistency of the commutation relations (15.9.11). Note that the terms in Eqs. (15.9.11) and (15.9.12) that arise from the terms in $S_{\min}$ quadratic in antifields are proportional to $\delta I[\phi]/\delta\chi$, so they vanish when the field equations are satisfied, and in this sense are characteristic of open symmetry algebras. Terms in the master equation of second or higher order in antifields involve terms in $S_{\min}$ that are of third and/or higher order in antifields. These provide consistency conditions for Eqs. (15.9.11) and (15.9.12), consistency conditions for these consistency conditions, and so on. It is a virtue of the Batalin–Vilkovisky formalism that all these consistency conditions are incorporated in the one master equation.

The master equation may be reinterpreted as a statement of invariance of S under a generalized BRST transformation. In order to see this, and for future purposes, it is useful to introduce a formal device known as the *antibracket*. Returning now to our previous notation, the antibracket of two general functionals $F[\chi,\chi^\ddagger]$ and $G[\chi,\chi^\ddagger]$ is defined by

$$(F,G) \equiv \frac{\delta_R F}{\delta\chi^n}\frac{\delta_L G}{\delta\chi^\ddagger_n} - \frac{\delta_R F}{\delta\chi^\ddagger_n}\frac{\delta_L G}{\delta\chi^n}\,. \quad (15.9.13)$$

Note that right- and left-functional derivatives of a bosonic functional like S with respect to a bosonic or fermionic field variable are equal to each other or negatives of each other, respectively. Since just one of either $\chi^\ddagger_n$ or χ^n is always fermionic and the other bosonic, it follows that for the antibracket (S,S) the second term on the right-hand side in Eq. (15.9.13) changes sign if we reverse left- and right-differentiation

$$\frac{\delta_R S}{\delta\chi^\ddagger_n}\frac{\delta_L S}{\delta\chi^n} = -\frac{\delta_L S}{\delta\chi^\ddagger_n}\frac{\delta_R S}{\delta\chi^n} = -\frac{\delta_R S}{\delta\chi^n}\frac{\delta_L S}{\delta\chi^\ddagger_n}\,.$$

(The last step is allowed because, since one of the factors here is bosonic, their order is immaterial.) We see that the second term on the right in Eq. (15.9.13) for (S,S) is the negative of the first. The master equation (15.9.2) may therefore be written as the requirement that the antibracket of S with itself vanishes:

$$(S,S) = 0\,. \quad (15.9.14)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

This is a non-trivial requirement, because the antibracket has the general symmetry property

$$(F,G) = \pm (G,F)\,, \tag{15.9.15}$$

the sign being $+1$ where F and G are both bosonic, and -1 otherwise. In particular, (F,F) automatically vanishes if F is fermionic, but not if F is bosonic.

The generalized BRST transformation is defined by

$$\hat{\delta}_\theta \chi^n = \theta \frac{\delta_R S}{\delta \chi^\ddagger_n} = -\theta(S,\chi^n)\,, \tag{15.9.16}$$

$$\hat{\delta}_\theta \chi^\ddagger_n = -\theta \frac{\delta_R S}{\delta \chi^n} = -\theta(S,\chi^\ddagger_n)\,, \tag{15.9.17}$$

where θ is a fermionic infinitesimal constant. (When S is of the form (15.9.1), the transformation of χ is the same as the original BRST transformation $\hat{\delta}_\theta \chi^n = \theta s \chi^n$.) To evaluate the effect of this transformation on general functionals, we note that the antibracket acts as a derivative, in the sense that

$$(F,GH) = (F,G)\,H \pm G\,(F,H)\,, \tag{15.9.18}$$

where the sign is -1 if G is fermionic and F is bosonic, and $+1$ otherwise. Hence if G and H are arbitrary functionals of χ and $\chi^\ddagger$ with $\hat{\delta}_\theta G = -\theta(S,G)$ and $\hat{\delta}_\theta H = -\theta(S,H)$, then

$$\hat{\delta}_\theta(GH) = -\theta(S,G)H - G\theta(S,H) = -\theta[(S,G)H \pm G(S,H)]\,,$$

where the sign is $+$ or $-$ if G is bosonic or fermionic. Taking F in Eq. (15.9.18) equal to the bosonic functional S, we see then that

$$\hat{\delta}_\theta(GH) = -\theta(S,GH)\,.$$

Together with Eqs. (15.9.16) and (15.9.17), this shows that for any functional F formed as a sum of products of fields and antifields

$$\hat{\delta}_\theta F = -\theta(S,F)\,. \tag{15.9.19}$$

The master equation (15.9.14) may be interpreted as the statement that these generalized BRST transformations leave S invariant

$$\hat{\delta}_\theta S = -\theta(S,S) = 0\,. \tag{15.9.20}$$

Like the original BRST transformation, this symmetry transformation is nilpotent. To see this, we use the Jacobi identity for the antibracket:

$$\pm (F,(G,H)) + \text{cyclic permutations} = 0\,, \tag{15.9.21}$$

where the sign in the first term is $-$ if F and H are bosonic and $+$ otherwise, with corresponding signs for the other two cyclic permutations

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

of F, G, and H. Taking $F = G = S$, Eq. (15.9.21) becomes

$$0 = \mp(S,(S,H)) \mp (H,(S,S)) + (S,(H,S) = \mp 2(S,(S,H)) \mp (H,(S,S)) ,$$

where the sign is $-$ or $+$ if H is bosonic or fermionic. The master equation (15.9.14) then yields the nilpotency condition

$$(S,(S,H)) = 0 . \tag{15.9.22}$$

Because of this symmetry, the solution of the master equation is not unique. For instance, Eq. (15.9.22) shows that for any given solution S we can find another solution given by the infinitesimal transformation

$$S' = S + (\delta F, S) , \tag{15.9.23}$$

with δF an infinitesimal functional of χ and $\chi^\ddagger$ that is arbitrary, except that it must be fermionic and have ghost number -1 in order that S' should be bosonic and have ghost number zero. In particular, taking δF to be a fermionic functional $\epsilon\Psi$ of χ^n alone gives

$$S'[\chi,\chi^\ddagger] = S[\chi,\chi^\ddagger] + \epsilon\frac{\delta\Psi[\chi]}{\delta\chi^n}\frac{\delta_R S}{\delta\chi^\ddagger_n} = S\left[\chi,\chi^\ddagger + \epsilon\frac{\delta\Psi}{\delta\chi}\right] . \tag{15.9.24}$$

These infinitesimal transformations may be trivially integrated, and show that the master equation is still satisfied if we shift the antifields to new variables $\chi^{\ddagger\prime}_n \equiv \chi^\ddagger_n - \delta\Psi/\delta\chi^n$.

The transformation (15.9.23) is a special case of what are usually called canonical transformations, and will here be called 'anticanonical transformations' to distinguish them from the canonical transformations of Chapter 7. An anticanonical transformation is any transformation of fields and antifields, finite or infinitesimal, that leaves unchanged the fundamental antibracket relations:

$$(\chi^n,\chi^\ddagger_m) = \delta^n_m , \qquad (\chi^n,\chi^m) = (\chi^\ddagger_n,\chi^\ddagger_m) = 0 . \tag{15.9.25}$$

For instance, consider the infinitesimal anticanonical transformation generated by an infinitesimal fermion generator δF, under which any bosonic or fermionic functional G is transformed into

$$G \to G' = G + (\delta F, G) . \tag{15.9.26}$$

It is easy to show that this does not change the fundamental antibrackets (15.9.25). For this purpose, note that the antibracket (G,H) of two functionals G and H is transformed into (G',H'), which to first order in infinitesimals is

$$(G',H') = (G,H) + ((\delta F),G),H) + (G,(\delta F,H)) .$$

Using the Jacobi identity (15.9.21), this is

$$(G',H') = (G,H) \pm (\delta F,(G,H)) , \tag{15.9.27}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

with the sign + if G and H are both bosonic, and − otherwise. In particular, if (G,H) is a c-number then it is unchanged by an anticanonical transformation. (This is another way of seeing that the transformation (15.9.23) leaves the master equation unchanged.) The fields and antifields have c-number antibrackets (15.9.25), so the same must be true for the transformed fields and antifields.

To calculate the S-matrix we must give some definite value to the antifields. As in the simple case of closed gauge algebras where S is of the linear form (15.9.5), we can do this by taking the antifields in the form (15.9.5); that is, we calculate the S-matrix using the 'gauge-fixed' action

$$I_\Psi[\chi] = S\left[\chi, \frac{\delta\Psi[\chi]}{\delta\chi}\right], \tag{15.9.28}$$

where $\Psi[\chi]$ is a fermionic functional of ghost number -1. According to the remarks following Eq. (15.9.24), this is the same as taking the canonically transformed antifields $\chi^{\ddagger'}$ equal to zero.

The gauge-fixed action is invariant under a BRST transformation that acts on fields χ^n alone:

$$\delta_\theta\chi^n = \theta s\chi^n, \qquad \text{where} \qquad s\chi^n = \left(\frac{\delta_R S[\chi,\chi^\ddagger]}{\delta\chi^\ddagger_n}\right)_{\chi^\ddagger=\delta\Psi/\delta\chi}. \tag{15.9.29}$$

To check this, note that

$$sI_\Psi[\chi] = \left(\frac{\delta_R S[\chi,\chi^\ddagger]}{\delta\chi^\ddagger_n}\,\frac{\delta_L S[\chi,\chi^\ddagger]}{\delta\chi^n}\right)_{\chi^\ddagger=\delta\Psi/\delta\chi} + \left(\frac{\delta_R S[\chi,\chi^\ddagger]}{\delta\chi^\ddagger_m}\,\frac{\delta_L^2\Psi[\chi]}{\delta\chi^m\,\delta\chi^n}\,\frac{\delta_L S[\chi,\chi^\ddagger]}{\delta\chi^\ddagger_n}\right)_{\chi^\ddagger=\delta\Psi/\delta\chi}.$$

The first term on the right-hand side vanishes as a result of the master equation, and the second because the summand is antisymmetric[‡] in m and n.

For closed algebras with S of the form (15.9.1), the transformation (15.9.29) is the original BRST transformation $\delta_\theta\chi^n = \theta s\chi^n$. But for general open algebras, the transformation (15.9.29) is not the same as the original BRST transformation, and in general is not even nilpotent unless the field equations are satisfied. Instead, from the terms in the master equation of

[‡] For χ^n and χ^m both bosonic this is because $\delta S/\delta\chi^n$ and $\delta S/\delta\chi^m$ anticommute. For one of χ^n and χ^m fermionic and the other bosonic, this is because the right- and left-derivatives of S with respect to whichever χ is fermionic have opposite signs. The terms with χ^n and χ^m both fermionic are antisymmetric because for these terms $\delta_L^2\Psi/\delta\chi^m\delta\chi^n$ is antisymmetric.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

first order in the shifted antifields $\chi_n^{\ddagger\prime} = \chi_n^{\ddagger} - \delta\Psi[\chi]/\delta\chi^n$, we find

$$s^2\chi^m = \mp\left(\frac{\delta_L}{\delta\chi_m^{\ddagger}}\frac{\delta_R}{\delta\chi_n^{\ddagger}}I[\chi,\chi^{\ddagger}]\right)_{\chi^{\ddagger}=\delta\Psi/\delta\chi}\frac{\delta I_\Psi[\chi]}{\delta\chi^n}\,, \tag{15.9.30}$$

with the sign $-$ or $+$ if χ^m is bosonic or fermionic, respectively. Again, we see that the dependence on the field equations characteristic of open gauge algebras is associated with terms in S quadratic in the antifields.

Until now we have been considering only the formulation of a classical field theory based on an open or closed gauge algebra. Now we must consider how quantum mechanical calculations are done in such theories. Physical matrix elements may be calculated by functional integrals weighted with $\exp(iI_\Psi[\chi])$, where as explained above, $I_\Psi[\chi]$ is obtained from $S[\chi,\chi^{\ddagger}]$ by setting $\chi_n^{\ddagger} = \delta\Psi[\chi]/\delta\chi^n$, or in other words $\chi_n^{\ddagger\prime} = 0$. We wish to evaluate the effect of a change in $\Psi[\chi]$ on these matrix elements. First consider the vacuum–vacuum amplitude

$$Z_\Psi = \int\left[\prod d\chi\right]\,\exp(iI_\Psi[\chi])\,. \tag{15.9.31}$$

Under a shift $\delta\Psi[\chi]$ in $\Psi[\chi]$, this changes by an amount

$$\delta Z = i\int\left[\prod d\chi\right]\,\exp(iI_\Psi[\chi])\left(\frac{\delta_R S[\chi,\chi^{\ddagger}]}{\delta\chi_n^{\ddagger}}\right)_{\chi^{\ddagger}=\delta\Psi/\delta\chi}\left(\frac{\delta\,(\delta\Psi[\chi])}{\delta\chi^n}\right)\,. \tag{15.9.32}$$

Integrating by parts in field space, this becomes

$$\delta Z = \int\left[\prod d\chi\right]\,\exp(iI_\Psi[\chi]) \times\left\{\frac{\delta_R S[\chi,\chi^{\ddagger}]}{\delta\chi_n^{\ddagger}}\frac{\delta_L I_\Psi[\chi]}{\delta\chi^n} - i\Delta S[\chi,\chi^{\ddagger}]\right\}_{\chi^{\ddagger}=\delta\Psi/\delta\chi}\delta\Psi[\chi]\,, \tag{15.9.33}$$

where

$$\Delta \equiv \frac{\delta_R}{\delta\chi_n^{\ddagger}}\frac{\delta_L}{\delta\chi^n}\,. \tag{15.9.34}$$

We see that the condition for the Ψ independence of the vacuum–vacuum amplitude is in general *not* the master equation (15.9.2), but rather what is called the *quantum master equation*

$$(S,S) - 2i\Delta S = 0 \quad \text{at} \quad \chi^{\ddagger} = \delta\Psi/\delta\chi \quad . \tag{15.9.35}$$

In cgs units a factor $1/\hbar$ would accompany each factor of $S[\chi,\chi^{\ddagger}]$, so the second term in Eq. (15.9.35) would have a coefficient $-2i\hbar$ in place of $-2i$. Thus whenever the quantum master equation (15.9.35) is satisfied, the term in S of zeroth order in $\hbar$ satisfies the original master equation (15.9.2). Usually it is easy to construct an action that satisfies the classical

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

master equation, so that the first term in Eq. (15.9.35) vanishes, and the question is then whether the second term also vanishes. The case where the quantum master equation is not satisfied by a local action is considered in Chapter 22, on anomalies.

Assuming the quantum master equation (15.9.35) to be satisfied, the change in the vacuum expectation value of an operator $\mathcal{O}[\chi]$ due to a change $\delta\Psi$ in Ψ is

$$\delta\langle\mathcal{O}\rangle = \frac{-i}{Z_\Psi}\int\left[\prod d\chi\right]\,\exp\left(iI_\Psi[\chi]\right)\frac{\delta_R\mathcal{O}[\chi]}{\delta\chi^n}\left(\frac{\delta_R S(\chi,\chi^\ddagger)}{\delta\chi^\ddagger_n}\right)_{\chi^\ddagger=\delta\Psi/\delta\chi}\delta\Psi[\chi]\,. \tag{15.9.36}$$

The coefficient of the exponential in the integrand in Eq. (15.9.36) is just $s\mathcal{O}[\chi]$. We see that expectation values of operators that are invariant[‡‡] under the generalized BRST transformation (15.9.16)–(15.9.17) are unaffected by a change in the gauge-fixing fermion Ψ. Corresponding results hold for vacuum expectation values of two or more operators.

Appendix A A Theorem Regarding Lie Algebras

In this appendix we consider a general Lie algebra $\mathcal{G}$, with generators t_α and structure constants $C^\alpha{}_{\beta\gamma}$, and prove the equivalence of three conditions:

a: There exists a real symmetric positive-definite matrix $g_{\alpha\beta}$ that satisfies the invariance condition

$$g_{\alpha\beta}C^\beta{}_{\gamma\delta} = -\,g_{\gamma\beta}C^\beta{}_{\alpha\delta}\,. \tag{15.A.1}$$

(This is the condition (15.2.4) shown in Section 15.2 to be necessary on physical grounds.)

b: There is a basis for the Lie algebra (that is, a set of generators $\tilde{t}_\alpha = \mathcal{S}_{\alpha\beta}t_\beta$, with $\mathcal{S}$ a real non-singular matrix) for which the structure constants $\tilde{C}^\alpha{}_{\beta\gamma}$ are antisymmetric not only in the lower indices β and γ but in all three indices α, β and γ.

c: The Lie algebra $\mathcal{G}$ is the direct sum of commuting compact simple and $U(1)$ subalgebras $\mathcal{H}_m$.

[‡‡] For open gauge theories there are generally no operators other than constants that are invariant under the transformations (15.9.16)–(15.9.17). Instead one should consider operators $O(\chi,\chi^\ddagger)$ that are invariant under the nilpotent 'quantum BRST operator' σ, defined by $\sigma O = (O,S) - i\Delta O$. Where O depends only on χ, this condition reduces to Eq. (15.9.36). If $\sigma O = 0$ then the expectation value of $O(\Psi,\delta\Psi/\delta\chi)$ is unaffected by a small change in the gauge-fixing fermion Ψ.[24]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

We shall prove the equivalence of statements **a**, **b**, and **c** by showing that **a** implies **b**, **b** implies **c**, and **c** implies **a**. As a by-product, we shall also show that if these conditions are satisfied, then it is possible to choose the generators of $\mathscr{G}$ as t_{ma}, with m labelling the simple or $U(1)$ subalgebra $\mathscr{H}_m$ to which t_{ma} belongs and a labelling the individual generators in this subalgebra, and with the matrix $g_{ma,nb}$ that satisfies Eq. (15.A.1) taking the form

$$g_{ma,nb} = g_m^{-2}\delta_{mn}\delta_{ab} , \tag{15.A.2}$$

where the g_m^{-2} are arbitrary real positive constants.

First, let us assume **a**, the existence of a real symmetric positive-definite $g_{\alpha\beta}$ satisfying the invariance condition (15.A.1). We may then define new generators

$$\tilde{t}_\alpha \equiv (g^{-1/2})_{\alpha\beta} t_\beta , \tag{15.A.3}$$

in which the existence of a real inverse square-root matrix $g^{-1/2}$ is guaranteed by the positive-definiteness of $g_{\alpha\beta}$. These satisfy a Lie algebra

$$\left[\tilde{t}_\alpha, \tilde{t}_\beta\right] = i\, \tilde{C}_{\alpha\beta\gamma}\tilde{t}_\gamma , \tag{15.A.4}$$

where

$$\tilde{C}_{\gamma\alpha\beta} \equiv (g^{-1/2})_{\alpha\alpha'}(g^{-1/2})_{\beta\beta'}(g^{+1/2})_{\gamma\gamma'} C^{\gamma'}{}_{\alpha'\beta'} . \tag{15.A.5}$$

(In this basis it is convenient to drop the distinction between upper and lower indices α, β, etc., and write $\tilde{C}_{\gamma\alpha\beta}$ in place of $\tilde{C}^{\gamma}{}_{\alpha\beta}$.) Then Eq. (15.A.1) tells us that $\tilde{C}_{\alpha\gamma\delta}$ is antisymmetric in α and γ as well as in γ and δ, and hence is totally antisymmetric, verifying **b**.

Next let us assume **b**, the existence of a basis for the Lie algebra for which the structure constants are totally antisymmetric. In this basis, the matrices $(\tilde{t}^{\mathrm{A}}{}_\alpha)_{\beta\gamma} \equiv -i\tilde{C}_{\beta\gamma\alpha}$ of the adjoint representation are imaginary and antisymmetric, and hence Hermitian. According to a general theorem about Hermitian matrices,* the $\tilde{t}^{\mathrm{A}}{}_\alpha$ are then either irreducible or totally reducible.

---

* If a set of matrices H_α is not irreducible, then by definition there must be a set of vectors u_n that span a subspace (other than the whole space) that is left invariant by the H_α: that is, for all α and n, $H_\alpha u_n = \sum_m (C_\alpha)_{mn} u_m$. In this case we can adopt a basis consisting of the vectors u_n together with a set v_k that span the space orthogonal to all the u_n. If the H_α are Hermitian then $(u_n, H_\alpha v_k) = \sum_m (C_\alpha)^*_{mn}(u_m, v_k) = 0$, so the space spanned by the v_k is also invariant under the H_α: $H_\alpha v_k = \sum_\ell (D_\alpha)_{\ell k} v_\ell$. In this basis the matrices H_α are simultaneously reduced to a block-diagonal form:

$$H_\alpha = \begin{pmatrix} C_\alpha & 0 \\ 0 & D_\alpha \end{pmatrix} .$$

Continuing in this way, we can completely reduce the matrices H_α to a block-diagonal form with irreducible matrices in the blocks.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

By an irreducible set of $N \times N$ matrices $\tilde{t}^{A}{}_{\alpha}$ is meant a set for which there exists no subspace of dimensionality $< N$ that is left invariant by all the $\tilde{t}^{A}{}_{\alpha}$ — that is, no set of less than N non-zero vectors $(u_r)_\beta$ for which $(\tilde{t}^{A}{}_{\alpha})_{\beta\gamma}(u_r)_\gamma$ is for each α and r a linear combination of vectors $(u_s)_\beta$. Since the matrices $(\tilde{t}^{A}{}_{\alpha})_{\beta\gamma}$ are proportional to the structure constants in this basis, this is equivalent to the statement that there is no set of linear combinations $\mathscr{T}_r \equiv (u_r)_\gamma \tilde{t}^{A}{}_{\gamma}$ that is closed under commutation with all the $\tilde{t}^{A}{}_{\alpha}$, that is, for which $[\tilde{t}^{A}{}_{\alpha}, \mathscr{T}_r]$ is for each α and r a linear combination of the $\mathscr{T}_s$. Such a set of matrices $\mathscr{T}_r$ would furnish the generators of an invariant subalgebra of the full Lie algebra; the absence of such a set means that the Lie algebra is simple.

By a totally reducible set of matrices $\tilde{t}^{A}{}_{\alpha}$, is meant one that by a suitable choice of basis may be written as block-diagonal supermatrices

$$(\tilde{t}^{A}{}_{\alpha})_{ma,nb} = [t^{A(m)}{}_{\alpha}]_{ab}\, \delta_{mn}\,, \tag{15.A.6}$$

where the submatrices $t^{A(m)}{}_{\alpha}$ are either irreducible or vanish.** Adopting this basis also for the Lie algebra itself, the structure constants are then

$$\tilde{C}_{\ell c,ma,nb} = i(\tilde{t}^{A}{}_{\ell c})_{ma,nb} = i\,(t^{A(m)}{}_{\ell c})_{ab}\, \delta_{mn}\,. \tag{15.A.7}$$

But since this is totally antisymmetric, and proportional to δ_{mn}, it must also be proportional to $\delta_{\ell n}$ and $\delta_{\ell m}$:

$$\tilde{C}_{\ell c,ma,nb} = \delta_{\ell n}\, \delta_{\ell m}\, C^{(\ell)}_{cab}\,. \tag{15.A.8}$$

In other words, for *any* representation $t^{(m)}{}_{a} \equiv t_{ma}$ of the Lie algebra in this basis, we have

$$[t^{(m)}{}_{a}, t^{(n)}{}_{b}] = i\,\delta_{mn} C^{(m)}_{cab} t^{(m)}{}_{c}\,, \tag{15.A.9}$$

with $C^{(m)}_{cab}$ real and totally antisymmetric in the indices a, b, c. The fact that it is possible to construct a basis in which the generators fall into sets $t^{(m)}$, with the commutators of the generators in one set with each other given by a linear combination of generators in that set, and with all members of one set commuting with all members of any other set, is what we mean when we say that the Lie algebra is a *direct sum* of the subalgebras $t^{(m)}$. For each m, the set of matrices of the adjoint representation $t^{(m)A}$ is either irreducible or zero, corresponding to a subalgebra that is either simple or consists of so-called $U(1)$ generators, which commute with all generators of the whole algebra.

We have thus shown that the most general Lie algebra with totally antisymmetric real structure constants is a direct sum of one or more

** Here m and n label the blocks along the main diagonal, and a and b label rows and columns within these blocks. Also, m is not summed, and the range of the indices a, b, etc. in general depends on m.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

simple and/or $U(1)$ Lie algebras. Furthermore, the simple subalgebras are compact, in the sense that each matrix $-C^{(m)}_{acd}C^{(m)}_{bdc}$ is positive-definite, because for any real vector u_a, $-C^{(m)}_{acd}C^{(m)}_{bdc}u_a u_b = \sum_{cd}[\sum_a u_a C^{(m)}_{acd}]^2$ is a sum of positive quantities, which cannot vanish unless $u_a = 0$, since for $u_a \neq 0$ the condition that $\sum_a u_a C^{(m)}_{acd} = 0$ would imply that $\sum_a u_a t^{(m)}{}_a$ is itself an invariant Abelian subalgebra, in contradiction with the fact that the $t^{(m)}{}_a$ form a simple Lie algebra. This completes the proof of **c**.

Finally, let us assume **c**, that we have a Lie algebra that is the direct sum of a set of simple or $U(1)$ Lie algebras; that is, in some basis $t^{(m)}_a = \mathscr{S}_{ma,\alpha}t_\alpha$ with real non-singular $\mathscr{S}$, we have

$$[t^{(m)}_a, t^{(n)}_b] = i\delta_{nm}C^{(m)c}{}_{ab}t^{(m)}_c ,$$

where each subalgebra $t^{(m)}$ is either simple or commutes with everything. We furthermore assume that the simple subalgebras are compact, in the sense that the matrices

$$g^{(m)}_{ab} \equiv -C^{(m)c}{}_{ad}C^{(m)d}{}_{bc} \tag{15.A.10}$$

are positive-definite. To construct a real positive-definite matrix $g_{\alpha\beta}$ satisfying Eq. (15.A.1), in the basis of generators $t_{ma} = t^{(m)}_a$, we take

$$g_{ma,nb} \equiv g^{(m)}_{ab}\delta_{mn} , \tag{15.A.11}$$

where $g^{(m)}_{ab}$ is taken as the matrix (15.A.10) when $t^{(m)}$ is a simple subalgebra, and is taken as an arbitrary real symmetric positive-definite matrix when $t^{(m)}$ is a direct sum of one or more $U(1)$ subalgebras. The matrix (15.A.11) is obviously real, symmetric, and positive-definite because each $g^{(m)}_{ab}$ is. To check the requirement (15.A.1), recall the Jacobi identity for the structure constants:

$$C^{(m)c}{}_{ad}C^{(m)d}{}_{be} + C^{(m)c}{}_{bd}C^{(m)d}{}_{ea} + C^{(m)c}{}_{ed}C^{(m)d}{}_{ab} = 0 \tag{15.A.12}$$

and contract with $C^{(m)e}{}_{fc}$. After renaming the summation indices in the third term so that $c \to d \to e \to c$ and using Eq. (15.A.10), the result for the simple subalgebras may be written:

$$g^{(m)}_{df}C^{(m)d}{}_{ab} = C^{(m)c}{}_{ad}C^{(m)d}{}_{be}C^{(m)e}{}_{cf} - C^{(m)c}{}_{fd}C^{(m)d}{}_{be}C^{(m)e}{}_{ca} .$$

The important point is that this shows that the left-hand side is antisymmetric in a and f:

$$g^{(m)}_{df}C^{(m)d}{}_{ab} = -g^{(m)}_{da}C^{(m)d}{}_{fb} . \tag{15.A.13}$$

The same result holds trivially for the $U(1)$ subalgebras, where the structure constants vanish. The symmetry condition (15.A.1) follows immediately from Eq. (15.A.13), thus completing the proof of **a**. This concludes our proof of the equivalence of statements **a**, **b**, and **c**.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Now we come back to the matrix $g_{\alpha\beta}$. With totally antisymmetric structure constants, the invariance condition (15.2.4) may be expressed as the statement that this matrix commutes with all the matrices in the adjoint representation of the Lie algebra

$$[g,\ t^{A}{}_{\gamma}] = 0\,. \tag{15.A.14}$$

We have seen that all $(t^{A}{}_{\gamma})_{\alpha\beta}$ can be put in block-diagonal form, with irreducible (or zero) submatrices along the main diagonal. A well-known theorem[25] tells us that then $g_{\alpha\beta}$ must also be block-diagonal, with blocks of the same size and position as in the $t^{A}{}_{\gamma}$, and with the submatrix in each block proportional to the unit matrix. (Where two of the submatrices in the $t^{A}{}_{\gamma}$ are equivalent, in the sense that they can be related by a similarity transformation, it may be necessary to make a suitable change of basis in order to bring the submatrices of $g_{\alpha\beta}$ into a form proportional to unit matrices.) In the notation of Eq. (15.A.11), the metric is then given by Eq. (15.A.2).

Appendix B The Cartan Catalog

We present here without proof the complete catalog of simple Lie algebras, worked out in its final form by E. Cartan.[26] These will be presented here in their 'compact' form — that is, with generators that can be faithfully represented by finite-dimensional Hermitian matrices. The Lie algebras will be labelled with a subscript $n \geq 1$ indicating their 'rank' — the number of independent commuting linear combinations of generators.

A_n: This is the algebra of the special unitary group $SU(n+1)$, the group of all unitary ($U^{\dagger} = U^{-1}$) unimodular ($\mathrm{Det}\, U = 1$) matrices in $n+1$ dimensions. Any such matrix that is infinitesimally close to the identity may be expressed as

$$U = 1 + iH\,,$$

with infinitesimal H satisfying the conditions

$$H^{\dagger} = H\,,\quad \mathrm{Tr}\, H = 0\,,$$

so A_n is the Lie algebra of all Hermitian traceless matrices in $n+1$ dimensions. Any set of commuting Hermitian matrices may be simultaneously diagonalized, and the maximum number of independent diagonal *traceless* matrices in $n+1$ dimensions is n, so this is the rank of A_n. Any Hermitian matrix in $n+1$ dimensions has $(n+1)^2$ independent real parameters ($n+1$ real numbers on the main diagonal, and $n(n+1)/2$ complex numbers above the main diagonal, equal to the complex conjugates of those below

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

it), and one of these is eliminated by the tracelessness condition, so the dimensionality of A_n is

$$d(A_n) = (n+1)^2 - 1 = n(n+2)\,.$$

All of the A_n are simple.

$\boldsymbol{B_n}$: This is the algebra of the unitary orthogonal group $O(2n+1)$, consisting of all unitary ($\mathcal{O}^\dagger = \mathcal{O}^{-1}$) and orthogonal ($\mathcal{O}^{\mathrm{T}} = \mathcal{O}^{-1}$) and hence real matrices in $2n+1$ dimensions. (The restriction to unitary matrices is sometimes indicated by calling the group $UO(2n+1)$.) Any such matrix $\mathcal{O}$ that is infinitesimally close to the identity may be expressed as

$$\mathcal{O} = 1 + iA\,,$$

where A is an infinitesimal matrix satisfying the conditions

$$A^* = -A = A^{\mathrm{T}}\,.$$

(It would make no difference here if we restricted ourselves to the subgroup $SO(2n+1)$, for which $\mathcal{O}$ is subject to the further condition $\mathrm{Det}\,\mathcal{O} = 1$, since any orthogonal matrix $\mathcal{O}$ that is close to the identity would have $\mathrm{Det}\,\mathcal{O} = 1$ anyway.) Any set of commuting imaginary antisymmetric $(2n+1)$-dimensional matrices may (by a common orthogonal transformation) be put in the form of a supermatrix

$$\begin{bmatrix} a_1\sigma_2 & & & 0 \\ & a_2\sigma_2 & & \\ & & \ddots & \\ 0 & & & a_n\sigma_2 \quad 0 \end{bmatrix}$$

where σ_2 is the usual 2×2 matrix

$$\sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

and $a_1, \cdots, a_n$ are real. The rank of B_n is thus evidently n. An imaginary antisymmetric matrix is completely specified by the imaginary numbers above the real diagonal, so its dimensionality is

$$d(B_n) = \frac{(2n+1)(2n)}{2} = n(2n+1)\,.$$

All of the B_n are simple.

There is an alternative definition of $O(N)$ that will help in understanding the motivation for the next large set of simple Lie algebras. Instead of defining $O(N)$ as the group of N-dimensional real matrices that satisfy the orthogonality condition $\mathcal{O}^{\mathrm{T}}\mathcal{O} = 1$, it can just as well be defined as the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

group of N-dimensional real matrices $\mathscr{M}$ that satisfy the condition

$$\mathscr{M}^{\mathrm{T}}\mathscr{P}\mathscr{M} = \mathscr{P} ,$$

where $\mathscr{P}$ is an arbitrary positive-definite real symmetric matrix. This is because any such $\mathscr{P}$ may always be expressed as $\mathscr{P} = \mathscr{R}^{\mathrm{T}}\mathscr{R}$ with $\mathscr{R}$ some real non-singular matrix, so there is a similarity transformation that takes any matrix $\mathscr{M}$ satisfying the above condition into a real orthogonal matrix, $\mathscr{R}\mathscr{M}\mathscr{R}^{-1}$. Thus we may let $\mathscr{P}$ be various different real symmetric positive-definite matrices without changing the group.

$\boldsymbol{C_n}$: This is the algebra of the unitary symplectic group $USp(2n)$, the group of unitary matrices $\mathscr{M}$ that leave an *antisymmetric* non-singular matrix $\mathscr{A}$ invariant:

$$\mathscr{M}^{\mathrm{T}}\mathscr{A}\mathscr{M} = \mathscr{A} ,$$

$$\mathscr{A}^{\mathrm{T}} = -\mathscr{A} \quad , \quad \mathrm{Det}\,\mathscr{A} \neq 0 .$$

(Note that in d dimensions, $\mathrm{Det}\,\mathscr{A} = \mathrm{Det}\,\mathscr{A}^{\mathrm{T}} = (-)^d \mathrm{Det}\,\mathscr{A}$, so if d is odd then $\mathrm{Det}\,\mathscr{A}$ must vanish. Thus there is no $USp(d)$ unless d is even.) Any such antisymmetric non-singular (perhaps complex) matrix $\mathscr{A}$ may be written in the standard form

$$\mathscr{A} = \mathscr{R}^{\mathrm{T}}\mathscr{A}_0\mathscr{R} ,$$

where $\mathscr{R}$ is unitary and $\mathscr{A}_0$ is the supermatrix:

$$\mathscr{A}_0 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} .$$

(Alternatively, we could take $\mathscr{A}_0$ as a block-diagonal supermatrix, with σ_2 s along the main diagonal.) Hence $USp(2n)$ may be described as the group of unitary matrices $\mathscr{S}$ satisfying

$$\mathscr{S}^{\mathrm{T}}\mathscr{A}_0\mathscr{S} = \mathscr{A}_0$$

since by a unitary transformation any such $\mathscr{S}$ can be transformed into a unitary matrix $\mathscr{M} = \mathscr{R}^{-1}\mathscr{S}\mathscr{R}$ that satisfies the previous condition $\mathscr{M}^{\mathrm{T}}\mathscr{A}\mathscr{M} = \mathscr{A}$. Any such matrix $\mathscr{S}$ that is infinitesimally far from the identity may be written

$$\mathscr{S} = 1 + i\mathscr{H} ,$$

where $\mathscr{H}$ is an infinitesimal matrix satisfying the conditions

$$\mathscr{H}^{\dagger} = \mathscr{H}, \quad \mathscr{H}^{\mathrm{T}}\mathscr{A}_0 + \mathscr{A}_0\mathscr{H} = 0 .$$

The most general $2n$-dimensional matrix $\mathscr{H}$ satisfying these conditions

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

may be written as a supermatrix

$$\mathscr{H} = \begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{B}^* & -\mathscr{A}^* \end{bmatrix},$$

where the n-dimensional complex submatrices $\mathscr{A}, \mathscr{B}$ satisfy

$$\mathscr{A}^\dagger = \mathscr{A}, \qquad \mathscr{B}^{\mathrm{T}} = \mathscr{B}.$$

A maximal set of commuting generators have $\mathscr{A}$ diagonal and $\mathscr{B}$ zero, and so are of the form

$$\mathscr{H} = \begin{bmatrix} a_1 & & & & 0 \\ & \ddots & & & \\ & & a_n & & \\ & & & -a_1 & \\ & & & & \ddots \\ 0 & & & & -a_n \end{bmatrix}$$

with $a_1, \cdots, a_n$ real. The rank of C_n is thus evidently n. The dimensionality of C_n is the number n^2 of independent real parameters in the Hermitian matrix $\mathscr{A}$, plus the number $2n(n+1)/2$ of independent real parameters in the complex symmetric matrix $\mathscr{B}$

$$d(C_n) = n^2 + 2 \times n(n+1)/2 = n(2n+1).$$

All of the C_n algebras are simple.

$\boldsymbol{D_n}$: This is the algebra of the unitary orthogonal group $O(2n)$, consisting of all unitary orthogonal matrices in $2n$ dimensions. The discussion of B_n can be carried over to D_n, except that here any set of commuting generators can be put in the form

$$\begin{bmatrix} a_1\sigma_2 & & & 0 \\ & a_2\sigma_2 & & \\ & & \ddots & \\ 0 & & & a_n\sigma_2 \end{bmatrix}$$

so the rank is still n. Also, the dimensionality here is

$$d(D_n) = \frac{(2n)(2n-1)}{2} = n(2n-1).$$

All of the D_n are simple except for D_1, which is the Abelian algebra with just one generator, and D_2, which is the direct sum $B_1 + B_1$.

Exceptional Lie Algebras: In addition to the above classical Lie algebras, there are just five special cases, the algebras $G_2(d = 14)$; $F_4(d = 52)$; $E_6(d = 78)$; $E_7(d = 133)$; $E_8(d = 248)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

Not all of the classical Lie algebras are really different. There are just four isomorphisms

$$A_1 = B_1 = C_1 , \qquad C_2 = B_2 , \qquad A_3 = D_3 .$$

These correspond to isomorphisms among Lie groups of which these are the algebras. However, isomorphisms like $B_1 = A_1$, $B_2 = C_2$, and $D_3 = A_3$ do not mean that $SO(3)$ is isomorphic to $SU(2)$, or that $SO(5)$ is isomorphic to $USp(4)$, or that $SO(6)$ is isomorphic to $SU(4)$. Instead, $SU(2), USp(4)$, and $SU(4)$ are the simply connected *covering groups* of $SO(3)$, $SO(5)$, and $SO(6)$. (Covering groups are discussed in Chapter 2.) Nevertheless, the isomorphisms of the Lie algebras make it especially easy to construct the double-valued fundamental spinor representations of $SO(3)$, $SO(5)$, and $SO(6)$; they are just the defining representations of $SU(2)$, $USp(4)$, and $SU(4)$, respectively. Also $SO(4)$ is isomorphic to $SO(3) \times SO(3)$, so its double-valued spinor representation is just the defining representation of $SU(2) \times SU(2)$. For $d \geq 7$ the double-valued spinor representations of $SO(d)$ must be constructed by other means. The simplest technique uses Clifford algebras, discussed in Section 5.4.

Problems

1. Derive the Bianchi identity

$$D_\mu F_{\alpha\nu\lambda} + D_\nu F_{\alpha\lambda\mu} + D_\lambda F_{\alpha\mu\nu} = 0 .$$

2. Suppose we use generalized Coulomb gauge in a non-Abelian gauge theory, taking the gauge-fixing function as $f_\alpha = \nabla \cdot \mathbf{A}_\alpha$. Derive the ghost Lagrangian. What is the ghost propagator? (Take $B[f] = \exp(-i \int d^4x \, f_\alpha f_\alpha / 2\xi)$.)

3. Suppose that in electrodynamics we use a gauge-fixing function $f = \partial_\mu A^\mu + cA_\mu A^\mu$, with c an arbitrary constant. Derive the ghost Lagrangian. (Take $B[f] = \exp(-i \int d^4x \, f^2 / 2\xi)$.) What is the ghost propagator?

4. Show that there is no simple Lie algebra with just four generators.

5. Show that if a field $\psi_\ell(x)$ belonging to a representation of a gauge group with generator matrices t_α varies along a path $x^\mu = x^\mu(\tau)$ according to the differential equation

$$\frac{d\psi(\tau)}{d\tau} = it_\alpha \psi(\tau) A_{\alpha\mu}(x(\tau)) \frac{dx^\mu(\tau)}{d\tau} ,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

(with both $\psi_\ell(x)$ and $A_{\alpha\mu}(x)$ classical c-number fields) then the change in ψ around any *small* closed path $\mathscr{P}$ around a point X^μ is proportional to

$$t_\alpha \psi(X) F_{\alpha\mu\nu}(X) \oint_{\mathscr{P}} x^\mu(\tau) \frac{dx^\nu(\tau)}{d\tau} d\tau .$$

Find the constant of proportionality. Use this result to show that if $F_{\alpha\mu\nu}(x)$ vanishes everywhere then it is possible by a gauge transformation to make $A_{\alpha\mu}(x)$ vanish throughout at least a finite region. (Hint: Follow the analogous argument in electrodynamics, or in general relativity, as in Ref. 6.)

6. Applying path-integral methods to a general non-Abelian gauge theory, calculate the propagator of the gauge field $A_{\alpha\mu}(x)$ if we choose a gauge-fixing functional $f_\alpha = n_\mu A^\mu_\alpha$, with n_μ arbitrary constants. (Take $B[f] = \exp(-i \int d^4x\, f_\alpha f_\alpha / 2\xi)$.) What is the ghost Lagrangian? What is the ghost propagator? What is the ghost interaction vertex?

7. Suppose we adopt BRST invariance instead of gauge invariance as a fundamental physical principle. Derive the most general Lagrangian density constructed from sums of products of fields and field derivatives of dimensionality (mass)^d with $d \leq 4$, constructed from $A_{\alpha\mu}$, ω_α, ω^*_α, h_α, and/or their derivatives, that is invariant (up to total derivatives) under Lorentz transformations, ghost number phase transformations ($\omega_\alpha \to e^{i\theta}\omega_\alpha$, $\omega^*_\alpha \to e^{-i\theta}\omega^*_\alpha$), global gauge transformations (ϵ_α constant), and BRST transformations.

8. Show that the antibracket satisfies the symmetry condition (15.9.15) and the Jacobi identity (15.9.21).

9. Show that if a functional O satisfies the condition $(O, S) = i\Delta S$ and the action S satisfies the quantum master equation then the quantum average $\langle O \rangle$ is independent of the gauge-fixing functional Ψ.

References

1. C. N. Yang and R. L. Mills, *Phys. Rev.* **96**, 191 (1954). O. Klein had come close to a formulation of the $SU(2)$ Yang–Mills theory in a talk at a 1938 conference at Warsaw, reported in *New Theories in Physics* (International Institute of Intellectual Cooperation, Paris, 1939). For a critical discussion of Klein's theory, see D. J. Gross, 'Oscar Klein and Gauge Theory,' talk at the 1994 Oscar Klein symposium at Stockholm, Princeton preprint PUPT–1508.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

2. R. Utiyama, *Phys. Rev.* **101**, 1597 (1956).

3. R. P. Feynman, *Acta Phys. Polonica* **24**, 697 (1963).

4. L. D. Faddeev and V. N. Popov, *Phys. Lett.* **25B**, 29 (1967).

5. B. S. De Witt, *Phys. Rev.* **162**, 1195, 1239 (1967).

6. The proof is easily adapted from the standard proof of the corresponding result in general relativity that, for the existence of a coordinate transformation to a flat metric in a finite simply connected region, it is necessary and sufficient that the Riemann–Christoffel curvature tensor should vanish. See, for example, S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Sections 6.3 and 6.4.

7. The proof that **a** (with $g_{\alpha\beta} = \delta_{\alpha\beta}$) implies **b** and that **b** implies **c** was given by M. Gell-Mann and S. L. Glashow, *Ann. Phys. (N.Y.)* **15**, 437 (1961).

8. See, for example, S. Weinberg, *ibid.*, Section 7.6.

9. V. Gribov, *Nucl. Phys.* **B139**, 1 (1978). Also see R. Jackiw, I. Muzinich, and C. Rebbi, *Phys. Rev.* **D17**, 1576 (1978); R. Jackiw, in *New Frontiers in High Energy Physics*, eds. B. Kursunoglu, A. Perlmutter, and L. Scott (Plenum, New York, 1978); N. Christ and T. D. Lee, *Phys. Rev.* **D22**, 939 (1980); R. Jackiw, in *Current Algebra and Anomalies* (World Scientific, Singapore, 1985): footnote 50.

10. C. Becchi, A. Rouet, and R. Stora, *Comm. Math. Phys.* **42**, 127 (1975); in *Renormalization Theory*, eds. G. Velo and A. S. Wightman (Reidel, Dordrecht, 1976); *Ann. Phys.* **98**, 287 (1976).

11. I. V. Tyutin, Lebedev Institute preprint N39 (1975).

11*a*. N. Nakanishi, *Prog. Theor. Phys.* **35**, 1111 (1966); B. Lautrup, *Mat. Fys. Medd. Kon. Dan. Vid.-Sel. Medd.* **35**, 29 (1967).

12. The original source of this argument is unknown to me. I learned it from J. Polchinski.

13. S. N. Gupta, *Proc. Phys. Soc.* **63**, 681 (1950); **64**, 850 (1951); K. Bleuler, *Helv. Phys. Acta* **3**, 567 (1950); K. Bleuler and W. Heitler, *Progr. Theor. Phys.* **5**, 600 (1950).

13*a*. G. Curci and R. Ferrari, *Nuovo Cimento* **35**, 273 (1976); T. Kugo and I. Ojima, *Prog. Theor. Phys.* **60**, 1869 (1978); *Prog. Theor. Phys. Suppl.* **66**, 1 (1979). Also see Ref. 11*a* for the case of electrodynamics.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

14. J. Thierry-Mieg, *J. Math. Phys.* **21**, 2834 (1980); R. Stora, in *Progress in Gauge Theories: Proceedings of a Symposium at Cargèse*, 1983, eds. G. 't Hooft, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora (Plenum, New York, 1984): p. 543; L. Bonora and P. Cotta-Ramusino, *Commun. Math. Phys.* **87**, 589 (1983); J. Mañes, R. Stora, and B. Zumino, *Commun. Math. Phys.* **102**, 157 (1985); A. S. Schwarz, *Commun. Math. Phys.* **155**, 249 (1993); P. M. Lavrov, P. Yu Moshhin and A. A. Reshetnyak, Tomsk preprint hep-th/9507104; P. M. Lavrov, Tomsk preprint hep-th/9507105.

15. W. Siegel, *Phys. Lett.* **93B**, 170 (1980); T. Kimura, *Prog. Theor. Phys.* **64**, 357 (1980); **65**, 338 (1981).

16. G. Barnich, F. Brandt, and M. Henneaux, *Commun. Math. Phys.* **174**, 57 (1995): Section 14.

17. G. Curci and R. Ferrari, *Nuovo Cimento* **32A**, 151 (1976); I. Ojima, *Prog. Theor. Phys. Supp.* **64**, 625 (1980); L. Baulieu and J. Thierry-Mieg, *Nucl. Phys.* **B197**, 477 (1982).

18. The anti-BRST transformation was extended to general local symmetries by L. Alvarez-Gaumè and L. Baulieu, *Nucl. Phys.* **B212**, 255 (1983). It is induced by the operator:

$$\bar{s} = \omega^{*A}\,\delta_A\phi^r\,\frac{\delta_L}{\delta\phi^r} - \frac{1}{2}\omega^{*B}\omega^{*C}\,f^A{}_{BC}\,\frac{\delta_L}{\delta\omega^{*A}} - \omega^{*B}\,h^C\,f^A{}_{BC}\,\frac{\delta}{\delta h^A} + \Big[-f^A{}_{BC}\omega^{*B}\omega^C + h^A\Big]\frac{\delta_L}{\delta\omega^A}\,.$$

It is straightforward to show that this transformation is nilpotent:

$$\bar{s}^2 = 0\,.$$

Also, the BRST and anti-BRST transformations anticommute. Furthermore, there is a cohomology theorem (unpublished) for the anti-BRST transformations, like that for the BRST transformations: the most general functional I of ϕ^r, ω^A, ω^{*A}, and h^A that satisfies the anti-BRST invariance condition $\bar{s}I = 0$ and has ghost number zero is of the form

$$I[\phi,\omega,\omega^*,h] = I_0[\phi] + \bar{s}\,\bar{\Psi}[\phi,\omega,\omega^*,h]\,.$$

(Just use the Hodge operator $\bar{t} \equiv \omega^A\,\delta/\delta h^A$ in place of t.) Anti-BRST symmetry can be an aid in enumerating possible terms in Lagrangians and Greens functions, but I do not know of any case where it is indispensable.

19. I. A. Batalin and G. A. Vilkovisky, *Phys. Lett.* **B102**, 27 (1981); *Nucl. Phys.* **B234**, 106 (1984); *J. Math. Phys.* **26**, 172 (1985). Also see B.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

L. Voronov and I. V. Tyutin, *Theor. Math. Phys.* **50**, 218, 628 (1982). For a lucid review, see J. Gomis, J. París and S. Samuel, *Phys. Rep.* **259**, 1 (1995).

20. E. S. Fradkin and G. A. Vilkovisky, *Phys. Lett.* **B55**, 224 (1975); CERN report TH2332 (1977); I. A. Batalin and G. A. Vilkovisky, *Phys. Lett.* **B69**, 309 (1977); I. S. Fradkin and T. E. Fradkina, *Phys. Lett.* **B72**, 334 (1977). Also see M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems* (Princeton University Press, Princeton, 1992).

21. I. A. Batalin and I. V. Tyutin, *Phys. Lett.* **B356**, 373 (1995).

22. J. Zinn-Justin, in *Trends in Elementary Particle Theory – International Summer Institute on Theoretical Physics in Bonn 1974* (Springer-Verlag, Berlin, 1975): p. 2.

23. D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, *Phys. Rev.*, **D13**, 3214 (1976); R. E. Kallosh, *Nucl. Phys.* **B141**, 141 (1978); B. de Wit and J. W. van Holten, *Phys. Lett.*, **B79**, 389 (1979); P. van Nieuwenhuizen *Phys. Rep.*, **68**, 189 (1981).

24. M. Henneaux and C. Teitelboim, Ref. 20: Section 18.1.4.

25. E. P. Wigner, *Group Theory* (Academic Press, New York, 1959): pp. 76–8.

26. The original reference is E. Cartan, *Sur la Structure des Groupes de Transformations Finis et Continue*, (Paris, 1894; 2nd edn 1933). For a textbook treatment, see, for example, R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974): Chapter 9.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.003

16
External Field Methods

It is often useful to consider quantum field theories in the presence of a classical external field. One reason is that in many physical situations, there really is an external field present, such as a classical electromagnetic or gravitational field, or a scalar field with a non-vanishing vacuum expectation value. (As we shall see in Chapter 19, such scalar fields can play an important role in the spontaneous breakdown of symmetries of the Lagrangian.) But even where there is no actual external field present in a problem, some calculations are greatly facilitated by considering physical amplitudes in the presence of a fictitious external field. This chapter will show that it is possible to take all multiloop effects into account by summing 'tree' graphs whose vertices and propagators are taken from a *quantum effective action*, which is nothing but the one-particle-irreducible connected vacuum–vacuum amplitude in the presence of an external field. It will turn out in the next chapter that this provides an especially handy way both of completing the proof of the renormalizabilty of non-Abelian gauge theories begun in Chapter 15, and of calculating the charge renormalization factors that we need in order to establish the crucial property of asymptotic freedom in quantum chromodynamics.

16.1 The Quantum Effective Action

Consider a quantum field theory with action $I[\phi]$, and suppose we 'turn on' a set of classical currents $J_r(x)$ coupled to the fields $\phi^r(x)$ of the theory. The complete vacuum–vacuum amplitude in the presence of these currents is then

$$\begin{aligned} Z[J] &\equiv \langle \text{VAC, out} | \text{VAC, in} \rangle_J \\ &= \int \left[\prod_{s,y} d\phi^s(y) \right] \exp\left(iI[\phi] + i \int d^4x \, \phi^r(x) J_r(x) + \epsilon \text{ terms} \right) . \end{aligned} \tag{16.1.1}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

(The fields $\phi^r(x)$ need not be scalars. They might even be fermionic, though until Section 16.4 we shall not bother to keep track of the signs that would appear in that case.) The Feynman rules for calculating $Z[J]$ are just the same as for calculating the vacuum–vacuum amplitude $Z[0]$ in the absence of the external current, except that the Feynman diagrams now contain vertices of a new kind, to which a *single* ϕ^r-line is attached. Such a vertex labelled with a coordinate x contributes a 'coupling' factor $iJ_r(x)$ to the integrand of the position-space Feynman amplitude. Equivalently, we could say that in the expansion of $Z[J]$ in powers of J, the coefficient of the term proportional to $iJ_r(x)\,iJ_s(y)\cdots$ is just the sum of diagrams with external lines (including propagators) corresponding to the fields $\phi^r(x)$, $\phi^s(y)$, etc. In particular, the first derivative gives the vacuum matrix element of the quantum mechanical operator $\Phi^r(x)$ corresponding to $\phi^r(x)$:

$$\left[\frac{\delta}{\delta J_r(y)}\,Z[J]\right]_{J=0} = i\int\left[\prod_{r,x} d\phi^r(x)\right]\,\phi^r(y)\,\exp\{iI[\phi]+\epsilon\text{ terms}\}$$
$$= i\,\langle \text{VAC, out}|\Phi^r(y)|\text{VAC, in}\rangle_{J=0}\,. \qquad (16.1.2)$$

We sometimes work with functionals $Z[J]$ defined by (16.1.1) where the $\phi^r(x)$ are not elementary fields (that is, fields appearing in the action) but products of such fields. Where $\phi^r(x)$ is a product of N elementary fields, the new vertices in the Feynman rules for $Z[J]$ have N lines attached. Some of the results of this chapter (including Eq. (16.1.2)) apply in this case, but where Feynman diagrams are involved, it will be tacitly assumed that the $\phi^r(x)$ are elementary.

Now, $Z[J]$ is given by the sum of *all* vacuum–vacuum amplitudes in the presence of the current J, including disconnected as well as connected diagrams, but not counting as different those diagrams that differ only by a permutation of vertices in the same or different connected subdiagrams. A general diagram that consists of N connected components will contribute to $Z[J]$ a term equal to the product of the contributions of these components, divided by the number $N!$ of permutations of vertices that merely permute all the vertices in one connected component with all the vertices in another.* Hence, the sum of all graphs is

$$Z[J] = \sum_{N=0}^{\infty}\frac{1}{N!}\,(iW[J])^N = \exp(iW[J])\,, \qquad (16.1.3)$$

* The contribution of a Feynman diagram with N connected components containing $n_1, n_2, \cdots n_N$ vertices will be proportional to a factor $1/(n_1+\cdots n_N)!$ from the Dyson expansion, and a factor $(n_1+\cdots n_N)!/N!$ equal to the number of permutations of these vertices, counting as identical those permutations that merely permute all the vertices in one component with all the vertices in another.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

where $iW[J]$ is the sum of all *connected* vacuum–vacuum amplitudes, again not counting as different those diagrams that differ only by a permutation of vertices.

For many purposes, it will be useful to go one step further, and in place of $W(J)$ work with the sum of all connected *one-particle-irreducible* graphs. (A one-particle-irreducible graph is one that cannot be disconnected by cutting through any one internal line.) We can give a formal expression for this sum as follows. First, define $\phi^r_J(x)$ as the vacuum expectation value of the operator $\Phi^r(x)$ in the presence of the current J:

$$\phi^r_J(x) \equiv \frac{\langle \text{VAC, out}|\Phi^r(x)|\text{VAC, in}\rangle_J}{\langle \text{VAC, out}|\text{VAC, in}\rangle_J} = -\frac{i}{Z[J]}\frac{\delta}{\delta J_r(x)} Z[J] \qquad (16.1.4)$$

or in terms of the sum of connected graphs

$$\phi^r_J(x) = \frac{\delta}{\delta J_r(x)} W[J] . \qquad (16.1.5)$$

This formula can be inverted. Define $J_{\phi r}(x)$ as the current for which (16.1.4) has a prescribed value $\phi^r(x)$:

$$\phi^r_J(x) = \phi^r(x) \quad \text{if} \quad J_r(x) = J_{\phi r}(x) .$$

The *quantum effective action*[1] $\Gamma[\phi]$ is defined (as a functional of ϕ, not J) by the Legendre transformation

$$\Gamma[\phi] \equiv -\int d^4x\, \phi^r(x) J_{\phi r}(x) + W[J_\phi] . \qquad (16.1.6)$$

We will soon show that $\Gamma[\phi]$ is the sum of all connected one-particle-irreducible graphs in the presence of the current J_ϕ. However, let us first take a look at another aspect of its physical significance.

Note that the variational derivative of $\Gamma[\phi]$ is

$$\frac{\delta\Gamma[\phi]}{\delta\phi^s(y)} = -\int d^4x\, \phi^r(x) \frac{\delta J_{\phi r}(x)}{\delta\phi^s(y)} - J_{\phi s}(y) + \int d^4x \left[\frac{\delta W[J]}{\delta J_r(x)}\right]_{J=J_\phi} \frac{\delta J_{\phi r}(x)}{\delta\phi^s(y)}$$

or, using Eq. (16.1.5),

$$\frac{\delta\Gamma[\phi]}{\delta\phi^s(y)} = -J_{\phi s}(y) . \qquad (16.1.7)$$

Thus $\Gamma[\phi]$ is the 'effective action', in the sense that the possible values for the external fields $\phi^s(y)$ in the *absence* of a current J are given by the stationary 'points' of Γ:

$$\frac{\delta\Gamma[\phi]}{\delta\phi^s(y)} = 0 \quad \text{for} \quad J = 0 . \qquad (16.1.8)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

This may be compared with the classical field equations, which just require that the actual action $I[\phi]$ be stationary. Hence Eq. (16.1.8) may be regarded as the equation of motion for the external field ϕ, taking quantum corrections into account.

Not only does $\Gamma[\phi]$ provide the quantum-corrected field equations; it is an effective action in the sense that $iW[J]$ may be calculated as a sum of connected *tree* graphs for the vacuum–vacuum amplitude, with vertices calculated as if the action were $\Gamma[\phi]$ instead of $I[\phi]$. By a tree graph is meant one that becomes disconnected if we cut any internal line. All the effects of loop diagrams are taken into account by using $\Gamma[\phi]$ in place of $I[\phi]$.

To see this,[2] let us consider the quantity $W_\Gamma[J,g]$ that we would get instead of $W[J]$, if we used an action $g^{-1}\Gamma[\phi]$ in place of $I[\phi]$:

$$\exp\{iW_\Gamma[J,g]\} \equiv \int \prod_{r,x} d\phi^r(x)\, \exp\left\{ig^{-1}\left[\Gamma[\phi] + \int d^4x\, \phi^r(x) J_r(x)\right]\right. \\ \left. + \epsilon \text{ terms}\right\}, \tag{16.1.9}$$

with arbitrary constant g. The propagator here is the inverse of the coefficient of the term in $g^{-1}\Gamma(\phi)$ quadratic in ϕ, and is hence proportional to g, while all vertices make a contribution proportional to $1/g$, so a graph with V vertices (including those produced by the current J) and I internal lines (including those attached to the J vertices) is proportional to g^{I-V}. For any connected graph, the number of loops is $L = I - V + 1$, so the L-loop term in $W_\Gamma[J,g]$ has the g dependence

$$\Big(W_\Gamma[J,g]\Big)_{L \text{ loops}} \propto g^{L-1}\,. \tag{16.1.10}$$

Equivalently, we may write (at least formally)

$$W_\Gamma[J,g] = \sum_{L=0}^{\infty} g^{L-1} W_\Gamma^{(L)}[J]\,, \tag{16.1.11}$$

where (as can be seen by setting $g = 1$), the quantity $W_\Gamma^{(L)}[J]$ is the L-loop contribution to the connected vacuum amplitude $W_\Gamma[J,1]$ that we would obtain if we used $\Gamma[\phi]$ (*without* a factor g) in place of the action $I[\phi]$.

Now, we are specially interested here in the sum of *tree* graphs, those without loops, calculated with vertices and propagators calculated as if the action were $\Gamma[\phi]$ instead of $I[\phi]$. In our present notation, this is $W_\Gamma^{(0)}[J]$. In order to isolate the $L = 0$ term in Eq. (16.1.11), consider the limit $g \to 0$. In this limit, the path integral (16.1.9) is dominated by the point of stationary phase,

$$\exp\left\{i\, W_\Gamma[J,g]\right\} \propto \exp\left\{ig^{-1}\left[\Gamma[\phi_J] + \int d^4x\, \phi_J^r(x) J_r(x)\right]\right\}\,, \tag{16.1.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

where, because of its definition as the field produced by the current J, the field ϕ_J is the stationary point of the exponent, in the sense that

$$\left.\frac{\delta\Gamma[\phi]}{\delta\phi^r(x)}\right|_{\phi=\phi_J} = -J_r(x)\,. \tag{16.1.13}$$

The proportionality factor in Eq. (16.1.12) is in general a functional of J, but it is a power series in g starting with terms of order g^0. Hence taking the logarithm of both sides and isolating the terms of order g^{-1} gives

$$W_\Gamma^{(0)}[J] = \Gamma[\phi_J] + \int d^4x\,\phi_J^r(x)\,J_r(x)\,. \tag{16.1.14}$$

By setting $\phi = \phi_J$ in Eq. (16.1.6), we see that the right-hand side of Eq. (16.1.14) is just $W[J]$:

$$W_\Gamma^{(0)}[J] = W[J]\,. \tag{16.1.15}$$

To recapitulate, this says that $W[J]$ may be calculated by using $\Gamma[\phi]$ in place of $I[\phi]$ (the subscript Γ) and keeping only tree (0-loop) graphs:

$$iW[J] = \int_{\substack{\text{CONNECTED}\\ \text{TREE}}} \left[\prod_{r,x} d\phi^r(x)\right] \exp\left\{i\,\Gamma[\phi] + i\int \phi^r(x)\,J_r(x)\,d^4x\right\}\,. \tag{16.1.16}$$

Now, any connected graph for $iW[J]$ can be regarded as a tree, whose vertices consist of one-particle-irreducible subgraphs. Thus in order for Eq. (16.1.16) to be correct, $i\Gamma[\phi]$ must be the sum of all one-particle-irreducible connected graphs with arbitrary numbers of external lines, each external line corresponding to a factor ϕ rather than a propagator or wave function. For this reason, the coefficients in an expansion of $\Gamma[\phi]$ in powers of fields and their derivatives around some fixed field ϕ_0 may be regarded as *renormalized* coupling constants, with the renormalization 'point' specified by ϕ_0 rather than by some set of momenta.

Equivalently, $i\Gamma[\phi_0]$ for some fixed field $\phi_0^r(x)$ may be expressed as the sum of one-particle-irreducible graphs for the vacuum–vacuum amplitude, calculated with a shifted action $I[\phi + \phi_0]$:

$$i\,\Gamma[\phi_0] = \int_{\substack{\text{1PI}\\ \text{CONNECTED}}} \left[\prod_{r,x} d\phi^r(x)\right] \exp\{i\,I[\phi + \phi_0]\}\,. \tag{16.1.17}$$

This is because any place where ϕ_0 appears in any of the vertices or propagators within the one-particle-irreducible graphs in Eq. (16.1.17) is also a place where an external ϕ-line could be attached. (The restriction to one-particle-irreducible graphs plays an essential role in Eq. (16.1.17); without this restriction we could shift the variable of integration, yielding

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

an integral that would be manifestly independent of ϕ_0.) In place of Eq. (16.1.17), it is often convenient to write

$$\exp\left[i\Gamma[\phi_0]\right] = \int_{1\text{PI}} \left[\prod_{rx} d\phi^r(x)\right] \exp\left\{iI[\phi+\phi_0]\right\}, \qquad (16.1.18)$$

in which we evaluate the path integral including all graphs, connected or not, in which each connected component is one-particle-irreducible.

* * *

This formalism provides a simple method of summing tree graphs. As one example, consider the relation between the complete two-point function $\Delta^{rx,sy}$ and its one-particle-irreducible part $\Pi_{rx,sy}$. From Eqs. (16.1.5) and (16.1.7), we find

$$\Delta^{rx,sy} \equiv \frac{\delta^2 W[J]}{\delta J_r(x)\,\delta J_s(y)} = \frac{\delta\phi^r_J(x)}{\delta J_s(y)}, \qquad (16.1.19)$$

$$\Pi_{rx,sy} \equiv \frac{\delta^2\Gamma[\phi]}{\delta\phi^r(x)\,\delta\phi^s(y)} = -\frac{\delta J_{\phi\, r}(x)}{\delta\phi^s(y)}. \qquad (16.1.20)$$

It follows immediately that the 'matrices' Δ and Π are related by

$$\Delta = -\Pi^{-1}. \qquad (16.1.21)$$

This is the counterpart of the familiar relation (10.3.15) between propagators and self-energy parts, with the extra term q^2+m^2 in the denominator in Eq. (10.3.15) representing the zeroth-order term in the one-particle-irreducible two-point function.

16.2 Calculation of the Effective Potential

To see how the formalism of the previous section works in practice, consider a simple example, the renormalizable theory of a single real scalar field $\phi(x)$, with action

$$I[\phi] = -\int d^4x \left[\lambda + \tfrac{1}{2}\partial_\rho\phi\partial^\rho\phi + \tfrac{1}{2}m^2\phi^2 + \tfrac{1}{24}g\phi^4\right]. \qquad (16.2.1)$$

(We are here including a 'cosmological constant' $-\lambda$ in the Lagrangian density, for reasons that will become apparent). Suppose for simplicity that we wish to calculate $\Gamma[\phi_0]$ for a position-independent field $\phi_0(x) = \phi_0$. Then every term in $\Gamma[\phi_0]$ will contain a factor of the volume of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004



Figure 16.1. Feynman diagrams with zero, one, or two loops for the quantum effective action of the theory of a neutral scalar field ϕ with interaction ϕ^4.

spacetime

$$\mathscr{V}_4 = \int d^4x = \delta^4(p-p)(2\pi)^4 \tag{16.2.2}$$

arising from the momentum conservation delta-function. We will therefore write, for ϕ_0 constant

$$\Gamma[\phi_0] = -\mathscr{V}_4 V(\phi_0)\,, \tag{16.2.3}$$

where $V(\phi_0)$ is an ordinary function, known as the *effective potential.* In this section we will calculate the effective potential to one-loop order. This was originally done by Coleman and E. Weinberg[3] in a study of spontaneous symmetry breaking, the subject of Chapters 19 and 21. The results were also used by them in one of the early applications of the renormalization group, to be described in Section 18.2.

Shifting ϕ by ϕ_0, the action in Eq. (16.1.18) is:*

$$\begin{aligned} I[\phi+\phi_0] = &-\mathscr{V}_4\left[\lambda + \tfrac{1}{2}m^2\phi_0^2 + \tfrac{1}{24}g\phi_0^4\right] - [m^2\phi_0 + \tfrac{1}{6}g\,\phi_0^3]\int d^4x\,\phi \\ &- \int d^4x\left[\tfrac{1}{2}\partial_\rho\phi\partial^\rho\phi + \tfrac{1}{2}\mu^2(\phi_0)\phi^2\right] - \int d^4x\left[\tfrac{1}{6}g\,\phi_0\phi^3 + \tfrac{1}{24}g\,\phi^4\right], \end{aligned} \tag{16.2.4}$$

where μ^2 is the field-dependent mass

$$\mu^2(\phi_0) = m^2 + \tfrac{1}{2}g\phi_0^2\ . \tag{16.2.5}$$

Note that there now appear new interactions proportional to ϕ (which have no effect on one-particle-irreducible graphs) and also ϕ^3, as well as terms with the same structure as those in the original action.

The Feynman diagrams for $\Gamma[\phi_0]$ with up to two loops are shown in Figure 16.1. The zero-loop term in the vacuum–vacuum amplitude is just given by the constant term in $I[\phi+\phi_0]$

$$i\,\Gamma^{(0\text{ loop})}[\phi_0] = -i\mathscr{V}_4\left(\lambda + \frac{1}{2}m^2\phi_0^2 + \frac{g}{24}\,\phi_0^4\right)\ . \tag{16.2.6}$$

* There are limitations on the applicability of perturbation theory for $m^2 < 0$, discussed in the following section.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

The one-loop term is given by

$$\exp\left(i\Gamma^{(1\text{ loop})}[\phi_0]\right) = \int \prod_x d\phi(x)\, \exp\left\{ -\tfrac{1}{2}i \int d^4x \left[\partial_\rho\phi\partial^\rho\phi + \mu^2(\phi_0)\phi^2\right] + \epsilon \text{ terms}\right\}. \tag{16.2.7}$$

We learned how to calculate such integrals in Chapter 9; the result is given by Eq. (9.A.18):

$$i\Gamma^{(1\text{ loop})}[\phi_0] = \ln \text{Det}\left(\frac{iK}{\pi}\right)^{-1/2} = -\frac{1}{2}\,\text{Tr}\,\ln\left(\frac{iK}{\pi}\right), \tag{16.2.8}$$

where here

$$K_{x,y} = \left[\frac{\partial^2}{\partial x^\lambda \partial y_\lambda} + \mu^2(\phi_0) - i\epsilon\right]\delta^4(x-y). \tag{16.2.9}$$

As usual, to calculate such traces it is helpful to diagonalize the 'matrix' K by passing to momentum space:

$$\begin{aligned} K_{p,q} &= \int \frac{d^4x}{(2\pi)^2}\, e^{-ip\cdot x} \frac{d^4y}{(2\pi)^2}\, e^{iq\cdot y} K_{x,y} \\ &= \left(p^2 + \mu^2(\phi_0) - i\epsilon\right)\delta^4(p-q). \end{aligned} \tag{16.2.10}$$

The logarithm of this diagonal matrix is just the diagonal matrix with logarithms along the main diagonal:

$$\left[\ln\left(\frac{iK}{\pi}\right)\right]_{p,q} = \ln\left[\frac{i}{\pi}\left(p^2 + \mu^2(\phi_0) - i\epsilon\right)\right]\delta^4(p-q) \tag{16.2.11}$$

and its trace is then

$$\begin{aligned} i\Gamma^{(1\text{ loop})}[\phi_0] &= -\frac{1}{2}\int d^4p \left[\ln\left(\frac{iK}{\pi}\right)\right]_{p,p} \\ &= -\frac{\mathscr{V}_4}{2(2\pi)^4}\int d^4p\, \ln\left(\frac{i}{\pi}\left(p^2 + \mu^2(\phi_0) - i\epsilon\right)\right). \end{aligned} \tag{16.2.12}$$

Putting together Eqs. (16.2.6) and (16.2.12), we have the effective potential to one-loop order

$$V(\phi_0) = \lambda + \frac{1}{2}m^2\phi_0^2 + \frac{g}{24}\phi_0^4 + \mathscr{I}\left(\mu^2(\phi_0)\right), \tag{16.2.13}$$

where

$$\mathscr{I}(\mu^2) \equiv \frac{-i}{2(2\pi)^4}\int d^4p\, \ln\left(\frac{i}{\pi}[p^2 + \mu^2 - i\epsilon]\right). \tag{16.2.14}$$

It is painfully obvious that this formula for the effective potential contains ultraviolet divergences. Fortunately, these are naturally absorbed into a renormalization of the parameters of the theory. Although the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

integral (16.2.14) is divergent, simple power-counting shows that it is made convergent by differentiating three times with respect to μ^2:

$$\mathcal{I}'''(\mu^2) = -\frac{i}{(2\pi)^4}\int d^4p\,(p^2+\mu^2-i\epsilon)^{-3}\,.$$

Once again, the $-i\epsilon$ term tells us that we must rotate the p^0 contour counterclockwise, so that $p^0 = ip_4$, with p_4 running from $-\infty$ to $+\infty$:

$$\mathcal{I}'''(\mu^2) = \frac{1}{(2\pi)^4}\int_0^\infty \frac{2\pi^2 k^3 dk}{(k^2+\mu^2)^3} = \frac{1}{32\pi^2\mu^2}\,.$$

Integrating thrice, we have then

$$\mathcal{I}(\mu^2) = \frac{\mu^4 \ln \mu^2}{64\pi^2} + A + B\mu^2 + C\mu^4\,.$$

The constants A, B, C are not determined by this method of calculation, which is hardly a serious problem since they are obviously infinite anyway. We eliminate these constants by defining 'renormalized' values for λ, m^2, and g:

$$\begin{aligned}\lambda_R &\equiv \lambda + A + Bm^2 + Cm^4\,,\\ m_R^2 &\equiv m^2 + gB + 2gm^2C\,,\\ g_R &\equiv g + 6g^2C\,.\end{aligned}$$

Our final result for the potential to one-loop order is then

$$V(\phi_0) = \lambda_R + \frac{1}{2}m_R^2\phi_0^2 + \frac{g_R}{24}\phi_0^4 + \frac{\mu^4(\phi_0)\ln\mu^2(\phi_0)}{64\pi^2}\,, \tag{16.2.15}$$

where $\mu(\phi_0)$ is the field-dependent mass defined by Eq. (16.2.5), which to this order can be calculated using m_R and g_R in place of m and g:

$$\mu^2(\phi) = m_R^2 + \tfrac{1}{2}g_R\phi^2\,.$$

Similar results hold if the theory contains a complex spin $\frac{1}{2}$ fermion field $\psi(x)$ that interacts with the scalar ϕ. For instance, if this interaction Hamiltonian density has the simple form $G\phi\bar{\psi}\psi$, then the mass $M(\phi_0)$ of the fermion in the presence of a constant scalar background field ϕ_0 is of the form:

$$M(\phi_0) = M(0) + G\phi_0\,.$$

It is easy to see that in this case the potential (16.2.15) then receives an additional term:

$$\begin{aligned}V(\phi_0) = \lambda_R + \frac{1}{2}m_R^2\phi_0^2 + \frac{g_R}{24}\phi_0^4 + \frac{\mu^4(\phi_0)\ln\mu^2(\phi_0)}{64\pi^2}\\ -\frac{M^4(\phi_0)\ln M^2(\phi_0)}{16\pi^2}\,.\end{aligned} \tag{16.2.16}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

The numerical coefficient in the new term is four times that of the $\mu^4 \ln \mu^2$ term in Eq. (16.2.15), because the fermion and antifermion described by ψ and $\bar{\psi}$ are assumed here to be different, and because each has two spin states, while the sign is opposite because (as shown in Chapter 9) fermionic path integrals of Gaussians give a result proportional to the determinant of the matrix coefficient in the exponent, in contrast to bosonic path integrals, which give a result proportional to the inverse of this determinant.

16.3 Energy Interpretation

The effective action $\Gamma[\phi]$ and potential $V[\phi]$ have an important interpretation in terms of the energy and energy density respectively.[4] To see this, suppose that we turn on a current $J^n(\mathbf{x}, t)$ that rises smoothly from zero at $t = -\infty$ to a finite value $\mathscr{J}^n(\mathbf{x})$, and remains at that value for a long time T, after which it drops smoothly again to zero at $t = +\infty$. The effect of this perturbation is to convert the vacuum to a state with a definite energy $E[\mathscr{J}]$ (a functional of $\mathscr{J}^n(\mathbf{x})$), in which it remains for a time T, after which it returns to the vacuum. However, although the 'out' vacuum is the same physical state as the 'in' vacuum, the state vectors differ by the phase $\exp(-iE[\mathscr{J}]T)$ accumulated during the time T:

$$\langle \text{VAC, out}|\text{VAC, in}\rangle_J = \exp\Big(-iE[\mathscr{J}]T\Big) . \tag{16.3.1}$$

Comparing with Eqs. (16.1.1) and (16.1.3), this gives

$$W[J] = -E[\mathscr{J}]T . \tag{16.3.2}$$

To see the connection between this energy and the effective action, suppose we seek the state Ω_ϕ that minimizes the energy expectation value

$$\langle H\rangle_\Omega = \frac{(\Omega, H\Omega)}{(\Omega, \Omega)} , \tag{16.3.3}$$

subject to the condition that the quantum fields $\Phi_n(\mathbf{x}, t)$ have the time-independent expectation value $\phi_n(\mathbf{x})$

$$\frac{(\Omega, \Phi_n(\mathbf{x}, t)\Omega)}{(\Omega, \Omega)} = \phi_n(\mathbf{x}) . \tag{16.3.4}$$

It is convenient also to impose the condition that Ω is normalized

$$(\Omega, \Omega) = 1 . \tag{16.3.5}$$

To minimize the expectation value (16.3.3) subject to the constraints (16.3.4) and (16.3.5), we use the method of Lagrange multipliers, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

instead minimize the quantity

$$(\Omega, H\Omega) - \alpha(\Omega,\Omega) - \int d^3x\, \beta^n(\mathbf{x}) \left(\Omega, \Phi_n(\mathbf{x})\Omega\right) \tag{16.3.6}$$

with no constraints on Ω. This gives

$$H\Omega = \alpha\Omega + \int d^3x\, \beta^n(\mathbf{x})\Phi_n(\mathbf{x})\Omega\,. \tag{16.3.7}$$

Both α and $\beta^n(\mathbf{x})$ are to be chosen in such a way as to satisfy the constraints (16.3.4) and (16.3.5) and therefore depend functionally on the prescribed expectation value $\phi_n(\mathbf{x})$.

Now, we have said that in the presence of a current $\mathscr{J}^n(\mathbf{x})$, the Hamiltonian $H - \int d^3x\, \mathscr{J}^n(\mathbf{x})\,\Phi_n(\mathbf{x})$ has an eigenvalue $E[\mathscr{J}]$:

$$\left[H - \int d^3x\, \mathscr{J}^n(\mathbf{x})\,\Phi_n(\mathbf{x})\right] \Psi_{\mathscr{J}} = E[\mathscr{J}]\Psi_{\mathscr{J}} \tag{16.3.8}$$

with a normalized eigenvector $\Psi_{\mathscr{J}}$. Furthermore, since slowly turning on this current converts the vacuum into this energy eigenstate, we can presume that $E[\mathscr{J}]$ is the lowest energy eigenstate in the presence of this current. Therefore Eqs. (16.3.4), (16.3.5), and (16.3.7) are satisfied by

$$\Omega = \Psi_{\mathscr{J}_\phi}\,, \tag{16.3.9}$$

$$\alpha = E[\mathscr{J}_\phi]\,, \tag{16.3.10}$$

$$\beta^n(\mathbf{x}) = \mathscr{J}^n_\phi(\mathbf{x})\,, \tag{16.3.11}$$

where $\mathscr{J}_\phi(\mathbf{x})$ is the current for which $\Phi(\mathbf{x})$ has an expectation value $\phi(\mathbf{x})$ in the state $\Psi_{\mathscr{J}}$.

Setting $\mathscr{J} = \mathscr{J}_\phi$ in Eq. (16.3.8) and taking the scalar product with $\Psi_{\mathscr{J}_\phi}$, the minimum energy of states in which the fields Φ_n are constrained to have the expectation values ϕ_n is seen to be

$$\langle H\rangle_\Omega = E[\mathscr{J}_\phi] + \int d^3x\, \mathscr{J}^n_\phi(\mathbf{x})\phi_n(\mathbf{x})\,. \tag{16.3.12}$$

Recalling Eq. (16.3.2) and the assumed form of $J(x)$, this is

$$\langle H\rangle_\Omega = \frac{1}{T}\Big[-W[J_\phi] + \int d^4x\, J^n_\phi(\mathbf{x})\phi_n(\mathbf{x})\Big] = -\frac{1}{T}\,\Gamma[\phi]\,. \tag{16.3.13}$$

As noted in the previous section, if the field $\phi(x)$ has a constant value ϕ over a large spacetime volume $\mathscr{V}_4 = \mathscr{V}_3 T$, then we may write the effective action in terms of an effective potential $V(\phi)$:

$$\Gamma[\phi] = -\mathscr{V}_3 T\, V(\phi)\,. \tag{16.3.14}$$

In this case Eq. (16.3.13) tells us that the energy density is

$$\langle H\rangle_\Omega/\mathscr{V}_3 = V(\phi)\,. \tag{16.3.15}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

This is the main result: $V(\phi)$ *is the minimum of the expectation value of the energy density for all states constrained by the condition that the scalar fields* Φ_n *have expectation values* ϕ_n. One consequence is that in the absence of external currents the vacuum state will relax to a state in which the potential $V(\phi)$ is not only *stationary*, which is required by the field equations (16.1.8), but also a *minimum*.

This result helps to resolve a problem in the interpretation of the quantum effective potential. The Euclidean version of the path-integral formalism (described in Appendix A of Chapter 23) makes it manifest that the two-point function Δ is positive (in the matrix sense), so according to Eq. (16.1.2) the same is true of $-\Pi = \Delta^{-1}$. Together with Eq. (16.2.3), this implies that for a single scalar field the effective potential $V(\phi)$ must have a positive (or zero) second derivative with respect to ϕ. More generally, the effective potential must be *convex*:[5]

$$V\Big(\lambda\phi_1 + (1-\lambda)\phi_2\Big) \leq \lambda V(\phi_1) + (1-\lambda)V(\phi_2) \text{ for } 0 \leq \lambda \leq 1 .$$

But inspection of Eq. (16.2.13) shows that for $m^2 < 0$ and $g > 0$ the zero-loop approximation to the effective potential for the scalar field theory with action (16.2.1) has a *negative-definite* second derivative when ϕ is between the two minima of the effective potential at $\pm\sqrt{6|m^2|/g}$. This contradiction arises because the derivation of perturbation theory implicitly relies on the existence of a stable vacuum, but when $V''(\phi) < 0$, the field ϕ is at a value of $\tilde{\phi}$ where $V(\tilde{\phi}) - J_\phi\tilde{\phi}$ is a *maximum* rather than a minimum, which means that the vacuum state in the presence of the current J_ϕ is unstable.

So what is the true effective potential for this scalar field theory when $m^2 < 0$ and ϕ lies between the two minima of the potential? The result of this section is that we must find the state of minimum energy in which the expectation value of the operator Φ equals ϕ. As long as ϕ is between the two minima of the potential, we can give Φ an expectation value ϕ by taking the state as a suitable linear combination of the two states where ϕ is at the minima $\phi \simeq \pm\sqrt{6|m^2|/g}$. The energy in this state is equal to the energy at the minima, so this state clearly minimizes the energy. (Interference terms here vanish in the limit of infinite volume, for reasons explained in Section 19.1.) Thus the effective potential between the two minima of the potential is a *constant*, satisfying the requirement that it have a nonpositive second derivative. The same argument shows that in more general theories where the potential has two local minima of unequal energy, the potential between these minima is linear.[6]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

16.4 Symmetries of the Effective Action

In some but not all cases, the symmetries of the action $I[\phi]$ are automatically also symmetries of the effective action $\Gamma[\phi]$. For instance, in the example of Section 16.2, the action (16.2.1) has a symmetry under the discrete transformation $\phi \to -\phi$. Thus it follows from their definition that $Z[J]$ and $W[J]$ are even under the corresponding reflection $J \to -J$. Eq. (16.1.5) then shows that $\phi_{-J} = -\phi_J$, and hence $J_{-\phi} = -J_\phi$, so Eq. (16.1.6) shows that $\Gamma[\phi]$ is even under $\phi \to -\phi$. This is borne out by the one-loop result (16.2.15). The fermion-loop contribution in (16.2.16) also exhibits the symmetry under $\phi \to -\phi$ in the special case where $M(0) = 0$, because in this case the action is invariant under the combined transformation $\phi \to -\phi$, $\psi \to \gamma_5 \psi$.

We encounter problems in establishing the renormalizability of a theory, unless we can show that the symmetries that we impose on the action also apply to the effective action. For instance, in the example above, if $I[\phi]$ were assumed to be even in ϕ but $\Gamma[\phi]$ turned out not to be, then the coefficients of the terms in Γ proportional to $\int d^4x\, \phi$ and $\int d^4x\, \phi^3$ would be divergent, but the symmetry of the action would not allow us to introduce counterterms to absorb these infinities.

With this motivation, let's now turn to the important class of symmetries generated by infinitesimal transformations

$$\chi^n(x) \to \chi^n(x) + \epsilon F^n[x;\chi]\,, \tag{16.4.1}$$

where F^n is a function of x^μ that depends functionally on χ^n. (For instance, $F^n[x;\chi]$ may be an ordinary function of the χ^n and their derivatives at the point x.) We are now using the symbol χ^n rather than ϕ^r to denote the different types of fields, to emphasize that these include not only ordinary gauge and matter fields (which in the next chapter will be denoted $\phi^r(x)$), but all other fields appearing in the gauge-fixed action, including ghost fields. We repeat that these $\chi^n(x)$ may be of any type, not necessarily scalars.

We assume that both the action and the measure are invariant under the symmetry transformation (16.4.1):

$$I[\chi + \epsilon F] = I[\chi]\,, \tag{16.4.2}$$

$$\prod_{n,x} d\Big(\chi^n(x) + \epsilon F^n[x;\chi]\Big) = \prod_{n,x} d\chi^n(x)\,. \tag{16.4.3}$$

(It is actually sufficient for only the product $(\prod_{n,x} d\chi^n(x))\exp(iI)$ to be invariant, but where this is true usually Eqs. (16.4.2) and (16.4.3) both apply.) Replacing the integration variables in Eq. (16.1.1) with $\chi^n(x) +$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

$\epsilon F^n[x;\chi]$, we have then

$$
\begin{aligned}
Z[J] &= \int \left[\prod_{n,x} d\Big(\chi^n(x) + \epsilon F^n[x;\chi]\Big)\right] \\
&\quad \times \exp\left\{iI[\chi + \epsilon F] + i\int d^4x \,\Big(\chi^n(x) + \epsilon F^n[x;\chi]\Big)J_n(x)\right\} \\
&= \int \left[\prod_{n,x} d\chi^n(x)\right] \exp\left\{iI[\chi] + i\int d^4x \,\Big(\chi^n(x) + \epsilon F^n[x;\chi]\Big)J_n(x)\right\} \\
&= Z[J] + i\epsilon \int \left(\prod_{n,x} d\chi^n(x)\right) \int F^n[y;\chi]J_n(y)\, d^4y \\
&\quad \times \exp\left\{iI[\chi] + i\int d^4x\, \chi^n(x)J_n(x)\right\}
\end{aligned}
$$

and hence

$$
\int d^4y\, \langle F^n(y)\rangle_J J_n(y) = 0\,, \tag{16.4.4}
$$

where $\langle\ \ \rangle_J$ denotes the quantum average in the presence of the current $J_n(x)$,

$$
\begin{aligned}
Z[J]\langle F^n(y)\rangle_J \equiv \int \left(\prod_{n,x} d\chi^n(x)\right) F^n[y;\chi] \\
\times \exp\left\{iI[\chi] + i\int d^4x\, \chi^n(x)J_n(x)\right\}\,,
\end{aligned} \tag{16.4.5}
$$

normalized so that $\langle 1\rangle_J = 1$. But recall that $J_n(y)$ is given in terms of the effective action $\Gamma[\chi]$ by Eq. (16.1.7)

$$
J_{n,\chi}(y) = -\frac{\delta\Gamma[\chi]}{\delta\chi^n(y)}\,.
$$

Therefore Eq. (16.4.4) may be written as

$$
0 = \int d^4y\, \langle F^n(y)\rangle_{J_\chi} \frac{\delta\Gamma[\chi]}{\delta\chi^n(y)}\,. \tag{16.4.6}
$$

In other words, $\Gamma[\chi]$ is invariant under the infinitesimal transformation

$$
\chi^n(y) \to \chi^n(y) + \epsilon\langle F^n(y)\rangle_{J_\chi}\,. \tag{16.4.7}
$$

Such symmetry conditions are known as *Slavnov–Taylor identities.*[7]

Is this the symmetry transformation with which we started? It is for one very important class of infinitesimal symmetry transformations: those that are *linear*. For such symmetries F is

$$
F^n[x;\chi] = s^n(x) + \int t^n{}_m(x,y)\chi^m(y)d^4y\,. \tag{16.4.8}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

(In the most common case $s^n(x)$ vanishes and $t^n{}_m(x,y)$ is a constant matrix times $\delta^4(x-y)$.) For any linear F, we have

$$\langle F^n(x)\rangle_J = s^n(x) + \int t^n{}_m(x,y)\langle\chi^m(y)\rangle_J\, d^4y\ .$$

But for any fixed χ, J_χ is defined as the value of the current J that makes $\langle\chi^m(y)\rangle_J$ equal to $\chi^m(y)$, so

$$\langle F^n(x)\rangle_{J_\chi} = s^n(x) + \int t^n{}_m(x,y)\chi^m(y)\, d^4y = F^n[x;\chi]\ . \tag{16.4.9}$$

Hence Eq. (16.4.6) requires that $\Gamma[\chi]$ be invariant under all the functional linear transformations $\chi^n \to \chi^n + \epsilon F^n$ that leave $I[\chi]$ and the measure invariant.

We occasionally have to deal with symmetry transformations that are not linear. One important example is provided by the BRST transformation discussed in Section 15.7. For non-linear transformations, the symmetry transformation (16.4.7) under which the effective action is invariant is *not* generally the same as the assumed symmetry transformation (16.4.1) that leaves the original action invariant, because the average of a non-linear functional of fields is not generally the same as the functional of the average fields. Indeed, the form of $\langle F\rangle_{J_\chi}$ as a functional of χ depends in general on the dynamics of the system, and is usually non-local. This complication will be dealt with in the next chapter by the method of antibrackets.

* * *

Up to now, we have tacitly been assuming that the fields χ^n and the corresponding transformation functions F^n and currents J_n are all bosonic. We will need to take note of the sign factors that appear when some of these are fermionic, as in particular for supersymmetry or BRST transformations, where ϵ is fermionic and χ^n and F^n have opposite statistics. With currents inserted to the right of fields, as in Eqs. (16.1.1) and (16.4.5), Eqs. (16.1.5) and (16.1.7) hold in the form

$$\frac{\delta_R W[J]}{\delta J_m(y)} = \chi^m_J(y)\ , \tag{16.4.10}$$

$$\frac{\delta_L \Gamma[\chi]}{\delta\chi^m(y)} = -J_{\chi,m}(y)\ , \tag{16.4.11}$$

with the subscripts R and L indicating that the derivative is to act from the right or left. In consequence, the Slavnov–Taylor identity (16.4.6) should be written

$$0 = \int d^4y\ \langle F^n(y)\rangle_{J_\chi}\ \frac{\delta_L\Gamma[\chi]}{\delta\chi^n(y)}\ . \tag{16.4.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

Problems

1. Consider a theory of real pseudoscalars $\phi(x)$ and complex Dirac fields $\psi(x)$, with masses M and m respectively, and interaction $g\bar{\psi}\gamma_5\psi\phi$. Evaluate the effective potential for $\phi =$ constant, $\psi = 0$, to one-loop order.

2. Derive general formulas for $\delta^3 W[J]/\delta J_n(x)\,\delta J_m(y)\,\delta J_\ell(z)$ and $\delta^4 W[J]/\delta J_n(x)\,\delta J_m(y)\,\delta J_\ell(z)\,\delta J_k(w)$ in terms of the variational derivatives of $\Gamma[\phi]$ with respect to ϕ. Show which Feynman diagrams correspond to each term in these formulas.

3. Calculate the effective potential to one-loop order for the theory of a neutral scalar field ϕ with interaction Lagrangian density $g\phi^3/6$ in six spacetime dimensions.

4. Suppose that the action $I[\phi]$ is invariant under a *finite* matrix transformation $\phi_n(x) \to \sum_m M_{nm}\phi_m(x)$. Under which transformation of the currents is $W[J]$ then invariant? Use this result to derive a symmetry property of $\Gamma[\phi]$.

References

1. The effective action $\Gamma[\phi]$ was introduced by J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127**, 965 (1962), who defined it perturbatively, as a sum over one-particle-irreducible connected diagrams. The non-perturbative definition (16.1.6) was first given independently by B. De Witt, in *Relativity, Groups, and Topology – Lectures Delivered at Les Houches during the 1963 Session of the Summer School of Theoretical Physics*, C. De Witt and B. De Witt, eds. (Gordon and Breach, New York, 1964); G. Jona-Lasinio, *Nuovo Cimento* **34**, 1790 (1964).

2. S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1985): pp. 135–6.

3. S. Coleman and E. Weinberg, *Phys. Rev.* **D7**, 1888 (1973).

4. K. Symanzik, *Comm. Math. Phys.* **16**, 48 (1970); S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1985): pp 139–42.

5. K. Symanzik, Ref. 4; J. Iliopoulos, C. Itzykson, and A. Martin, *Rev. Mod. Phys.* **47**, 165 (1975).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

6. Y. Fujimoto, L. O'Raifeartaigh, and G. Parravicini, *Nucl. Phys.* **B212**, 268 (1983); R. W. Haymaker and J. Perez-Mercader, *Phys. Rev.* **D27**, 1948 (1983); C. M. Bender and F. Cooper, *Nucl. Phys.* **B224**, 403 (1983); M. Hindmarsh and D. Johnston, *J. Math. Phys.* **A19**, 141 (1986); V. Branchina, P. Castorina, and D. Zappalà, *Phys. Rev.* **D41**, 1948 (1990); K. Cahill, *Phys. Rev.* **D52**, 4704 (1995).

7. A. A. Slavnov, *Theor. Math. Phys.* **10**, 152 (1972) [English translation: *Theor. and Math. Phys.* **10**, 99 (1972)]; J. C. Taylor, *Nucl. Phys.* **B33**, 436 (1971).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.004

17

Renormalization of Gauge Theories

We now return to gauge theories, and use the external field formalism described in the previous chapter to study the renormalizability of these theories and to carry out an important calculation.

17.1 The Zinn-Justin Equation

In this section the BRST symmetry described in Section 15.7 will be used to demonstrate a fundamental property of the quantum effective action, first derived by Zinn-Justin.[1] According to the general rules outlined in Section 16.4, the BRST invariance of the action $I[\chi]$ imposes on the effective action $\Gamma[\chi]$ the condition

$$\int d^4x \, \langle \Delta^n(x) \rangle_{J_\chi} \frac{\delta_L \Gamma[\chi]}{\delta \chi^n(x)} = 0 \,, \tag{17.1.1}$$

where the change in $\chi^n(x)$ under a BRST transformation with infinitesimal fermionic parameter θ is

$$\delta_\theta \chi^n(x) = \theta \Delta^n(x) \tag{17.1.2}$$

and $\langle \cdots \rangle$ here denotes a vacuum expectation value taken in the presence of a current J_χ that makes the vacuum expectation value of the operator fields $X^n(x)$ equal to the c-number functions $\chi^n(x)$. The implied sum over n runs over all of the fields in the BRST formalism; that is, over ω_α, ω^*_α, and h_α as well as the gauge and matter fields that in Section 15.8 we have collectively called ϕ^r. Because $\Delta^n(x)$ is quadratic in the fields when χ^n is a gauge or matter field or ω_α, Eq. (17.1.1) does *not* in general tell us that the effective action is invariant under the same BRST transformations as the action itself.

To handle this complication, we employ a trick that proves useful in dealing with any sort of nilpotent symmetry transformation. First, we introduce a set of c-number external fields $K_n(x)$, and define a new

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

effective action by

$$\Gamma[\chi,K] \equiv W[J_{\chi,K},K] - \int d^4x\, \chi^n(x) J_{\chi,K\,n}(x)\,, \tag{17.1.3}$$

where the connected vacuum persistence amplitude W is here calculated with the gauge-fixed action* $I + \int d^4x\, \Delta^n K_n$:

$$e^{iW[J,K]} \equiv \int \left[\prod_{n,x} d\chi^n(x)\right] \exp\left(iI + i\int d^4x\, \Delta^n K_n + i\int d^4x\, \chi^n J_n\right) \tag{17.1.4}$$

and $J_{\chi,K}$ is the current required to give the fields the expectation values χ in the presence of the external fields K:

$$\left.\frac{\delta_R W[J,K]}{\delta J_n(x)}\right|_{J=J_{\chi,K}} \equiv \chi^n(x)\,. \tag{17.1.5}$$

(The K_n must have the same fermionic or bosonic statistics as Δ^n, which is opposite to that of χ^n.) Since the BRST transformation is nilpotent, the quantities $\Delta^n(x)$ are BRST-invariant, so in the same way as in Section 16.4 we can show that the new effective action $\Gamma[\chi,K]$ satisfies a BRST-invariance condition:

$$\int d^4x\, \langle \Delta^n(x)\rangle_{J_{\chi,K},K} \frac{\delta_L \Gamma[\chi,K]}{\delta\chi^n(x)} = 0\,, \tag{17.1.6}$$

where $\langle\cdots\rangle_{J,K}$ denotes a vacuum expectation value calculated in the presence of the current J and the external fields K:

$$\langle \mathcal{O}[\chi]\rangle_{J,K} = \frac{\int \left[\prod_{n,x} d\chi^n(x)\right] \mathcal{O}[\chi] \exp\left(iI + i\int d^4x\, \Delta^n K_n + i\int d^4x\, \chi^n J_n\right)}{\int \left[\prod_{n,x} d\chi^n(x)\right] \exp\left(iI + i\int d^4x\, \Delta^n K_n + i\int d^4x\, \chi^n J_n\right)}\,. \tag{17.1.7}$$

It is convenient to express the expectation value of Δ^n as a variational derivative of the effective action. Taking the right variational derivative of Eq. (17.1.3) with respect to K gives

$$\frac{\delta_R \Gamma[\chi,K]}{\delta K_n(x)} = \left.\frac{\delta_R W[J,K]}{\delta K_n(x)}\right|_{J=J_{\chi,K}} + \int d^4y\, \left.\frac{\delta_R W[J,K]}{\delta J_m(y)}\right|_{J=J_{\chi,K}} \frac{\delta_R J_{\chi,K\,m}(y)}{\delta K_n(x)} - \int d^4y\, \chi^m(y) \frac{\delta_R J_{\chi,K\,m}(y)}{\delta K_n(x)}\,.$$

Using Eq. (17.1.5) we see that the last two terms cancel, and using the

* Here I is the action $I_{\rm NEW}$ modified as described in Section 15.7 to depend on the ghost and antighost fields ω_α and ω_α^* and on the auxiliary field h_α, with the subscript 'NEW' dropped from now on.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

definitions (17.1.4) and (17.1.7) gives us then our desired relation

$$\frac{\delta_R \Gamma[\chi,K]}{\delta K_n(x)} = \frac{\delta_R W[J,K]}{\delta K_n(x)}\bigg|_{J=J_{\chi,K}} = \langle \Delta^n(x)\rangle_{J_{\chi,K},K} \,. \tag{17.1.8}$$

The BRST symmetry condition (17.1.6) may now be written as a simple condition, the *Zinn-Justin equation*, involving the effective action alone:

$$\int d^4x \, \frac{\delta_R \Gamma[\chi,K]}{\delta K_n(x)} \frac{\delta_L \Gamma[\chi,K]}{\delta \chi^n(x)} = 0 \,. \tag{17.1.9}$$

As remarked after Eq. (15.9.3), the interchange of fields and antifields (or in this case χ^n and K_n) results simply in a change of sign of the left-hand side of Eq. (17.1.9), so this can be written as

$$(\Gamma,\Gamma) = 0 \,, \tag{17.1.10}$$

where the antibracket is calculated here with K_n in place of the antifield of χ^n:

$$(F,G) \equiv \int d^4x \, \frac{\delta_R F[\chi,K]}{\delta \chi^n(x)} \frac{\delta_L G[\chi,K]}{\delta K_n(x)} - \int d^4x \, \frac{\delta_R F[\chi,K]}{\delta K_n(x)} \frac{\delta_L G[\chi,K]}{\delta \chi^n(x)} \,. \tag{17.1.11}$$

This is formally the same as the Batalin–Vilkovisky 'master equation' discussed in Section 15.9, but appears here as a constraint on the quantum effective action $\Gamma[\chi,K]$ rather than on the fundamental action $S[\chi,\chi^\ddagger]$. The Zinn-Justin equation (17.1.10) will be used in the next two sections to show how to renormalize gauge theories, and in Section 22.6 to study anomalies in these theories.

17.2 Renormalization: Direct Analysis

The simplest non-Abelian gauge theories are renormalizable in the 'Dyson' sense that the operators in the Lagrangian density all have dimensionality (in powers of mass) four or less. As we saw in Chapter 12, this guarantees that the infinities in the quantum effective action only appear in terms that could be cancelled by the counterterms in interactions of dimensionality four or less. But there is more to renormalizability than this. The Lagrangian density is constrained by gauge invariance and other symmetries. For a theory to be renormalizable, it is necessary that the infinities in the quantum effective action satisfy the same constraints, up to possible renormalizations of the fields.[2]

The effective action $\Gamma[\chi,K]$ is a complicated functional of both χ and K, about which the symmetry condition (17.1.9) says complicated things, but fortunately matters are much simpler for the infinite terms in Γ. We will write the action $S[\chi,K] \equiv I[\chi] + \int d^4x \, \Delta^n K_n$ as the sum of a term $S_R[\chi,K]$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

in which masses and coupling constants are set equal to their renormalized values, plus a correction $S_\infty[\chi,K]$, which contains the counterterms that we intend to cancel the infinities from loop graphs. Both S_R and S_∞ must be taken to have the symmetries of the original action $S[\chi,K]$, so the question is whether the infinite parts of the higher-order contributions to Γ share the same symmetries, so that they can be cancelled by the counterterms in S_∞.

We may expand Γ in a series of terms Γ_N that arise both from diagrams with just N loops and also from graphs with $N-M$ loops (where $1 \leq M \leq N$) involving various counterterms in $S_\infty[\chi,K]$ that will be used to cancel infinities in graphs with a total of M loops:

$$\Gamma[\chi,K] = \sum_{N=0}^{\infty} \Gamma_N[\chi,K] \,. \tag{17.2.1}$$

The symmetry condition (17.1.10) then reads,* for each N,

$$\sum_{N'=0}^{N} (\Gamma_{N'}, \Gamma_{N-N'}) = 0 \,. \tag{17.2.2}$$

In the sum (17.2.1) the leading term is just $\Gamma_0[\chi,K] = S_R[\chi,K]$, which of course is finite. Suppose that, for all $M \leq N-1$, all infinities arising from M-loop graphs have been cancelled by counterterms in S_∞ . Then infinities can appear in Eq. (17.2.2) only in the $N'=0$ and $N'=N$ terms, which are equal, and the infinite part of this condition tells us that the infinite part $\Gamma_{N,\infty}$ of Γ_N is subject to the condition that

$$(S_R, \Gamma_{N,\infty}) = 0 \,. \tag{17.2.3}$$

This is a symmetry principle generated by S_R, just like that described by Eqs. (15.9.16) and (15.9.17). Note in particular that the transformation $X \mapsto (S_R, X)$ acts on the external fields K_n as well as the fields χ^n.

Up to this point, we have used none of the special properties of a renormalizable Yang–Mills theory. Now note that, according to the general power-counting rules of renormalization theory, with all infinities cancelled in subgraphs of Γ_N, the infinite part $\Gamma_{N,\infty}[\chi,K]$ of $\Gamma_N[\chi,K]$ can only be a sum of products of fields (including K) and their derivatives of

* Such order-by-order relations may be derived formally by repeating the reasoning of Section 16.1: When the action S_R is replaced with $g^{-1}S_R$, the contribution of a connected L-loop graph with I internal lines and V vertices is multiplied by a factor $g^{V-I} = g^{L-1}$. If the counterterms in S_∞ associated with N-loop diagrams are also provided with factors g^N, then Γ_L is the value for $g=1$ of the term in Γ of order g^{L-1}. Eq. (17.2.2) then follows by requiring Eq. (17.1.10) to hold in each order in g. In cgs units the action has the same dimensions as $\hbar$, and so appears in the path integral multiplied with a factor $1/\hbar$, so we can also use $\hbar$ as a loop-counting parameter in place of g.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

dimensionality (in powers of mass) four or less. Finally, the arguments of Section 16.4 show that $\Gamma[\chi, K]$ and hence also $\Gamma_{N,\infty}[\chi, K]$ are invariant under all of the *linearly* realized symmetry transformations under which the action is invariant. (As described below, these are: Lorentz transformations, *global* gauge transformations, antighost translations, and the ghost phase transformations associated with ghost number conservation. Of course, the auxiliary fields K_n must be assigned suitable transformation properties under these symmetry transformations.) These conditions together with Eq. (17.2.3) will suffice to tell us all we need to know about the structure of $\Gamma_{N,\infty}[\chi, K]$.

To implement these conditions, we need to know the dimensionalities of the external fields K_n. If a field χ^n has dimensionality d_n (that is, d_n powers of mass), then Δ^n correspondingly has dimensionality $d_n + 1$ (as can be seen by inspection of the BRST transformation rules (15.7.7)–(15.7.11)), so in order for $\int d^4x K_n \Delta^n$ to be dimensionless, K_n must have dimensionality $3 - d_n$. The fields $A^{\alpha\mu}$, ω^α, and $\omega^{\alpha*}$ all have dimensionalities $d_n = +1$, so the corresponding K_n all have dimensionalities $+2$. (We do not introduce any external field corresponding to h^α, because this field is BRST-invariant.) Any spin 1/2 matter fields ψ_ℓ have dimensionalities 3/2, and the corresponding K_n thus also have dimensionalities 3/2. Thus a dimension four quantity like $\Gamma_{N,\infty}[\chi, K]$ is at most quadratic in any of the K_n. Furthermore terms that are of second order in the K_n cannot involve any other fields, except that a term of second order in the K_n for spin 1/2 matter fields may involve at most one additional field of dimensionality unity.

We can now use ghost number conservation to show that in fact $\Gamma_{N,\infty}[\chi, K]$ does not contain any terms of second order in the K_n. For this purpose, we also need the ghost quantum numbers of the K_n. If χ^n has ghost quantum number γ_n, then Δ^n has ghost quantum number $\gamma_n + 1$, so K_n must be assigned ghost quantum number $-\gamma_n - 1$. The ghost quantum numbers of the fields $A^{\alpha,\mu}$, ψ^ℓ, ω^α, and $\omega^{\alpha*}$ are respectively 0, 0, $+1$, and -1, so the corresponding external fields K_n have ghost quantum numbers $-1, -1, -2$, and 0, respectively. This rules out any terms in $\Gamma_{N,\infty}[\chi, K]$ that are of second order in the K_n, with the one possible exception of a term of second order in the external fields K^*_α associated with $\omega^{\alpha*}$ (and involving no other fields). However, these last terms are also forbidden, for a different reason. The BRST transformation of $\omega^{\alpha*}$ is linear in the fields, with

$$\Delta^{\alpha*} = -h^\alpha \,, \tag{17.2.4}$$

so here

$$\frac{\delta_L \Gamma_{N,\infty}[\chi, K]}{\delta K^*_\alpha} = \langle \Delta^{\alpha*} \rangle_{J_\chi, K} = -h^\alpha$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

is independent of K^*_α. It follows that $\Gamma_{N,\infty}[\chi, K]$ is linear in $K^{\alpha,*}$, and depends on $K^{\alpha,*}$ only through a term $-\int d^4x\, K^*_\alpha h^\alpha$. (Both K^*_α and h^α are bosonic, so their order is immaterial.) In particular, for $N > 0$, $\Gamma_{N,\infty}[\chi, K]$ is independent of $K^{\alpha *}$.

We have seen that $\Gamma_{N,\infty}[\chi, K]$ is at most linear in all of the K_n. We will write it as

$$\Gamma_{N,\infty}[\chi, K] = \Gamma_{N,\infty}[\chi, 0] + \int d^4x\, \mathscr{D}^n_N[\chi; x]\, K_n(x)\,. \tag{17.2.5}$$

We also recall that S_R has the K dependence:

$$S_R[\chi, K] = S_R[\chi] + \int d^4x\, \Delta^n[\chi; x] K_n(x)\,.$$

The terms in Eq. (17.2.3) of zeroth and first order in K therefore give**

$$\int d^4x \left[\Delta^n[\chi; x] \frac{\delta_L \Gamma_{N,\infty}[\chi, 0]}{\delta \chi^n(x)} + \mathscr{D}^n_N[\chi; x] \frac{\delta_L S_R[\chi]}{\delta \chi^n(x)}\right] = 0 \tag{17.2.6}$$

and

$$\int d^4x \left[\Delta^n(\chi; x) \frac{\delta_L \mathscr{D}^m_N(\chi; y)}{\delta \chi^n(x)} + \mathscr{D}^n_N(\chi; x) \frac{\delta_L \Delta^m(\chi; y)}{\delta \chi^n(x)}\right] = 0\,, \tag{17.2.7}$$

respectively. These relations may be made more perspicuous by introducing the quantities

$$\Gamma^{(\epsilon)}_N[\chi] \equiv S_R[\chi] + \epsilon \Gamma_{N,\infty}[\chi, 0]\,, \tag{17.2.8}$$

and

$$\Delta^{(\epsilon)n}_N(x) \equiv \Delta^n(x) + \epsilon \mathscr{D}^n_N(x)\,, \tag{17.2.9}$$

with ϵ infinitesimal. Then (17.2.6) (together with the BRST invariance of S_R) just says that $\Gamma^{(\epsilon)}_N[\chi]$ is invariant under the transformation

$$\chi^n(x) \to \chi^n(x) + \theta \Delta^{(\epsilon)n}_N(x)\,, \tag{17.2.10}$$

while Eq. (17.2.7) (together with the nilpotence of the original BRST transformation) tells us that this transformation is *nilpotent*.

We must now consider what form this nilpotent transformation may take. As already mentioned, $\Gamma_{N,\infty}$ consists only of terms of dimensionality four or less, so $\mathscr{D}^n_N$ and hence $\Delta^{(\epsilon)n}_N(x)$ have at most the dimensionality

** The second terms in Eqs. (17.2.6) and (17.2.7) have been put into the form shown here by recalling that, because χ^n and K_n have opposite statistics, for any bosonic functionals A and B,

$$\frac{\delta_R A}{\delta \chi^n} \frac{\delta_L B}{\delta K_n} = -\frac{\delta_L A}{\delta \chi^n} \frac{\delta_R B}{\delta K_n} = -\frac{\delta_R B}{\delta K_n} \frac{\delta_L A}{\delta \chi^n}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

of the original BRST transformation function $\Delta^n(x)$. Also, $\mathscr{D}_N^n$ and hence also $\Delta_N^{(\epsilon)n}(x)$ must have the same Lorentz transformation properties and ghost quantum numbers as $\Delta^n(x)$. Hence the most general form of the transformation (17.2.10) is

$$\begin{aligned}
\psi &\to \psi + i\theta\omega^\alpha T_\alpha \psi \,, \\
A_{\alpha\mu} &\to A_{\alpha\mu} + \theta\left[B_{\alpha\beta}\partial_\mu\omega_\beta + D_{\alpha\beta\gamma}A_{\beta\mu}\omega_\gamma\right] \,, \\
\omega_\alpha &\to \omega_\alpha - \tfrac{1}{2}\theta E_{\alpha\beta\gamma}\omega_\beta\omega_\gamma \,,
\end{aligned}$$

where T_α is some matrix acting on the spinor fields, and $B_{\alpha\beta}$, $D_{\alpha\beta\gamma}$, and $E_{\alpha\beta\gamma}$ are constants, with $E_{\alpha\beta\gamma}$ antisymmetric in β and γ. Also, the transformations of ω_α^* and h_α are linear, and are therefore unchanged:

$$\omega_\alpha^* \to \omega_\alpha^* - \theta h_\alpha \,, \qquad h_\alpha \to h_\alpha \,.$$

Next we impose the condition of nilpotence. The most important requirement is that $E_{\alpha\beta\gamma}\omega_\beta\omega_\gamma$ should be invariant. This yields the requirement that $E_{\alpha\beta\gamma}E_{\beta\delta\epsilon}\omega_\delta\omega_\epsilon\omega_\gamma$ should vanish, so that the part of $E_{\alpha\beta\gamma}E_{\beta\delta\epsilon}$ that is totally antisymmetric in δ, ϵ, γ vanishes:

$$E_{\alpha\beta\gamma}E_{\beta\delta\epsilon} + E_{\alpha\beta\epsilon}E_{\beta\gamma\delta} + E_{\alpha\beta\delta}E_{\beta\epsilon\gamma} = 0 \,.$$

But this just tells us that $E_{\alpha\beta\gamma}$ is the structure constant of some Lie algebra $\mathscr{E}$. Because $E_{\alpha\beta\gamma}$ goes to the structure constant $C_{\alpha\beta\gamma}$ of the original gauge Lie algebra $\mathscr{A}$ for $\epsilon \to 0$, $\mathscr{E}$ must be the same as $\mathscr{A}$, and the structure constants $E_{\alpha\beta\gamma}$ can differ from the original $C_{\alpha\beta\gamma}$ only by a multiplicative factor:†

$$E_{\alpha\beta\gamma} = \mathscr{Z}C_{\alpha\beta\gamma} \,.$$

(This is for simple gauge groups; in the general case we would have a separate factor $\mathscr{Z}$ for each simple subgroup.)

Next we turn to the condition that the transformation (17.2.10) be nilpotent when acting on the gauge fields. The requirement that $B_{\alpha\beta}\partial_\mu\omega_\beta + D_{\alpha\beta\gamma}A_{\beta\mu}\omega_\gamma$ be invariant tells us that

$$D_{\alpha\beta\gamma}D_{\beta\delta\epsilon} - D_{\alpha\beta\epsilon}D_{\beta\delta\gamma} = E_{\beta\epsilon\gamma}D_{\alpha\delta\beta} = \mathscr{Z}C_{\beta\epsilon\gamma}D_{\alpha\delta\beta}$$

and

$$B_{\alpha\beta}E_{\beta\gamma\delta} = D_{\alpha\beta\delta}B_{\beta\gamma} \,.$$

† The requirement of global gauge invariance rules out any non-trivial similarity transformation in the relation between $E_{\alpha\beta\gamma}$ and $C_{\alpha\beta\gamma}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

The first condition has the unique solution[††]

$$D_{\alpha\beta\gamma} = \mathscr{Z} C_{\alpha\beta\gamma} .$$

The second condition tells us that the matrix $B_{\alpha\beta}$ commutes with the adjoint representation of the gauge group, and hence (since we have chosen the structure constants totally antisymmetric) must be proportional to a Kronecker delta, with a coefficient we shall call $\mathscr{Z}\mathscr{N}$:

$$B_{\alpha\beta} = \mathscr{Z}\mathscr{N}\delta_{\alpha\beta} .$$

Finally, the condition that the transformation (17.2.10) be nilpotent when acting on the fermion fields (if any) requires that $\omega^\alpha T_\alpha \psi$ is invariant. This tells us that

$$[T_\beta, T_\gamma] = iE_{\alpha\beta\gamma} T_\alpha ,$$

so T_α differs from the generator t_α in the original Lagrangian only by a factor $\mathscr{Z}$:

$$T_\alpha = \mathscr{Z} t_\alpha .$$

We have thus seen that, apart from the appearance of the new constants $\mathscr{Z}$ and $\mathscr{N}$, the transformation (17.2.10) is just the BRST transformation with which we started:

$$\psi \to \psi + i\mathscr{Z}\theta\omega^\alpha t_\alpha \psi , \tag{17.2.11}$$

$$A_{\alpha\mu} \to A_{\alpha\mu} + \mathscr{Z}\theta \left[\mathscr{N}\partial_\mu \omega_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu}\omega_\gamma\right] , \tag{17.2.12}$$

$$\omega_\alpha \to \omega_\alpha - \tfrac{1}{2}\mathscr{Z}\,\theta\, C_{\alpha\beta\gamma}\omega_\beta\omega_\gamma , \tag{17.2.13}$$

$$\omega^*_\alpha \to \omega^*_\alpha - \theta h_\alpha , \tag{17.2.14}$$

$$h_\alpha \to h_\alpha . \tag{17.2.15}$$

Now we must use this symmetry to constrain the structure of the corrected action (17.2.8). Since this contains only the original renormalized action plus the infinite part of the N-loop contribution, it must be the integral of a Lagrangian density

$$\Gamma^{(\epsilon)}_N = \int d^4x\, \mathscr{L}^{(\epsilon)}_N \tag{17.2.16}$$

with $\mathscr{L}^{(\epsilon)}_N$ a local function of fields and field derivatives of dimensionality (in powers of mass) no greater than 4. Furthermore, as we found in Section 16.4, $\mathscr{N}^{(\epsilon)}_N$ must be invariant under all the symmetries of the

[††] The matrix $(D_\gamma)_{\alpha\beta} \equiv D_{\gamma\alpha\beta}/\mathscr{Z}$ satisfies the commutation relations of the gauge Lie algebra, $[D_\gamma, D_\epsilon] = C_{\beta\epsilon\gamma} D_\beta$. But the only representation of a simple Lie algebra with the same dimensionality and $\mathscr{A}$ transformation properties as the adjoint representation of $\mathscr{A}$ is the adjoint representation itself.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

original Lagrangian that act *linearly* on the fields. To identify these symmetries, recall that in generalized ξ-gauge, the 'new' Lagrangian density in Eq. (15.7.6) takes a form given by replacing the term $-(\partial_\mu A^\mu_\alpha)(\partial_\nu A^\nu_\alpha)/2\xi$ in Eq. (15.6.16) with the terms $h_\alpha f_\alpha + \frac{1}{2}\xi h_\alpha h_\alpha$ in Eq. (15.7.6):

$$\mathscr{L}_{\rm NEW} = \mathscr{L}_{\rm M} - \tfrac{1}{4}F^{\mu\nu}_\alpha F_{\alpha\mu\nu} - \partial_\mu\omega^*_\alpha\,\partial^\mu\omega_\alpha + C_{\alpha\beta\gamma}(\partial_\mu\omega^*_\alpha)\,A^\mu_\gamma\,\omega_\beta + h_\alpha\partial_\mu A^\mu_\alpha + \tfrac{1}{2}\xi h_\alpha h_\alpha\,. \quad (17.2.17)$$

Inspection of this formula reveals the following linear symmetries:
(1) **Lorentz invariance**.
(2) **Global gauge invariance** – that is, invariance under the transformations

$$\delta\psi_\ell(x) = i\,\epsilon^\alpha(t_\alpha)_\ell{}^m\psi_m(x)\,, \quad (17.2.18)$$

$$\delta A^\beta{}_\mu(x) = C_{\beta\gamma\alpha}\epsilon^\alpha A^\gamma{}_\mu(x)\,, \quad (17.2.19)$$

$$\delta\omega_\beta(x) = C_{\beta\gamma\alpha}\epsilon^\alpha\omega_\gamma(x)\,, \quad (17.2.20)$$

$$\delta\omega^*_\beta(x) = C_{\beta\gamma\alpha}\epsilon^\alpha\omega^*_\gamma(x)\,, \quad (17.2.21)$$

$$\delta h_\beta(x) = C_{\beta\gamma\alpha}\epsilon^\alpha h_\gamma(x)\,, \quad (17.2.22)$$

with constant infinitesimal parameters ϵ^α.
(3) **Antighost translation invariance** — that is, invariance under the transformation

$$\omega^*_\alpha(x) \to \omega^*_\alpha(x) + c_\alpha\,, \quad (17.2.23)$$

with arbitrary constant parameters c_α.
(4) **Ghost number conservation** — that is, the conservation of a ghost number equal to +1 for ω_α, −1 for ω^*_α, and 0 for all other fields.

We shall now proceed to work out the structure of the most general Lagrangian density that is renormalizable, in the sense that it consists only of terms of dimensionality +4 or less, that has these linearly acting symmetries, and that is invariant under the modified BRST transformations (17.2.11)–(17.2.15).

From Eq. (17.2.17) we may conclude that the fields A^μ_α, ω_α, ω^*_α, and h_α have the dimensionalities +1, +1, +1, and +2, respectively (in powers of mass). Note also that ghost number conservation requires that ω and ω^* come in pairs, while antighost translation invariance dictates that ω^* always appears as a derivative. Each pair of ω and $\partial_\mu\omega^*$ fields adds +3 to the dimensionality, so renormalizability rules out any term with more than one such pair. With one such pair we can have at most one more derivative or one additional gauge field, and Lorentz invariance dictates that we must have one or the other. The only renormalizable allowed interactions involving ghost fields are then linear combinations of terms of the form $\partial_\mu\omega^*_\alpha\,\partial^\mu\omega_\beta$ or $\partial_\mu\omega^*_\alpha\,A^\mu_\gamma\,\omega_\beta$.

Next, let us consider the terms that involve the field h_α and possibly other fields but not ω or ω^*. This field has dimensionality +2, so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

renormalizability and Lorentz invariance allows this field to appear only‡ multiplied with another h_β or $\partial_\mu A^\mu_\beta$ or $A^\mu_\beta A_{\mu\gamma}$.

Finally, the Lagrangian will contain renormalizable terms involving only the matter and gauge fields. We will call the sum of these terms $\mathscr{L}_{\psi A}$. Putting this all together and using global gauge invariance, the most general renormalizable interaction allowed by the assumed symmetries (aside from BRST invariance) takes the form:

$$\begin{aligned}\mathscr{L}_N^{(\epsilon)} = \mathscr{L}_{\psi A} + \tfrac{1}{2}\xi' h_\alpha h_\alpha + c h_\alpha \partial_\mu A^\mu_\alpha - e_{\alpha\beta\gamma} h_\alpha A^\mu_\beta A_{\gamma\mu} \\ - Z_\omega(\partial_\mu \omega^*_\alpha)(\partial_\mu \omega_\alpha) - d_{\alpha\beta\gamma}(\partial_\mu \omega^*_\alpha)\omega_\beta A^\mu_\gamma \,, \end{aligned} \tag{17.2.24}$$

where ξ', Z_ω, c, $d_{\alpha\beta\gamma}$, and $e_{\alpha\beta\gamma}$ are unknown constants, with no constraints except obvious symmetry properties such as global gauge invariance and $e_{\alpha\beta\gamma} = e_{\alpha\gamma\beta}$. (As mentioned earlier, we are assuming for simplicity that the gauge group is simple, but the extension to a direct sum of simple and $U(1)$ gauge groups would be trivial; for instance, instead of one term proportional to $h_\alpha h_\alpha$ we would have a sum of such terms, one for each simple subgroup of the gauge group.)

Now we impose BRST invariance. The cancellation of terms in $\delta\mathscr{L}_N^{(\epsilon)}$ proportional to $\theta \partial_\mu h_\alpha \partial^\mu \omega_\alpha$ tells us that

$$c = Z_\omega / \mathscr{Z}\mathscr{N} \,. \tag{17.2.25}$$

The cancellation of terms in $\delta\mathscr{L}_N^{(\epsilon)}$ proportional to $\theta\,\partial_\mu h_\alpha\, \omega_\beta A^\mu_\gamma$ (or $\theta \partial_\mu \omega^*_\alpha \omega_\beta \partial^\mu \omega_\gamma$) requires that

$$d_{\alpha\beta\gamma} = -(Z_\omega/\mathscr{N})\, C_{\alpha\beta\gamma} \,. \tag{17.2.26}$$

The terms in $\delta\mathscr{L}_N^{(\epsilon)}$ proportional to $\theta \partial_\mu \omega^*_\alpha\, \omega_\beta\, \omega_\gamma\, A^\mu_\delta$ automatically then cancel by virtue of the Jacobi identity for the structure constants. The cancellation of terms in $\delta\mathscr{L}_N^{(\epsilon)}$ proportional to $\theta h_\alpha \partial_\mu \omega_\beta A^\mu_\gamma$ (or $\theta h_\alpha A^\mu_\beta A_{\gamma\mu}\omega_\delta$) yields

$$e_{\alpha\beta\gamma} = 0 \,. \tag{17.2.27}$$

Finally, the effect of the infinitesimal transformation (17.2.10) on matter and gauge fields is the same as a local gauge transformation with gauge parameters $\epsilon_\alpha = \mathscr{Z}\mathscr{N}\theta\omega_\alpha$ and gauge couplings renormalized by a factor $1/\mathscr{N}$ (that is, with t_α and $C_{\alpha\beta\gamma}$ replaced with $\tilde{t}_\alpha \equiv t_\alpha/\mathscr{N}$ and $\tilde{C}_{\alpha\beta\gamma} \equiv C_{\alpha\beta\gamma}/\mathscr{N}$), so the cancellation of terms in $\delta\mathscr{L}_N^{(\epsilon)}$ with only one factor of ω_α and no factors h_α or ω^*_α simply tells us that the Lagrangian $\mathscr{L}_{\psi,A}$ for these fields is *gauge-invariant*, with a renormalized gauge coupling. We

‡ In a theory with scalar fields, we could also have renormalizable terms with h_α multiplied with one or two scalar fields. Such terms cause no trouble, but for brevity they will not be considered here.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

conclude then that the most general renormalizable Lagrangian density allowed by our assumed symmetry principles is

$$\mathscr{L}_N^{(\epsilon)} = -\tfrac{1}{4}Z_A\tilde{F}_\alpha^{\mu\nu}\tilde{F}_{\alpha\mu\nu} - Z_\psi\bar{\psi}\gamma^\mu[\partial_\mu - i\tilde{t}_\alpha A_{\alpha\mu}]\psi - \bar{\psi}m_\psi\psi + \tfrac{1}{2}\xi' h_\alpha h_\alpha + (Z_\omega/\mathscr{N}\mathscr{Z})h_\alpha\partial_\mu A_\alpha^\mu - Z_\omega(\partial_\mu\omega_\alpha^*)(\partial_\mu\omega_\alpha) + Z_\omega\tilde{C}_{\alpha\beta\gamma}(\partial_\mu\omega_\alpha^*)\omega_\beta A_\gamma^\mu , \tag{17.2.28}$$

where the tilde on $\tilde{F}_\alpha^{\mu\nu}$ indicates that the field strength is to be calculated using the renormalized structure constant $\tilde{C}_{\alpha\beta\gamma} \equiv C_{\alpha\beta\gamma}/\mathscr{N}$. But apart from the appearance of a number of new constant coefficients, this is the same Lagrangian with which we started. The new constants in this Lagrangian (including the gauge coupling constant) may be freely shifted by adjusting the Nth-order terms in the corresponding constants in the original unrenormalized Lagrangian. In particular, we can adjust these terms to make $\Gamma_N^{(\epsilon)} = S_R$, in which case $\Gamma_{N,\infty} = 0$, completing the proof.

* * *

In the above proof we made important use of the accidental invariance of the gauge-fixed Lagrangian (17.2.17) under the antighost translation transformation (17.2.23). This symmetry would not be present for gauge-fixing functions other than $f_\alpha = \partial_\mu A_\alpha^\mu$, which are not spacetime derivatives. One frequently cited example which preserves Lorentz invariance and global gauge invariance is $f_\alpha = \partial_\mu A_\alpha^\mu + a_{\alpha\beta\gamma}A_\beta^\mu A_{\gamma\mu}$, where $a_{\alpha\beta\gamma}$ is a constant matrix, symmetric in β and γ, which transforms as a tensor under global gauge transformations. (Such constant tensors exist for all $SU(N)$ groups with $N \geq 3$.) Another more important example is the background gauge-fixing functional to be introduced in Section 17.4.

The absence of antighost translation invariance does not affect our argument that $\mathscr{L}_N^{(\epsilon)}$ is a Lorentz- and global gauge-invariant local function of fields and field derivatives of dimensionality no greater than 4, that is invariant under the renormalized BRST transformation (17.2.11)–(17.2.15). But without antighost translation invariance there are new terms in $\mathscr{L}_N^{(\epsilon)}$ that satisfy these conditions. Since the transformation (17.2.11)–(17.2.15) is nilpotent, we can construct such terms as $s'F$, where the transformation (17.2.11)–(17.2.15) is written as $\chi^n \to \chi^n + \theta s'\chi^n$, and F is an arbitrary Lorentz- and global gauge-invariant function of ghost number -1. One such term is

$$a_{\alpha\beta\gamma}s'\left(\omega_\alpha^* A_{\beta\mu}A_\gamma^\mu\right) = -a_{\alpha\beta\gamma}\left[h_\alpha A_{\beta\mu}A_\gamma^\mu + 2\mathscr{Z}\omega_\alpha^*(\mathscr{N}\partial_\mu\omega_\beta + C_{\beta\delta\epsilon}A_{\delta\mu}\omega_\delta)A_\gamma^\mu\right] .$$

This causes no trouble: it is just a renormalized version of the usual ghost and gauge-fixing terms arising from terms $a_{\alpha\beta\gamma}A_\beta^\mu A_{\gamma\mu}$ in the gauge-fixing function f_α. But there is also another possible term of the form

$$b_{\alpha\beta\gamma}s'\left(\omega_\alpha^*\omega_\beta^*\omega_\gamma\right) = -b_{\alpha\beta\gamma}\left[2h_\alpha\omega_\beta^*\omega_\gamma + \tfrac{1}{2}\mathscr{Z}C_{\gamma\delta\epsilon}\omega_\alpha^*\omega_\beta^*\omega_\delta\omega_\epsilon\right] ,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

where $b_{\alpha\beta\gamma}$ is a constant, antisymmetric in α and β, which transforms as a tensor under global gauge transformations. (Such tensors exist for any Lie group; for instance, we could take $b_{\alpha\beta\gamma}$ to be proportional to $C_{\alpha\beta\gamma}$.) But the Faddeev–Popov–De Witt method cannot yield four-ghost interactions in the Lagrangian, so there are no counterterms available to absorb ultraviolet divergences in this term. This is not just a technical obstacle to proving renormalizability; for gauge-fixing functions like $f_\alpha = \partial_\mu A^\mu_\alpha + a_{\alpha\beta\gamma}A^\mu_\beta A_{\gamma\mu}$, one-loop graphs actually do yield divergences in four-ghost amplitudes that cannot be cancelled by counterterms in the Faddeev–Popov–De Witt Lagrangian.

Aside from avoiding gauge-fixing functions other than $f_\alpha = \partial_\mu A^\mu_\alpha$, the only solution to this problem seems to be the one mentioned in Section 15.7. We must give up the Faddeev–Popov–De Witt approach, and instead take the action from the beginning as the most general renormalizable function of the gauge, matter, ghost, and auxiliary fields that is invariant under BRST and the other symmetries of the theory. According to the arguments of Section 15.8, the action can be written in the form $I_0 + s\Psi$ with I_0 ghost-free, and the S-matrix is independent of Ψ, so we can justify this procedure by quantizing the gauge theory in axial gauge, where ghosts decouple, and then taking Ψ to be anything we like. In particular, we can include $s(\omega^*\omega^*\omega)$ terms in the action that can serve as counterterms to the divergences in four-ghost vertices.

17.3 Renormalization: General Gauge Theories*

The proof of the renormalizability of non-Abelian gauge theories in the previous section relied on a 'brute force' analysis of the possible terms in the action of dimensionality four or less. But as we saw in Chapter 12, this limitation on the dimensionality of terms in the action can be at best a good approximation. The successful renormalizable quantum field theories that are used to describe the strong, weak, and electromagnetic interactions are almost certainly effective field theories, accompanied with terms of dimensionality $d > 4$; these terms are normally not observed because they are suppressed by $4 - d$ powers of some very large mass, perhaps of order 10^{16}– 10^{18} GeV. Gravitation too can be described by an effective field theory, in which the Lagrangian density contains not only the Einstein–Hilbert term $-\sqrt{g}R/16\pi G$, but also all scalars constructed from four or more derivatives of the gravitational field. We need to

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

show that gauge theories of this sort, which are not renormalizable in the power-counting sense, are nonetheless renormalizable in the modern sense that the ultraviolet divergences are governed by the gauge symmetries in such a way that there is a counterterm available to cancel every infinity.[3]

For this purpose, let us return to the action $S[\chi, \chi^\ddagger]$ introduced in Section 15.9, taken as a function of independent fields χ^n (including gauge and matter fields ϕ^r and ghost fields ω^A as well as the non-minimal fields ω^{A*} and h^A), together with their antifields $\chi^\ddagger_n$. In theories like quantum gravity or Yang–Mills theories, that are based on a closed gauge algebra with structure constants $f^C{}_{AB}$, this action is constrained to be of the form

$$S = I[\phi] + \omega^A f^r_A[\phi]\, \phi^\ddagger_r + \tfrac{1}{2}\omega^A \omega^B f^C{}_{AB}[\phi]\, \omega^\ddagger_C - h^A\, \omega^{*\ddagger}_A\,, \qquad (17.3.1)$$

where $I[\phi]$ is invariant under the infinitesimal gauge transformations $\phi^r \to \phi^r + \epsilon^A f^r_A[\phi]$. (As in Section 15.9, the indices r, A, etc. include a spacetime coordinate, over which we integrate in sums over these indices.)

We shall not limit ourselves here to actions of this form, but we shall suppose that the local symmetries of the theory are imposed by requiring that the action must obey some 'structural constraints' on its antifield dependence, of which Eq. (17.3.1) provides just one example. The structural constraints are assumed to be linear, in the sense that if they are satisfied for $S + \mathscr{S}_1$ and for $S + \mathscr{S}_2$, then for any constants α_1, α_2 they are satisfied for $S + \alpha_1 \mathscr{S}_1 + \alpha_2 \mathscr{S}_2$. We also impose the quantum master equation (15.9.35)

$$(S, S) - 2i\hbar \Delta S = 0\,, \qquad (17.3.2)$$

with ΔS defined by Eq. (15.9.34), and the factor $\hbar$ now made explicit as a 'loop-counting' parameter, in the sense described in the footnote of the previous section.

The action is taken as a power series in $\hbar$

$$S = S_R + \hbar S_1 + \hbar^2 S_2 + \cdots\,, \qquad (17.3.3)$$

where S_R is an action of the same general form as S, but with all coupling parameters replaced with finite renormalized values, and the S_N are a set of infinite counterterms. The action S is supposed to satisfy the quantum master equation (17.3.2) for all $\hbar$, so S_R satisfies the classical master equation

$$(S_R, S_R) = 0\,, \qquad (17.3.4)$$

while the counterterms satisfy

$$\Big(S_R, S_N\Big) = -\tfrac{1}{2} \sum_{M=1}^{N-1} \Big(S_M, S_{N-M}\Big) + i\Delta S_{N-1}\,. \qquad (17.3.5)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

The counterterms S_N are not by themselves sufficient to cancel the ultraviolet divergences in loop graphs. As a generalization of the conventional renormalization of fields, we also have to introduce a set of renormalized fields and antifields, defined in terms of the original fields and antifields by an arbitrary anticanonical transformation. An infinitesimal anticanonical transformation may be defined in terms of an infinitesimal generating functional δF by Eq. (15.9.26), so a sequence of anticanonical transformations $G(t) \to G(t+\delta t) = G(t) + (F(t)\delta t, G(t))$ (where $G(t)$ is any functional of fields and antifields, and $F(t)$ is the generating functional) leads to a finite canonical transformation $G \to \tilde{G} \equiv G(1)$, with

$$\frac{d}{dt}G(t) = (F(t), G(t)) , \qquad G(0) = G . \tag{17.3.6}$$

If $F(t)$ is given by a power series

$$F(t) = \hbar F_1 + \tfrac{1}{2}\hbar^2 t F_2 + \cdots , \tag{17.3.7}$$

then Eqs. (17.3.3), (17.3.6), and (17.3.7) yield a transformed action

$$\tilde{S} = S_R + \hbar\Big[S_1 + (F_1, S_R)\Big] + \hbar^2\Big[S_2 + (F_1, S_1) + (F_2, S_R) + \tfrac{1}{2}(F_1, (F_1, S_R))\Big] + \cdots . \tag{17.3.8}$$

The question is whether we can use what freedom we have to choose the F_N and S_N so as to cancel all infinities arising from loop graphs.

As already noted, the first term S_R in Eq. (17.3.3) is automatically finite. Suppose that by cancellation of infinities with S_M and F_M for $M < N$ it has been possible to eliminate all infinities in the terms Γ_M of order $\hbar^M$ in the quantum effective action with $M < N$. As we saw in the previous section, the Zinn-Justin equation (derived here by setting $\chi_n^\ddagger = K_n + \delta\Psi/\delta\chi^n$) tells us in this case that the infinite part $\Gamma_{N,\infty}$ of the term in the quantum effective action of order $\hbar^N$ satisfies the condition

$$\Big(S_R, \Gamma_{N,\infty}\Big) = 0 . \tag{17.3.9}$$

The field and antifield variables χ^n and $\chi_n^\ddagger$ are related to the variables χ^n and K_n by an anticanonical transformation, which preserves all antibrackets, so the antibracket in Eq. (17.3.9) may be calculated in terms of χ^n and $\chi_n^\ddagger$ instead of χ^n and K_n.

The condition (17.3.5) satisfied by the counterterm S_N is not the same as the condition (17.3.9) satisfied by $\Gamma_{N,\infty}$. However, given any S_N^0 that satisfies Eq. (17.3.5), we can find a class of other solutions

$$S_N = S_N^0 + S_N' , \tag{17.3.10}$$

where S_N' is arbitrary, except for the condition that $S_R + S_N'$ like $S_R + S_N^0$ satisfies the same symmetry conditions as S_R, and that

$$\Big(S_R, S_N'\Big) = 0 \tag{17.3.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

so as not to invalidate Eq. (17.3.5). The infinite part of the Nth-order term in the quantum effective action may therefore be written

$$\Gamma_{N,\infty} = S'_{N,\infty} + (F_{N,\infty}, S_R) + X_{N,\infty} , \tag{17.3.12}$$

where X_N consists of terms from loop graphs, as well as from the term S_N^0 and various terms in Γ that involve S_M and F_M for $M < N$. For instance, for $N = 2$ Eq. (17.3.8) gives

$$X_2 = S_2^0 + 2(F_1, S_1) + (F_1, (F_1, S_R)) + \text{two-loop terms involving only } S_R + \text{ one-loop terms involving } S_R,\ S_1 \text{ and } F_1 .$$

For our purposes the only thing we need to know about X_N is that it does not involve S'_N or F_N, and that it is invariant under any linearly realized global symmetries of S_R.

Now, because $(S_R, S_R) = 0$, the operation $F \mapsto (S_R, F)$ is nilpotent; for all F,

$$\Big(S_R, \big(S_R, F\big)\Big) = 0 . \tag{17.3.13}$$

Hence it follows from Eqs. (17.3.9) and (17.3.11)–(17.3.13) that

$$\Big(S_R, X_{N,\infty}\Big) = 0 . \tag{17.3.14}$$

Any term in $X_{N,\infty}$ of the form (S_R, Y) may be cancelled in Eq. (17.3.12) by choosing $F_{N,\infty}$ equal to Y. Thus the space of possible remaining infinite terms in $X_{N,\infty}$ that need to be cancelled by the counterterm $S'_{N,\infty}$ consists of those functionals X that satisfy $(S_R, X) = 0$, counting as equivalent functionals that differ only by terms of the form (S_R, Y). In other words, the infinities in Γ_N that need to be cancelled by the counterterms S'_N belong to the *cohomology* of the mapping $X \mapsto (S_R, X)$.

The possible form of the counterterm S'_N is limited by the requirement that $S_R + S'_N$ must satisfy whatever structural constraints are imposed on the action S. Thus we can complete the proof of renormalizability if we can show that the cohomology of the mapping $X \mapsto (S_R, X)$ consists only of functionals that satisfy this structural constraint.

In the case of quantum gravity coupled to the Yang–Mills fields of a semisimple gauge symmetry, the symmetries of the theory are implemented by the structural constraint (17.3.1). In this case there is a theorem[4] that states that the cohomology of the mapping $X \mapsto (S_R, X)$ (on the space of local functionals, rather than spacetime-dependent functions, of ghost number zero) consists** of functionals $A[\phi]$ that are invariant under the

** Strictly speaking, this is valid if one requires that the constant coefficients in S do not take special values for which S would be invariant under a larger group of local symmetries. For instance, this excludes the case $S = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

gauge transformation $\phi^r \to \phi^r + \epsilon^A f^r_A[\phi]$ with structure constants $f^C{}_{AB}$. Any Nth-order infinity of this sort may be cancelled with a counterterm S'_N of the same form, so although these theories are not renormalizable in the conventional power-counting sense, they are renormalizable in the sense that all infinities can be eliminated by a choice of parameters in the original bare action $I[\phi]$ and by a suitable renormalization of fields and antifields.

In other theories the cohomology of the map $X \mapsto (S_R, X)$ contains additional terms. This does not necessarily require a weakening of the structural constraints, because the additional terms in the cohomology may not correspond to actual ultraviolet divergences. For instance, in gauge theories with $U(1)$ factors the cohomology contains terms[4] corresponding to a redefinition of the action of the $U(1)$ gauge symmetry on the various fields of the theory, which if infinite would require us to weaken the structural constraints by leaving the normalization of the transformation functions $f^r_A[\phi]$ in Eq. (17.3.1) arbitrary. But this infinity is forbidden by the same soft-photon theorems that tell us that ratios of $U(1)$ couplings to various fields (like the ratios of the various lepton charges in quantum electrodynamics) are unaffected by radiative corrections. (See Section 10.4.) Where the extra terms in the cohomology do contain ultraviolet divergences, it is necessary to weaken the structural constraint imposed on S in order that there should be a counterterm for every possible ultraviolet divergence. It is not known whether this will always be possible; if not, some theories may have to be rejected because of their unremovable ultraviolet divergences.

17.4 Background Field Gauge

We next turn to a method of calculation that explicitly preserves a sort of gauge invariance, and that therefore proves extremely convenient, especially in one-loop calculations. We consider the effective action $\Gamma[A, \psi, \omega, \omega^*]$ as a functional* of classical external gauge, matter, ghost and antighost fields: $A_{\alpha\mu}(x)$, $\psi_\ell(x)$, $\omega_\alpha(x)$, $\omega^*_\alpha(x)$. Even though ghosts and antighosts never appear in initial or final states, we are considering background ghost and antighost fields as well as gauge and matter fields in order to deal with parts of diagrams that have external ghost or antighost lines.

* We are now returning to the specific choice of the gauge-fixing functional $B[f]$ as the Gaussian (15.7.4), and we are integrating out the auxiliary field h_α, so that the gauge-fixing term in the modified Lagrangian is just $-f_\alpha f_\alpha/2\xi$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

As described in Section 16.1, $\Gamma[A, \psi, \omega, \omega^*]$ is the sum of connected one-particle-irreducible graphs for the vacuum–vacuum amplitude, calculated in a theory in which the quantum fields $A', \psi', \omega', \omega'^*$ over which we integrate are replaced in the action with shifted fields $A + A', \psi + \psi', \omega + \omega', \omega^* + \omega'^*$, the path integral being taken over primed fields with the unprimed fields held fixed. We are free to choose the gauge-fixing function $f_\alpha(x)$ pretty much any way we like; instead of our previous choice $f_\alpha = \partial_\mu A'^\mu_\alpha$ (or $\partial_\mu[A'^\mu_\alpha + A^\mu_\alpha]$) we shall now take[5]

$$f_\alpha = \partial_\mu A'^\mu_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu} A'^\mu_\gamma \,. \tag{17.4.1}$$

The reason for this choice is that it makes the gauge-fixing term $f_\alpha f_\alpha$ invariant under a formal transformation, in which the background field A^μ_α transforms as a gauge field, while the quantum field A'^μ_α transforms homogeneously, like an ordinary matter field that happens to belong to the adjoint representation of the gauge group

$$\delta A^\mu_\alpha = \partial^\mu \epsilon_\alpha - C_{\alpha\beta\gamma}\, \epsilon_\beta\, A^\mu_\gamma \,, \tag{17.4.2}$$

$$\delta A'^\mu_\alpha = -C_{\alpha\beta\gamma}\, \epsilon_\beta\, A'^\mu_\gamma \,. \tag{17.4.3}$$

The transformation properties of f_α can be seen most easily by writing it as a new sort of covariant derivative

$$f_\alpha \equiv \bar{D}_\mu A'^\mu_\alpha \,, \tag{17.4.4}$$

where for any field ϕ_α in the adjoint representation

$$\bar{D}_\mu \phi_\alpha \equiv \partial_\mu \phi_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu} \phi_\gamma \,. \tag{17.4.5}$$

We see that under the transformation (17.4.2), (17.4.3), the function (17.4.1) transforms just like A'_α:

$$\delta f_\alpha = -C_{\alpha\beta\gamma}\, \epsilon_\beta f_\gamma \,, \tag{17.4.6}$$

so the term $f_\alpha f_\alpha$ in the modified Lagrangian is invariant

$$\delta(f_\alpha f_\alpha) = -2C_{\alpha\beta\gamma}\, f_\alpha\, \epsilon_\beta\, f_\gamma = 0 \,. \tag{17.4.7}$$

Also, the original Lagrangian density $\mathscr{L}$ depends on A and A' only through the sum $A + A'$, which under the combined transformation (17.4.2), (17.4.3) undergoes an ordinary gauge transformation

$$\delta(A^\mu_\alpha + A'^\mu_\alpha) = \partial^\mu \epsilon_\alpha - C_{\alpha\beta\gamma}\, \epsilon_\beta (A^\mu_\gamma + A'^\mu_\gamma) \,. \tag{17.4.8}$$

If we transform the background and quantum matter fields by

$$\delta\psi = i\, t_\alpha \epsilon_\alpha \psi \,, \tag{17.4.9}$$

$$\delta\psi' = i\, t_\alpha \epsilon_\alpha \psi' \,, \tag{17.4.10}$$

then also

$$\delta(\psi + \psi') = i\, t_\alpha \epsilon_\alpha (\psi + \psi') \,. \tag{17.4.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

The original Lagrangian $\mathscr{L}$ is invariant under the original gauge transformations (17.4.8), (17.4.11), and only depends on $A + A'$ and $\psi + \psi'$, so it is also invariant under the new formal transformations (17.4.2), (17.4.3), (17.4.9), (17.4.10).

It will be useful to make this invariance property more explicit, by writing $\mathscr{L}$ in terms of the background-covariant derivative $\bar{D}_\mu$. In general, we have

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}\Big(\partial_\mu[A_{\alpha\nu} + A'_{\alpha\nu}] - \partial_\nu[A_{\alpha\mu} + A'_{\alpha\mu}] + C_{\alpha\beta\gamma}[A_{\beta\mu} + A'_{\beta\mu}][A_{\gamma\nu} + A'_{\gamma\nu}]\Big)^2 \\ &+ \mathscr{L}_M\Big(\psi + \psi',\ \partial_\mu(\psi + \psi') - i t_\alpha(A_{\alpha\mu} + A'_{\alpha\mu})(\psi + \psi')\Big) \\ = &-\tfrac{1}{4}\Big(F_{\alpha\mu\nu} + \bar{D}_\mu A'_{\alpha\nu} - \bar{D}_\nu A'_{\alpha\mu} + C_{\alpha\beta\gamma}A'_{\beta\mu}A'_{\gamma\nu}\Big)^2 \\ &+ \mathscr{L}_M\Big(\psi + \psi',\ \bar{D}_\mu(\psi + \psi') - i t_\alpha A'_{\alpha\mu}(\psi + \psi')\Big)\ , \end{aligned} \tag{17.4.12}$$

where, as in Eq. (17.4.5),

$$\bar{D}_\mu A'_{\alpha\nu} \equiv \partial_\mu A'_{\alpha\nu} + C_{\alpha\beta\gamma}A_{\beta\mu}A'_{\gamma\nu}\ , \tag{17.4.13}$$

$$\bar{D}_\mu \psi \equiv \partial_\mu \psi - i t_\alpha A_{\alpha\mu}\psi\ , \tag{17.4.14}$$

and $F_{\alpha\mu\nu}$ is the background field strength

$$F_{\alpha\mu\nu} \equiv \partial_\mu A_{\alpha\nu} - \partial_\nu A_{\alpha\mu} + C_{\alpha\beta\gamma}A_{\beta\mu}A_{\gamma\nu}\ . \tag{17.4.15}$$

(The square in the first term of $\mathscr{L}$ is intended to imply obvious index contractions.) Clearly $\mathscr{L}$ is invariant under the new transformations (17.4.2), (17.4.3), (17.4.9), (17.4.10), because it involves $A_{\alpha\mu}$ only in the field strength $F_{\alpha\mu\nu}$ and in background covariant derivatives $\bar{D}_\mu$ of 'matter' fields $A'_{\alpha\mu}, \psi'$ and ψ.

This new transformation should be carefully distinguished from a true gauge transformation. Such a transformation can have no effect on A or ψ, which are just prescribed classical background fields, and induces an ordinary gauge transformation on $A + A'$ and $\psi + \psi'$, so

$$\delta_{\text{TRUE}}A^\mu_\alpha = 0\ , \tag{17.4.16}$$

$$\begin{aligned}\delta_{\text{TRUE}}A'^\mu_\alpha &= \partial^\mu\epsilon_\alpha - C_{\alpha\beta\gamma}\epsilon_\beta(A^\mu_\gamma + A'^\mu_\gamma) \\ &= \bar{D}_\mu\epsilon_\alpha - C_{\alpha\beta\gamma}\epsilon_\beta A'^\mu_\gamma\end{aligned} \tag{17.4.17}$$

and

$$\delta_{\text{TRUE}}\ \psi = 0\ , \tag{17.4.18}$$

$$\delta_{\text{TRUE}}\ \psi' = i t_\alpha\epsilon_\alpha(\psi + \psi')\ . \tag{17.4.19}$$

Of course, this is the same as the formal transformations (17.4.2), (17.4.3), (17.4.9), (17.4.10) in its effect on $A + A'$ and $\psi + \psi'$, and therefore also leaves the original Lagrangian $\mathscr{L}$ invariant. However, for our new choice

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

(17.4.1) of f_α, the term $f_\alpha f_\alpha$ does *not* depend only on $A + A'$, and is *not* invariant under (17.4.16) and (17.4.17). Instead,

$$\delta_{\text{TRUE}} f_\alpha = \bar{D}_\mu(\bar{D}^\mu \epsilon_\alpha - C_{\alpha\beta\gamma}\epsilon_\beta A'^\mu_\gamma) \tag{17.4.20}$$

with $\bar{D}_\mu$ given by (17.4.5).

Finally, let us consider the ghost Lagrangian in this new gauge. The quantity (15.7.3) in the ghost action is given in general by just replacing ϵ_α with the ghost field $\omega_\alpha + \omega'_\alpha$ in $\delta_{\text{TRUE}} f_\alpha$:

$$\Delta_\alpha = \bar{D}_\mu\left[\bar{D}^\mu(\omega_\alpha + \omega'_\alpha) - C_{\alpha\beta\gamma}(\omega_\beta + \omega'_\beta)A'^\mu_\gamma\right] . \tag{17.4.21}$$

The ghost Lagrangian in Eq. (15.6.2) is therefore

$$\mathscr{L}_{\text{GH}} = (\omega^*_\alpha + \omega'^*_\alpha)\,\bar{D}_\mu\left[\bar{D}^\mu(\omega_\alpha + \omega'_\alpha) - C_{\alpha\beta\gamma}(\omega_\beta + \omega'_\beta)\,A'^\mu_\gamma\right] \tag{17.4.22}$$

or, integrating by parts,

$$\mathscr{L}_{\text{GH}} = -\left(\bar{D}_\mu(\omega^*_\alpha + \omega'^*_\alpha)\right)\left(\bar{D}^\mu(\omega_\alpha + \omega'_\alpha) - C_{\alpha\beta\gamma}(\omega_\beta + \omega'_\beta)A'^\mu_\gamma\right) . \tag{17.4.23}$$

This is manifestly invariant under the joint transformations (17.4.2), (17.4.3), supplemented now with transformations on ω and ω':

$$\delta\omega_\alpha = -C_{\alpha\beta\gamma}\epsilon_\beta\omega_\gamma , \tag{17.4.24}$$

$$\delta\omega'_\alpha = -C_{\alpha\beta\gamma}\epsilon_\beta\omega'_\gamma , \tag{17.4.25}$$

and likewise

$$\delta\omega^*_\alpha = -C_{\alpha\beta\gamma}\epsilon_\beta\omega^*_\gamma , \tag{17.4.26}$$

$$\delta\omega'^*_\alpha = -C_{\alpha\beta\gamma}\epsilon_\beta\omega'^*_\gamma . \tag{17.4.27}$$

We see that the formal combined transformation (17.4.2), (17.4.3), (17.4.9), (17.4.10), and (17.4.24)–(17.4.27) leave invariant the complete Lagrangian in the modified action (15.6.4):

$$\mathscr{L}_{\text{MOD}} = \mathscr{L} - \frac{1}{2\xi} f_\alpha f_\alpha + \mathscr{L}_{GH} . \tag{17.4.28}$$

We are integrating over A', ψ', ω', and ω'^* with a measure that is presumed to be invariant under the simple matrix transformations (17.4.3), (17.4.10), (17.4.25), and (17.4.27), so the effective action $\Gamma[A, \psi, \omega, \omega^*]$ is invariant under the remaining transformations (17.4.2), (17.4.9), (17.4.24), and (17.4.26). In other words, it is gauge-invariant in the same sense as the original action $I[A, \psi, \omega, \omega^*]$.

This formal gauge invariance sets powerful constraints on the infinities that can occur in the effective action. The ultraviolet divergences in Γ appear in the coefficients of terms whose dimensionalities are [mass]^d with $d \leq 4$, but here these terms are invariant under the background gauge transformations (17.4.2), (17.4.9), (17.4.24), and (17.4.26). For instance, in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

a gauge theory based on a simple gauge group, with spin 1/2 fermions belonging to an irreducible representation of this group, the only such terms are of the form

$$\Gamma_\infty = \int d^4x\, \mathscr{L}_\infty\,, \tag{17.4.29}$$

$$\begin{aligned}\mathscr{L}_\infty = &-\tfrac{1}{4} L_A\, F_{\alpha\mu\nu} F_\alpha^{\mu\nu} - L_\psi \bar\psi \gamma^\mu \bar D_\mu \psi \\ &- m\, L_m \bar\psi \psi - L_\omega (\bar D_\mu \omega_\alpha^*)(\bar D^\mu \omega_\alpha)\,,\end{aligned} \tag{17.4.30}$$

where here $F_{\alpha\mu\nu}$, $\bar D_\mu \psi$, $\bar D_\mu \omega_\alpha$, and $\bar D_\mu \omega_\alpha^*$ are constructed entirely from background fields:**

$$F_{\alpha\mu\nu} \equiv \partial_\mu A_{\alpha\nu} - \partial_\nu A_{\alpha\mu} + C_{\alpha\beta\gamma} A_{\beta\mu} A_{\gamma\nu}\,, \tag{17.4.31}$$

$$\bar D_\mu \psi \equiv \partial_\mu \psi - i\, t_\alpha A_{\alpha\mu} \psi\,, \tag{17.4.32}$$

$$\bar D_\mu \omega_\alpha \equiv \partial_\mu \omega_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu} \omega_\gamma\,, \tag{17.4.33}$$

$$\bar D_\mu \omega_\alpha^* \equiv \partial_\mu \omega_\alpha^* + C_{\alpha\beta\gamma} A_{\beta\mu} \omega_\gamma^*\,. \tag{17.4.34}$$

Dimensional analysis leads us to expect that the constants L_A, L_ψ, L_m, and L_ω are logarithmically divergent.

To deal with these infinities, we note that the Lagrangian (17.4.12) contains a purely classical piece

$$\mathscr{L}_{\rm CLASS} = -\tfrac{1}{4} F_{\alpha\mu\nu} F_\alpha^{\mu\nu} - \bar\psi(\gamma^\mu \bar D_\mu + m)\psi - (\bar D_\mu \omega_\alpha^*)(\bar D^\mu \omega_\alpha) \tag{17.4.35}$$

obtained by dropping all terms in $\mathscr{L}_{\rm MOD}$ that involve the quantum fields $A', \psi', \omega', \omega'^*$. We define renormalized fields

$$A_{\alpha\mu}^R \equiv \sqrt{1 + L_A}\, A_{\alpha\mu}\,, \tag{17.4.36}$$

$$\psi_\ell^R \equiv \sqrt{1 + L_\psi}\, \psi_\ell\,, \tag{17.4.37}$$

$$\omega_\alpha^R \equiv \sqrt{1 + L_\omega}\, \omega_\alpha\,, \tag{17.4.38}$$

$$\omega_\alpha^{R*} \equiv \sqrt{1 + L_\omega}\, \omega_\alpha^*\,, \tag{17.4.39}$$

so that the sum of the terms (17.4.30) and (17.4.35) takes the form

$$\begin{aligned}\mathscr{L}_{\rm CLASS} + \mathscr{L}_\infty = &-\tfrac{1}{4} F_{\alpha\mu\nu}^R F_\alpha^{R\mu\nu} - \bar\psi^R \gamma^\mu D_\mu^R \psi^R \\ &- m^R \bar\psi^R \psi^R - (D_\mu^R \omega_\alpha^{*R})(D^{\mu R} \omega_\alpha^R)\,,\end{aligned} \tag{17.4.40}$$

where m^R is the renormalized mass

$$m^R \equiv m(1 + L_m)\big/(1 + L_\psi)\,, \tag{17.4.41}$$

** The condition of a simple gauge group insures that there is just a single kinetic term for A and for ω, proportional to $F_{\alpha\mu\nu} F_\alpha^{\mu\nu}$ and $\bar D_\mu \omega_\alpha^* \bar D^\mu \omega_\alpha$, respectively, while the condition that ψ transforms irreducibly insures that there is just a single kinetic term and a single mass term for ψ. It would be easy to treat more general possibilities, at the cost of a slight complication in notation. We are also implicitly using the conservation of ghost number.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

and

$$F^R_{\alpha\mu\nu} \equiv \partial_\mu A^R_{\alpha\nu} - \partial_\nu A^R_{\alpha\mu} + C^R_{\alpha\beta\gamma} A^R_{\beta\mu} A^R_{\gamma\nu} , \tag{17.4.42}$$

$$D^R_\mu \psi^R \equiv \partial_\mu \psi^R - i\, t^R_\alpha A^R_{\alpha\mu} \psi^R , \tag{17.4.43}$$

$$D^R_\mu \omega^R_\alpha \equiv \partial_\mu \omega^R_\alpha + C^R_{\alpha\beta\gamma} A^R_{\beta\mu} \omega^R_\gamma , \tag{17.4.44}$$

$$D^R_\mu \omega^{*R}_\alpha \equiv \partial_\mu \omega^{*R}_\alpha + C^R_{\alpha\beta\gamma} A^R_{\beta\mu} \omega^{*R}_\gamma . \tag{17.4.45}$$

The renormalized structure constants and group generators here are just

$$C^R_{\alpha\beta\gamma} \equiv (1 + L_A)^{-1/2} C_{\alpha\beta\gamma} , \tag{17.4.46}$$

$$t^R_\alpha \equiv (1 + L_A)^{-1/2}\, t_\alpha . \tag{17.4.47}$$

Because we assumed that the Lie algebra here is simple, the structure constant $C_{\alpha\beta\gamma}$ and group generator t_α are fixed by the group structure except for a single common factor, the unrenormalized gauge coupling constant g. Eqs. (17.4.46) and (17.4.47) thus simply tell us that the gauge coupling constant g^R factor in C^R and t^R is renormalized by

$$g^R = g(1 + L_A)^{-1/2} . \tag{17.4.48}$$

This result exhibits the particular virtue of the background field gauge. In a general gauge we would encounter independent renormalization factors for the gauge field and the gauge coupling constant, and we would have to calculate two separate amplitudes (say, the vacuum polarization and three-gauge-field vertex function) in order to be able to sort these out. In background field gauge, background field gauge invariance ties these two renormalizations together by requiring that the infinite terms in the effective Lagrangian involve the field strength in its original form (17.4.31), and so we can calculate the charge renormalization factor by studying just one gauge field amplitude.

17.5 A One-Loop Calculation in Background Field Gauge

As an exercise, we are now going to calculate the one-loop renormalization factor for the gauge coupling constant in a general non-Abelian gauge theory. As we will see in the next chapter, this provides an essential input in so-called 'renormalization group' calculations of physical processes at high energy; the results we obtain here will be used there to demonstrate the asymptotic freedom of non-Abelian gauge theories.

The method to be employed here is somewhat novel. Usually, one considers the effective action in a spacetime-dependent background gauge field, and calculates the terms quadratic in this field, extracting a factor $(q_\mu A_{\alpha\nu} - q_\nu A_{\alpha\mu})^2$ (where q is the gauge field four-momentum), and only

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

then isolating the logarithmic divergence by setting $q = 0$ in the coefficient of this factor. Instead we shall follow the much simpler course of taking the gauge field to be spacetime-independent from the beginning. In this case the terms in the effective action that are quadratic or cubic in the gauge field of course vanish, but there is a non-vanishing quartic term that is ultraviolet divergent, and which can be used to calculate the coupling constant renormalization factor $(1 + L_A)^{-1/2}$. In this way our one-loop calculation becomes a matter of simple matrix algebra. Note that this procedure can only work in background field gauge; otherwise there would be independent logarithmic divergences in the parts of the effective action that are quartic and quadratic in the background gauge field.

With this motivation, we turn to the calculation of the one-loop effective action in a background field for which $A_{\alpha\mu}$ is constant and $\psi = \omega = \omega^* = 0$. For such a background field, the full modified Lagrangian is*

$$\mathscr{L}_{\text{MOD}} = \mathscr{L} + \mathscr{L}_f + \mathscr{L}_{GH} , \tag{17.5.1}$$

$$\mathscr{L} = -\tfrac{1}{4}\left(F_{\alpha\mu\nu} + \bar{D}_\mu A'_{\alpha\nu} - \bar{D}_\nu A'_{\alpha\mu} + C_{\alpha\beta\gamma} A'_{\beta\mu} A'_{\gamma\nu}\right)^2 - \bar{\psi}'(\bar{\not{D}} - i t_\alpha \not{A}'_\alpha + m)\psi' , \tag{17.5.2}$$

$$\mathscr{L}_f \equiv -\frac{1}{2\xi} f_\alpha f_\alpha = -\frac{1}{2\xi}(\bar{D}_\mu A'^{\mu}_{\alpha})^2 , \tag{17.5.3}$$

$$\mathscr{L}_{GH} = -(\bar{D}_\mu \omega'^{*}_{\alpha})(\bar{D}^\mu \omega'_\alpha - C_{\alpha\beta\gamma}\omega'_\beta A'^{\mu}_{\gamma}) . \tag{17.5.4}$$

One-loop graphs for vacuum–vacuum amplitudes are calculated from the part of the action that is quadratic in the quantum fields $A', \psi', \omega', \omega'^*$ over which one integrates. Keeping only such quadratic terms, we have

$$\mathscr{L}_{\text{QUAD}} = -\frac{1}{4}(\bar{D}_\mu A'_{\alpha\nu} - \bar{D}_\nu A'_{\alpha\mu})^2 - \frac{1}{2} F^{\mu\nu}_{\alpha} C_{\alpha\beta\gamma} A'_{\beta\mu} A'_{\gamma\nu} - \bar{\psi}'(\bar{\not{D}} + m)\psi' - \frac{1}{2\xi}(\bar{D}_\mu A'^{\mu}_{\alpha})^2 - (\bar{D}_\mu \omega'^{*}_{\alpha})(\bar{D}^\mu \omega'_\alpha) . \tag{17.5.5}$$

The corresponding action may be put in the general quadratic form:

$$I_{\text{QUAD}} \equiv \int d^4x\, \mathscr{L}_{\text{QUAD}} = -\tfrac{1}{2}\int d^4x\, d^4y\, A'^{\mu}_{\alpha}(x) A'^{\nu}_{\beta}(y)\, \mathscr{D}^{A}_{x\alpha\mu, y\beta\nu} - \int d^4x\, d^4y\, \bar{\psi}'_k(x)\psi'_\ell(y)\, \mathscr{D}^{\psi}_{xk, y\ell} - \int d^4x\, d^4y\, \omega'^{*}_{\alpha}(x)\omega'_\beta(y)\, \mathscr{D}^{\omega}_{x\alpha, y\beta} , \tag{17.5.6}$$

* See Eqs. (17.4.12), (17.4.4), and (17.4.23). We are here specializing to the case of matter fields forming a multiplet of spin $\frac{1}{2}$ fermions. The squares in $\mathscr{L}$ and $\mathscr{L}_f$ include obvious index contractions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

with

$$\begin{aligned}\mathscr{D}^{A}_{x\alpha\mu,y\beta\nu} \equiv \eta_{\mu\nu}&\Big(-\delta_{\gamma\alpha}\frac{\partial}{\partial x_\lambda}+C_{\gamma\delta\alpha}A^{\lambda}_{\delta}(x)\Big)\Big(-\delta_{\gamma\beta}\frac{\partial}{\partial y^\lambda}+C_{\gamma\epsilon\beta}A_{\epsilon\lambda}(y)\Big)\delta^4(x-y)\\ &-\Big(-\delta_{\gamma\alpha}\frac{\partial}{\partial x^\nu}+C_{\gamma\delta\alpha}A_{\delta\nu}(x)\Big)\Big(-\delta_{\gamma\beta}\frac{\partial}{\partial y^\mu}+C_{\gamma\epsilon\beta}A_{\epsilon\mu}(y)\Big)\delta^4(x-y)\\ &+F_{\gamma\mu\nu}(x)C_{\gamma\alpha\beta}\delta^4(x-y)\\ &+\frac{1}{\xi}\Big(-\delta_{\gamma\alpha}\frac{\partial}{\partial x^\mu}+C_{\gamma\delta\alpha}A_{\delta\mu}(x)\Big)\Big(-\delta_{\gamma\beta}\frac{\partial}{\partial y^\nu}+C_{\gamma\epsilon\beta}A_{\epsilon\nu}(y)\Big)\delta^4(x-y)\end{aligned} \tag{17.5.7}$$

$$\mathscr{D}^{\psi}_{xk,y\ell} \equiv \left(-\gamma^\mu\frac{\partial}{\partial y^\mu}-it_\alpha A_\alpha(y)+m\right)_{k\ell}\delta^4(x-y)\,, \tag{17.5.8}$$

$$\mathscr{D}^{\omega}_{x\alpha,y\beta} = \left(-\delta_{\gamma\alpha}\frac{\partial}{\partial x^\lambda}+C_{\gamma\delta\alpha}A_{\delta\lambda}(x)\right)\left(-\delta_{\gamma\beta}\frac{\partial}{\partial y_\lambda}+C_{\gamma\epsilon\beta}A^{\lambda}_{\epsilon}(y)\right)\delta^4(x-y)\,. \tag{17.5.9}$$

(The minus signs in front of $\partial/\partial x$ and $\partial/\partial y$ drop out when we integrate by parts.)

The one-loop contribution to the effective action is given (as in Section 16.2) by

$$\begin{aligned}\exp\left(i\Gamma^{1\text{ loop}}[A]\right) &\propto \int_{1PI}(\prod dA')(\prod d\psi')(\prod d\bar{\psi}')(\prod d\omega')(\prod d\omega^{*\prime})\\ &\quad\times\exp\left(iI_{\text{QUAD}}[A',\psi',\bar{\psi}',\omega',\omega'^{*};A]\right)\\ &\propto (\text{Det}\mathscr{D}^{A})^{-1/2}(\text{Det}\mathscr{D}^{\psi})^{+1}(\text{Det}\mathscr{D}^{\omega})^{+1}\,.\end{aligned} \tag{17.5.10}$$

(The exponents $-1/2$ and $+1$ appear because A' is a real boson field, while ψ', $\bar{\psi}'$, ω', and ω'^{*} are distinct fermionic fields.) The calculation of such determinants is generally not an easy task. However, it becomes much simpler in the case of constant external fields, where the $\mathscr{D}$s can be diagonalized by passing to momentum space.

Let's therefore now consider the case of constant background field $A_{\alpha\mu}$. For non-Abelian gauge theories such a constant field *cannot* be removed by a gauge transformation, as shown by the non-zero values of various gauge-covariant fields

$$F_{\alpha\mu\nu} = C_{\alpha\beta\gamma}A_{\beta\mu}A_{\gamma\nu}\,, \tag{17.5.11}$$

$$D_\lambda F_{\delta\mu\nu} = C_{\delta\epsilon\alpha}C_{\alpha\beta\gamma}A_{\epsilon\lambda}A_{\beta\mu}A_{\gamma\nu}\,, \tag{17.5.12}$$

and so on. Lorentz and gauge invariance tell us how to express the part of $\Gamma[A]$ of a given dimensionality as the integral of a finite number of local functions of $F_{\alpha\mu\nu}$, $D_\lambda F_{\alpha\mu\nu}$, etc.; the coefficients of the terms in this expression can be inferred by comparing the contribution that these terms

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

make to $\Gamma[A]$ for constant background field $A_{\alpha\mu}$ with the results of a perturbative expansion.

We transform each of the 'matrices' $\mathscr{D}^A$, $\mathscr{D}^\psi$, and $\mathscr{D}^\omega$ to a momentum basis by the usual normalized Fourier transform

$$\mathscr{D}_{q\cdots,p\cdots} = \int \frac{d^4x}{(2\pi)^2}\, e^{-iq\cdot x} \int \frac{d^4y}{(2\pi)^2}\, e^{ip\cdot y}\, \mathscr{D}_{x\cdots,y\cdots}\,. \tag{17.5.13}$$

With A constant, this gives

$$\mathscr{D}_{q\cdots,p\cdots} = \delta^4(p-q)\,\mathscr{M}_{\cdots,\cdots}(q)\,, \tag{17.5.14}$$

where $\cdots$ denotes discrete indices, and the $\mathscr{M}$ are finite q-dependent matrices

$$\begin{aligned}\mathscr{M}^A_{\alpha\mu,\beta\nu}(q) = {}& \eta_{\mu\nu}(-iq_\lambda\delta_{\gamma\alpha} + A_{\delta\lambda}C_{\gamma\delta\alpha})(iq^\lambda\delta_{\gamma\beta} + A_\epsilon{}^\lambda C_{\gamma\epsilon\beta}) \\ & -(-iq_\nu\delta_{\gamma\alpha} + A_{\delta\nu}C_{\gamma\delta\alpha})\,(iq_\mu\delta_{\gamma\beta} + A_{\epsilon\mu}C_{\gamma\epsilon\beta}) \\ & +F_{\gamma\mu\nu}C_{\gamma\alpha\beta} \\ & +(-iq_\mu\delta_{\gamma\alpha} + A_{\delta\mu}C_{\gamma\delta\alpha})\,(iq_\nu\delta_{\gamma\beta} + A_{\epsilon\nu}C_{\gamma\epsilon\beta})/\xi \\ & +\epsilon \text{ terms}\,, \end{aligned} \tag{17.5.15}$$

$$\mathscr{M}^\psi_{k\ell}(q) = (i\,\not{q} - it_\alpha \not{A}_\alpha + m)_{k\ell} + \epsilon \text{ terms}\,, \tag{17.5.16}$$

$$\begin{aligned}\mathscr{M}^\omega_{\alpha\beta}(q) = {}& (-iq_\lambda\delta_{\gamma\alpha} + A_{\delta\lambda}C_{\gamma\delta\alpha})\,(iq^\lambda\delta_{\gamma\beta} + A_\epsilon{}^\lambda C_{\gamma\epsilon\beta}) \\ & +\epsilon \text{ terms}\,, \end{aligned} \tag{17.5.17}$$

with $F_{\gamma\mu\nu}$ given by Eq. (17.5.11). From Eq. (17.5.10) we have then

$$\begin{aligned} i\,\Gamma^{(1\text{ loop})}[A] &= -\tfrac{1}{2}\ln \mathrm{Det}\mathscr{D}^A + \ln \mathrm{Det}\mathscr{D}^\psi + \ln \mathrm{Det}\mathscr{D}^\omega \\ &= -\tfrac{1}{2}\mathrm{Tr}\ln\,\mathscr{D}^A + \mathrm{Tr}\ln\,\mathscr{D}^\psi + \mathrm{Tr}\ln\,\mathscr{D}^\omega \\ &= \delta^4(p-p)\int d^4q\Big[-\tfrac{1}{2}\mathrm{tr}\ln\,\mathscr{M}^A(q) + \mathrm{tr}\ln\,\mathscr{M}^\psi(q) \\ &\quad +\mathrm{tr}\ln\,\mathscr{M}^\omega(q)\Big]\,. \end{aligned} \tag{17.5.18}$$

We denote traces by 'tr' instead of 'Tr' in the last line of Eq. (17.5.18) (and from now on in this section) to indicate that these are the usual traces of finite matrices rather than of integral operators.

Since we are aiming here at a calculation of the infinite factor L_A multiplying FF terms in the effective action, let's isolate the term in (17.5.18) that is of fourth order in the background field A. For this purpose, it is convenient to divide each $\mathscr{M}$ into terms $\mathscr{M}_n$ containing $n = 0, 1,$ or 2 factors of A:

$$\mathscr{M} = \mathscr{M}_0 + \mathscr{M}_1 + \mathscr{M}_2\,. \tag{17.5.19}$$

It is then elementary algebra to show that the term in (17.5.18) of fourth

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

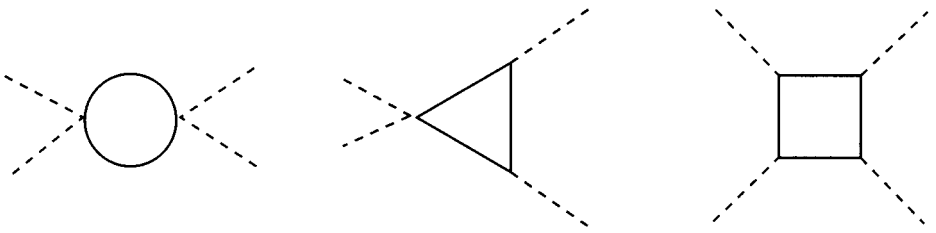

Figure 17.1. One-loop Feynman diagrams for the term in the quantum effective action that is quartic in a constant background gauge field A^μ_α. Here solid lines represent internal gauge, ghost, or matter lines; dashed lines indicate factors of A^μ_α. These three diagrams correspond to the three terms in Eq. (17.5.20).

order in $A_{\alpha\mu}$ is

$$[\text{tr ln } \mathscr{M}]_{A^4} = \text{tr}\left\{ -\tfrac{1}{2}[\mathscr{M}_0^{-1}\mathscr{M}_2]^2 + [\mathscr{M}_0^{-1}\mathscr{M}_1]^2\mathscr{M}_0^{-1}\mathscr{M}_2 - \tfrac{1}{4}[\mathscr{M}_0^{-1}\mathscr{M}_1]^4 \right\} . \quad (17.5.20)$$

(To see this, insert factors ϵ and ϵ^2 multiplying $\mathscr{M}_1$ and $\mathscr{M}_2$ in Eq. (17.5.19), differentiate tr ln $\mathscr{M}$ four times with respect to ϵ, divide by 4! and set $\epsilon = 0$.) The $\mathscr{M}_0^{-1}$ factors here are just the usual propagators; for $\xi = 1$, these are

$$[\mathscr{M}_0^A(q)]^{-1}_{\alpha\mu,\beta\nu} = \delta_{\alpha\beta}\eta_{\mu\nu}(q^2 - i\epsilon)^{-1} , \quad (17.5.21)$$

$$[\mathscr{M}_0^\psi(q)]^{-1}_{k\ell} = [i\not{q} + m]^{-1}_{k\ell} , \quad (17.5.22)$$

$$[\mathscr{M}_0^\omega(q)]^{-1}_{\alpha,\beta} = \delta_{\alpha\beta}(q^2 - i\epsilon)^{-1} . \quad (17.5.23)$$

Indeed, the three terms in Eq. (17.5.20) just correspond to the three Feynman diagrams shown in Figure 17.1; the present method of calculation saves us from having to think about signs and combinatoric factors.

For the A loop, Eq. (17.5.15) gives for $\xi = 1$:

$$[\mathscr{M}_1^A(q)]_{\alpha\mu,\beta\nu} = -2\eta_{\mu\nu}q^\lambda[\mathscr{A}_\lambda]_{\alpha\beta} ,$$

$$[\mathscr{M}_2^A(q)]_{\alpha\mu,\beta\nu} = +\left[\eta_{\mu\nu}\mathscr{A}_\lambda\mathscr{A}^\lambda - \mathscr{A}_\nu\mathscr{A}_\mu - \mathscr{A}_\mu\mathscr{A}_\nu\right]_{\alpha\beta} + F_{\gamma\mu\nu}C_{\gamma\alpha\beta} ,$$

where $\mathscr{A}_\lambda$ is the matrix

$$[\mathscr{A}_\lambda]_{\alpha\beta} \equiv -iC_{\alpha\beta\gamma}A_{\gamma\lambda}$$

for which

$$\begin{aligned}[\mathscr{A}_\lambda, \mathscr{A}_\rho]_{\alpha\beta} &= -C_{\alpha\delta\gamma}C_{\delta\beta\epsilon}(A_{\gamma\lambda}A_{\epsilon\rho} - A_{\gamma\rho}A_{\epsilon\lambda}) \\ &= -(C_{\alpha\delta\gamma}C_{\delta\beta\epsilon} + C_{\alpha\delta\epsilon}C_{\delta\beta\gamma})A_{\gamma\lambda}A_{\epsilon\rho} \\ &= +C_{\alpha\delta\beta}C_{\delta\epsilon\gamma}A_{\gamma\lambda}A_{\epsilon\rho} = C_{\alpha\delta\beta}F_{\delta\rho\lambda} .\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

The integrals have the structure

$$\int d^4q\, q^\mu q^\nu f(q^2) = \tfrac{1}{4}\eta^{\mu\nu}\int d^4q\, q^2 f(q^2)\,,$$

$$\int d^4q\, q^\mu q^\nu q^\rho q^\sigma f(q^2) = \tfrac{1}{24}[\eta^{\mu\nu}\eta^{\rho\sigma}+\eta^{\mu\rho}\eta^{\nu\sigma}+\eta^{\mu\sigma}\eta^{\nu\rho}]\int d^4q\,(q^2)^2 f(q^2)\,.$$

We then find, for $\xi = 1$:

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^A(q)^{-1}\mathscr{M}_2^A(q)\right]^2\right\} = 4\mathscr{I}\ \mathrm{tr}[\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta] + 4\mathscr{I}\, C_{\gamma\alpha\beta}C_{\delta\alpha\beta}F_{\gamma\mu\nu}F_\delta^{\mu\nu}\,,$$

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^A(q)^{-1}\mathscr{M}_1^A(q)\right]^2\mathscr{M}_0^A(q)^{-1}\mathscr{M}_2^A(q)\right\} = 4\mathscr{I}\ \mathrm{tr}[\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta]\,,$$

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^A(q)^{-1}\mathscr{M}_1^A(q)\right]^4\right\} = \tfrac{8}{3}\mathscr{I}\ \mathrm{tr}\left[2\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta + \mathscr{A}^\lambda\mathscr{A}^\eta\mathscr{A}_\lambda\mathscr{A}_\eta\right],$$

where $\mathscr{I}$ is the divergent integral

$$\mathscr{I} \equiv \int d^4q\,[q^2 - i\epsilon]^{-2}\,, \tag{17.5.24}$$

whose significance is discussed below. Putting this together in Eq. (17.5.20), we have

$$\int d^4q\left[\mathrm{tr}\ln\mathscr{M}^A(q)\right]_{A^4} = \tfrac{2}{3}\mathscr{I}\ \mathrm{tr}\left[\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta - \mathscr{A}^\lambda\mathscr{A}^\eta\mathscr{A}_\lambda\mathscr{A}_\eta\right] - 2\mathscr{I}\, C_{\gamma\alpha\beta}C_{\delta\alpha\beta}\,F_{\gamma\mu\nu}\,F_\delta^{\mu\nu}\,.$$

Both terms are actually of the same form, and when combined yield

$$\int d^4q\left[\mathrm{tr}\ln\mathscr{M}^A(q)\right]_{A^4} = -\tfrac{5}{3}\mathscr{I}\, C_{\gamma\alpha\beta}C_{\delta\alpha\beta}F_{\gamma\mu\nu}F_\delta^{\mu\nu}\,. \tag{17.5.25}$$

Now jumping ahead to the ghost loop, we see from Eq. (17.5.17) that

$$[\mathscr{M}_1^\omega(q)]_{\alpha\beta} = -2[\mathscr{A}^\lambda]_{\alpha\beta}\,q_\lambda\,, \tag{17.5.26}$$

$$[\mathscr{M}_2^\omega(q)]_{\alpha\beta} = [\mathscr{A}^\lambda\mathscr{A}_\lambda]_{\alpha\beta}\,. \tag{17.5.27}$$

We then find

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^\omega(q)^{-1}\mathscr{M}_2^\omega(q)\right]^2\right\} = \mathscr{I}\ \mathrm{tr}[\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta]\,,$$

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^\omega(q)^{-1}\mathscr{M}_1^\omega(q)\right]^2\mathscr{M}_0^\omega(q)^{-1}\mathscr{M}_2^\omega(q)\right\} = \mathscr{I}\ \mathrm{tr}[\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta]\,,$$

$$\int d^4q\ \mathrm{tr}\left\{\left[\mathscr{M}_0^\omega(q)^{-1}\mathscr{M}_1^\omega(q)\right]^4\right\} = \tfrac{2}{3}\mathscr{I}\ \mathrm{tr}[2\mathscr{A}^\lambda\mathscr{A}_\lambda\mathscr{A}^\eta\mathscr{A}_\eta + \mathscr{A}^\lambda\mathscr{A}^\eta\mathscr{A}_\lambda\mathscr{A}_\eta]\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

Thus for the ghost loop the integral of the quantity (17.5.20) is

$$\int d^4q\ [\mathrm{tr}\ \ln \mathscr{M}^{\omega}(q)]_{A^4} = \tfrac{1}{6}\mathscr{I}\ \mathrm{tr}[\mathscr{A}^{\lambda}\mathscr{A}_{\lambda}\mathscr{A}^{\eta}\mathscr{A}_{\eta} - \mathscr{A}^{\lambda}\mathscr{A}^{\eta}\mathscr{A}_{\lambda}\mathscr{A}_{\eta}]$$
$$= \tfrac{1}{12}\mathscr{I}C_{\gamma\alpha\beta}C_{\delta\alpha\beta}F_{\gamma\mu\nu}F_{\delta}^{\mu\nu}\ . \qquad (17.5.28)$$

Finally, the vertices in the matter loop are

$$[\mathscr{M}_1^{\psi}(q)]_{k\ell} = -i(t_{\alpha}\not{A}_{\alpha})_{k\ell}\ , \qquad [\mathscr{M}_2^{\psi}(q)]_{k\ell} = 0\ ,$$

so here there is only one term in Eq. (17.5.20):

$$\int d^4q\ [\mathrm{tr}\ \ln \mathscr{M}^{\psi}(q)]_{A^4} = -\tfrac{1}{4}\int d^4q\ \mathrm{tr}\left\{\left[(i\not{q}+m)^{-1}t_{\alpha}\not{A}_{\alpha}\right]^4\right\}\ .$$

We are interested in the ultraviolet-divergent part of this integral, so we may drop the mass (which is negligible for large q^{ν}) and write

$$\int d^4q\ [\mathrm{tr}\ \ln \mathscr{M}^{\psi}(q)]_{A^4} = -\tfrac{1}{4}\,\mathrm{tr}\{t_{\alpha}t_{\beta}t_{\gamma}t_{\delta}\}\,A_{\alpha\mu}A_{\beta\nu}A_{\gamma\rho}A_{\delta\sigma}$$
$$\times \int d^4q \frac{\mathrm{tr}\{\not{q}\gamma^{\mu}\not{q}\gamma^{\nu}\not{q}\gamma^{\rho}\not{q}\gamma^{\sigma}\}}{(q^2-i\epsilon)^4}$$
$$= -\frac{\mathscr{I}}{96}\,\mathrm{tr}\{t_{\alpha}t_{\beta}t_{\gamma}t_{\delta}\}A_{\alpha\mu}A_{\beta\nu}A_{\gamma\rho}A_{\delta\sigma}$$
$$\times \mathrm{tr}\left\{2\gamma_{\lambda}\gamma^{\mu}\gamma^{\lambda}\gamma^{\nu}\gamma_{\eta}\gamma^{\rho}\gamma^{\eta}\gamma^{\sigma} + \gamma_{\lambda}\gamma^{\mu}\gamma_{\eta}\gamma^{\nu}\gamma^{\lambda}\gamma^{\rho}\gamma^{\eta}\gamma^{\sigma}\right\}, \qquad (17.5.29)$$

where $\mathscr{I}$ is the same divergent integral as in Eq. (17.5.24). To calculate the traces of Dirac matrices, we use the anticommutation relations of these matrices to write

$$\mathrm{tr}\left\{2\gamma_{\lambda}\gamma^{\mu}\gamma^{\lambda}\gamma^{\nu}\gamma_{\eta}\gamma^{\rho}\gamma^{\eta}\gamma^{\sigma} + \gamma_{\lambda}\gamma^{\mu}\gamma_{\eta}\gamma^{\nu}\gamma^{\lambda}\gamma^{\rho}\gamma^{\eta}\gamma^{\sigma}\right\}$$
$$= 8\mathrm{tr}\{\gamma^{\mu}\gamma^{\nu}\gamma^{\rho}\gamma^{\sigma}\} - 4\mathrm{tr}\{\gamma^{\nu}\gamma^{\mu}\gamma^{\rho}\gamma^{\sigma}\} - 4\mathrm{tr}\{\gamma^{\mu}\gamma^{\rho}\gamma^{\nu}\gamma^{\sigma}\}$$
$$= -64\eta^{\mu\rho}\eta^{\nu\sigma} + 32\eta^{\mu\nu}\eta^{\rho\sigma} + 32\eta^{\mu\sigma}\eta^{\nu\rho}\ .$$

Eq. (17.5.29) then gives

$$\int d^4q\left[\mathrm{tr}\ \ln \mathscr{M}^{\psi}(q)\right]_{A^4} = \tfrac{1}{3}\mathscr{I}\ \mathrm{tr}\left\{[t_{\alpha},t_{\beta}][t_{\gamma},t_{\delta}]\right\}A_{\alpha\mu}A_{\beta\nu}A_{\gamma}^{\mu}A_{\delta}^{\nu}$$
$$= -\tfrac{1}{3}\mathscr{I}F_{\gamma\mu\nu}F_{\delta}^{\mu\nu}\mathrm{tr}\{t_{\gamma}t_{\delta}\}\ . \qquad (17.5.30)$$

Using Eqs. (17.5.25), (17.5.28), and (17.5.30) in Eq. (17.5.18) gives at last

$$\Gamma_{A^4}^{(1\ \mathrm{loop})} = \frac{-i\mathscr{I}}{(2\pi)^4}\int d^4x\, F_{\gamma\mu\nu}F_{\delta}^{\mu\nu}\left[\left(\frac{5}{6}+\frac{1}{12}\right)C_{\gamma\alpha\beta}C_{\delta\alpha\beta} - \frac{1}{3}\,\mathrm{tr}\{t_{\gamma}t_{\delta}\}\right], \qquad (17.5.31)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

where we have expressed the momentum-space delta-function in (17.5.14) as

$$\delta^4(p-p) = (2\pi)^{-4}\int d^4x\, 1\ . \tag{17.5.32}$$

It is important that the result turns out to depend on $A_{\alpha\mu}$ only through the field strength (17.5.11), as required by background gauge invariance.

Let us now (for the first time in this section) use the assumed simplicity of the gauge group and irreducibility of the matter field multiplet. In this case

$$C_{\gamma\alpha\beta}C_{\delta\alpha\beta} = g^2 C_1\,\delta_{\gamma\delta}\ , \tag{17.5.33}$$

$$\mathrm{tr}\{t_\gamma t_\delta\} = g^2 C_2 \delta_{\gamma\delta}\ , \tag{17.5.34}$$

where g is the common gauge coupling constant appearing as a factor in $C_{\gamma\alpha\beta}$ and t_γ, and C_1 and C_2 are numerical constants that characterize the gauge group and the representation of this group provided by the matter multiplet. For instance, in the original Yang–Mills theory the gauge group is $SU(2)$ (or equivalently $SO(3)$) and the structure constants are

$$C_{\gamma\alpha\beta} = g\epsilon_{\alpha\beta\gamma}$$

with α, β, and γ running over the values 1, 2, 3. Comparing with Eq. (17.5.33), we see that here

$$C_1 = 2\ .$$

Also in this theory the matter field forms a doublet with t_α given by $g/2$ times the usual Pauli matrix σ_α, so

$$C_2 = 1/2\ .$$

Somewhat more generally, for the group $SU(N)$ with n_f fermions in the defining representation, with a conventional normalization of generators we have**

$$C_1 = N\ , \qquad C_2 = n_f/2\ . \tag{17.5.35}$$

Returning now to the general case, Eqs. (17.5.33) and (17.5.34) give

$$\Gamma^{(1\ \mathrm{loop})}_{A^4} = \frac{-ig^2\mathscr{J}}{(2\pi)^4}\int d^4x F_{\gamma\mu\nu}F_\gamma^{\mu\nu}\left[\frac{11}{12}C_1 - \frac{1}{3}C_2\right]\ . \tag{17.5.36}$$

That is, the infinite constant L_A in Eq. (17.4.30) is

$$L_A = \frac{4ig^2\mathscr{J}}{(2\pi)^4}\left(\frac{11}{12}C_1 - \frac{1}{3}C_2\right)\ . \tag{17.5.37}$$

** For $SU(3)$ the t_α are taken as $g/2$ times the Gell-Mann matrices λ_α used in Section 19.7, so that $C_{\alpha\beta\gamma} = (g/2)f_{\alpha\beta\gamma}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

It remains to say a word about the interpretation of the divergent integral $\mathscr{I}$. First, before we try to integrate over the three-momentum $\mathbf{q}$, we can rotate the contour of integration of q^0 in Eq. (17.5.24) to the imaginary axis; as usual, the $-i\epsilon$ in the denominator forces us to rotate counterclockwise, so that $q^0 = iq^4$, with q^4 running from $-\infty$ to $+\infty$. The integral is then

$$\mathscr{I} = i \int_0^\infty \frac{2\pi^2 q^3 dq}{q^4} , \tag{17.5.38}$$

where q is now the magnitude of the Euclidean four-vector (q^1, q^2, q^3, q^4). To go further, we evidently need some method of regulating the integral. The simplest way of dealing with the ultraviolet divergence is just to cut off the integral for q above a scale Λ. However, we also need a lower cut-off to deal with the infrared divergence. This is provided by the physics of the situation. If the momenta of the four vector particles is not zero, then a momentum flows through the internal lines of the diagrams, providing an infrared cut-off at the scale μ of these momenta. Similarly, if we evaluate the fourth variational derivative of $\Gamma[A]$ with respect to A not at $A = 0$ but at a finite A, then the propagators of the internal lines do not blow up at zero momenta, and we have an infrared cut-off at a scale $\mu \sim gA$. Either way, $\mathscr{I}$ takes the form

$$\mathscr{I} = 2\pi^2 i \int_\mu^\Lambda \frac{dq}{q} = 2\pi^2 i \ln\left(\frac{\Lambda}{\mu}\right) \tag{17.5.39}$$

and hence

$$L_A = -\frac{g^2}{2\pi^2}\left(\frac{11}{12}C_1 - \frac{1}{3}C_2\right)\ln\left(\frac{\Lambda}{\mu}\right) + O(g^4) . \tag{17.5.40}$$

Eq. (17.4.48) then gives the renormalized coupling as

$$g_R = g\left[1 + \frac{g^2}{4\pi^2}\ln\left(\frac{\Lambda}{\mu}\right)\left(\frac{11}{12}C_1 - \frac{1}{3}C_2\right) + O(g^4)\right] . \tag{17.5.41}$$

We note that while in quantum electrodynamics the radiative corrections discussed in Section 11.2 decrease the physical coupling g_R relative to the bare coupling g, in non-Abelian gauge theories they *increase* the physical coupling over the bare coupling, provided that the fermion multiplet is small enough so that $C_2 < 11C_1/4$. The importance of this point will be explored in Chapter 18.

Alternatively, we can deal with the ultraviolet divergence by the methods of dimensional regularization, discussed in Section 11.2. Here, in place of Eq. (17.5.39), we write

$$\mathscr{I} = i \int_0^\infty \frac{2\pi^2 q^{d-1} dq}{(q^2 + \mu^2)^2} , \tag{17.5.42}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

where d is a complex dimensionality, allowed to approach 4 at the end of the calculation, and μ is an infrared cut-off, again taken of the order of the external momenta (or of the background fields times g). As long as d is complex with Re $d < 0$ and $\mu^2 > 0$, this has the finite value

$$\mathscr{I} = -i\pi^2 \left(\frac{d}{2} - 1\right) \mu^{d-4} \pi \Big/ \sin\left[\left(\frac{d}{2} - 2\right)\pi\right] .$$

Analytically continuing to $d \to 4$, this is

$$\mathscr{I} \to -2i\pi^2 \left[\frac{1}{d-4} + \ln \mu + \cdots\right] , \tag{17.5.43}$$

where $\cdots$ denotes finite μ-independent terms. Here we have

$$L_A = \frac{g^2}{2\pi^2}\left(\frac{11}{12} C_1 - \frac{1}{3} C_2\right)\left(\frac{1}{d-4} + \ln \mu + \cdots\right) + O(g^4)$$

and so

$$g_R = g\left[1 - \frac{g^2}{4\pi^2}\left(\frac{11}{12} C_1 - \frac{1}{3} C_2\right)\left(\frac{1}{d-4} + \ln \mu + \cdots\right) + O(g^4)\right] . \tag{17.5.44}$$

Note that the ultraviolet divergence here takes a different form, but the dependence on the infrared cut-off μ is the same. Eq. (17.5.44) will provide an important input to our discussion of asymptotic freedom in Section 18.7.

Problems

1. Carry out the proof of renormalizability given in Section 17.2, including elementary scalar fields in the Lagrangian.

2. Carry out the quantization of a non-Abelian gauge theory in background field gauge, using the BRST method of quantization discussed at the end of Section 17.2.

3. Derive the relation (17.5.44) between the renormalized and unrenormalized gauge couplings by calculating the terms in $\Gamma^{(1\text{ loop})}$ that are quadratic in a spacetime-dependent gauge field.

4. Calculate the one-loop relation between the renormalized and unrenormalized gauge couplings in a gauge theory containing elementary scalar fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

References

1. J. Zinn-Justin, in *Trends in Elementary Particle Theory — International Summer Institute on Theoretical Physics in Bonn 1974* (Springer-Verlag, Berlin, 1975).

2. Before the advent of BRST symmetry, proofs of the renormalizability of non-Abelian gauge theories were based directly on the Slavnov–Taylor identities for gauge transformations; see B. W. Lee and J. Zinn-Justin, *Phys. Rev.* **D5**, 3121, 3137 (1972); *Phys. Rev.* **D7**, 1049 (1972); G. 't Hooft and M. Veltman, *Nucl. Phys.* **B50**, 318 (1972); B. W. Lee, *Phys. Rev.* **D9**, 933 (1974). The original proof of renormalizability based on BRST symmetry was given by C. Becchi, A. Rouet, and R. Stora, *Commun. Math. Phys.* **42**, 127 (1975); in *Renormalization Theory — Proceedings of the International School of Mathematical Physics at Erice, August 1975*, eds. G. Velo and A.S. Wightman (D. Reidel, Dordrecht, 1976): pp. 269–97, 299–343. The proof given here follows the general outline of J. Zinn-Justin, Ref. 1; B. W. Lee, in *Methods in Field Theory*, eds. R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1976): pp. 79–139.

3. The point of view and presentation here is based on that of J. Gomis and S. Weinberg, *Nucl. Phys.* **B469**, 475 (1996). For earlier use of these methods, see B. L. Voronov and I. V. Tyutin, *Theor. Math. Phys.* **50**, 218 (1982); **52**, 628 (1982); B. L. Voronov, P. M. Lavrov, and I. V. Tyutin, *Sov. J. Nucl. Phys.* **36**, 292 (1982); P. M. Lavrov and I. V. Tyutin *Sov. J. Nucl. Phys.* **41**, 1049 (1985); D. Anselmi, *Class. and Quant. Grav.* **11**, 2181 (1994); **12**, 319 (1995); M. Harada, T. Kugo, and K. Yamawaki, *Prog. Theor. Phys.* **91**, 801 (1994).

4. G. Barnich and M. Henneaux, *Phys. Rev. Lett.* **72**, 1588 (1994); G. Barnich, F. Brandt, and M. Henneaux, *Phys. Rev.* **51**, R143 (1995); *Commun. Math. Phys.* **174**, 57, 93 (1995); *Nucl. Phys.* **B455**, 357 (1995).

5. The background field gauge was introduced by B. S. De Witt, *Phys. Rev.* **162**, 1195, 1239 (1967). For the treatment of multi-loop effects, see G 't Hooft, in *Functional and Probabilistic Methods in Quantum Field Theory: Proceedings of the 12th Karpacz Winter School of Theoretical Physics* (Acta Universitatis Wratislavensis no. 38, 1975); B. S. De Witt, in *Quantum Gravity* II, eds. C. Isham, R. Penrose, and D. Sciama (Oxford University Press, Oxford, 1982); L. F. Abbott, *Nucl. Phys.* **B185**, 189 (1981).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.005

18

Renormalization Group Methods

The method of the renormalization group was originally introduced by Gell-Mann and Low[1] as a means of dealing with the failure of perturbation theory at very high energies in quantum electrodynamics. An n-loop contribution to an amplitude involving momenta of order q, such as the vacuum polarization $\Pi_{\mu\nu}(q)$, is found to contain up to n factors of $\ln(q^2/m_e^2)$ as well as a factor α^n, so perturbation theory will break down when $\alpha|\ln(q^2/m_e^2)|$ is large, even though the fine structure constant α is small. Even in a massless theory like a non-Abelian gauge theory we must introduce some scale μ to specify a renormalization point at which the renormalized coupling constants are to be defined, and in this case we encounter logarithms $\ln(E/\mu)$, so that perturbation theory may break down if $E \gg \mu$ or $E \ll \mu$, even if the coupling constant is small.

Fortunately, there is a modified version of perturbation theory that can often be used in such cases. The key idea of this approach consists in the introduction of coupling constants g_μ defined at a sliding renormalization scale μ — that is, a scale that is not related to particle masses in any fixed way. By then choosing μ to be of the same order of magnitude as the energy E that is typical of the process in question, the factors $\ln(E/\mu)$ are rendered harmless. We can then do perturbation theory as long as g_μ remains small. In particular, given the coupling constants defined at scale μ, we can use perturbation theory to calculate physical amplitudes at an energy $\mu + d\mu$, and use these to calculate the coupling constants defined at a renormalization scale $\mu + d\mu$. By integrating the resulting differential equation we can then relate the coupling constants at the scale of interest to the coupling constants as conventionally defined. (The name 'renormalization group' arose originally because one is concerned here with equations that describe how the appearance of a theory changes under a redefinition of the renormalized coupling constants, but it really has nothing to do with group theory.) The method of the renormalization group can also provide qualitative guidance regarding asymptotic behavior at very high or (in massless theories) at very low energy, even where the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

coupling constants at the scale of interest are too large to allow the use of perturbation theory.

Although the method of the renormalization group arose originally in connection with changes in the prescription used to define renormalized coupling constants, it has come to have a wider meaning. When we replace bare couplings and fields with renormalized couplings and fields defined in terms of matrix elements evaluated at a characteristic energy scale μ, the integrals over virtual momenta will be effectively cut off at energy and momentum scales of order μ. Thus as we change μ, we are in effect changing the scope of the degrees of freedom taken into account in our calculations. The lesson of the renormalization group, that in order to avoid large logarithms we should take μ to be of the order of the energy E typical of the process being studied, is a special case of a broader principle, that in order to do calculations at a given energy we should first get rid of the degrees of freedom of much higher energy.

There are various other ways to accomplish this. As we saw in Section 12.4, in the approach to the renormalization group pioneered by Wilson[2] one introduces a *finite* explicit cut-off accompanied by a change in the parameters of the theory designed to keep physical quantities cut-off-independent. This approach requires introduction of an infinite number of interaction types, all those allowed by the symmetries of the theory, and is therefore not particularly convenient in dealing with theories that are actually renormalizable, like quantum electrodynamics (although, as discussed in Section 12.3, quantum electrodynamics is today regarded as only a very good approximation to a non-renormalizable theory in which the higher-dimensional interactions are suppressed by negative powers of some very large mass.) Where the cut-off is imposed by quantizing a gauge theory on a finite spacetime lattice, the Wilson approach has the advantage that calculations can be done while maintaining manifest gauge invariance (the volume of the gauge group equaling the volume of the global symmetry group times the number of lattice sites), but it has the disadvantage of not maintaining manifest Lorentz or rotational invariance. In any case much of the formalism of the renormalization group remains the same whatever approach is used to eliminate the high-energy degrees of freedom.

18.1 Where do the Large Logarithms Come From?

Let us first consider how large logarithms can arise at very high energies. Consider a physical amplitude or cross section or other rate parameter $\Gamma(E, x, g, m)$, that depends on an over-all energy scale E, on various angles and energy ratios collectively called x, on various dimensionless coupling

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

constants collectively called g, and on various masses collectively called m. If Γ has dimensionality $[\text{mass}]^D$ (as, for instance, a cross section would have $D = -2$) then simple dimensional analysis tells us that

$$\Gamma(E, x, g, m) = E^D \Gamma\left(1, x, g, \frac{m}{E}\right). \tag{18.1.1}$$

We might expect that in the limit $E \to \infty$ such an amplitude would behave as a simple power

$$\Gamma(E, x, g, m) \longrightarrow E^D \Gamma(1, x, g, 0).$$

But this is not what is found. Instead, in perturbation theory calculations the factor E^D is found to be accompanied by powers of $\ln(E/m)$, which invalidate this simple power-law behavior.

Clearly, powers of $\ln(E/m)$ can enter as $E \to \infty$ with fixed m only if the amplitude Γ at *fixed energy* E becomes singular as $m \to 0$. There are two classes of such mass singularities, one which is simply eliminated by calculating the right sort of amplitude or rate constant, the other of which requires a change in our renormalization procedure.

Zero-mass singularities of the first sort arise from a confluence of poles of propagators on the mass shells of the corresponding particles. For instance, suppose that a Feynman diagram has an incoming line with total four-momentum p^μ, attached at a vertex to internal lines of mass $m_1, m_2, \ldots, m_n$. According to the arguments of Chapter 10, the corresponding Feynman diagram will have a cut running along the negative real p^2 axis, from $p^2 = -(m_1 + \ldots + m_n)^2$ to $-\infty$. This does not lead to singularities if the external line is for a stable particle, with a mass $M < m_1 + \ldots + m_n$, because then $p^2 = -M^2$ is off the cut. However, when $M, m_1, \ldots, m_n$ all go to zero, the value of $-p^2$ on the mass shell and the branch point at the tip of this cut move together to join at $p^2 = 0$, producing a singularity.

This suggests that we can avoid the infrared divergences at $m = 0$ by simply staying off the mass shell, as, for instance, by letting p^2 for all external lines go to $+\infty$ along with all energy variables. We would then have to employ dispersion relations or some other technique of analytic continuation to use the results for the behavior of Feynman amplitudes in this limit to tell us anything about S-matrix elements. Often this continuation is unnecessary because we are interested not in on-shell S-matrix elements but rather in the matrix elements of currents carrying momenta q unrelated to any masses. For instance, the vacuum polarization function $\pi(q^2)$ defined in Section 10.5 is free of zero-mass singularities of the first sort except for $q^2 < 0$.

Another approach to the elimination of mass singularities of the first type is suggested by the observation that zero-mass singularities typically occur if we try to calculate a cross section that becomes unmeasurable

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

in the limit $m \to 0$. For instance, in quantum electrodynamics the cross section for any process involving definite numbers of electrons and photons becomes infrared-divergent in the limit $m_e \to 0$, even if we sum over unlimited numbers of soft photons, because for $m_e \to 0$ it is impossible to distinguish an electron from a jet of electrons, positrons, and photons with total charge $-e$, all moving in the same direction at the same speed. As shown in Chapter 13, such infrared divergences can be cured by considering only suitably integrated cross sections, which *would* be measurable for $m_e \to 0$. For instance, instead of trying to calculate the cross section for a specific Compton scattering process, we would calculate the cross section for the scattering of a jet with total charge $-e$ with another of total charge zero into two other such jets, plus soft photons. Such inclusive rates or cross sections, which remain finite when all masses vanish, are known as 'infrared safe.'

Our troubles are not over. Even where we avoid infrared divergences by integrating over cross sections or staying off the mass shell, the resulting integrated cross sections or off-shell amplitudes for energies E contain mass singularities of a second type, leading to factors $\ln(E/m)$ that invalidate the naive power-law behavior suggested by dimensional analysis. The reason can be traced to the fact that renormalized coupling constants are conventionally defined in terms of amplitudes that become infrared-divergent when all masses vanish. For instance, consider the theory of a real scalar field with Lagrangian density

$$\mathscr{L} = -\frac{1}{2}\,\partial_\lambda\phi\partial^\lambda\phi - \frac{1}{2}\,m^2\phi^2 - \frac{1}{24}\,g\,\phi^4. \tag{18.1.2}$$

To one-loop order, the invariant elastic scattering amplitude for a scattering process with initial four-momenta p_1, p_2, and final four-momenta p_1', p_2', is given by Eq. (12.2.24) as

$$A = g - \frac{g^2}{32\pi^2}\int_0^1 dx\left\{\ln\left(\frac{\Lambda^2}{m^2 - sx(1-x)}\right) + \ln\left(\frac{\Lambda^2}{m^2 - tx(1-x)}\right) + \ln\left(\frac{\Lambda^2}{m^2 - ux(1-x)}\right) - 3\right\} + O(g^3), \tag{18.1.3}$$

where s, t, and u are the Mandelstam variables

$$s = -(p_1 + p_2)^2\,, \quad t = -(p_1 - p_1')^2\,, \quad u = -(p_1 - p_2')^2$$

and Λ is an ultraviolet cut-off. This has no zero-mass singularities as long as we keep s, t, and u away from the positive real axis, and in particular if they all go to $-\infty$ (which violates the mass shell condition $s+t+u = 4m^2$.) Of course, the amplitude depends on the cut-off Λ as well as on m, so even though there are no zero-mass singularities, we do not find the result $A \to$ constant that would be found on the basis of naive scaling arguments

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

in the limit as s, t, and u go to $-\infty$. The dependence on the cut-off can be buried by renormalization; we replace the bare coupling g with a renormalized coupling g_R, defined as the value of A at some convenient renormalization point. For instance, we might take

$$\begin{aligned} g_R &\equiv A(s = t = u = 0) \\ &= g - \frac{3g^2}{32\pi^2}\left\{\ln\frac{\Lambda^2}{m^2} - 1\right\} + O(g^3). \end{aligned} \tag{18.1.4}$$

Then (18.1.3) becomes

$$\begin{aligned} A = g_R + \frac{g_R^2}{32\pi^2}\int_0^1 dx\Big\{ &\ln\left(1 - \frac{sx(1-x)}{m^2}\right) \\ &+ \ln\left(1 - \frac{tx(1-x)}{m^2}\right) + \ln\left(1 - \frac{ux(1-x)}{m^2}\right)\Big\} + O(g_R^3). \end{aligned} \tag{18.1.5}$$

(We can freely replace g^2 with g_R^2 in the second term, because the difference is only of order g^3.) This is free of ultraviolet divergences, but it now has a singularity at $m = 0$, even where s, t, and u are all kept negative. In consequence, where s, t, and u all go to $-\infty$, we again find an asymptotic behavior in disagreement with expectations based on naive scaling

$$A \to g_R + \frac{g_R^2}{32\pi^2}\left\{\ln\left(\frac{-s}{m^2}\right) + \ln\left(\frac{-t}{m^2}\right) + \ln\left(\frac{-u}{m^2}\right) - 6\right\}. \tag{18.1.6}$$

(Much the same happens with any other 'natural' definition of the renormalized coupling; for instance, we might define g_R as the value of A at the on-shell symmetrical point $s = t = u = 4m^2/3$, and would recover the same asymptotic behavior as in Eq. (18.1.6), except that -6 would be replaced with some other numerical constant.) It is clear that the zero-mass singularity that is encountered when A is expressed in terms of g_R arises entirely from the $\ln m^2$ term in the formula (18.1.4) for the renormalized coupling g_R in terms of the bare coupling g.

There are in addition other zero-mass singularities that are encountered when we calculate matrix elements of operators (such as off-shell Feynman amplitudes) rather than integrals of cross sections. These are due to the necessity of renormalizing these operators as well as coupling constants. For instance, suppose that in the scalar field theory with Lagrangian (18.1.2), we wish to calculate some matrix element $\langle\beta|\mathcal{O}(p)|\alpha\rangle$ of the operator

$$\mathcal{O}(p) \equiv \int d^4x\, e^{-ip\cdot x}\phi^2(x). \tag{18.1.7}$$

In Feynman diagram terms, this corresponds to inserting a vertex in which two internal ϕ-lines come together, and through which flows a total four-momentum p in diagrams for the transition $\alpha \to \beta$. (See Figure 18.1.)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

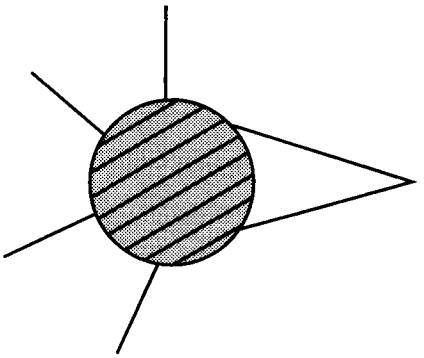

Figure 18.1. Momentum-space Feynman diagrams for the matrix element of the operator $\int d^4x \exp(-ip \cdot x)\ \phi^2(x)$ in the theory of an elementary scalar field $\phi(x)$. The cross-hatched disk represents the sum of diagrams with the indicated external lines. Apart from the pair of external lines that meet in a ϕ^2 vertex, the other external lines attached to the disk represent the particles in the initial and final states between which the matrix element is evaluated.

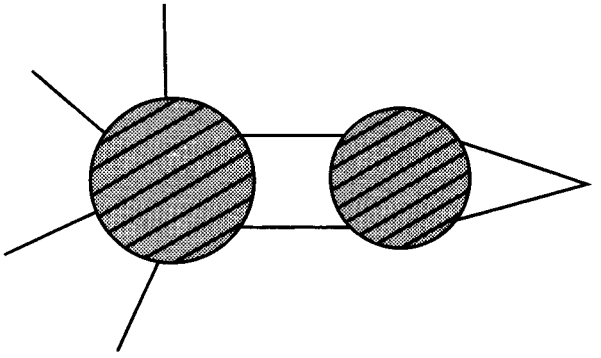

Figure 18.2. A class of Feynman diagrams for the matrix element of the operator $\int d^4x \exp(-ip \cdot x)\ \phi^2(x)$ that exhibit an ultraviolet divergence. The notation is the same as in Figure 18.1.

Ultraviolet divergences arise from a class of diagrams in which this new vertex is part of a subdiagram that is connected to the rest of the graph by just two ϕ-lines. (See Figure 18.2.) Dimensional analysis shows that the subgraph would be convergent if connected to the rest of the diagram by more than two ϕ-lines, because the interaction ϕ^2 has dimensionality $+2$, and hence a subdiagram in which this vertex is connected to the rest of the graph by $n > 2$ lines has dimensionality $4 - 2 - n < 0$. The divergent part of this subdiagram is just a logarithmically divergent constant, so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

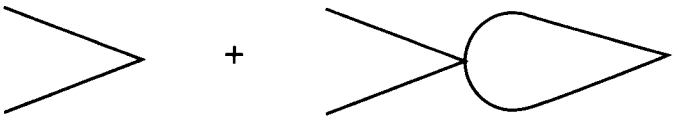

Figure 18.3. The divergent part of the diagram in Figure 18.2, to one-loop order.

the matrix elements of ϕ^2 can be made finite* by multiplying ϕ^2 with a suitable divergent constant Z_{ϕ^2}. To order g^2, the relevant subdiagram is given by the diagrams of Figure 18.3, and hence contributes to matrix elements of $\mathcal{O}(p)$ a divergent factor

$$F(p) = 1 + \frac{1}{2}\left[-i(2\pi)^4 g\right]\left[\frac{-i}{(2\pi)^4}\right]^2 \int \frac{d^4k}{\left[k^2 + m^2 - i\epsilon\right]\left[(p-k)^2 + m^2 - i\epsilon\right]}. \tag{18.1.8}$$

Combining denominators, rotating the k^0 integration contour, and imposing an ultraviolet cut-off Λ gives a result, for $\Lambda \to \infty$

$$F(p) = 1 - \frac{g}{32\pi^2}\int_0^1 dx \left[\ln\left(\frac{\Lambda^2}{m^2 + p^2 x(1-x)}\right) - 1\right] + O(g^2). \tag{18.1.9}$$

This has no zero-mass singularity (as long as we keep p^2 positive), but of course it does depend on the cut-off. This logarithmic divergence is eliminated by defining a renormalized ϕ^2 operator

$$(\phi^2)_R = N^{(\phi^2)}\phi^2 \tag{18.1.10}$$

with $N^{(\phi^2)}$ chosen so that $N^{(\phi^2)}F(p)$ has some definite finite value at some definite renormalization point. For instance, we could define the renormalized ϕ^2 operator so that

$$N^{(\phi^2)}F(0) = 1\,, \tag{18.1.11}$$

in which case

$$N^{(\phi^2)} = 1 + \frac{g}{32\pi^2}\left(\ln\frac{\Lambda^2}{m^2} - 1\right) + O(g^2). \tag{18.1.12}$$

The matrix elements of the renormalized operator $(\phi^2)_R$ then contain a

* In making this argument it was assumed that any divergences arising from subsubdiagrams containing the ϕ^2 vertex which are attached to the rest of the subdiagram by just two ϕ-lines are eliminated in the same way.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

factor

$$F_R(p) \equiv N^{(\phi^2)}F(p) = 1 + \frac{g}{32\pi^2}\int_0^1 dx\ \ln\left(1 + \frac{p^2x(1-x)}{m^2}\right) + O(g^2). \tag{18.1.13}$$

This is finite for all $p^2 > 0$ and $m^2 > 0$, but it now contains an infrared singularity for $m \to 0$, corresponding to large logarithms in the asymptotic behavior when $p^2 \to +\infty$. Of course in order to eliminate the cut-off in higher-order calculations we would have to introduce a renormalized coupling constant as well as a renormalized ϕ^2 operator, and we would encounter logarithms arising from both sources.

Similar renormalization factors are needed for any sort of operator, not just $\phi^2(x)$. In particular, taking a matrix element of one of the elementary fields ψ of a theory introduces ultraviolet divergences that arise from radiative corrections to the corresponding propagator. As we saw in Chapter 12, these infinities can be cancelled by working with a renormalized field ψ_R:

$$\psi_R = N^{(\psi)}\psi \tag{18.1.14}$$

with $N^{(\psi)}$ chosen to make the matrix element of ψ_R between a one-particle state and the vacuum the same as for a conventionally normalized field in the absence of interactions. This is related to the usual Z-factor of renormalization theory by

$$Z^{(\psi)} = |N^{(\psi)}|^{-2}. \tag{18.1.15}$$

For instance, let's recall our earlier results for the renormalization of the photon field in spinor quantum electrodynamics.** In this case, the renormalized electromagnetic field A_R^μ is conventionally written in terms of the 'bare' field A_B^μ as

$$A_R^\mu = Z_3^{-1/2}A_B^\mu$$

with Z_3 given by Eq. (11.2.21) as

$$Z_3 = 1 - \frac{e^2}{12\pi^2}\ln\left(\frac{\Lambda^2}{m^2}\right) + O(e^4). \tag{18.1.16}$$

This has a zero-mass singularity, which will affect the asymptotic behavior of matrix elements of the renormalized photon field. In particular,

** In the scalar field theory with ϕ^4 interaction used as an example above, the lowest-order terms in $N^{(\phi)}$ arise from a two-loop graph, so it will not be convenient to use this theory to illustrate the calculation of the N-factors.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Eq. (11.2.22) gives the self-energy function of the renormalized electromagnetic field as

$$\pi(q^2) = \frac{e_R^2}{2\pi^2} \int_0^1 dx\, x(1-x) \ln \left[1 + \frac{q^2 x(1-x)}{m^2} \right] + O(e_R^4) . \quad (18.1.17)$$

This has a singularity at $m = 0$, and so has large logarithms in its asymptotic behavior: for $q^2 \to +\infty$

$$\pi(q^2) \longrightarrow \frac{e_R^2}{2\pi^2} \left[\frac{1}{6} \ln \frac{q^2}{m^2} - \frac{5}{18} \right] + O(e_R^4) . \quad (18.1.18)$$

Electrodynamics has the special feature that the constant Z_3 that appears in the renormalization of the electromagnetic field also appears in the renormalization of the electric charge:

$$e_R = Z_3^{1/2} e_{\text{BARE}} , \quad (18.1.19)$$

but this is not generally the case. Renormalization group techniques were first applied in quantum electrodynamics, but the scalar field theory discussed here gives a more typical illustration of these methods, with separate renormalizations for fields and couplings.

18.2 The Sliding Scale

We have seen in the previous section that the large logarithms that appear at high energy in suitably integrated cross sections or off-shell Feynman amplitudes can be traced to the prescription used to define renormalized coupling constants and operators. The central idea of the renormalization group method is to change this prescription.

Suppose we find some way of defining a new kind of renormalized coupling constant $g(\mu)$, that depends on a sliding energy scale μ, but that (at least for $\mu \gg m$) has no dependence on the scale m of the masses of the theory. Then suitably integrated cross sections or other infrared-safe rate parameters may be expressed as functions of g_μ and μ instead of g_R. By dimensional analysis such functions may be written as

$$\Gamma(E, x, g_\mu, m, \mu) = E^D \Gamma \left(1, x, g_\mu, \frac{m}{E}, \frac{\mu}{E} \right) . \quad (18.2.1)$$

(Our notation here is the same as in Section 18.1; in particular, x stands for all dimensionless angles, energy ratios, etc. on which Γ may depend.) Since μ is a completely arbitrary renormalization scale, we can choose $\mu = E$, in which case Eq. (18.2.1) reads

$$\Gamma(E, x, g_\mu, m, \mu) = E^D \Gamma \left(1, x, g_E, \frac{m}{E}, 1 \right) . \quad (18.2.2)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

This now has no zero-mass singularities because g_E does not depend on m for $m \ll E$, so there are no large logarithms, and we can use perturbation theory to calculate Γ in terms of g_E as long as g_E itself remains sufficiently small. In particular, in any finite order of perturbation theory Γ has the asymptotic behavior, for $E \gg m$,

$$\Gamma(E, x, g_\mu, m, \mu) \longrightarrow E^D \Gamma(1, x, g_E, 0, 1) . \tag{18.2.3}$$

(Non-perturbative corrections are considered in Section 18.4.)

It remains to calculate g_E. For instance, in the scalar field theory with Lagrangian (18.1.2), we may define g_μ in terms of the value of the scattering amplitude at a renormalization point $s = t = u = -\mu^2$:

$$\begin{aligned} g_\mu &\equiv A(s = t = u = -\mu^2) \\ &= g - \frac{3g^2}{32\pi^2} \int_0^1 dx \left\{ \ln\left(\frac{\Lambda^2}{m^2 + \mu^2 x(1-x)} \right) - 1 \right\} + O(g^3) \end{aligned} \tag{18.2.4}$$

or, in terms of the conventional renormalized coupling (18.1.4),

$$g_\mu = g_R + \frac{3g_R^2}{32\pi^2} \int_0^1 dx \ln\left(1 + \frac{\mu^2 x(1-x)}{m^2} \right) + O(g_R^3) . \tag{18.2.5}$$

But this formula is reliable only if the correction term is smaller than g_R; that is, only if $|g_R \ln(\mu/m)| \ll 1$. If this were the case for $\mu \simeq E$, then we would not need the methods of the renormalization group; ordinary perturbation theory would be good enough.

Instead of using formulas like (18.2.5) directly for large μ, we must instead proceed in stages: g_μ may be calculated in terms of g_R as long as μ/m is not much larger than unity; then $g_{\mu'}$ may be calculated in terms of g_μ as long as μ'/μ is not much larger than unity; and so on, up to g_E. Instead of discrete stages, this may also be done continuously. Dimensional analysis tells us that the relation between $g_{\mu'}$ and g_μ takes the form

$$g_{\mu'} = G(g_\mu, \mu'/\mu, m/\mu) . \tag{18.2.6}$$

Differentiating with respect to μ' and then setting $\mu' = \mu$ yields the differential equation

$$\mu \frac{d}{d\mu} g_\mu = \beta\left(g_\mu, \frac{m}{\mu} \right) , \tag{18.2.7}$$

where

$$\beta\left(g_\mu, \frac{m}{\mu} \right) \equiv \left[\frac{\partial}{\partial z} G(g_\mu, z, m/\mu) \right]_{z=1} . \tag{18.2.8}$$

There are no zero-mass singularities here, so for $\mu \gg m$ the differential

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

equation becomes simply

$$\mu \frac{d}{d\mu} g_\mu = \beta(g_\mu, 0) \equiv \beta(g_\mu), \tag{18.2.9}$$

which is often known as the *Callan–Symanzik equation.*[1] We are to calculate g_E by integrating the differential equation (18.2.9), with an initial value g_M at some scale $\mu = M$, chosen in practice large enough so that for $\mu \geq M$ we can neglect the masses m compared with μ, but small enough so that large logarithms $\ln(M/m)$ do not prevent us from using perturbation theory to calculate g_M in terms of the conventional renormalized coupling constant g_R. The solution may be formally written

$$\ln(E/M) = \int_{g_M}^{g_E} dg/\beta(g) \tag{18.2.10}$$

as long as $\beta(g)$ does not vanish between g_M and g_E.

The results of the previous paragraph do not rely on perturbation theory, but we usually need to use perturbation theory to calculate the functions G and β. As an example, suppose we calculate $g_{\mu'}$ in the scalar field theory with interaction Hamiltonian density $g\phi^4/24$, renormalizing by expressing g in terms of g_μ rather than g_R. Following the same procedure that led to Eq. (18.2.5), this gives

$$g_{\mu'} = g_\mu - \frac{3g_\mu^2}{32\pi^2} \int_0^1 dx \ln\left(\frac{m^2 + \mu^2 x(1-x)}{m^2 + \mu'^2 x(1-x)}\right) + O(g_\mu^3).$$

Then Eq. (18.2.8) gives

$$\beta\left(g_\mu, \frac{m}{\mu}\right) = +\frac{3g_\mu^2}{16\pi^2} \int_0^1 dx \frac{\mu^2 x(1-x)}{m^2 + \mu^2 x(1-x)} + O(g_\mu^3). \tag{18.2.11}$$

For $\mu \gg m$, this is

$$\beta(g_\mu) = \frac{3g_\mu^2}{16\pi^2} + O(g_\mu^3). \tag{18.2.12}$$

In next order, the beta-function for $\mu \gg m$ is[3]

$$\beta(g_\mu) = g_\mu \left[3\left(\frac{g_\mu}{16\pi^2}\right) - \frac{17}{3}\left(\frac{g_\mu}{16\pi^2}\right)^2 + \cdots\right].$$

If we are content with the one-loop approximation the calculation of $\beta(g)$ can be done even more easily. In order to avoid large radiative corrections in matrix elements at energies of order μ, we must write the bare coupling g in terms of a finite renormalized coupling g_μ, as

$$g = g_\mu + B(g_\mu) \ln \frac{\Lambda}{\mu} + \cdots. \tag{18.2.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

For instance, from Eq. (18.1.3) we could have immediately read off that the $\ln \Lambda$ terms in g have coefficient

$$B(g) = -\frac{3}{2}\left[-i(2\pi)^4 g^2\right]\left[\frac{-i}{(2\pi)^4}\right]^2 2\pi^2 i = \frac{3g^2}{16\pi^2}. \tag{18.2.14}$$

The unrenormalized coupling is of course independent of μ, so to lowest order

$$\mu \frac{d}{d\mu} g_\mu - B(g_\mu) = 0$$

and so to lowest order

$$\beta(g) = B(g). \tag{18.2.15}$$

With $B(g)$ given by Eq. (18.2.14), this agrees with our previous result (18.2.12) for $\beta(g)$ in the ϕ^4 scalar field theory.

Instead of using a simple ultraviolet cut-off, which can interfere with gauge invariance, it is often more convenient to deal with ultraviolet divergences by use of dimensional regularization. For a spacetime dimensionality $d < 4$, we find in place of $\ln(\Lambda/\mu)$ the convergent integral

$$\int_\mu^\infty k^{d-4}\frac{dk}{k} = \frac{\mu^{d-4}}{4-d} \underset{d \to 4}{\longrightarrow} \left[\frac{1}{4-d} - \ln \mu\right].$$

Thus instead of eliminating the cut-off dependence by writing the unrenormalized coupling constant as in (18.2.13), we instead write

$$g = g_\mu + B(g_\mu)\left[\frac{1}{4-d} - \ln \mu\right] \tag{18.2.16}$$

with the same function $B(g_\mu)$ as before. Thus in order to calculate $\beta(g_\mu)$, all we need to do is to pick out the coefficient of the singular factor $1/(4-d)$ in the renormalized coupling. This argument is extended to all orders of perturbation theory in Section 18.6.

As long as g_μ is sufficiently small in the scalar field theory with Lagrangian (18.1.2), the solution of Eqs. (18.2.9) and (18.2.12) can be well approximated by

$$g_\mu = -\frac{16\pi^2}{3\ln(\mu/M)}, \tag{18.2.17}$$

where M is an integration constant. This expression illustrates a common aspect of renormalization group calculations, that dimensionless couplings like g_R become replaced with parameters like M that have the dimensionality of mass. The value of M may be related to g_R by comparing the solution (18.2.17) with the behavior of the coupling for values of μ that are large enough to allow us to use approximations based on $\mu \gg m$, but

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

small enough so that $|g_R \ln(\mu/m)| \ll 1$, where (18.2.5) gives

$$g_\mu \simeq g_R + \frac{3g_R^2}{16\pi^2} \ln\left(\frac{\mu}{m}\right) . \tag{18.2.18}$$

In this way, we find

$$M \simeq m \exp\left(\frac{16\pi^2}{3g_R}\right) , \tag{18.2.19}$$

so that Eq. (18.2.17) may be put in a more conventional form

$$g_\mu = g_R \left[1 - \frac{3g_R}{16\pi^2} \ln \frac{\mu}{m}\right]^{-1} . \tag{18.2.20}$$

To repeat, this is valid provided g_μ is small, even if $|g_R \ln(\mu/m)|$ is of order unity, so it represents a significant improvement over the perturbative result (18.2.18). Of course, the condition that g_μ should be small will become violated when $g_R \ln(\mu/m)$ is sufficiently close to the critical value $16\pi^2/3$. But at least Eq. (18.2.20) makes the unequivocal prediction that g_E becomes large enough to invalidate perturbation theory at some energy E *below* the critical value (18.2.19).

In calculating off-shell matrix elements of operators instead of integrated cross sections, we also need to take into account the N-factors that appear in the definition of the renormalized operators whose matrix elements are finite. We saw in the previous section that if these N-factors are defined in a conventional way (say, so that the correction factors produced by the divergent subgraphs are cancelled when the operator carries zero four-momentum, or is a field on its mass shell) then the formula for the N-factor involves zero-mass singularities as in Eq. (18.1.12) or Eq. (18.1.16), resulting in large logarithms at energies $E \gg m$. The cure is to define renormalization constants $N_\mu^{(\mathcal{O})}$ at a sliding scale μ, so that in matrix elements of the renormalized operator

$$\mathcal{O}_\mu = N_\mu^{(\mathcal{O})} \mathcal{O} \tag{18.2.21}$$

the correction factor produced by divergent subgraphs containing operator $\mathcal{O}$ are cancelled at a renormalization point characterized by four-momenta of order μ. If M_R is a matrix element of operators that are conventionally renormalized, and M is one in which the operators are renormalized as in Eq. (18.2.21), then for any μ

$$M_R = \left[\prod_{\mathcal{O}} \left(N^{(\mathcal{O})} \Big/ N_\mu^{(\mathcal{O})}\right)\right] M(E, x, g_\mu, m, \mu) . \tag{18.2.22}$$

We can again use dimensional analysis (assuming M has dimensionality

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

D) and set $\mu = E$, to write this as

$$M_R = E^D \left[\prod_{\mathcal{O}} \left(N^{(\mathcal{O})} \Big/ N_E^{(\mathcal{O})}\right)\right] M\left(1, x, g_E, \frac{m}{E}, 1\right) . \tag{18.2.23}$$

Thus to find the high energy behavior of an off-shell amplitude M_R, we need to know how N_μ varies with the renormalization scale μ.

For any two renormalization scales μ and μ', the renormalized operators $N_\mu^{(\mathcal{O})}\mathcal{O}$ and $N_{\mu'}^{(\mathcal{O})}\mathcal{O}$ both have finite matrix elements, so the ratio $N_{\mu'}^{(\mathcal{O})}/N_\mu^{(\mathcal{O})}$ must be cut-off independent. On dimensional grounds, this ratio must take the form

$$N_{\mu'}^{(\mathcal{O})} \Big/ N_\mu^{(\mathcal{O})} = G^{(\mathcal{O})}(g_\mu, \mu'/\mu, m/\mu) . \tag{18.2.24}$$

Differentiating with respect to μ' and then setting $\mu' = \mu$ gives

$$\mu \frac{d}{d\mu} N_\mu^{(\mathcal{O})} = \gamma^{(\mathcal{O})}(g_\mu, m/\mu) N_\mu^{(\mathcal{O})} , \tag{18.2.25}$$

where

$$\gamma^{(\mathcal{O})}(g_\mu, m/\mu) \equiv \left[\frac{\partial}{\partial z} G^{(\mathcal{O})}(g_\mu, z, m/\mu)\right]_{z=1} . \tag{18.2.26}$$

The solution is

$$N_E^{(\mathcal{O})} \propto \exp\left[\int^E \gamma^{(\mathcal{O})}\left(g_\mu, \frac{m}{\mu}\right) \frac{d\mu}{\mu}\right] . \tag{18.2.27}$$

This is a useful result because the introduction of the sliding scale prevents the appearance of zero-mass singularities in $N_\mu^{(\mathcal{O})}$ and $N_{\mu'}^{(\mathcal{O})}$, and hence also in $G^{(\mathcal{O})}$ and $\gamma^{(\mathcal{O})}$. Hence as long as g_μ is small, there are no large logarithms that prevent the application of perturbation theory to calculate $\gamma^{(\mathcal{O})}$. Also, for $\mu \gg m$, $\gamma^{(\mathcal{O})}(g_\mu, m/\mu)$ has a smooth limit

$$\gamma^{(\mathcal{O})}(g_\mu) \equiv \gamma^{(\mathcal{O})}(g_\mu, 0) . \tag{18.2.28}$$

As an example, consider the operator $\mathcal{O} = \phi^2$ in the scalar field theory with interaction $g\phi^4/24$. Instead of renormalizing it so that the correction factor (18.1.9) is cancelled at $p^2 = 0$, we cancel it at a sliding scale $p^2 = \mu^2$, by introducing a new renormalized ϕ^2 operator $N_\mu^{(\phi^2)}\phi^2$, with

$$N_\mu^{(\phi^2)} \equiv F^{(\phi^2)}(\mu^2)^{-1} = 1 + \frac{g}{32\pi^2} \int_0^1 dx \left[\ln\left[\frac{\Lambda^2}{m^2 + \mu^2 x(1-x)}\right] - 1\right] + O(g^2) .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Then the function (18.2.24) for this operator is

$$G^{(\phi^2)}\left(g_\mu, \frac{\mu'}{\mu}, \frac{m}{\mu}\right) \equiv \frac{N_{\mu'}^{(\phi^2)}}{N_\mu^{(\phi^2)}} = 1 + \frac{g_\mu}{32\pi^2}\int_0^1 dx \ln\left[\frac{m^2+\mu^2 x(1-x)}{m^2+\mu'^2 x(1-x)}\right] + O(g_\mu^2).$$

(We can use g_μ here instead of g or $g_{\mu'}$ because the difference only affects the terms of order g_μ^2 or higher.) From Eq. (18.2.26), we have then

$$\gamma^{(\phi^2)}\left(g_\mu, \frac{m}{\mu}\right) = -\frac{g_\mu}{16\pi^2}\int_0^1 \frac{\mu^2 x(1-x)}{m^2+\mu^2 x(1-x)}dx + O(g_\mu^2)$$

or for $\mu \gg m$,

$$\gamma^{(\phi^2)}(g_\mu) = -\frac{g_\mu}{16\pi^2} + O(g_\mu^2). \tag{18.2.29}$$

Another good example is provided by the N-factor associated with the renormalization of the electromagnetic field in quantum electrodynamics. Recall that the photon propagator can be made finite for all momenta by evaluating it for renormalized electromagnetic fields, or equivalently by multiplying the propagator of the unrenormalized field by Z_3^{-1}:

$$\tilde{\Delta}_{\rho\sigma}(q) = Z_3^{-1}\Delta_{\rho\sigma}(q). \tag{18.2.30}$$

Eq. (10.5.17) shows that this renormalized propagator may be written

$$\tilde{\Delta}_{\rho\sigma}(q) = \frac{\eta_{\rho\sigma}}{[q^2 - i\epsilon][1-\pi(q^2)]} + q_\rho q_\sigma\text{-terms}. \tag{18.2.31}$$

Suppose we instead define a renormalized field $N_\mu^{(A)}A^\rho$, whose propagator has a term proportional to $\eta_{\rho\sigma}/[q^2 - i\epsilon]$ with a coefficient that is equal to unity at a sliding renormalization scale $q^2 = \mu^2$. For this purpose, we must clearly take

$$N_\mu^{(A)} = Z_3^{-1/2}[1-\pi(\mu^2)]^{1/2}. \tag{18.2.32}$$

Using Eq. (11.2.22), the function (18.2.24) is then

$$\begin{aligned} G^{(A)}(g_\mu, \mu'/\mu, m/\mu) &= \left[\frac{1-\pi(\mu'^2)}{1-\pi(\mu^2)}\right]^{1/2} \\ &= 1 - \frac{e_\mu^2}{4\pi^2}\int_0^1 dx\, x(1-x)\ln\left[\frac{m^2+\mu'^2 x(1-x)}{m^2+\mu^2 x(1-x)}\right] \\ &\quad + O(e_\mu^4) \end{aligned} \tag{18.2.33}$$

and so Eq. (18.2.26) gives

$$\gamma^{(A)}(e_\mu, m/\mu) = -\frac{e_\mu^2}{2\pi^2}\int_0^1 dx \frac{x^2(1-x)^2\mu^2}{m^2+\mu^2 x(1-x)} + O(e_\mu^4). \tag{18.2.34}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

As promised, this has a smooth limit for $\mu \gg m$

$$\gamma^{(A)}(e_\mu) \equiv \gamma^{(A)}(e_\mu, 0) = -\frac{e_\mu^2}{12\pi^2} + O(e_\mu^4). \tag{18.2.35}$$

As already mentioned, electrodynamics is a special case, because the renormalization constant $Z_3^{-1/2}$ in the definition of the renormalized electromagnetic field is just the reciprocal of the constant used to define the renormalized electric charge of the electron: $e_R = Z_3^{1/2} e$. The natural definition of the renormalized electric charge at a sliding scale μ is then

$$e_\mu = N_\mu^{(A)-1} e = Z_3^{-1/2} N_\mu^{(A)-1} e_R\,, \tag{18.2.36}$$

so that e_μ times the field $N_\mu^{(A)} A^\rho$ renormalized at scale μ is independent of μ. From Eq. (18.2.25) we see that the function $\beta(e)$ which gives the μ dependence of e_μ according to Eq. (18.2.9) for $\mu \gg m$ has the value

$$\beta(e) = -e\gamma^{(A)}(e) = \frac{e^3}{12\pi^2} + O(e^5)\,. \tag{18.2.37}$$

Using an earlier calculation[3a] of the fourth-order term in the vacuum polarization function $\pi(q^2)$, Gell-Mann and Low were able also to give the term in $\beta(e)$ of next order in e:

$$\beta(e) = \frac{e^3}{12\pi^2} + \frac{e^5}{64\pi^2} + O(e^7)\,. \tag{18.2.38}$$

In other words, the electric charge at a sliding scale μ satisfies the renormalization group equation

$$\mu \frac{d}{d\mu} e_\mu = \frac{e_\mu^3}{12\pi^2} + \frac{e_\mu^5}{64\pi^2} + O(e_\mu^7)\,. \tag{18.2.39}$$

This shows, for small e_μ, that e_μ *increases* with increasing μ.

We also need an initial condition. This is provided by the known value of the conventionally renormalized charge $e_R = Z_3^{1/2} e$, for which $\alpha \equiv e_R^2/4\pi = 1/137.036\ldots$. Eqs. (18.2.32) and (18.2.36) yield

$$\begin{aligned} e_R/e_\mu = Z_3^{1/2} N_\mu^{(A)} &= \sqrt{1 - \pi(\mu^2)} \\ &= 1 - \frac{e_R^2}{4\pi^2} \int_0^1 dx\, x(1-x) \ln\left[1 + \frac{\mu^2 x(1-x)}{m_e^2}\right] + O(e_R^4)\,. \end{aligned} \tag{18.2.40}$$

We need to match this with the solution of Eq. (18.2.39) at a value of μ which is large enough to justify the approximation $\mu \gg m_e$ in (18.2.39) but small enough so that the logarithm in Eq. (18.2.40) is still small enough compared with $4\pi^2/e_R^2$ to justify the use of perturbation theory. (For

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

instance, we might take μ to be of order 100 MeV.) For such values of μ, Eq. (18.2.40) gives

$$e_\mu \simeq e_R + \frac{e_R^3}{12\pi^2}\left[\ln\frac{\mu}{m_e} - \frac{5}{6}\right] . \tag{18.2.41}$$

On the other hand, the solution of Eq. (18.2.39) for e_μ small (keeping only the leading term on the right-hand side) is

$$e_\mu = \left[\text{constant} - \frac{\ln\mu}{6\pi^2}\right]^{-1/2} . \tag{18.2.42}$$

Comparing Eqs. (18.2.41) and (18.2.42) gives the solution

$$e_\mu = e_R\left[1 - \frac{e_R^2}{6\pi^2}\left(\ln\left(\frac{\mu}{m_e}\right) - \frac{5}{6}\right)\right]^{-1/2} . \tag{18.2.43}$$

Unlike Eq. (18.2.41), Eq. (18.2.43) is valid as long as $e_\mu^2/6\pi^2$ is small, whether or not $(e_R^2/6\pi^2)\ln(\mu/m_e)$ is small.

For instance, we have already seen in Section 11.3 that the leading fourth-order radiative correction to the magnetic moment of the muon can be obtained by multiplying the second-order (Schwinger) term (11.3.16) by the vacuum polarization function $\pi_e(k^2)$ at $k^2 \approx m_\mu^2$, which according to Eq. (18.2.40) is the same (to this order) as using $e_{m_\mu}^2/4\pi$ in place of α in the Schwinger term.

Another example: Experiments at high energy electron–positron colliders such as LEP at CERN or SLC at SLAC now study physical processes at energies of the order of the mass of the Z^0 particle, or 91 GeV. Eq. (18.2.43) shows that at these energies, radiative corrections in pure quantum electrodynamics should be calculated using a value for the fine structure constant which is not $\alpha = 1/137.036$ but rather

$$\frac{e^2(91\,\text{GeV})}{4\pi} = \frac{\alpha}{1 - 2(11.25)\alpha/3\pi} = \frac{1}{134.6} . \tag{18.2.44}$$

This is for a theory in which electrons are the only charged particles with masses below m_Z. In the real world there are many such particle types, and the effective fine structure constant[4] at m_Z is $(128.87 \pm 0.12)^{-1}$.

* * *

The sliding scale at which the coupling parameters of a theory are calculated may be the value of an external *field* rather than the momentum of an external line. In one of the early applications of the renormalization group method, Coleman and E. Weinberg[4a] considered the effective potential $V(\phi)$ for a spacetime-independent external scalar field ϕ. In the simple case where the field interacts with itself alone, their one-loop result

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

is given by Eq. (16.2.15). It is particularly interesting to consider this potential in the case where the renormalized mass m_R vanishes, where to one-loop order we may set $\mu^2(\phi)$ in the last term equal to $g_R\phi^2/2$, so that Eq. (16.2.15) here reads (with g a slightly redefined coupling):

$$V(\phi) = \lambda_R + \frac{g}{24}\phi^4 + \frac{g^2\phi^4 \ln \phi^2}{256\pi^2} \,. \tag{18.2.45}$$

This looks like at first sight as if for $g > 0$ the potential becomes less than λ_R for very small ϕ, so that the point $\phi = 0$ is a local maximum instead of a minimum, but for such small values of ϕ the third term is larger than the second, and the perturbation theory is obviously untrustworthy. Also, we would like to be able to argue that g must be positive in order for the potential to be bounded below at large fields, but Eq. (18.2.45) shows that however small g is, there is some sufficiently large ϕ where perturbation theory breaks down, and hence Eq. (18.2.45) cannot be trusted to tell us whether the potential is bounded below at large fields.

We can do much better by using a coupling constant defined at a sliding scale μ of field strength. Suppose we define a coupling g_μ by the condition that

$$V(\mu) = \lambda_R + \frac{g_\mu}{24}\mu^4 \,. \tag{18.2.46}$$

If we had used g_μ as the coupling parameter from the beginning, then in place of Eq. (18.2.45) we would have obtained

$$V(\phi) = \lambda_R + \frac{g_\mu}{24}\phi^4 + \frac{g_\mu^2\phi^4}{256\pi^2} \ln\left(\frac{\phi^2}{\mu^2}\right) \,, \tag{18.2.47}$$

which obviously satisfies Eq. (18.2.46).* The renormalization group equation for g_μ can be obtained from the condition that this effective potential

* Eq. (18.2.47) differs from what we would get from Eq. (18.2.45) by simply using Eq. (18.2.46) to express g in terms of g_μ, in that the last term is proportional to g_μ^2 rather than g^2. The difference is of higher order in g, but can become significant if $V(\phi)$ is evaluated at ϕ very different from μ, where large logarithms can compensate for powers of coupling. If we use g_μ as the coupling parameter from the beginning and take μ to be of order ϕ then no such large logarithms occur, and the approximation Eq. (18.2.47) is valid as long as g_μ remains small.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

is independent of μ,**

$$\mu \frac{dg_\mu}{d\mu} = \frac{3g_\mu^2}{16\pi^2} . \tag{18.2.48}$$

(Terms involving the derivative of g_μ^2 are dropped here, because they are of higher order in g_μ, and are hence negligible as long as g_μ is sufficiently small.) It is not a coincidence that this takes the same form as the renormalization group equation (18.2.9), (18.2.12), where μ was a renormalization momentum, because as we shall see in the next section the first two terms in the renormalization group equation are always independent of the way that we define the sliding scale. The solution of this equation is given by Eq. (18.2.17), in general with a different integration constant M. Hence by taking $\mu = \phi$ in Eq. (18.2.17), we now have

$$V(\phi) = \lambda_R - \frac{4\pi^2\phi^4}{9\ln(\phi^2/M^2)} , \tag{18.2.49}$$

$$g_\phi = -\frac{32\pi^2}{3\ln(\phi^2/M^2)} . \tag{18.2.50}$$

This result should be used with some care, because it is only valid where the coupling constant g_ϕ is small. The problem is not just that Eq. (18.2.49) loses its validity for ϕ near M; we also cannot integrate the renormalization group equation through the singularity at $\phi = M$, so a knowledge of g_ϕ on one side of this singularity tells us nothing about its behavior on the other side.

If g_ϕ is found to have a small *positive* value for some ϕ_0 then from Eq. (18.2.50) we know that $M > |\phi_0|$, so g_ϕ remains small and Eq. (18.2.49) is valid for $|\phi| < |\phi_0|$. This shows that the point $\phi = 0$ is a local *minimum* of $V(\phi)$, contrary to what Eq. (18.2.45) might have led us to suppose. This means that the vacuum which is invariant under the symmetry transformation $\phi \to -\phi$ is stable in this model, aside from the possibility of quantum mechanical barrier penetration. On the other hand, we do not know that Eq. (18.2.49) becomes valid for $|\phi|$ sufficiently large compared with M. The coupling might remain too large to use perturbation theory for all $|\phi| > M$, and even if it becomes small for some $|\phi| > M$, the potential might be given for such ϕ by Eq. (18.2.49) with a renormalization

** To eliminate all cut-off dependence, ϕ should be written in terms of a renormalized field $\phi_\mu = N_\mu^\phi \phi$, which gives $V(\phi_\mu)$ a μ dependence arising from the μ dependence of the renormalization constant N_μ^ϕ. This point is ignored here, because in the scalar field theory with interaction $\propto \phi^4$ the lowest-order graphs contributing to the μ dependence of N_μ^ϕ have two loops, so that to the order of the calculations presented here we can take $N_\mu^\phi = 1$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

scale $M' > \phi$, producing a second singularity. Thus in this case we cannot conclude that $V(\phi) \to -\infty$ for $|\phi| \to \infty$.

Similarly, if g_ϕ is found to have a small *negative* value for some ϕ_0 then we know that $M < |\phi_0|$, so g_ϕ remains small and Eq. (18.2.49) is valid for $|\phi| > |\phi_0|$. In this case we cannot conclude anything about the behavior of the potential for $|\phi| < M$, but here we *can* use Eq. (18.2.49) to see that $V(\phi) \to -\infty$ for $|\phi| \to \infty$, ruling out the possibility of any stable vacuum. Because we are here considering the limit $|\phi| \to \infty$, this conclusion holds also for scalar fields of mass $m > 0$, provided $m \ll |\phi_0|$. This is why it is necessary to assume that the ϕ^4 coupling (renormalized at any scale much larger than the scalar mass) is positive.

18.3 Varieties of Asymptotic Behavior

The renormalization group method provides useful insight into the types of possible asymptotic behavior that are encountered in quantum field theories, even in cases where the running coupling g_μ does not remain small enough to allow the use of perturbation theory. We will distinguish four different ways that g_μ may behave for $\mu \to \infty$, that correspond to four different shapes of the function $\beta(g)$ in theories with a single coupling constant. In the next section we will take up the case of theories with several independent couplings.

Let's first recall the results for $\beta(g)$ obtained for the two examples considered in the previous section. One of these is the scalar field theory with interaction $g\phi^4/24$, for which the beta-function for small g is

$$\beta(g) = \frac{3g^2}{16\pi^2} - \frac{17}{3}\frac{g^3}{(16\pi^2)^2} + O(g^4)\,. \tag{18.3.1}$$

The other is quantum electrodynamics. Instead of writing the beta-function here in the form (18.2.38), we will emphasize the similarities between this theory and the scalar field theory by writing

$$g \equiv e^2\,, \tag{18.3.2}$$

with $\beta(g_\mu)$ now understood to be $\mu\, dg_\mu/d\mu$, so that for small g

$$\begin{aligned}\beta(g) &= 2e\left[\frac{e^3}{12\pi^2} + \frac{e^5}{64\pi^2} + O(e^7)\right] \\ &= \frac{g^2}{6\pi^2} + \frac{g^3}{32\pi^2} + O(g^4)\,.\end{aligned} \tag{18.3.3}$$

Note that in both cases the physically allowed coupling constants fall in the range $g \geq 0$, where $\beta(g) \geq 0$ for small g. In electrodynamics this is simply

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

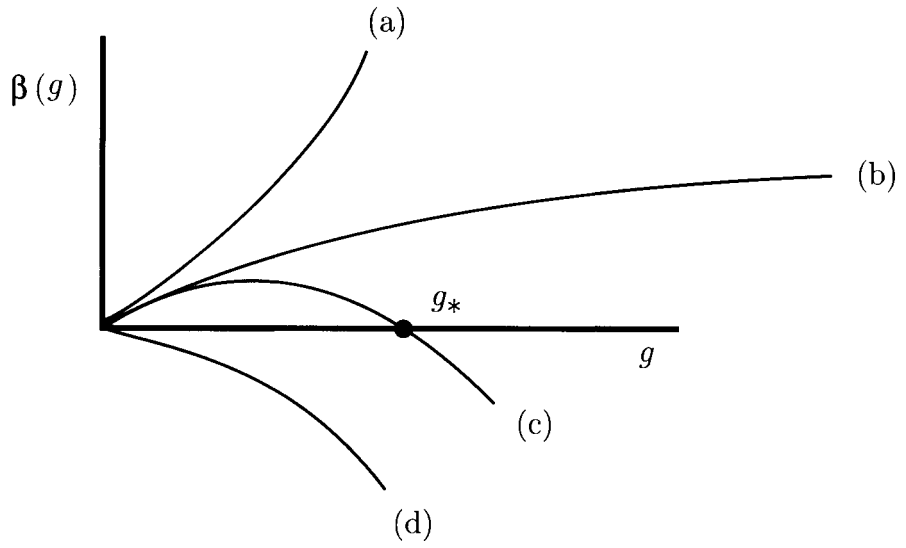

Figure 18.4. Schematic representation of four possible forms of the function $\beta(g)$. For such forms of $\beta(g)$, the running coupling g_μ would: (a) approach infinity at a finite value of μ; (b) continue to grow as μ increases; (c) approach a finite limit g_* for $\mu \to \infty$; (d) approach zero for $\mu \to \infty$.

because the reality of the Lagrangian requires e to be real. In the scalar field theory, as we saw at the end of the previous section, it is necessary to have $g > 0$ in order to have any stable vacuum state. However, there are other examples that have $\beta(g) \leq 0$ for $g \geq 0$. For instance, we can consider a scalar field theory with interaction Hamiltonian $-g\phi^4/24$, with g taken positive. This may be unphysical, but stability problems will not bother us as long as we stick to perturbation theory. Eq. (18.2.9) shows that if we redefine $g \to -g$ the beta-function undergoes the change $\beta(g) \to -\beta(-g)$, so our previous result that $\beta(g) = 3g^2/16\pi^2 + O(g^3)$ for an interaction $g\phi^4/24$ now gives

$$\beta(g) = -3g^2/16\pi^2 + O(g^3) \tag{18.3.4}$$

for an interaction Hamiltonian density $-g\phi^4/24$. Of greater relevance to physics, we shall see in Section 18.7 that non-Abelian gauge theories with not too many spinor fields have $\beta \leq 0$ for small positive gauge coupling constants. In what follows we will always define the coupling g so that $g \geq 0$, but we shall consider the cases of $\beta(g)$ either positive or negative for small g.

Now let us turn to our list of possibilities. (See Figure 18.4.)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

(a) **Singularity at Finite Energy**

Suppose that $\beta(g) > 0$ for small positive g (as is the case for Eqs. (18.3.1) and (18.3.3)), and that $\beta(g)$ remains positive and continues to rise sufficiently rapidly with increasing g, so that the integral $\int^{\infty} dg/\beta(g)$ converges:

$$\int^{\infty} \frac{dg}{\beta(g)} < \infty . \tag{18.3.5}$$

Then g_μ will move steadily away from $g_\mu = 0$, and Eq. (18.2.10) shows that g_E must become infinite at a *finite* value of E:

$$E_\infty = \mu \exp\left(\int_{g_\mu}^{\infty} \frac{dg}{\beta(g)} \right) , \tag{18.3.6}$$

where μ is any renormalization scale with $\mu \gg m$. We saw an example of this phenomenon in the previous section; if the lowest-order formula $\beta(g) = 3g^2/16\pi^2$ for the beta-function in the scalar field theory were taken as exact for all values of g, then the running coupling (18.2.17) would become infinite at the energy (18.2.19). Similarly, if the lowest order formula $\beta(g) = g^2/6\pi^2$ (with $g \equiv e^2$) for the beta-function in spinor quantum electrodynamics were taken as exact for all g, then the energy (18.3.6) at which g_E and e_E become infinite would be

$$E_\infty = \mu \exp(6\pi^2/g_\mu) . \tag{18.3.7}$$

Using Eq. (18.2.43), we may express this in terms of the conventional renormalized charge:

$$E_\infty \simeq m_e \exp\left(\frac{6\pi^2}{e_R^2} + \frac{5}{6} + O(e_R^2) \right) = e^{646.6} m_e . \tag{18.3.8}$$

Of course, the approximation that $\beta(g) = g^2/6\pi^2$ will break down before this energy is reached, so all we can say with confidence is that e_E will become large enough to invalidate perturbation theory at some energy E below E_∞.

(b) **Continued Growth**

Now suppose that in some theory $\beta(g)$ remains positive definite for $g \to \infty$, but rises slowly enough (or decreases) so that $\int^{\infty} dg/\beta(g)$ is divergent. The coupling constant g_E then continues to increase as $E \to \infty$, but becomes infinite only for $E = \infty$. Furthermore, the leading term in the asymptotic behavior of g_E for $E \to \infty$ is independent of the conventionally renormalized coupling. For instance, if $\beta(g)$ behaves for large g like bg^k with $b > 0$ and $k < 1$, then the solution of Eq. (18.2.9) is

$$g_E = \left[1 + (1-k) b\, g_\mu^{k-1} \ln \frac{E}{\mu} \right]^{1/(1-k)} g_\mu . \tag{18.3.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

If g_μ is small for some μ (say, of order m) then the growth of g_E is seen only at energies which are exponentially large compared with this μ. However, in the extreme high energy limit the coupling grows according to

$$g_E \to [(1-k)b \ln E]^{1/(1-k)} , \tag{18.3.10}$$

a limiting behavior independent of g_μ!

(c) **Fixed Point at Finite Coupling**

Suppose next that $\beta(g)$ remains positive-definite for $0 < g < g_*$, but drops to zero at $g = g_*$ and is negative thereafter. Then Eq. (18.2.9) dictates that as μ increases g_μ will increase for $g_\mu < g_*$ and decrease for $g_\mu > g_*$, in either case approaching the fixed point g_* for $\mu \to \infty$. If the zero of $\beta(g)$ at g_* is simple, then in the neighborhood of this point we have

$$\beta(g) \to a(g_* - g) \quad \text{for} \quad g \to g_* \tag{18.3.11}$$

with $a > 0$. The solution of Eq. (18.2.9) is then

$$g_* - g_\mu \propto \mu^{-a} . \tag{18.3.12}$$

(The behavior of type (b) described above may be regarded as the special case where the fixed point g_* is at infinity.) Also, $\gamma(g)$ for a general operator $\mathcal{O}$ may be expected to behave smoothly near g_*:

$$\gamma(g) = \gamma(g_*) + c(g_* - g) + O\left((g_* - g)^2\right) . \tag{18.3.13}$$

(We are here dropping the label $\mathcal{O}$ on γ and c.) Hence in matrix elements of this (and perhaps other) operators, we encounter a factor (see Eq. (18.2.27))

$$N_E^{-1} \propto \exp\left[-\int^E \gamma(g_\mu)\frac{d\mu}{\mu}\right] \propto E^{-\gamma(g_*)}\left[1 + O(E^{-a})\right] . \tag{18.3.14}$$

The product of the factors $E^{-\gamma(g_*)}$ can be lumped together with the factor E^D in Eq. (18.2.23), with the result that the whole matrix element goes as

$$M_R \propto E^{D_*} , \tag{18.3.15}$$

where the dimensionality D_* is calculated adding an 'anomalous dimension' $-\gamma(g_*)$ to the actual dimensionality of each operator appearing in the matrix element.

(d) **Asymptotic Freedom**

In the examples that have been discussed so far, $\beta(g)$ was positive for small positive g, so that g_μ is driven away from $g = 0$ as μ increases. Suppose that for some other theory $\beta(g)$ is negative for small positive g.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Then

$$\beta(g) \to -b\,g^n\,, \tag{18.3.16}$$

where $b > 0$. Here n is the order of the lowest-order diagrams that contribute to $\beta(g)$, and hence is always an integer greater than unity. (In the theories used as examples here, $n = 2$.) The solution of Eq. (18.2.9) here is

$$g_E = g_\mu \left[1 + b(n-1)g_\mu^{n-1} \ln \frac{E}{\mu}\right]^{-1/(n-1)}. \tag{18.3.17}$$

For $E \to \infty$, this has a limit independent of g_μ:

$$g_E \to [b(n-1)\ln E]^{-1/(n-1)}\,. \tag{18.3.18}$$

Since this gives a vanishing g_E for $E \to \infty$, we can trust perturbation theory in this limit, provided only that g_E for some finite E is within the region around $g = 0$ where g and $\beta(g)$ have opposite signs. The anomalous dimensions $\gamma^{(\mathcal{O})}$ of various operators $\mathcal{O}$ are expected to have the weak-coupling behavior (dropping the label $\mathcal{O}$)

$$\gamma(g) \to c\,g^m\,, \tag{18.3.19}$$

where m is the order of the lowest-order diagrams that contribute to the renormalization of the operator, and c is a real constant that can be positive or negative. The asymptotic behavior at high energies of the factor introduced into matrix elements by renormalization of this operator is then

$$\begin{aligned} N_E^{-1} &\propto \exp\left[-\int^E \gamma(g_\mu)\frac{d\mu}{\mu}\right] \\ &\to \exp\left[-c\int^E [b(n-1)\ln\mu]^{-m/(n-1)}\,\frac{d\mu}{\mu}\right] \\ &\propto \exp\left[-\frac{c[b(n-1)]^{-m/(n-1)}}{(1-m/(n-1))}(\ln E)^{1-m/(n-1)}\right] \end{aligned} \tag{18.3.20}$$

except that for $m = n - 1$

$$N_E^{-1} \propto (\ln E)^{-c/b(n-1)}\,. \tag{18.3.21}$$

We see that in the case of asymptotic freedom there are no corrections to those effective dimensionalities that determine the powers of energy appearing in the asymptotic behavior of the matrix element, but this asymptotic behavior is instead modified by powers of $\ln E$.

For a toy model that exhibits asymptotic freedom, we can use the scalar field theory with interaction Hamiltonian density $-g\phi^4/24$, with g taken

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

positive. Eq. (18.3.4) here gives the parameters of Eq. (18.3.16) as:

$$b = 3/16\pi^2\,, \qquad n = 2\,, \tag{18.3.22}$$

so Eq. (18.3.17) gives, for $E \to \infty$,

$$g_E \to \left[\frac{3 \ln E}{16\pi^2}\right]^{-1}\,. \tag{18.3.23}$$

Also, the operator ϕ^2 in this theory has an anomalous dimension given by Eq. (18.2.29) as

$$\gamma(g) = -\frac{g}{16\pi^2} + O(g^2)\,. \tag{18.3.24}$$

Hence Eq. (18.3.19) applies here, with

$$c = -1/16\pi^2\,, \qquad m = 1\,. \tag{18.3.25}$$

Each ϕ^2 operator in a matrix element therefore contributes a factor given by Eq. (18.3.21) as

$$N_E^{-1} \propto (\ln E)^{1/3}\,. \tag{18.3.26}$$

The scalar field ϕ itself in this theory has $\gamma(g) \propto g^2$, and hence $m = 2$, so each ϕ operator in a matrix element therefore contributes a factor given by Eq. (18.3.20) as

$$N_E^{-1} \propto 1 + O\left(\frac{1}{\ln E}\right)\,. \tag{18.3.27}$$

We will see another, more physical, example of asymptotic freedom when we take up quantum chromodynamics in Section 18.7.

In all cases where g_E can be extended to infinite energy, its behavior in this limit turns out to be independent of the renormalized coupling g_R. However, this does not necessarily mean that the theory involves no arbitrary dimensionless parameters. In all cases, in order to describe how g_E approaches its limit for $E \to \infty$, we need to specify a free parameter λ with the dimensions of energy. For case (b), Eq. (18.3.10) may be written

$$g_E \to \left[(1-k)b \ln(E/\lambda)\right]^{1/(1-k)}\,.$$

For case (c), Eq. (18.3.12) may be written

$$g_E \to g_* \left[1 - \left(\frac{\lambda}{E}\right)^b\right]\,.$$

Finally, for case (d), Eq. (18.3.18) may be written

$$g_E \to \left[b(n-1)\ln(E/\lambda)\right]^{-1/(n-1)}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Such theories in general do have a free dimensionless parameter: the ratio of λ to the mass m. Coupling constants like e_R that are renormalized at scales tied to m may be expressed as functions of m/λ. It is only when all masses in a theory vanish that we can say that the theory has no free dimensionless parameters.

Of the four types of asymptotic behavior described here, types (a) and (b) lead to the apparently unphysical behavior that the running coupling g_E becomes infinite, either at a finite energy (case (a)) or for $E \to \infty$ (case (b)). This does not in itself mean disaster: we have to look at how the coupling is defined. For instance, if g_μ drops smoothly from a finite value g_m at $\mu = m$ to zero for $\mu \to \infty$, and we define a new coupling $\tilde{g}_\mu \equiv g_\mu/[1 - g_\mu/g_{2m}]$, then $\tilde{g}_\mu$ becomes infinite at $\mu = 2m$, but this is just an artifact of this particular choice of coupling parameter. However, the conventional renormalized couplings g_μ in both the ϕ^4 scalar field theory and quantum electrodynamics have been defined here in terms of the values of matrix elements at energies of order μ. Specifically, g_μ in ϕ^4 scalar field theory is defined as the invariant Feynman amplitude A for scalar–scalar scattering at $s = t = u = \mu^2$, where A is supposed to be analytic. Also, $g_\mu = e_\mu^2$ in spinor electrodynamics is given by

$$e_\mu^2/e_R^2 = Z_3 \Big/ N_\mu^{(A)2} = \left[1 - \pi(\mu^2)\right]^{-1} . \tag{18.3.28}$$

An infinity in e_μ^2 at a point μ_∞ would therefore produce a pole or other singularity in the renormalized photon propagator at a positive value of p^2, that is, at $p^2 = \mu_\infty^2$, where the propagator is supposed to be analytic. Thus, with g_μ as defined here, the type of asymptotic behavior described in case (a) *is* ruled out physically.

How then do our various quantum field theories behave? Years ago, Landau[4b] argued that in quantum electrodynamics the increasing powers of $\ln(E/M)$ encountered at each order of perturbation theory would add up to give singularities (so-called 'Landau ghosts') at finite values of E. In modern terms, Landau could be said to have discovered possibility (a) above, but he did not give any argument against cases (b) or (c).

Nevertheless, there is today a widespread view that interacting quantum field theories that are not asymptotically free, like quantum electrodynamics or the scalar field theory with ϕ^4 interaction, are not mathematically consistent. In quantum electrodynamics there is some evidence against case (c), the existence of a finite fixed point e_*. Such a fixed point would only be possible[4c] if non-perturbative effects changed the qualitative nature of the operator product expansion, the subject of Chapter 20, or if there were a non-perturbative renormalization of the triangle anomaly discussed in Chapter 22. But even if case (c) is indeed ruled out in quan-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

tum electrodynamics, there is still the possibility of case (b), a fixed point at infinite coupling.

Most of the evidence against the consistency of interacting non-asymptotically free quantum field theories comes from the study of the scalar field theory in four spacetime dimensions with ϕ^4 interaction, quantized on a finite spacetime lattice. There are rigorous theorems[4d] to the effect that this theory (with arbitrary dependence of the parameters of the theory on lattice spacing) does not have an interacting continuum theory as its limit for zero lattice spacing unless the theory is asymptotically free, which of course is contrary to what is found for this theory in perturbation theory. This argument also seems inconclusive. It is true that if there were a consistent continuum scalar field theory that is not asymptotically free, then it would be possible to construct a lattice theory by integrating out the values of the scalar field at all points except on a spacetime lattice. *But this would not be the lattice theory that is considered in these theorems.* It would be a lattice theory with every possible coupling allowed by symmetry principles — not just a term proportional to ϕ^4, but also terms proportional to ϕ^6, $\phi^3\Box\phi$, etc, with coefficients having a dependence on the cut-off (the inverse lattice spacing) governed by the Wilson renormalization group equations discussed in Section 12.4. No one has proved anything about the continuum limit of such a theory.

If it is really true that there is no interacting continuum scalar field theory in the limit of zero lattice spacing, then we must encounter some obstruction when we try to solve the Wilson renormalization group equations, which for weak renormalized couplings would have to be at very small lattice spacings. Such a theory would appear like an interacting continuum field theory unless examined at very short distances. The renormalized coupling constant in this approximate continuum theory presumably has a singularity at finite energy, as in case (a) above, so that it too breaks down at short distances. (But for strong couplings there is no direct connection between the forms of the Wilson and Gell-Mann–Low renormalization group equations, so the existence of a singularity in the bare couplings at finite lattice spacing does not necessarily imply a singularity in the renormalized coupling constants at a finite renormalization scale.)

Theories of this sort are sometimes called *trivial*, either because, under various assumptions about the bare couplings of the theory quantized on a lattice, the continuum limit turns out to be a free field theory, or because the only way to make a continuum theory of type (a) physically satisfactory at all energies is to adopt the solution $g_\mu = 0$ of the renormalization group equation (18.2.9). Even if a field theory is trivial in either sense, there is no reason not to include it as part of a realistic theory of physical phenomena. The existence of an obstruction to the solution of the Wilson

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

renormalization group equations for a field theory at very small lattice spacings is not important if in the real world there are other fields that must also be taken into account at such short distances. Similarly, the fact that a given quantum field theory has unphysical singularities at some large energy E_∞ is not a physical problem if the theory in question is only a low-energy approximation to a larger theory, an approximation valid only at energies far below E_∞. In particular, long before we reach the energies near (18.3.8) where quantum electrodynamics could be expected to become singular, it becomes necessary to take even gravitation into account, and no one knows how to calculate the effects of strong gravitational forces at such energies.

Despite these reassuring remarks, it is possible that in order to avoid unphysical singularities, all our separate quantum field theories like spinor quantum electrodynamics will eventually have to be integrated into an asymptotically free theory. Fortunately, the question of whether a theory is asymptotically free for *some* finite range of coupling constants can be settled by perturbative calculations: if $\beta(g)$ is negative as $g \to 0+$, then the theory is asymptotically free for all renormalized couplings g_μ lying between zero and the first zero of $\beta(g)$.

* * *

In this connection, it is worth noting that although the detailed form of $\beta(g)$ depends on the gauge and on precisely how the running coupling is defined, the first two terms in the power series for $\beta(g)$ do not. Suppose we have two definitions g_μ and $\tilde{g}_\mu$ of the running coupling, perhaps employing different definitions of the renormalization scale μ or different gauges. Since both g_μ and $\tilde{g}_\mu$ are dimensionless and cut-off independent, there is no way that $\tilde{g}_\mu$ for $\mu \gg m$ can depend on anything but g_μ:

$$\tilde{g}_\mu = \tilde{g}(g_\mu)\,.$$

We then have

$$\tilde{\beta}(\tilde{g}_\mu) \equiv \mu \frac{d}{d\mu} \tilde{g}_\mu = \frac{d\tilde{g}(g_\mu)}{dg_\mu}\, \beta(g_\mu)$$

and so

$$\tilde{\beta}(\tilde{g}) = \frac{d\tilde{g}(g)}{dg}\, \beta(g)\,. \tag{18.3.29}$$

As long as we are sticking to the same definition of the unrenormalized coupling, all renormalized couplings are equal in lowest order, so the power series for $\tilde{g}$ in terms of g may be written

$$\tilde{g}(g) = g + a\, g^2 + O(g^3)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

or, equivalently,

$$g = \tilde{g} - a\,\tilde{g}^2 + O(\tilde{g}^3)\,.$$

The derivative is

$$\frac{d\tilde{g}}{dg} = 1 + 2\,a\,g\, + O(g^2) = 1 + 2\,a\,\tilde{g} + O(\tilde{g}^2)\,.$$

Also, for the couplings we have been considering here (including $g = e^2$) the power series for $\beta(g)$ takes the form

$$\beta(g) = b\,g^2 + b'g^3 + O(g^4)$$

or in terms of $\tilde{g}$

$$\beta(\tilde{g}) = b\tilde{g}^2 + (b' - 2ab)\,\tilde{g}^3 + O(\tilde{g}^4)\,.$$

From (18.3.29), we have then

$$\begin{aligned}\tilde{\beta}(\tilde{g}) &= \left[1 + 2a\,\tilde{g} + O(\tilde{g}^2)\right]\left[b\tilde{g}^2 + (b' - 2ab)\tilde{g}^3 + O(\tilde{g}^4)\right] \\ &= b\tilde{g}^2 + b'\tilde{g}^3 + O(\tilde{g}^4)\,. \qquad (18.3.30)\end{aligned}$$

We see that the first two terms in the power series for $\tilde{\beta}$ in terms of $\tilde{g}$ have the same coefficients as in the power series for β in terms of g. However, this is definitely not the case for the higher-order terms. In fact, it is always possible to choose the function $\tilde{g}(g)$ so that all terms in $\tilde{\beta}(\tilde{g})$ of higher than third order in $\tilde{g}$ vanish, so we can describe the asymptotic behavior of $\tilde{g}_E$ for $E \to \infty$ by inspection of the first two terms in the perturbation series for $\beta(g)$. But this is of little value, since we would need to carry our calculations to all orders to determine how g depends on $\tilde{g}$, and without this we cannot use our knowledge of the asymptotic behavior of $\tilde{g}$ to say anything about the asymptotic behavior of g, or of physical quantities.

The same argument that led to Eq. (18.3.30) shows that, *at small coupling*, the Wilson renormalization group equation for the bare coupling constant as a function of lattice spacing for inverse lattice spacings greater than particle masses is the same as the Gell-Mann–Low renormalization group equation for the renormalized coupling constant as a function of renormalization scale for scales greater than particle masses. Hence if a continuum theory is asymptotically free, then there will be no obstruction in passing to the continuum limit of the theory quantized on a lattice.

18.4 Multiple Couplings and Mass Effects

Up to now, we have considered theories with only one dimensionless coupling g. It is easy to extend the formalism to incorporate several such

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

couplings g^ℓ: for each g^ℓ, we have a renormalization group equation that for $\mu \gg m$ takes the form

$$\mu \frac{d}{d\mu} g^\ell(\mu) = \beta^\ell(g(\mu)) , \tag{18.4.1}$$

with each β^ℓ depending in general on all the gs. There are now many more possibilities for the asymptotic behavior of the $g^\ell(\mu)$ as $\mu \to \infty$; in a given theory we may have some trajectories in g-space that go off to infinity, at finite or infinite values of μ, other trajectories that approach fixed points, and yet other trajectories that approach closed curves known as 'limit cycles'. To get a taste of some of the various possibilities, let's consider the behavior of $g^\ell(\mu)$ near a fixed point.

Eq. (18.4.1) has a fixed-point solution $g^\ell(\mu) = g^\ell_*$ if

$$\beta^\ell(g_*) = 0 . \tag{18.4.2}$$

In the neighborhood of this point, Eq. (18.4.1) becomes

$$\mu \frac{d}{d\mu} \left[g^\ell(\mu) - g^\ell_* \right] = \sum_k M^\ell{}_k \left[g^k(\mu) - g^k_* \right] , \tag{18.4.3}$$

where M is the matrix

$$M^\ell{}_k \equiv \left[\frac{\partial \beta^\ell(g)}{\partial g^k} \right]_{g=g_*} . \tag{18.4.4}$$

The solution can be expanded in eigenvectors of this matrix

$$g^\ell(\mu) = g^\ell_* + \sum_m c_m V_m{}^\ell \mu^{\lambda_m} , \tag{18.4.5}$$

where V_m is a eigenvector of M with eigenvalue λ_m (normalized in any convenient way):

$$\sum_k M^\ell{}_k V_m{}^k = \lambda_m V_m{}^\ell , \tag{18.4.6}$$

and the c_m are a set of expansion coefficients.* (The summation convention is suspended in this section.)

Eq. (18.4.5) shows that the coupling constants approach the fixed point as $\mu \to \infty$ if and only if $c_m = 0$ for all eigenvectors with $\lambda_m > 0$. (For simplicity we are assuming here that none of the eigenvalues vanish.) Thus

* We are assuming here that the eigenvectors V_m form a complete set. This is not always so, but it is the generic case; the eigenvectors of a finite matrix M will form a complete set if all of the roots of the secular equation $\mathrm{Det}(M - \lambda 1) = 0$ are different. A matrix whose eigenvectors do not form a complete set can be regarded as a limiting case of a matrix with a complete set of eigenvectors when some of its eigenvalues become degenerate.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

in general the trajectories that are attracted to the fixed point lie on an N_--dimensional surface, where N_- is the number of negative eigenvalues of M; the tangents to this surface at g_* are the eigenvectors with negative eigenvalues. Trajectories that are not on this surface may approach close to the fixed point, but are eventually repelled, predominantly in the direction of eigenvectors with the largest positive eigenvalues. Of course, if all eigenvalues are negative then there is a finite region around the fixed point within which all trajectories converge on this point.

Since the eigenvalues λ_m are evidently important in learning the asymptotic behavior of trajectories that approach a fixed point, it is useful to note that these eigenvalues are independent of the definition of the couplings. Suppose we introduce a new set of couplings $\tilde{g}^\ell$, defined as functions of the gs. These satisfy renormalization group equations

$$\mu \frac{d}{d\mu} \tilde{g}^\ell(\mu) = \sum_m \left. \frac{\partial \tilde{g}^\ell(g)}{\partial g^m} \right|_{g=g(\mu)} \beta^m(g(\mu)) \equiv \tilde{\beta}^\ell(\tilde{g}(\mu)) ,$$

so

$$\tilde{\beta}^\ell(\tilde{g}) = \sum_m \frac{\partial \tilde{g}^\ell(g)}{\partial g^m} \beta^m(g) . \tag{18.4.7}$$

(That is, β transforms as a contravariant vector in coupling-constant space.) Differentiating, we have

$$\sum_m \frac{\partial \tilde{\beta}^\ell(\tilde{g})}{\partial \tilde{g}^m} \frac{\partial \tilde{g}^m}{\partial g^k} = \sum_m \frac{\partial^2 \tilde{g}^\ell}{\partial g^m \partial g^k} \beta^m(g) + \sum_m \frac{\partial \tilde{g}^\ell}{\partial g^m} \frac{\partial \beta^m(g)}{\partial g^k} .$$

At a fixed point g_* the first term on the right vanishes, so this gives the matrix equation

$$\tilde{M} S = S M , \tag{18.4.8}$$

where

$$\tilde{M}^\ell{}_k \equiv \left[\frac{\partial \tilde{\beta}^\ell}{\partial \tilde{g}^k} \right]_{\tilde{g}=\tilde{g}(g_*)} , \tag{18.4.9}$$

$$S^\ell{}_k \equiv \left[\frac{\partial \tilde{g}^\ell}{\partial g^k} \right]_{g=g_*} . \tag{18.4.10}$$

As long as the transformation $g \to \tilde{g}$ is non-singular, Eq. (18.4.8) is a similarity transformation, and hence M and $\tilde{M}$ have the same eigenvalues λ_m.

The renormalization group formalism may be extended to non-renormalizable as well as renormalizable theories. As explained in Section 12.3, the infinities in non-renormalizable theories are eliminated by a suitable renormalization of coupling constants and masses, just as in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

renormalizable theories; the only difference is that in non-renormalizable theories the Lagrangian must be supposed to contain all possible interactions allowed by the symmetries of the theory. If g_B^ℓ is the unrenormalized coupling constant multiplying an operator of dimensionality D_ℓ in the Lagrangian (that is, a product of fields and spacetime derivatives of fields whose dimensionality in powers of mass or energy is D_ℓ), then g_B^ℓ will have a dimensionality $\Delta_\ell = 4 - D_\ell$. We may then re-express the bare couplings in terms of a set of *dimensionless* renormalized couplings $g^\ell(\mu)$ and a cut-off Λ, through relations of the general form

$$g_B^\ell \equiv \mu^{\Delta_\ell} \left[g^\ell(\mu) + \sum_{k,m} b^\ell{}_{km}\, g^k(\mu) g^m(\mu) \ln\left(\frac{\Lambda}{\mu}\right) + O(g(\mu)^3) \right] , \quad (18.4.11)$$

with the dimensionless numerical coefficients $b^\ell{}_{km}$ and similar coefficients in higher-order terms chosen to cancel the cut-off dependence of physical quantities. (In some theories the leading term might be trilinear or even higher order in couplings; the modifications that would be needed here are obvious.) From the requirement that g_B be cut-off-independent, we obtain the renormalization group equation (18.4.1), with

$$\beta^\ell(g) = -\Delta_\ell g^\ell - \sum_{k,m} b^\ell_{km} g^k g^m + O(g^3) . \quad (18.4.12)$$

Non-renormalizable interactions are those with $D_\ell > 4$, or $\Delta_\ell < 0$, so as long as the $g^\ell(\mu)$ all remain sufficiently small we expect the non-renormalizable renormalized couplings to have positive β^ℓ and hence to grow with μ, but no one knows what happens when the couplings become large enough to invalidate perturbation theory.

However, as explained in the next section, even theories with infinite number of independent parameters commonly have fixed points g_* at which the number N_- of negative eigenvalues of the matrix (18.4.6) is *finite*, just as it is at zero coupling. (In particular, often $N_- = 1$.) Where $N_- \neq 0$, the fixed point lies on an N_--dimensional critical surface, consisting of trajectories that are attracted into the fixed point as $\mu \to \infty$. A non-renormalizable theory with coupling parameters on such a critical surface, although not of course asymptotically free, is said to be asymptotically *safe*,[5] because the renormalized couplings remain finite for large values of μ. The condition of asymptotic safety in such a theory would play the role that used to be associated with the principle of renormalizability, of eliminating all but a finite number of free parameters, the coordinates of the critical surface.

In a renormalizable theory, all physical quantities are made cut-off-independent by adjusting the cut-off dependence of a *finite* number of bare couplings. These bare couplings may be expressed in terms of an

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

equal number of μ-dependent renormalized couplings, and the condition that the bare couplings are μ-independent yields renormalization group equations relating only these renormalized couplings. From the broader point of view which allows non-renormalizable as well as renormalizable couplings, a renormalizable theory just corresponds to a finite-dimensional *invariant surface* in the infinite-dimensional space of all renormalizable and non-renormalizable theories; that is, it is a surface for which $\beta^\ell(g)$ at any point g on the surface is tangent to the surface at that point.

So far in this section we have tacitly assumed that $\mu \gg m$, so that we could neglect the dependence of β^ℓ on m/μ. However, this is not necessary; we can if we like treat a mass as just another coupling parameter.[6] That is, all renormalized couplings can be defined as before in terms of various Greens functions at off-mass-shell momenta of order μ, but now evaluated with all bare masses zero. The dimensionless renormalized mass parameters for Dirac fields ψ or scalar fields ϕ may be defined as

$$m_\psi(\mu) \equiv N^{(\bar\psi\psi)}(\Lambda/\mu)^{-1}\, m_{\psi,\mathrm{BARE}}(\Lambda)/\mu\,, \tag{18.4.13}$$

$$m_\phi^2(\mu) \equiv N^{(\phi^2)}(\Lambda/\mu)^{-1}\, m_{\phi,\mathrm{BARE}}^2(\Lambda)/\mu^2\,, \tag{18.4.14}$$

where $N^{(\mathcal{O})}(\Lambda/\mu)$ are the dimensionless constants which, when multiplied into corresponding operators $\mathcal{O}$, cancel the infinities in the matrix elements of these operators, also evaluated with all bare masses zero. (See Section 18.1.) These new renormalized masses and couplings have no direct physical significance, but the true physical masses and all physical matrix elements can be expressed in terms of them. These matrix elements take the form of sums of matrix elements for zero bare mass, with any number of insertions of the renormalized mass operators $N^{(\phi^2)}\phi^2$ and $N^{(\bar\psi\psi)}\bar\psi\psi$, times the corresponding renormalized mass parameters.

In this renormalization scheme the beta-functions for the various couplings are obviously mass-independent, and the beta-functions for the mass parameters are proportional to these parameters, with coefficients that depend on all the various couplings; using Eq. (18.2.25), we have

$$\mu\frac{d}{d\mu}m_\psi(\mu) = [-1 - \gamma_{\bar\psi\psi}(g_\mu)]\, m_\psi(\mu)\,, \tag{18.4.15}$$

$$\mu\frac{d}{d\mu}m_\phi^2(\mu) = [-2 - \gamma_{\phi^2}(g_\mu)]\, m_\phi^2(\mu)\,. \tag{18.4.16}$$

For instance, we noted in Section 18.2 that in the scalar field theory with Lagrangian (18.1.2), the mass operator ϕ^2 has anomalous dimension (18.2.29) for $m = 0$, so here

$$\mu\frac{d}{d\mu}m_\phi^2(\mu) = \left[-2 + \frac{g_\mu}{16\pi^2} + O(g_\mu^2)\right] m_\phi^2(\mu)\,. \tag{18.4.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Also, Eq. (11.4.3) shows that the effect of higher-order corrections to the electron propagator is to replace the electron mass by $m_e - \Sigma^*(p, m_e)$, so the effect of these corrections on matrix elements of the operator $\bar{\psi}_e \psi_e$ between one-electron states of four-momentum p^μ is to multiply them by a factor

$$F(p) = 1 - \left(\frac{\partial \Sigma^*(p, m_e)}{\partial m_e} \right)_{m_e=0} .$$

The renormalization constant $N^{\bar{\psi}\psi}$ for the operator $\bar{\psi}_e \psi_e$ is therefore equal to $F^{-1}(p)$, evaluated with p^2 equal to some renormalization scale, say $+\mu^2$. According to Eq. (11.4.8), to one-loop order this is

$$\begin{aligned} N^{(\bar{\psi}\psi)} &= 1 + \left(\frac{\partial \Sigma^*_{1\text{ loop}}(p, m_e)}{\partial m_e} \right)_{m_e=0,\, p^2=\mu^2} \\ &= 1 - \frac{4\pi^2 e^2}{(2\pi)^4} \int_0^1 dx \, \ln \left[1 + \left(\frac{\Lambda^2}{\mu^2} \right) (1-x) \right] \\ &\to 1 - \frac{e^2}{4\pi^2} \left[\ln \left(\frac{\Lambda^2}{\mu^2} \right) - 1 \right] , \end{aligned} \tag{18.4.18}$$

where Λ is an ultraviolet cut-off,** and we take the limit $\Lambda \gg \mu$. The anomalous dimension of the operator $\bar{\psi}_e \psi_e$ is therefore given by (18.2.25) as

$$\gamma^{(\bar{\psi}\psi)} = \mu \frac{d}{d\mu} \ln N^{(\bar{\psi}\psi)} = \frac{e_\mu^2}{2\pi^2} + O(e_\mu^4) , \tag{18.4.19}$$

so Eq. (18.4.15) here reads

$$\mu \frac{d}{d\mu} m_e(\mu) = \left[-1 - \frac{e_\mu^2}{2\pi^2} + O(e_\mu^4) \right] m_e(\mu) . \tag{18.4.20}$$

The same formula holds in general gauge theories, with e^2 replaced with the value of $\sum_\alpha (t_\alpha)^2$ for the particular species of fermion in question.

The important difference between the $m(\mu)$ and the other renormalized parameters of the theory is of course that bare masses have positive dimensionality, so as long as the couplings remain small the $m(\mu)$ all decrease in magnitude. Our previous assumption that masses may be neglected as $\mu \to \infty$ is justified if in fact $m(\mu)$ does vanish for $\mu \to \infty$. However this is only known to be the case in asymptotically free theories, where the couplings all do remain small for $\mu \to \infty$; in all other cases this assumption is just an educated guess.

** This notation is different from that of Eq. (11.4.8), where the ultraviolet cut-off was called μ.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

18.5 Critical Phenomena*

For some purposes we may be interested in the limit of very low rather than high energies or wave numbers. The arguments of Section 18.2 can be repeated to study this limit, except that here we must examine the case $\mu \to 0$ rather than $\mu \to \infty$. This limit is of course simplest if there are no masses in the theory, as, for instance, in quantum electrodynamics with a symmetry under the chiral transformation $\psi \to \gamma_5 \psi$ which forbids an electron mass. In this particular case the only renormalizable coupling $eA^\mu \bar{\psi} \gamma_\mu \psi$ as well as all non-renormalizable couplings have $\beta^\ell > 0$ for sufficiently small couplings, so all trajectories in at least a finite region around the origin are attracted into the point $g^\ell = 0$ as $\mu \to 0$.

The same considerations may be applied even to theories with very small but non-zero masses if we include these masses among the coupling parameters of the theory, as described in the previous section. The coefficient Δ in Eq. (18.4.12) is positive for a mass parameter, so in this case the trajectories can never reach the point $g = 0$, but they may come close if the masses are small.

Of course, even if we can regard some degree of freedom like the electron field as having zero or very small mass, in the real world there are many other degrees of freedom whose masses are not small. The renormalization group should properly be applied not to the true theory that encompasses all these heavy degrees of freedom, but to an 'effective' field theory, in which only massless or nearly massless degrees of freedom appear explicitly, with interactions that include the effects of internal heavy particle lines. (We shall have more to say about effective field theories in Chapter 19.)

The low wave number limit is of particular interest in the study of critical phenomena, such as long-range correlations at or near a second-order phase transition (a smooth phase transition, with no latent heat) in condensed matter. Because we are interested in the limit $\mu \to 0$, the important eigenvectors of the matrix (18.4.4) are those with eigenvalues $\lambda < 0$, which are called *relevant*. The eigenvectors with $\lambda = 0$ and $\lambda > 0$ are called *marginal* and *irrelevant*, respectively.

Suppose that there is a non-trivial fixed point g_* with just one negative eigenvalue λ_0, perhaps corresponding approximately to a mass operator. The set of trajectories of $g^\ell(\mu)$ that are attracted into this fixed point for $\mu \to 0$ therefore forms a critical surface of codimension one; that is, a surface defined by a single condition on the couplings, the condition that

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

for $g \to g_*$, the tangents $g^\ell - g^\ell_*$ have no components in the direction of the eigenvector with negative eigenvalue. There is a phase transition as the physical value of the couplings at any fixed characteristic scale approaches this surface. Because the critical surface has codimension one, the phase transition can be reached by adjusting any one parameter on which the couplings depend, such as the pressure or temperature. The fact that a wide variety of substances do exhibit phase transitions of this sort shows that it is common to encounter fixed points for which the matrix (18.4.4) has a single negative eigenvalue, as already mentioned in the previous section.

To be specific, as the temperature T approaches its critical value T_c we expect the coefficient c_0 of the growing term in Eq. (18.4.5) to become proportional to $T - T_c$, because there is no reason why it should be singular or why it should vanish faster than this. Hence for $\mu \to 0$ and then $T \to T_c$, the couplings go as

$$g^\ell(\mu) \to (T - T_c)\, V_0{}^\ell\, \mu^{\lambda_0}\,, \tag{18.5.1}$$

where λ_0 is the only negative eigenvalue at g_*, and $V_0{}^\ell$ is the corresponding eigenvector.** Applying our renormalization group arguments to wave numbers instead of energies, the N-point function (the Nth partial derivative of the effective action with respect to a field ϕ of dimensionality [wave number]$^{D_\phi}$) at a small characteristic wave number scale κ has the form†

$$\Gamma_N(\kappa) \to \kappa^{d-N(D_\phi+\gamma_\phi(g_*))} F_N((T - T_c)\kappa^{\lambda_0})\,, \tag{18.5.2}$$

where $\gamma_\phi(g)$ is the anomalous dimension associated with the field ϕ, and d is the spacetime dimensionality, or in classical statistical mechanics the spatial dimensionality. It is convenient to rewrite this in the equivalent form

$$\Gamma_N(\kappa) \to (T - T_c)^{-[d-N(D_\phi+\gamma_\phi(g_*))]/\lambda_0} G_N(\kappa(T - T_0)^{1/\lambda_0})\,. \tag{18.5.3}$$

This shows for one thing that the correlation length ξ (the characteristic length that determines the scale over which the Fourier transform of Γ_N varies) increases as T approaches T_c like

$$\xi \propto (T - T_c)^{-\nu} \tag{18.5.4}$$

where ν is a conventionally defined positive 'critical exponent', given by Eq. (18.5.3) as

$$\nu = -1/\lambda_0\,. \tag{18.5.5}$$

** Other contributions to the couplings will go as $(T - T_0)^0 \mu^{\lambda_1}$ with $\lambda_1 > 0$. Thus Eq. (18.5.1) is valid here provided $T - T_0$ does not go to zero as fast as $\mu^{\lambda_1 - \lambda_0}$.

† The function F_N also depends on dimensionless angles and ratios of wave numbers. Note that Γ_N has 'naive' dimensionality $d - ND_\phi$ because $\delta^d(\sum \kappa)\Gamma$ must be dimensionless.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

Also, the zero-field effective action Γ_0 (or in statistical physics, the free energy) must be κ-independent because it corresponds to graphs with no external lines. It follows that Eq. (18.5.3) here becomes for $T \to T_c$:

$$\Gamma_0 - F_0 \propto (T - T_c)^{\nu d} , \tag{18.5.6}$$

where the constant F_0 is the effective action or free energy due to the heavy degrees of freedom which have been integrated out. Thus the exponent ν also governs the behavior of the part of the free energy which is not analytic in temperature for T near T_c.

In 1972 Wilson and Fisher[7] used an expansion in powers of $d-4$ both to show that the theory of a scalar field actually fits the above description, and also to carry out an approximate calculation of critical exponents such as ν. Consider a theory with a single 'light' degree of freedom, a scalar field ϕ, such as the magnetization in a ferromagnet, with a symmetry under $\phi \to -\phi$ that rules out interactions odd in ϕ. In addition to the 'mass' term $-g_2\phi^2/2$, the Lagrangian density of the effective field theory will contain interactions $-g_4\phi^4/4!$, $-g_6\phi^6/6!$, etc. The dimensionality of the field ϕ in powers of wave number is $(d-2)/2$ (so that $\int d^dx\,(\nabla\phi)^2$ should be dimensionless) so the dimensionalities of the couplings g_2, g_4, g_6, etc. in d dimensions are $+2$, $4-d$, $6-2d$, etc. For the fixed point at zero coupling in three dimensions, there are two relevant couplings, g_2 and g_4, but this conclusion is changed by interactions at non-trivial fixed points. Let's examine the surface in coupling constant space in which only g_2 and g_4 are non-zero, and take g_4 to be small.[††] Eq. (18.2.12) gives $\beta(g_4) = 3g_4^2/16\pi^2 + O(g_4^3)$ for $d = 4$, and Eq. (18.4.12) tells us that for $d = 4 - \epsilon$ dimensions we must add to this a term $-\epsilon g_4$, so

$$\mu\frac{d}{d\mu}g_4(\mu) = -\epsilon\, g_4(\mu) + \frac{3g_4^2(\mu)}{16\pi^2} + O(g_4^3(\mu)) . \tag{18.5.7}$$

Also, Eq. (18.4.17) gives

$$\mu\frac{d}{d\mu}g_2(\mu) = \left[-2 + \frac{g_4(\mu)}{16\pi^2} + O(g_4(\mu))\right] g_2(\mu) . \tag{18.5.8}$$

Therefore for small ϵ there is a non-trivial fixed point at

$$g_{4*} = \frac{16\pi^2\epsilon}{3} , \qquad g_{2*} = 0 . \tag{18.5.9}$$

[††] These are the only renormalizable couplings for $3 \le d \le 4$, so for such d this is an invariant surface. Note that we do not include the coefficient of $(\nabla\phi)^2$ among the couplings here, because this is a redundant coupling, in the sense described in Section 7.7.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

The matrix (18.4.4) at this fixed point is diagonal, with eigenvalues

$$\lambda_4 = M^4{}_4 = -\epsilon + \frac{3g_{4*}}{8\pi^2} + O(g_{4*}^2) = +\epsilon + O(\epsilon^2)\,, \tag{18.5.10}$$

$$\lambda_2 = M^2{}_2 = -2 + \frac{g_{4*}}{16\pi^2} + O(g_{4*}^2) = -2 + \frac{\epsilon}{3} + O(\epsilon^2)\,. \tag{18.5.11}$$

From Eq. (18.5.10) we see that the coupling g_4 is actually irrelevant, so that there is just one relevant coupling here, signalling the presence of a second-order phase transition. From Eq. (18.5.11) we see that the anomalous exponent (18.5.5) is

$$\nu = -\frac{1}{\lambda_2} = \frac{1}{2} + \frac{\epsilon}{12} + O(\epsilon^2)\,. \tag{18.5.12}$$

For the physical value $\epsilon = 1$ the first two terms give $\nu \simeq 0.58$. Three-loop calculations[8a] give this critical exponent to order ϵ^3 as

$$\nu = \frac{1}{2} + \frac{\epsilon}{12} + \frac{7\epsilon^2}{162} - 0.01904\epsilon^3\,, \tag{18.5.13}$$

which for $\epsilon = 1$ gives $\nu = 0.61$.

In the calculation presented here nothing was assumed about the system under study except that there is a second-order phase transition, near which the only long-wavelength degree of freedom is a single scalar field. There are a number of different physical systems that fit this description, such as the spontaneous appearance of magnetization (represented here by ϕ) in ferromagnetic and antiferromagnetic materials, and also second-order phase transitions between liquids and gases and in binary fluids. All of these systems are therefore expected to have the same value of ν. This is confirmed by experiment, which gives a value[8] $\nu = 0.63 \pm 0.04$, in good agreement with the three-loop result (18.5.13), and in fair agreement even with the one-loop result (18.5.12). It is fortunate though still somewhat mysterious that an expansion in powers of 1 should work so well.

More generally, all the systems that are described by the same set of long-wavelength degrees of freedom near their second-order phase transitions are said to belong to the same *universality class.* All the critical exponents are the same for all systems in a given universality class.

18.6 Minimal Subtraction

We saw in Section 11.2 that dimensional regularization provides a particularly convenient method for calculating radiative corrections in quantum electrodynamics, because it preserves the conservation laws associated with gauge invariance. For the same reason, dimensional regularization

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

also turns out to provide a very convenient alternative definition for the sliding scale of the renormalization group in general gauge theories.[9]

In calculations using dimensional regularization, ultraviolet divergences arise as poles in physical amplitudes when the spacetime dimensionality d approaches the physical value $d = 4$. (For an example, see Eq. (11.2.13).) To cancel these poles, the bare coupling constants $g_B^\ell(d)$ (including masses) must themselves have such poles, with residues fixed by the condition that physical amplitudes be regular as $d \to 4$. These bare couplings in general have non-zero dimensionalities $\Delta_\ell(d)$ that depend on the spacetime dimensionality d, so it is convenient to consider the dimensionless quantity $g_B^\ell(d)\mu^{-\Delta_\ell(d)}$, where μ is a sliding scale with the dimensions of energy or mass. This rescaled bare coupling may be expressed as a sum of terms proportional to positive-definite powers v of $1/(d-4)$, with coefficients b_v fixed by the requirement of cancellation of singularities as $d \to 4$ in physical amplitudes, plus a remainder that is analytic in d at $d = 4$. This remainder is identified as the dimensionless renormalized coupling constant $g^\ell(\mu, d)$, so

$$g_B^\ell(d)\mu^{-\Delta_\ell(d)} = g^\ell(\mu, d) + \sum_{v=1}^{\infty}(d-4)^{-v}\, b_v^\ell(g(\mu, d))\,. \tag{18.6.1}$$

We are free to give the bare couplings any d dependence we like, as long as the singularities at $d = 4$ in physical amplitudes are cancelled; we shall remove this ambiguity by requiring that $g^\ell(\mu, d)$ be analytic in d not only at $d = 4$, but for all d.

To calculate the renormalization group equation satisfied by $g^\ell(\mu, d)$, first differentiate Eq. (18.6.1) with respect to μ:

$$-\Delta^\ell(d)\left[g^\ell + \sum_{v=1}^{\infty}(d-4)^{-v} b_v^\ell(g)\right] = \beta^\ell(g,d) + \sum_{v=1}^{\infty}\sum_m b_{vm}^\ell(g)\,\beta^m(g,d)(d-4)^{-v} \tag{18.6.2}$$

where

$$b_{vm}^\ell(g) \equiv \frac{\partial}{\partial g^m} b_v^\ell(g) \tag{18.6.3}$$

and as before

$$\mu\frac{d}{d\mu} g^\ell(\mu, d) = \beta^\ell(g(\mu, d), d)\,. \tag{18.6.4}$$

Note that β^ℓ is a function of all of the $g^m(\mu, d)$ and also of d, but it cannot depend separately on μ because, with rescaled masses included among the dimensionless coupling parameters, there are no other dimensionful parameters besides μ.

As we have seen, the dimensionalities $\Delta_\ell(d)$ are always linear functions

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

of d, which we shall now write as

$$\Delta_\ell(d) = \Delta_\ell + \rho_\ell(d-4)\,. \tag{18.6.5}$$

We rewrite the left-hand side of Eq. (18.6.2) as

$$-\rho_\ell g^\ell(d-4) - \left[\Delta_\ell g^\ell + b_1^\ell(g)\rho^\ell\right] - \sum_{\nu=1}^{\infty}(d-4)^{-\nu}\left[\rho_\ell\, b_{\nu+1}^\ell(g) + \Delta_\ell b_\nu^\ell(g)\right]\,.$$

The highest power of d in the analytic part here is of first order, so the same must be true on the right-hand side of Eq. (18.6.2), and therefore $\beta(g,d)$ must be linear in d:

$$\beta^\ell(g,d) = \beta^\ell(g) + (d-4)\alpha^\ell(g)\,. \tag{18.6.6}$$

Equating terms of first and zeroth order in (18.6.2) gives then

$$\alpha^\ell(g) = -\rho_\ell\, g^\ell \tag{18.6.7}$$

and, more importantly,

$$\beta^\ell(g) = -\Delta_\ell g^\ell - b_1^\ell(g)\rho_\ell + \sum_{\nu=1}^{\infty}\sum_m b_{1m}^\ell(g)\,\rho_m g^m\,. \tag{18.6.8}$$

It is noteworthy that the beta-function depends only on the coefficients of the simple pole in the bare couplings. In fact, these coefficients also determine the coefficients of all the higher poles; equating the pole terms on the right and left of Eq. (18.6.2) yields the recursion relation

$$\rho_\ell b_{\nu+1}^\ell(g) - \sum_m \rho_m g^m b_{\nu+1\,m}^\ell(g) = -\Delta_\ell b_\nu^\ell(g) - \sum_m b_{\nu m}^\ell(g)\beta^m(g)\,. \tag{18.6.9}$$

For instance, in order for $\int d^dx\, F_{\mu\nu}F^{\mu\nu}$ to be dimensionless, any gauge field A^μ must have dimensions (in powers of mass) $(d-2)/2$, and since $g_B A^\mu$ must have the same dimensions as $\partial/\partial x^\mu$, g_B must have dimensions $(4-d)/2$, so that for gauge couplings $\Delta = 0$ and $\rho = -1/2$. Eq. (18.6.8) gives then for a gauge theory with a single coupling constant:

$$\beta(g) = \tfrac{1}{2}\left[b_1(g) - g b_1'(g)\right]\,. \tag{18.6.10}$$

In particular, Eq. (11.2.20) shows that in quantum electrodynamics in one-loop order the bare electric charge has a pole at $d \to 4$ with

$$e_B = Z_3^{-1/2} e \to e - \frac{e^3}{12\pi^2}\frac{1}{d-4}\,. \tag{18.6.11}$$

Setting $b_1(e) = -e^3/12\pi^2$ in Eq. (18.6.10) gives then

$$\beta(e) = \frac{e^3}{12\pi^2}\,, \tag{18.6.12}$$

in agreement with the previous result (18.2.37).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

The coupling constants $g^\ell(\mu)$ introduced in this section are said to be defined by *minimal subtraction.* There is a slightly different scheme that is somewhat more convenient. The simple poles $(d-4)^{-1}$ typically arise from functions $(4\pi)^{d/2-2}\Gamma(2-d/2)$ (as in Eq. (11.2.13)) which for $d \to 4$ have the limit

$$(4\pi)^{2-d/2}\,\Gamma\left(2-\frac{d}{2}\right) \to \frac{1}{2-d/2} - \gamma + \ln 4\pi\,, \tag{18.6.13}$$

where γ is the Euler constant, $\gamma = 0.5772157$. It is therefore convenient to make the replacement everywhere in Eq. (18.6.1):

$$\frac{1}{d-4} \to \frac{1}{d-4} + \frac{\gamma}{2} - \frac{1}{2}\ln 4\pi\,. \tag{18.6.14}$$

With this prescription, the coupling constants are said to be defined by *modified minimal subtraction.*

One of the distinguishing characteristics of the definition of couplings by minimal subtraction (or modified minimal subtraction) is that, since no factors of an ultraviolet cut-off ever appear in any calculation, loop diagrams have poles at $d = 4$ that correspond only to logarithmic ultraviolet divergences, not divergences that are linear, quadratic, etc. Hence a residue function $b_1^\ell(g)$ can contain a term of order $g^a g^b g^c \cdots$ only if at $d = 4$ the dimensionality of g_B^ℓ equals the total dimensionality of the couplings g_B^a, g_B^b, g_B^c, etc.:

$$\Delta_\ell = \Delta_a + \Delta_b + \Delta_c + \dots\,, \tag{18.6.15}$$

and it follows then from (18.6.8) that the same is true of $\beta^\ell(g)$. In particular, in a theory with no superrenormalizable couplings (like a gauge theory with massless spinors and no scalars, such as quantum electrodynamics with vanishing electron mass) all couplings have $\Delta_\ell \leq 0$, so *the renormalization group equations for the renormalizable couplings (with $\Delta_\ell = 0$) are unaffected by the presence of any non-renormalizable interactions.*[5] Also, in such a theory the beta-functions for the non-renormalizable couplings are polynomials of finite order in the non-renormalizable couplings, with each coefficient in each polynomial given by an infinite series of powers of the renormalizable couplings. For instance, in the theory of photons and massless electrons (assuming invariance under $\psi \to \gamma_5\psi$ and P), there are no nonrenormalizable interactions of dimensionality +5, and several interactions of dimensionality +6 (four-fermion interactions as well as the purely photonic interaction $F_{\mu\nu}\Box F^{\mu\nu}$) with couplings f_i of dimensionality −2. The beta-function for f_i is of the form $\sum_j b_{ij}(e)f_j$, with the coefficients $b_{ij}(e)$ given by a power series in e.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

18.7 Quantum Chromodynamics

Quantum chromodynamics is the modern theory of strong interactions. It is a non-Abelian gauge theory, based on the gauge group $SU(3)$. In addition to the gauge fields, quantum chromodynamics involves fields of spin $\frac{1}{2}$ particles known as quarks. There are quarks of six types, or 'flavors', the u, c, and t quarks having charge $2e/3$, and the d, s, and b quarks having charge $-e/3$. Quarks of each flavor come in three 'colors' which furnish the defining representation **3** of the $SU(3)$ gauge group.* Baryons like the protons and neutron may be approximately regarded as color-neutral bound states of three quarks, totally antisymmetric in quark colors, while mesons like the rho meson behave approximately like color-neutral bound states of quarks and antiquarks.[11]

In the approximation where the quark masses may be regarded as negligible compared with the energies of interest, the inversion of Eq. (17.5.44) shows that the bare coupling constant in a general gauge theory has a pole at spacetime dimensionality $d \to 4$ with residue given by

$$g_B \to \left[\frac{g^3}{4\pi^2} \left(\frac{11}{12} C_1 - \frac{1}{3} C_2 \right) + O(g^5) \right] \frac{1}{d-4} ,$$

where C_1 and C_2 are defined by Eqs. (17.5.33) and (17.5.34). That is, in the notation of the previous section,

$$b_1(g) = \frac{g^3}{4\pi^2} \left(\frac{11}{12} C_1 - \frac{1}{3} C_2 \right) + O(g^5) . \tag{18.7.1}$$

Using this in Eq. (18.6.10) gives

$$\beta(g) = -\frac{g^3}{4\pi^2} \left(\frac{11}{12} C_1 - \frac{1}{3} C_2 \right) + O(g^5) . \tag{18.7.2}$$

For an $SU(3)$ theory with n_f massless quarks in the defining representation **3** of $SU(3)$, Eq. (17.5.35) gives

$$C_1 = 3 , \qquad C_2 = n_f/2 . \tag{18.7.3}$$

Because we are taking the quarks as massless here, this formula may be applied only in the effective field theory obtained by integrating out all quarks heavier than the typical energy E under consideration, so that n_f is the number of quark flavors with masses much less than E. With this

* Before the final formulation of quantum chromodynamics several authors had speculated that there might be three varieties of quarks of each flavor,[10] both in order to account for the rate of decays like $\pi^0 \to \gamma + \gamma$ (see Sections 22.1 and 22.2) and to introduce an additional degree of freedom that would explain how the wave function of fermionic quarks in a baryon could be symmetric in spin, space, and flavor coordinates.[11]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

understanding, Eqs. (18.7.2) and (18.7.3) yield

$$\beta(g) = -\frac{g^3}{4\pi^2}\left(\frac{11}{4} - \frac{1}{6}n_f\right) + O(g^5)\,. \tag{18.7.4}$$

We see that the theory is asymptotically free as long as there are no more than 16 quark flavors with masses below the energy scale of interest. Since in fact there seem to be only six quark flavors of any mass, the theory of strong interactions based on the gauge group $SU(3)$ is asymptotically free.

It was the 1973 discovery of asymptotic freedom in non-Abelian gauge theories of this sort by Gross and Wilczek[12] and Politzer[13] that convinced theoretical physicists that this is the correct theory of strong interactions. Their calculation immediately explained the puzzling result of a famous 1968 experiment[14] at SLAC on deep-inelastic electron–nucleon scattering, that strong interactions seem to get weaker at high energies.** (This experiment will be discussed further in Section 20.6.) But the historical importance of the discovery of asymptotic freedom in Yang–Mills theories is not just that it explained an old experimental result; it for the first time opened up the prospect of doing reliable perturbative calculations of strong interaction processes, at least at high energy.

Asymptotic freedom was soon found to have another important implication. At first after the discovery of asymptotic freedom it was widely assumed that the gauge bosons in a realistic Yang–Mills theory of strong interactions would have to be quite heavy, to explain why these strongly-interacting bosons had not been discovered long before. Following the precedent of the theory of weak and electromagnetic interactions (discussed in Chapter 21), it was supposed that the masses of the gauge bosons arose from a spontaneous breakdown of the color $SU(3)$ gauge group, triggered by the vacuum expectation values of scalar fields in a nontrivial representation of this group. But these strongly interacting scalars would contribute positive terms to $\beta(g)$, which could destroy asymptotic freedom. Even worse, in a theory with strongly interacting scalar fields, radiative corrections involving weak interactions would introduce large violations of various symmetries like charge conjugation invariance and flavor conservation which, as we shall see, would not be violated without the scalars.[17] Then it was suggested to drop the strongly-interacting

** Zee[15] and perhaps other theorists had already understood that this experimental result could be understood in a theory with a beta-function that becomes negative for small positive coupling, but calculations of $\beta(g)$ in all renormalizable field theories *except* non-Abelian gauge theories gave $\beta(g) > 0$. On the other hand, by 1972 't Hooft had developed techniques for calculating $\beta(g)$ in Yang–Mills theories, and in June 1972 he announced at a conference on gauge theory at Marseilles[16] that $\beta(g) < 0$, but he waited to publish this result and work out its implications while he was doing other things, so his result did not attract much attention.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

scalars, and accept the consequence that the gluons, the $SU(3)$ gauge bosons, have zero mass.[18] The decrease of the strong coupling constant at high energy or short distance of course implies an increase at low energy or large distance, and it was suggested that this might explain why massless gluons and quarks had not been detected. According to this hypothesis, only color-neutral particles like baryons or mesons will ever appear in isolation.[19] This is unfortunately still a hypothesis rather than a theorem, but after two decades there seems to be little doubt that it is correct.

Even though quarks cannot materialize as free particles, they are in a sense observed as jets produced in high energy collision processes. For instance, in many events in electron–positron annihilation, the final state consists of two narrowly collimated hadron jets, with a distribution in the angle θ between the colliding lepton momenta and the jet directions (in the center-of-mass system) given by $1 + \sin^2\theta$, just as expected from the tree graph for electron–positron annihilation into quark–antiquark final states.[20] This can be understood[21] in terms of the general analysis of infrared divergences in Section 13.4. At extremely high energies we would expect the rate for a physical process to be given by lowest-order perturbation theory, provided it is 'infrared-safe,' in the sense of not becoming infrared divergent when all masses are taken to zero. The total rate for electron-positron annihilation into hadrons is infrared-safe, since we sum over all hadronic final states. (We are ignoring higher-order electromagnetic effects here.) Therefore we can rely on perturbation theory, which immediately tells us that the ratio R of this rate to the rate for $e^+ + e^- \to \mu^+ + \mu^-$ is $R = 3\sum_q Q_q^2$, where the sum runs over all quark flavors, Q_q is their charge in units of e, and the factor 3 is the number of colors. (For instance, in the wide energy range between $m_b \approx 4.5$ GeV and $m_t \approx 180$ GeV, $R \simeq 3(2(2/3)^2 + 3(-1/3)^2) = 11/3$.) On the other hand, the rate for electron–positron annihilation into some definite state of quarks and gluons is not infrared-safe, and its rate therefore cannot be calculated in perturbation theory at all; in fact, it is zero. In between these two extremes is the rate for electron–positron annihilation into a definite number of jets, each jet carrying a definite total momentum and charge, together with a set of unobserved hadrons with limited total energy outside the jets. As discussed in Section 13.4, this rate is infrared-safe. It can therefore be calculated at high energy in the tree approximation of perturbation theory, identifying jets (in this order of perturbation theory) with the outgoing quarks, antiquarks, and gluons. We can even calculate the rate for three-jet events, arising from tree diagrams in which a gluon is emitted from the outgoing quark or antiquark, and use the comparison of the results with experiment to measure the value of $\alpha_s(\mu)$.[22] But we cannot use perturbation theory to predict the distribution of momenta within a jet, because such a differential rate is not infrared-safe. Similar remarks

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

apply to the production of jets in deep inelastic lepton–hadron collisions, to be discussed in Section 20.6, but the presence of hadrons in the initial state makes the analysis more complicated.

Following the same reasoning as in Section 12.5, with no scalar fields the most general renormalizable Lagrangian for quantum chromodynamics can be put in the form

$$\mathscr{L} = -\tfrac{1}{4}F_\alpha^{\mu\nu}F_{\alpha\mu\nu} - \sum_n \bar{\psi}_n[\not{\partial} - ig\,\not{A}_\alpha t_\alpha + m_n]\psi_n\,, \tag{18.7.5}$$

where A_α^μ is the color gauge vector potential; $F_\alpha^{\mu\nu}$ is the color gauge-covariant field strength tensor; g is the strong coupling constant; t_α are a complete set of generators of color $SU(3)$ in the **3** representation (that is, Hermitian traceless 3×3 matrices with rows and columns labelled by the three quark colors), normalized so that $\mathrm{Tr}(t_\alpha t_\beta) = \frac{1}{2}\delta_{\alpha\beta}$; and the subscript n labels quark flavors, with quark color indices suppressed. Just as we found for electrodynamics in Section 12.5, this Lagrangian has important accidental symmetries: it conserves space parity,[†] charge conjugation parity, and the numbers of quarks of each flavor (minus the number of the corresponding antiquarks), including the long-established 'strangeness' quantum number, which counts the numbers of 's' quarks. Thus quantum chromodynamics immediately explained the mysterious fact that the strong interactions respect various symmetries that are not symmetries of all interactions. This argument also makes it clear why, as mentioned earlier, in this theory the weak interactions do not introduce large violations of parity, charge conjugation, strangeness, etc. Since all renormalizable interactions among quarks and gluons conserve these symmetries, at energies E much less than the masses m_W of the particles that carry the weak interactions these symmetries could be violated only by non-renormalizable terms in the effective field theory, such as $\bar{\psi}\psi\bar{\psi}\psi$ interactions, which as discussed in Section 12.3 would be suppressed by negative powers of m_W as well as by the coupling constants of the weak interactions.

Of course, it is possible that the quarks and gluons exhibit some new kind of *strong* interaction at an energy scale Λ' much larger than the scale Λ characteristic of quantum chromodynamics. For instance, as discussed in Section 22.5, the quarks might be bound states of more fundamental fermions, which interact with gauge fields whose asymptotically free couplings become strong at energies of order Λ', trapping them into the quarks. In that case the effective Lagrangian density for quarks at en-

[†] Although it was not known in 1973, we shall see in Section 23.6 that non-perturbative effects can violate parity in quantum chromodynamics. Various ways of avoiding strong parity violation have been suggested, but it is not yet clear which is correct.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

ergies $E \ll \Lambda'$ would contain non-renormalizable interactions such as $\bar{\psi}\psi\bar{\psi}\psi$, which are suppressed only by powers of E/Λ'. These interactions could show up, not only in small violations of symmetries like parity and quark flavor conservation at ordinary energies, but also in departures[23] from the quantitative predictions of quantum chromodynamics at energies approaching Λ'.

Now let us consider the behaviour of the coupling constant of quantum chromodynamics in greater detail. In lowest order, the renormalization group equation is given by Eq. (18.7.4) as

$$\mu\frac{d}{d\mu}g(\mu) = -\frac{g^3(\mu)}{4\pi^2}\left(\frac{11}{4} - \frac{1}{6}n_f\right) . \tag{18.7.6}$$

The solution is

$$\alpha_s(\mu) \equiv \frac{g^2(\mu)}{4\pi} = \frac{12\pi}{(33-2n_f)\ln(\mu^2/\Lambda^2)} , \tag{18.7.7}$$

where Λ is an integration constant. This formula exhibits a characteristic property of theories of massless (or, for the quarks, approximately massless) particles: in such theories one of the dimensionless couplings in the Lagrangian is exchanged for a free dimensionful parameter. Eq. (18.7.7) involves no free dimensionless parameters, but it does involve one free parameter with the dimensions of mass, the integration constant Λ.

These calculations have been carried to three-loop order. The renormalization group equations to this order are[24]

$$\mu\frac{d}{d\mu}g(\mu) = -\beta_0\frac{g^3(\mu)}{16\pi^2} - \beta_1\frac{g^5(\mu)}{128\pi^4} - \beta_2\frac{g^7(\mu)}{8192\pi^6} , \tag{18.7.8}$$

where β_n are the numerical coefficients:

$$\beta_0 = 11 - \frac{2}{3}n_f , \tag{18.7.9}$$

$$\beta_1 = 51 - \frac{19}{3}n_f , \tag{18.7.10}$$

$$\beta_2 = 2857 - \frac{5033}{9}n_f - \frac{325}{27}n_f^2 . \tag{18.7.11}$$

The solution is

$$\begin{aligned}\alpha_s(\mu) &\equiv \frac{g^2(\mu)}{4\pi} \\ &= \frac{4\pi}{\beta_0\ln(\mu^2/\Lambda^2)}\left[1 - \frac{2\beta_1}{\beta_0^2}\frac{\ln[\ln(\mu^2/\Lambda^2)]}{\ln(\mu^2/\Lambda^2)}\right. \\ &\quad \left. + \frac{4\beta_1^2}{\beta_0^4\ln^2(\mu^2/\Lambda^2)}\left(\left(\ln[\ln(\mu^2/\Lambda^2)] - \tfrac{1}{2}\right)^2 + \frac{8\beta_2\beta_0}{\beta_1^2} - \frac{5}{4}\right)\right] \end{aligned} \tag{18.7.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

It should be recalled that n_f in the above results is the number of quark flavors with masses below the energies of interest. In each energy range between any two successive quark masses we have a different value of n_f, and also a different Λ, chosen to make $g(\mu)$ continuous at each quark mass. In particular, experiments on the deep inelastic scattering of electrons typically involve energies above only the first four quark flavors (u, d, s, and c), so here we must take $n_f = 4$. On the other hand, experiments at electron–positron colliders like PEP, PETRA, TRISTRAN, and LEP are at energies well above the fifth (b) quark mass, so in these experiments we must take $n_f = 5$. But these results may be expressed in terms of those for $n_f = 4$ by matching the solutions of the renormalization group equations at the b quark mass. In this way it is found[25] (using the modified minimum subtraction prescription in calculating β_2) that the strong coupling extrapolated to $m_Z = 91.2$ GeV is $\alpha_s(m_Z) \equiv g_s^2(m_Z)/4\pi = 0.118 \pm 0.006$, corresponding to $\Lambda \approx 250$ MeV for energies μ with $m_b \ll \mu \ll m_t$, where $n_f = 5$. A more recent study[26] of hadronic production in e^+–e^- annihilation at the Z resonance has given a directly measured value $\alpha_s(m_Z) = 0.1200 \pm 0.0025$, with a theoretical uncertainty of ± 0.0078, corresponding to $\Lambda = 253^{+130}_{-96}$ MeV.

18.8 Improved Perturbation Theory*

The ground-breaking paper[1] of Gell-Mann and Low was in large part directed to the problem of 'improving' perturbation theory — that is, of using the ideas of the renormalization group and the results of perturbation theory to a given order to say something about the next order of perturbation theory. To illustrate this, let's return to the specific case studied by Gell-Mann and Low: vacuum polarization in quantum electrodynamics.

Recall that the renormalized electric charge e_μ at a sliding scale μ is given by Eq. (18.2.36) in terms of the bare charge e_B as

$$e_\mu = N_\mu^{(A)-1} e_B \,, \tag{18.8.1}$$

where $N_\mu^{(A)}$ is the constant which, when multiplied into the unrenormalized electromagnetic field, gives a field renormalized at scale μ. (See Eq. (18.2.21).) Thus we can define a renormalized (and hence cut-off-independent) complete photon propagator $\Delta'_{\rho\sigma}(q, \mu, e_\mu)$ in terms of the complete propagator $\Delta'_{B\rho\sigma}(q, e_B)$ of the unrenormalized field as

$$\Delta'_{\rho\sigma}(q, \mu, e_\mu) = N_\mu^{(A)2} \Delta'_{B\rho\sigma}(q, e_B) \tag{18.8.2}$$

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

in such a way that the function $e_\mu^2 \Delta'_{\rho\sigma}(q, \mu, e_\mu)$ is both independent of μ, because it equals $e_B^2 \Delta'_{B\rho\sigma}(q, e_B)$, and independent of the cut-off, because both e_μ and $\Delta'_{\rho\sigma}(q, \mu, e_\mu)$ are renormalized quantities. (We are not explicitly displaying cut-off dependence here.) But Lorentz invariance and dimensional analysis tell us that this function must take the form:

$$e_\mu^2 \Delta'_{\rho\sigma}(q, \mu, e_\mu) = \frac{\eta_{\rho\sigma} d(q^2/\mu^2, e_\mu)}{q^2} + q_\rho q_\sigma \text{ terms} . \tag{18.8.3}$$

Since Eq. (18.8.3) is μ-independent, we can set $\mu = \sqrt{q^2} \equiv q$ here, so that

$$d(q^2/\mu^2, e_\mu) = d(1, e_q) . \tag{18.8.4}$$

Now let us see what this tells us about the structure of the perturbation series for $d(q^2/\mu^2, e_\mu)$. The beta-function for e has the expansion:

$$\beta(e) = b_1 e^3 + b_2 e^5 + b_3 e^7 + \dots . \tag{18.8.5}$$

The renormalization group equation for e_μ then has a power series solution

$$\begin{aligned} e_q^2 = e_\mu^2 &- b_1 e_\mu^4 \ln \frac{q^2}{\mu^2} - b_2 e_\mu^6 \ln \frac{q^2}{\mu^2} \\ &- \left(\frac{b_1 b_2}{2} \ln^2 \frac{q^2}{\mu^2} + b_3 \ln \frac{q^2}{\mu^2} \right) e_\mu^8 + \dots \quad . \end{aligned} \tag{18.8.6}$$

If we also expand d:

$$d(1, e) = e^2 + d_1 e^4 + d_2 e^6 + d_3 e^8 + \dots \tag{18.8.7}$$

then

$$\begin{aligned} d(q^2/\mu^2, e_\mu) = d(1, e_q) = e_\mu^2 &- \left(b_1 \ln \frac{q^2}{\mu^2} - d_1 \right) e_\mu^4 - \left(b_2 \ln \frac{q^2}{\mu^2} - d_2 \right) e_\mu^6 \\ &- \left(\frac{1}{2} b_1 b_2 \ln^2 \frac{q^2}{\mu^2} + (b_3 - b_1 d_2) \ln \frac{q^2}{\mu^2} - d_3 \right) e_\mu^8 + \dots \quad . \end{aligned} \tag{18.8.8}$$

Note that the leading powers of $\ln(q^2/\mu^2)$ in each order of $d(q^2/\mu^2, e_\mu)$ are respectively $0, 1, 1, 2, 3, \dots$. Also, if we calculate $d(q^2/\mu^2, e_\mu)$ to order e_μ^6 and thus determine b_1 and b_2, we can immediately write down the coefficient of the leading logarithm in order e_μ^8, as $-\frac{1}{2} b_1 b_2$. None of this would be easy to infer without using the method of the renormalization group.

Problems

1. Consider an $SU(N)$ gauge theory with a scalar field in the defining representation of $SU(N)$. Calculate the beta-function for the gauge

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

coupling to one-loop order, including the contribution of a scalar loop. (Recommendation: Use the background field gauge, with a constant background field.)

2. Suppose that the beta-function $\beta(g)$ for a theory with positive coupling constant g has a simple zero at $g = g_*$, where $\beta(g) \to a(g_* - g)$ with $a > 0$. What is the asymptotic behavior of the correction to the leading term $\propto E^{-\gamma^{\mathcal{O}}(g_*)}$ in the factor $N_E^{\mathcal{O}-1}$ associated with the inclusion of an operator $\mathcal{O}$ in a vacuum expectation value.

3. Show that in a theory with $\beta(g) = bg^2 + b'g^3 + b''g^4 + \cdots$, it is possible by a redefinition of the coupling constant to make the coefficient b'' anything we want.

4. Calculate the effective electric charge that should be used in studying processes at energy 100 GeV, taking account of all known charged quarks and leptons with masses below 100 GeV.

5. Calculate the asymptotic behavior for large four-momentum of the electron propagator in quantum electrodynamics. (You may use the one-loop value of Z_2 calculated elsewhere, for instance in Section 11.4.)

6. Calculate the anomalous exponent ν to first order in the expansion in $\epsilon = 4 - d$ for an $O(N)$-invariant theory of scalar fields $\phi_n(x)$ with $n = 1, \cdots, N$ belonging to the vector representation of $O(N)$, and an interaction $\frac{1}{4}g(\sum_n \phi_n^2)^2$.

References

1. M. Gell-Mann and F. E. Low, *Phys. Rev.* **95**, 1300 (1954). The freedom of choosing the definition of renormalized coupling constants had been discussed a little earlier by E. C. G. Stueckelberg and A. Peterman, *Helv. Phys. Acta* **26**, 499 (1953) (who also introduced the unfortunate term 'renormalization group') but without an explanation of its relevance to the calculation of physical processes at very high energy or very low energy. After the work of Gell-Mann and Low renormalization group methods were further developed by N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959): Chapter VIII, and references quoted therein. Interest in renormalization group methods in particle physics was revived in 1970 by C. G. Callan, *Phys. Rev.* **D2**, 1541 (1970); K. Symanzik, *Commun. Math. Phys.* **18**, 227 (1970);

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

C. G. Callan, S. Coleman, and R. Jackiw, *Ann. of Phys.* (New York), **59**, 42 (1970).

2. K. G. Wilson, *Phys. Rev.* **B4**, 3174, 3184 (1971); *Rev. Mod. Phys.* **47**, 773 (1975).

3. J. C. Collins, *Phys. Rev.* **D10**, 1213 (1974).

3*a*. R. Jost and J. M. Luttinger, *Helv. Phys. Acta* **23**, 201 (1950).

4. This is the value quoted by the Particle Data Group in 1994. More recent calculations have been summarized by G. Altarelli, CERN preprint CERN-TH-95/203, to be published in *Proceedings of the Workshop on Physics at DAΦNE, April 1995.* These more recent values of $\alpha^{-1}(m_Z)$ range from 128.89 to 129.08.

4*a*. S. Coleman and E. Weinberg, *Phys. Rev.* **D7**, 1888 (1973).

4*b*. L. D. Landau, in *Niels Bohr and the Development of Physics* (Pergamon Press, New York, 1955): p. 52; and earlier works cited therein.

4*c*. S. L. Adler, C. G. Callan, D. J. Gross, and R. Jackiw, *Phys. Rev.* **D6**, 2982 (1972); M. Baker and K. Johnson, *Physica* **96A**, 120 (1979).

4*d*. For a discussion and references, see J. Glimm and A. Jaffe, *Quantum Physics – A Functional Integral Point of View*, Second Edition (Springer-Verlag, New York, 1987): Section 21.6; R. Fernandez, J. Fröhlich, and A. D. Sokal, *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory* (Springer-Verlag, Berlin, 1992): Chapter 15.

5. S. Weinberg, in *General Relativity*, eds. S. W. Hawking and W. Israel, eds. (Cambridge University Press, Cambridge, 1979): p. 790.

6. S. Weinberg, *Phys. Rev.* **D8**, 3497 (1973).

7. The original calculation is by K. G. Wilson and M. E. Fisher, *Phys. Rev. Lett.* **28**, 240 (1972); K. G. Wilson, *Phys. Rev. Lett.* **28**, 548 (1972). For reviews, see K. G. Wilson and J. Kogut, *Phys. Rep.* **12C**, No. 2 (1974); M. E. Fisher, *Rev. Mod. Phys.* **46**, 597 (1974); E. Brézin, J. C. Le Guillou, and J. Zinn-Justin, in *Phase Transitions and Critical Phenomena*, eds. C. Domb and M. S. Green (Academic Press, London, 1975).

8. See, e. g., P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995): p. 231.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

8*a*. E. Brézin, J. C. Le Guillou, J. Zinn-Justin, and B. G. Nickel, *Phys. Lett.* **44A**, 227 (1973); K. G. Wilson and J. Kogut, Ref. 7, Section 8.

9. G. 't Hooft, *Nucl. Phys.* **B61**, 455 (1973); *Nucl. Phys.* **B82**, 444 (1973). The derivation presented here is a somewhat simplified version of 't Hooft's.

10. O. W. Greenberg, *Phys. Rev. Lett.*, **13**, 598 (1964); M. Y. Han and Y. Nambu, *Phys. Rev.* **139**, B1006 (1965); W. A. Bardeen, H. Fritzsch, and M. Gell-Mann, in *Scale and Conformal Invariance in Hadron Physics*, ed. R. Gatto (Wiley, New York, 1973).

11. M. Gell-Mann, *Phys. Lett.* **8**, 214 (1964); G. Zweig, CERN preprint TH401 (1964).

12. D. J. Gross and F. Wilczek, *Phys. Rev. Lett.*, **30**, 1343 (1973).

13. H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973)

14. E. D. Bloom *et al.*, *Phys. Rev. Lett.* **23**, 930 (1969); M. Breidenbach *et al.*, *Phys. Rev. Lett.* **23**, 933 (1969); J. L. Friedman and H. W. Kendall, *Annual Reviews of Nuclear Science* **22**, 203 (1972).

15. A. Zee, unpublished.

16. G. 't Hooft, unpublished.

17. S. Weinberg, *Phys. Rev.* **D8**, 605 (1973).

18. D. J. Gross and F. Wilczek, *Phys. Rev.* **D8**, 3633 (1973); S. Weinberg, *Phys. Rev. Lett.* **31**, 494 (1973).

19. A similar idea had been suggested before the discovery of asymptotic freedom by H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Phys. Lett.* **47B**, 365 (1973).

20. G. Hanson *et al.*, *Phys. Rev. Lett.* **35**, 1609 (1975); R. F. Schwitters, in *Proceedings of the International Conference on Lepton and Photon Interactions at High Energy at Stanford, 1975*, ed. W. T. Kirk (Stanford Linear Accelerator Center, Stanford, 1975): p. 5; G. Hanson, Stanford Linear Accelerator Center Report SLAC–PUB–1814 (1976), unpublished.

21. G. Sterman and S. Weinberg, *Phys. Rev. Lett.* **39**, 1436 (1977).

22. J. Ellis, M. K. Gaillard, and G. G. Ross, *Nucl. Phys.* **B111**, 253 (1976).

23. E. Eichten, K. Lane, and M. Peskin, *Phys. Rev. Lett.* **50**, 811 (1983).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

24. For a review, see I. Hinchliffe, in 'Review of Particle Properties,' *Phys. Rev.* **D50**, 1177 (1994): Section 25.

25. G. Altarelli, in *Proceedings of the Rencontres de Hanoi*, CERN preprint CERN-PPE/94-71 (1994).

26. K. Abe *et al.* (SLD collaboration), *Phys. Rev.* **D51**, 962 (1995). A summary of earlier data on Z^0 decay into hadrons by M. Shifman, Minnesota preprint hep-ph/9501222 (1995), gave a value $\alpha_s(m_Z) = 0.125 \pm 0.005$, corresponding to $\Lambda \approx 500$ MeV.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.006

19

Spontaneously Broken Global Symmetries

Much of the physics of this century has been built on principles of symmetry: first the spacetime symmetries of Einstein's 1905 special theory of relativity, and then internal symmetries, such as the approximate $SU(2)$ isospin symmetry of the 1930s. It was therefore exciting when in the 1960s it was discovered that there are more internal symmetries than could be guessed by inspection of the spectrum of elementary particles. There are exact or approximate symmetries of the underlying theory that are 'spontaneously broken,' in the sense that they are not realized as symmetry transformations of the physical states of the theory, and in particular do not leave the vacuum state invariant. The breakthrough was the discovery of a broken approximate global $SU(2) \times SU(2)$ symmetry of the strong interactions, which will be discussed in detail in Section 19.3. This was soon followed by the discovery of an exact but spontaneously broken local $SU(2) \times U(1)$ symmetry of the weak and electromagnetic interactions, which will be taken up along with more general broken local symmetries in Chapter 21. In this chapter we shall begin with a general discussion of broken global symmetries, and then move on to physical examples.

19.1 Degenerate Vacua

We do not have to look far for examples of spontaneous symmetry breaking. Consider a chair. The equations governing the atoms of the chair are rotationally symmetric, but a solution of these equations, the actual chair, has a definite orientation in space. Here we will be concerned not so much with the breaking of symmetries by *objects* like chairs, but rather with the symmetry breaking in the ground state of any realistic quantum field theory, the vacuum.

A spontaneously broken symmetry in field theory is always associated with a degeneracy of vacuum states. For instance, consider a symmetry transformation of the action, and of the measure used in integrating over

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

fields, that acts linearly on a set of scalar fields $\phi_n(x)$:

$$\phi_n(x) \to \phi'_n(x) = \sum_m L_{nm}\phi_m(x) \,. \tag{19.1.1}$$

(The ϕ_n need not be elementary fields; they can be composite objects, like $\bar{\psi}\Gamma_n\psi$.) As we saw in Section 16.4, the quantum effective action $\Gamma[\phi]$ will then have the same symmetry

$$\Gamma[\phi] = \Gamma[L\phi] \,. \tag{19.1.2}$$

For the vacuum, the expectation value of $\phi(x)$ must be at a minimum of the vacuum energy $-\Gamma[\phi]$, say at $\phi(x) = \bar{\phi}$ (a constant). But if $L\bar{\phi} \neq \bar{\phi}$, then this vacuum is not unique; $-\Gamma[\phi]$ has the same value at $\phi = L\bar{\phi}$ as it does at $\bar{\phi}$. In the simple special case where the symmetry transformation (19.1.1) is a reflection, $\phi \to -\phi$, if $-\Gamma(\phi)$ has a minimum at a non-zero value $\bar{\phi}$ of ϕ, then it has two minima, at $\bar{\phi}$ and $-\bar{\phi}$, each corresponding to a state of broken symmetry.

We are not yet ready to conclude that in such cases the symmetry is broken, because we have not yet ruled out the possibility that the true vacuum is a linear superposition of vacuum states in which ϕ_m has various expectation values, which would respect the assumed symmetry. For instance, in a theory with a symmetry $\phi \to -\phi$, even if $\Gamma(\phi)$ has a minimum for some non-zero value $\bar{\phi}$ of ϕ, how do we know that the true vacuum is one of the states $|\text{VAC}, \pm\rangle$ for which Φ has expectation values $\bar{\phi}$ and $-\bar{\phi}$, and not some linear combination like $|\text{VAC}, +\rangle + |\text{VAC}, -\rangle$ that *would* respect the symmetry under $\phi \to -\phi$? The assumed symmetry under the transformation $\phi \to -\phi$ tells us that the vacuum matrix elements of the Hamiltonian are

$$\langle \text{VAC}, +|H|\text{VAC}, +\rangle = \langle \text{VAC}, -|H|\text{VAC}, -\rangle \equiv a$$

(with a real) and

$$\langle \text{VAC}, +|H|\text{VAC}, -\rangle = \langle \text{VAC}, -|H|\text{VAC}, +\rangle \equiv b \,,$$

(with b real), so the eigenstates of the Hamiltonian are $|\text{VAC}, +\rangle \pm |\text{VAC}, -\rangle$, with energies $a \pm |b|$. These energy eigenstates are invariant (or invariant up to a sign) under the symmetry $\phi \to -\phi$. In fact, the same issue also arises for chairs. The quantum mechanical ground state of an isolated chair is actually rotationally invariant; it is a state with zero angular momentum quantum numbers, and hence with no definite orientation in space.

Spontaneous symmetry breaking actually occurs only for idealized systems that are infinitely large. The appearance of broken symmetry for a chair arises because it has a macroscopic moment of inertia I, so that its ground state is part of a tower of rotationally excited states whose

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

energies are separated by only tiny amounts, of order $\hbar^2/I$. This gives the state vector of the chair an exquisite sensitivity to external perturbations; even very weak external fields will shift the energy by much more than the energy difference of these rotational levels. In consequence, any rotationally asymmetric external field will cause the ground state or any other state of the chair with definite angular momentum quantum numbers rapidly to develop components with other angular momentum quantum numbers. The states of the chair that are relatively stable with respect to small external perturbations are not those with definite angular momentum quantum numbers, but rather those with a definite orientation, in which the rotational symmetry of the underlying theory is broken.

For the vacuum also, the possibility of spontaneous symmetry breaking is again related to the large size of the system, specifically to the large volume of space. In the above example of a reflection symmetry, the off-diagonal matrix element b of the Hamiltonian involves an integration over field configurations that tunnel from the minimum at $\phi = \bar{\phi}$ to the one at $\phi = -\bar{\phi}$, so it is smaller than the diagonal matrix element a by a barrier penetration factor that for a spatial volume $\mathcal{V}$ is of the form $\exp(-C\mathcal{V})$, where C is a positive constant* depending on the microscopic parameters of the theory. The two energy eigenstates $|\text{VAC},+\rangle \pm |\text{VAC},-\rangle$ are thus essentially degenerate for any macroscopic volume, and so are strongly mixed by any perturbation that is an odd functional of ϕ. Even if such a perturbation H' is very weak, its diagonal elements $\langle \text{VAC},\pm|H'|\text{VAC},\pm\rangle$ will differ by much more than the exponentially suppressed off-diagonal elements of either H or the perturbation. Thus the vacuum eigenstates of the perturbed Hamiltonian will be very close to either one of the broken symmetry states $|\text{VAC},\pm\rangle$ which diagonalize the perturbation, and *not* to the invariant states $|\text{VAC},+\rangle \pm |\text{VAC},-\rangle$. Which one of the states $|\text{VAC},\pm\rangle$ is the true vacuum for very small perturbations? This depends on the perturbation, but since these two states are related by a symmetry transformation of the original Hamiltonian, it doesn't matter; if the perturbation is sufficiently small, no observer will be able to tell the difference.

The vanishing of matrix elements between vacuum states with different field expectation values becomes exact in a space of infinite volume.[1] For infinite volume, a general vacuum state $|v\rangle$ may be defined as a state with

* For instance, by analogy with the classic wave mechanical problem of barrier penetration, for a Lagrangian density of the form $-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)$, we have $C = \int_{-\bar{\phi}}^{+\bar{\phi}} \sqrt{2V(\phi)}\, d\phi$. We will not bother to calculate the off-diagonal matrix element b here, because we shall soon give a general argument that shows that it vanishes for infinite volume.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

zero momentum

$$\mathbf{P}|v\rangle = 0 \tag{19.1.3}$$

for which this is a *discrete* momentum eigenvalue. (This excludes single-particle or multiparticle states, for which the momentum value zero is always part of a continuum of momentum values in a space of infinite volume.) In general there may be a number of such states. They can usually be expanded in a discrete set, and our notation will treat them as if they were discrete. They will be chosen to be orthonormal

$$\langle u|v\rangle = \delta_{uv} \,. \tag{19.1.4}$$

Any matrix element of a product of local Hermitian operators at equal times between these states may be expressed as a sum over states:

$$\langle u|A(\mathbf{x})\,B(0)|v\rangle = \sum_w \langle u|A(0)|w\rangle\ \langle w|B(0)|v\rangle + \int d^3p \sum_N \langle u|A(0)|N,\mathbf{p}\rangle\ \langle N,\mathbf{p}|B(0)|v\rangle e^{-i\mathbf{p}\cdot\mathbf{x}} \,, \tag{19.1.5}$$

where $|N,\mathbf{p}\rangle$ are a set of orthonormalized continuum states of definite three-momentum $\mathbf{p}$ that together with the $|v\rangle$ span the whole physical Hilbert space. (Here N may include continuous as well as discrete labels. Also, we are dropping time arguments.) We assume without proof that because the $|N,\mathbf{p}\rangle$ belong to the continuous spectrum of the momentum operator $\mathbf{P}$, the dependence of matrix elements on $\mathbf{p}$ is smooth enough (that is, Lebesgue integrable) to allow the use of the Riemann–Lebesgue theorem,[2] so that the integral over $\mathbf{p}$ vanishes as $|\mathbf{x}| \to \infty$. In this limit, we have then

$$\langle u|A(\mathbf{x})\,B(0)|v\rangle \xrightarrow[|\mathbf{x}|\to\infty]{} \sum_w \langle u|A(0)|w\rangle\ \langle w|B(0)|v\rangle \,. \tag{19.1.6}$$

Likewise,

$$\langle u|B(0)\,A(\mathbf{x})|v\rangle \xrightarrow[|\mathbf{x}|\to\infty]{} \sum_w \langle u|B(0)|w\rangle\ \langle w|A(0)|v\rangle \,. \tag{19.1.7}$$

But causality tells us that the equal-time commutator $[A(\mathbf{x}), B(0)]$ vanishes for $\mathbf{x} \neq 0$ (see Section 5.1), so the matrix elements (19.1.6) and (19.1.7) are equal, and thus the Hermitian matrices $\langle u|A(0)|v\rangle$, $\langle u|B(0)|v\rangle$, etc., must all commute with one another. It follows that they can all be simultaneously diagonalized. Changing if necessary to this basis, we have then for every Hermitian local operator $A(\mathbf{x})$ of the theory

$$\langle u|A(0)|v\rangle = \delta_{uv}\, a_v \tag{19.1.8}$$

with a_v a real number, the expectation value of A in the state $|v\rangle$. So for infinite volume any Hamiltonian constructed from local operators will

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

have vanishing matrix elements between the different vacua $|v\rangle$. In the absence of off-diagonal terms in the Hamiltonian, any two $|v\rangle$s connected by a symmetry operation will be degenerate. A symmetry-breaking perturbation built out of such local operators will be diagonal in the same basis, and will therefore yield a ground state that is one of the $|v\rangle$s, rather than a linear combination of them.

It is reassuring that the vacuum states $|v\rangle$ which are stable against small field-dependent perturbations are also vacuum states in which the cluster decomposition condition (see Chapter 4) is satisfied. This principle requires that for the physical vacuum state $|\text{VAC}\rangle$

$$\langle \text{VAC}|A(\mathbf{x})B(0)|\text{VAC}\rangle \xrightarrow[\mathbf{x}\to\infty]{} \langle \text{VAC}|A(\mathbf{x})|\text{VAC}\rangle \, \langle \text{VAC}|B(0)|\text{VAC}\rangle \, . \quad (19.1.9)$$

This condition is satisfied if we take the vacuum state $|\text{VAC}\rangle$ to be any one of the states $|v\rangle$ in the basis defined by Eq. (19.1.8), but not if we take it to be a general linear combination of several of the $|v\rangle$s.

19.2 Goldstone Bosons

We now specialize to the case of a spontaneously broken *continuous* symmetry. In this case there is a theorem, that (with one important exception, to be considered in Chapter 21) the spectrum of physical particles must contain one particle of zero mass and spin for each broken symmetry. Such particles, known as Goldstone bosons (or Nambu–Goldstone bosons) were first encountered in specific models by Goldstone[3] and Nambu[4]; two general proofs of their existence were then given by Goldstone, Salam, and myself.[5] This section will present both of these proofs, and then go on to consider the properties of the Goldstone bosons.

Suppose that the action and measure are invariant under a continuous symmetry, under which a set of Hermitian scalar fields $\phi_n(x)$ (either elementary or composite) are subjected to the linear infinitesimal transformations

$$\phi_n(x) \to \phi_n(x) \; + \; i\epsilon \sum_m t_{nm}\phi_m(x) \, , \quad (19.2.1)$$

with it_{nm} a finite real matrix. As we found in Section 16.4, the effective action is then also invariant under this transformation

$$\sum_{n,m} \int \frac{\delta\Gamma[\phi]}{\delta\phi_n(x)} \, t_{nm}\phi_m(x) \, d^4x = 0 \, . \quad (19.2.2)$$

We shall specialize to the case of a translationally invariant theory with constant fields ϕ_n, where as we saw in Section 16.2, the effective action

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

takes the form

$$\Gamma[\phi] = -\mathscr{V} V(\phi)\,, \tag{19.2.3}$$

where $\mathscr{V}$ is the spacetime volume and $V(\phi)$ is known as the effective potential. Eq. (19.2.2) may then be written

$$\sum_{n,m} \frac{\partial V(\phi)}{\partial \phi_n}\, t_{nm}\, \phi_m = 0\,. \tag{19.2.4}$$

We will use this symmetry requirement in a form obtained by differentiating again with respect to ϕ_ℓ:

$$\sum_n \frac{\partial V(\phi)}{\partial \phi_n}\, t_{n\ell} + \sum_{n,m} \frac{\partial^2 V(\phi)}{\partial \phi_n \partial \phi_\ell}\, t_{nm}\phi_m = 0\,. \tag{19.2.5}$$

Now specialize to the case where ϕ_n is at the minimum of $V(\phi)$, that is, at the vacuum expectation value $\bar{\phi}_n$. Since $V(\phi)$ is stationary at its minimum, the first term in Eq. (19.2.5) vanishes, so

$$\sum_{n,m} \left.\frac{\partial^2 V(\phi)}{\partial \phi_n \partial \phi_\ell}\right|_{\phi=\bar{\phi}} t_{nm}\, \bar{\phi}_m = 0\,. \tag{19.2.6}$$

The general results of Section 16.1 show that the second derivative in Eq. (19.2.6) is just the sum of all connected one-particle-irreducible momentum-space Feynman diagrams with external lines labelled n and ℓ and carrying zero four-momentum. As shown at the end of Section 16.1, it is related to the reciprocal of the momentum-space propagator by

$$\frac{\partial^2 V(\phi)}{\partial \phi_n\, \partial \phi_\ell} = \Delta^{-1}_{n\ell}(0)\,, \tag{19.2.7}$$

so Eq. (19.2.6) gives

$$\sum_{nm} \Delta^{-1}_{n\ell}(0) t_{nm} \bar{\phi}_m = 0\,. \tag{19.2.8}$$

Thus, if the symmetry is broken, so that $\sum_m t_{nm}\bar{\phi}_m$ is non-zero, then this is an eigenvector of $\Delta^{-1}_{n\ell}(0)$ with eigenvalue zero. The existence of such an eigenvector means that $\Delta_{n\ell}(q)$ has a pole at $q^2 = 0$. The rank of the residue of the pole at $q^2 = 0$ is equal to the dimensionality of the space of the vectors $t\bar{\phi}$, with t running over all the generators of continuous symmetries of the theory. Roughly speaking, there is one massless boson for every independent broken symmetry.

In the classic example of a broken symmetry, the Lagrangian involves a set of N real scalar fields ϕ_n, and takes the form

$$\mathscr{L} = -\frac{1}{2} \sum_n \partial_\mu \phi_n \partial^\mu \phi_n - \frac{\mathscr{M}^2}{2} \sum_n \phi_n \phi_n - \frac{g}{4} \left(\sum_n \phi_n \phi_n \right)^2 . \tag{19.2.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

This is invariant under the group $O(N)$, consisting of rotations of the N-vector with components ϕ_n. For constant fields the effective potential in the tree approximation is given simply by minus the non-derivative terms in the Lagrangian density

$$V(\phi) \simeq \frac{\mathscr{M}^2}{2} \sum_n \phi_n \phi_n + \frac{g}{4} \left(\sum_n \phi_n \phi_n \right)^2 . \tag{19.2.10}$$

As usual we suppose that g is positive-definite. (Otherwise the minimum, if any, of $V(\phi)$ lies outside the range of validity of perturbation theory.) If $\mathscr{M}^2$ is also positive, the minimum of $V(\phi)$ is at the point $\phi = 0$, which is invariant under $O(N)$. On the other hand, for $\mathscr{M}^2 < 0$ the minimum is at points $\bar{\phi}_n$ at which

$$\sum_n \bar{\phi}_n \bar{\phi}_n = -\mathscr{M}^2/g . \tag{19.2.11}$$

The mass matrix in the tree approximation is then

$$\begin{aligned} M_{nm}^2 &= \left. \frac{\partial^2 V(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\bar{\phi}} \\ &= \mathscr{M}^2 \delta_{nm} + g \delta_{nm} \sum_\ell \bar{\phi}_\ell \bar{\phi}_\ell + 2 g \bar{\phi}_n \bar{\phi}_m \\ &= 2 g \, \bar{\phi}_n \, \bar{\phi}_m . \end{aligned} \tag{19.2.12}$$

This has one eigenvector $\bar{\phi}_n$ with a non-zero eigenvalue:

$$m^2 = 2 g \sum_n \bar{\phi}_n \bar{\phi}_n = 2|\mathscr{M}^2| , \tag{19.2.13}$$

and $N - 1$ eigenvectors perpendicular to $\bar{\phi}$ with eigenvalue zero. The reason for the appearance of only $N - 1$ Goldstone bosons is just that $O(N)$ is broken down to $O(N - 1)$ (the subgroup of $O(N)$ that leaves $\bar{\phi}$ invariant), and so the number of independent broken symmetries is the dimensionality of $O(N)$ minus the dimensionality of $O(N - 1)$, or

$$\frac{1}{2} N(N-1) - \frac{1}{2}(N-1)(N-2) = N - 1 . \tag{19.2.14}$$

Here is another proof of the existence of Goldstone bosons, one that makes no use of the effective action formalism. As we saw in Chapter 7, any continuous symmetry of the action leads to the existence of a conserved current J^μ:

$$\frac{\partial J^\mu(x)}{\partial x^\mu} = 0 , \tag{19.2.15}$$

with a charge Q that induces the associated symmetry transformation

$$Q = \int d^3x \, J^0(\mathbf{x}, 0) , \tag{19.2.16}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$$\left[Q, \phi_n(x)\right] = -\sum_m t_{nm}\phi_m(x)\,. \tag{19.2.17}$$

Operator relations like Eqs. (19.2.15)–(19.2.17) are unaffected by spontaneous symmetry breaking, which is manifested in the properties of physical *states*. Now, consider the vacuum expectation value of the commutator of the current and field. Summing over intermediate states, this is

$$\left\langle\left[J^\lambda(y),\, \phi_n(x)\right]\right\rangle_{\text{VAC}} = (2\pi)^{-3}\int d^4p\left[\rho_n^\lambda(p)\,e^{ip\cdot(y-x)} \;-\; \tilde\rho_n^\lambda(p)\,e^{ip\cdot(x-y)}\right], \tag{19.2.18}$$

where, using translation invariance,

$$(2\pi)^{-3}i\rho_n^\lambda(p) = \sum_N \langle \text{VAC}|J^\lambda(0)|N\rangle\,\langle N|\phi_n(0)|\text{VAC}\rangle\delta^4(p-p_N)\,, \tag{19.2.19}$$

$$(2\pi)^{-3}i\tilde\rho_n^\lambda(p) = \sum_N \langle \text{VAC}|\phi_n(0)|N\rangle\,\langle N|J^\lambda(0)|\text{VAC}\rangle\delta^4(p-p_N)\,. \tag{19.2.20}$$

We usually take J^λ as well as ϕ_n to be Hermitian operators, in which case Eqs. (19.2.19) and (19.2.20) are complex conjugates

$$\rho_n^\lambda(p) = -\tilde\rho_n^{\lambda *}(p)\,, \tag{19.2.21}$$

but this will not be assumed here.

Lorentz invariance tells us that ρ and $\tilde\rho$ must take the forms

$$\rho_n^\lambda(p) = p^\lambda\rho_n(-p^2)\theta(p^0)\,, \tag{19.2.22}$$

$$\tilde\rho_n^\lambda(p) = p^\lambda\tilde\rho_n(-p^2)\theta(p^0)\,. \tag{19.2.23}$$

(The factor $\theta(p^0)$, which is $+1$ for $p^0 > 0$ and zero otherwise, is required by the fact that p_N is the four-momentum of a physical state.) This gives

$$\left\langle\left[J^\lambda(y), \phi_n(x)\right]\right\rangle_{\text{VAC}} = \frac{\partial}{\partial y_\lambda}\int d\mu^2\left[\rho_n(\mu^2)\Delta_+(y-x;\mu^2) + \tilde\rho_n(\mu^2)\Delta_+(x-y;\mu^2)\right], \tag{19.2.24}$$

where Δ_+ is the familiar function

$$\Delta_+(z;\mu^2) = (2\pi)^{-3}\int d^4p\,\theta(p^0)\,\delta(p^2+\mu^2)\,e^{ip\cdot z}\,. \tag{19.2.25}$$

As remarked in Chapter 5, Lorentz invariance allows $\Delta_+(z;\mu^2)$ to depend only on z^2, μ^2, and $\theta(z^0)$, and for $z^2 > 0$ only on z^2 and μ^2. Hence $\Delta_+(x-y;\mu^2)$ and $\Delta_+(y-x;\mu^2)$ are equal for $x-y$ spacelike, and so in this case

$$\left\langle\left[J^\lambda(y),\, \phi_n(x)\right]\right\rangle_{\text{VAC}} = \frac{\partial}{\partial y_\lambda}\int d\mu^2\left[\rho_n(\mu^2)+\tilde\rho_n(\mu^2)\right]\Delta_+(y-x;\mu^2)\,. \tag{19.2.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

But for $x - y$ spacelike all commutators must vanish, so

$$\rho_n(\mu^2) = -\tilde{\rho}_n(\mu^2)\,, \tag{19.2.27}$$

and therefore Eq. (19.2.24) gives for general x and y:

$$\left\langle \left[J^\lambda(y),\, \phi_n(x) \right] \right\rangle_{\text{VAC}} = \frac{\partial}{\partial y_\lambda} \int d\mu^2 \rho_n(\mu^2) \left[\Delta_+(y-x;\mu^2) - \Delta_+(x-y;\mu^2) \right]. \tag{19.2.28}$$

Where Eq. (19.2.21) applies, Eq. (19.2.27) also shows that $\rho_n(\mu^2)$ is *real*.

Now let us use the fact that $J^\lambda(y)$ is conserved. Applying the derivative $\partial/\partial y^\lambda$ to both sides of Eq. (19.2.28), and using the familiar equation

$$(\Box_y - \mu^2)\, \Delta_+(y-x;\mu^2) = 0\,, \tag{19.2.29}$$

we find, for *all* x and y

$$0 = \int d\mu^2\, \mu^2 \rho_n(\mu^2) \left[\Delta_+(y-x;\mu^2) - \Delta_+(x-y;\mu^2) \right], \tag{19.2.30}$$

and so (since $\Delta_+(x-y)$ is *not* even for $x - y$ timelike or lightlike)

$$\mu^2 \rho_n(\mu^2) = 0\,. \tag{19.2.31}$$

Normally we would conclude from this that $\rho_n(\mu^2)$ vanishes for all μ^2. However, this is not possible in the case of broken symmetry. Set $\lambda = 0$ and $x^0 = y^0 = t$ in Eq. (19.2.28):

$$\begin{aligned} \left\langle \left[J^0(\mathbf{y},t), \phi_n(\mathbf{x},t) \right] \right\rangle_{\text{VAC}} &= 2i(2\pi)^{-3} \int d\mu^2 \rho_n(\mu^2) \\ &\quad \times \int d^4p \sqrt{\mathbf{p}^2 + \mu^2}\; e^{i\mathbf{p}\cdot(\mathbf{y}-\mathbf{x})} \delta(p^2 + \mu^2) \\ &= i\delta^3(\mathbf{y}-\mathbf{x}) \int d\mu^2 \rho_n(\mu^2)\,. \end{aligned}$$

Integrating and using Eqs. (19.2.16) and (19.2.17) gives

$$-\sum_m t_{nm} \langle \phi_m \rangle_{\text{VAC}} = i \int d\mu^2 \rho_n(\mu^2)\,. \tag{19.2.32}$$

Eqs. (19.2.31) and (19.2.32) can be reconciled only if

$$\rho_n(\mu^2) = i\,\delta(\mu^2) \sum_m t_{nm} \langle \phi_m(0) \rangle_{\text{VAC}}\,. \tag{19.2.33}$$

(This is real for Hermitian fields ϕ_n because in this case Eq. (19.2.1) requires t_{nm} to be imaginary.) Thus as long as the symmetry is broken, $\rho_n(\mu^2)$ cannot vanish, but rather consists entirely of a term proportional to $\delta(\mu^2)$. Such a term can obviously only arise in a theory that has massless particles, because otherwise the spectrum of center-of-mass squared energies $-p_N^2$ would not extend down to zero. Furthermore a delta-function $\delta(\mu^2)$ can only arise from single particle states of zero mass; multiparticle states

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

would contribute a continuum extending down to $\mu^2 = 0$. The state $\phi_n(0)|\text{VAC}\rangle$ is rotationally invariant, so $\langle N|\phi_n(0)|\text{VAC}\rangle$ must vanish for any state N of non-zero helicity. Also $\langle \text{VAC}|J^0|N\rangle$ vanishes for any state N that has different intrinsic parity or (unbroken) internal quantum numbers from J^0. We conclude then that *a broken symmetry with* $t_{nm}\langle\phi_n(0)\rangle_{\text{VAC}} \neq 0$ *requires the existence of a massless particle of spin zero and the same parity and internal quantum numbers as* J^0. These are our Goldstone bosons.

The above argument breaks down when the spontaneously broken symmetry is a local rather than a global symmetry. Either we choose a Lorentz-invariant gauge, such as the Landau gauge with $\partial_\mu A^\mu = 0$, in which case as shown in Section 15.7 the positivity assumptions of quantum mechanics are violated, or we adopt a gauge like the axial gauge with $A_3^\mu = 0$, in which case the ordinary rules of quantum mechanics apply but manifest Lorentz invariance is lost. As we shall see in Chapter 21, this exception is not just a technicality; spontaneously broken local symmetries do not lead to Goldstone bosons.

It will be useful to look in a little more detail at how the coefficient of the delta-function in $\rho_n(\mu^2)$ is related to the properties of the Goldstone boson. For a spin zero boson B of four-momentum p_B^μ, Lorentz invariance requires the matrix element of the current between the vacuum and one-particle states to take the form

$$\langle \text{VAC}|J^\lambda(x)|B\rangle = i\frac{F\,p_B^\lambda\, e^{ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}}\,, \tag{19.2.34}$$

where $\mathbf{p}_B$ is the momentum of B, $p_B^0 \equiv |\mathbf{p}_B|$ and F is a constant coefficient with the dimensions of energy. (This is consistent with current conservation because $p_{B\lambda}p_B^\lambda = 0$.) Also, the matrix element of the scalar field $\phi_n(y)$ between a single-particle state and the vacuum is of the form

$$\langle B|\phi_n(y)|\text{VAC}\rangle = \frac{Z_n\, e^{-ip_B\cdot y}}{(2\pi)^{3/2}\sqrt{2p_B^0}}\,, \tag{19.2.35}$$

where Z_n is a dimensionless constant. From Eqs. (19.2.34) and (19.2.35), we have then

$$\begin{aligned}(2\pi)^{-3}i\rho_n(-p^2)p^\lambda\theta(p^0) &\equiv \int d^3p_B\langle \text{VAC}|J^\lambda(0)|B\rangle\,\langle B|\phi_n(0)|\text{VAC}\rangle\,\delta^4(p-p_B)\\ &= \delta(p^0-|\mathbf{p}|)\,(2\pi)^{-3}(2p^0)^{-1}p^\lambda iFZ_n\\ &= \theta(p^0)\,\delta(-p^2)\,(2\pi)^{-3}p^\lambda iFZ_n\,,\end{aligned}$$

so

$$\rho_n(\mu^2) = FZ_n\,\delta(\mu^2)\,. \tag{19.2.36}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Comparing with Eq. (19.2.33), this gives

$$iFZ_n = -\sum_m t_{nm}\langle\phi_m(0)\rangle_{\rm VAC} \,. \tag{19.2.37}$$

More generally, we may have several broken symmetries with generators t_a and currents J_a^μ, which we can take independent in the sense that no linear combination of the t_a is unbroken. For each of these, there is a Goldstone boson $|B_a\rangle$, and we define Z_{an} and F_{ab} by

$$\langle {\rm VAC}|J_a^\lambda(x)|B_b\rangle = i\frac{F_{ab}p_B^\lambda e^{ip_B\cdot x}}{(2\pi)^{3/2}\sqrt{2p_B^0}} \,, \tag{19.2.38}$$

$$\langle B_a|\phi_n(y)|{\rm VAC}\rangle = \frac{Z_{an}\,e^{-ip_B\cdot y}}{(2\pi)^{3/2}\sqrt{2p_B^0}} \,. \tag{19.2.39}$$

Eq. (19.2.37) applies for each a, so

$$i\sum_b F_{ab}Z_{bn} = -\sum_m [t_a]_{nm}\langle\phi_m(0)\rangle_{\rm VAC} \,. \tag{19.2.40}$$

For instance, in the $O(N)$ example discussed earlier, we can adapt our basis so that the vacuum expectation value points in the one–direction

$$\bar\phi_m \equiv \langle\phi_m(0)\rangle_{\rm VAC} = v\delta_{m1} \,. \tag{19.2.41}$$

The $N-1$ broken symmetry generators t_a (with $a = 2\cdots N$) can be defined as those for infinitesimal rotations in the 1–a plane. With a convenient choice of normalization, these have the non-zero elements

$$[t_a]_{1a} = -[t_a]_{a1} = i \tag{19.2.42}$$

(with no summation over a). The unbroken $O(N-1)$ symmetry, under which the $N-1$ Goldstone bosons transform according to the vector representation, tells us that

$$F_{ab} = \delta_{ab}F \,, \qquad Z_{a1} = 0 \,, \qquad Z_{ab} = Z\delta_{ab} \,. \tag{19.2.43}$$

Then Eq. (19.2.40) requires

$$FZ = v \,. \tag{19.2.44}$$

It is conventional to adopt the field renormalization prescription that $Z = 1$, so $F = v$. Thus F is a measure of the strength of the symmetry breaking. As we shall now see, the parameter $1/F$ determines the strength with which the Goldstone bosons interact with each other and with other particles.

A broken symmetry tells us more about the Goldstone bosons than just that they have zero mass; it also tightly constrains their interactions at low energy. To see this in the simplest case, consider the matrix element of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

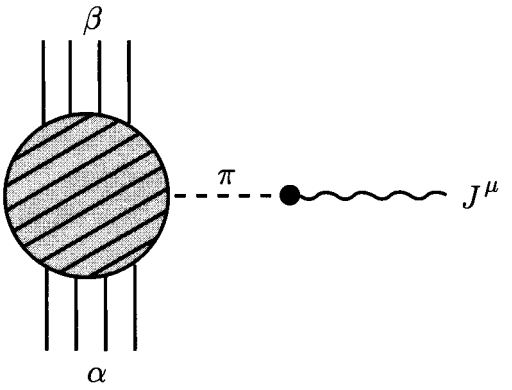

Figure 19.1. Feynman diagrams for pole terms in the matrix element of a symmetry current $J^\mu(x)$ between general states α and β, due to a Goldstone boson internal line, denoted π.

the current $J^\mu(x)$ associated with a broken symmetry. Between arbitrary states α, β:

$$\left\langle \beta | J^\mu(x) | \alpha \right\rangle = e^{iq\cdot x} \left\langle \beta | J^\mu(0) | \alpha \right\rangle , \tag{19.2.45}$$

with

$$q^\mu \equiv p_\alpha^\mu - p_\beta^\mu . \tag{19.2.46}$$

We know that $J^\mu(x)$ has a non-zero matrix element between the vacuum and a one-Goldstone-boson state $|B, \mathbf{q}\rangle$, given by Eq. (19.2.34). It follows by the usual rules of pology (see Chapter 10) that the matrix element (19.2.45) has a pole at $q^2 \to 0$, with*

$$\langle \beta | J^\mu(0) | \alpha \rangle \to \frac{iF\, q^\mu}{q^2}\, M_{\beta\alpha} , \tag{19.2.47}$$

where $i(2\pi)^4\delta^4(p_\alpha - p_\beta - q)M_{\beta\alpha}/(2\pi)^{3/2}(2q^0)^{1/2}$ is the S-matrix element for emitting a Goldstone boson of four-momentum q in the transition $\alpha \to \beta$. Let us therefore write

$$\langle \beta | J^\mu(0) | \alpha \rangle \equiv N^\mu_{\beta\alpha} + \frac{iFq^\mu}{q^2}\, M_{\beta\alpha} , \tag{19.2.48}$$

where $N^\mu_{\beta\alpha}$ is defined as the non-pole contributions to the matrix element of the current. From Eq. (19.2.45), we see that the conservation law

* Where the Goldstone boson corresponds to an elementary field, Eq. (19.2.47) may be obtained by inspection of the class of Feynman diagrams shown in Figure 19.1. The factor $i(2\pi)^4$ in the S-matrix element for $\alpha \to \beta + B$ is cancelled by the factor $-i(2\pi)^{-4}$ associated with the B propagator. The general rules of pology tell us that the same holds even where the Goldstone boson is a composite particle.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

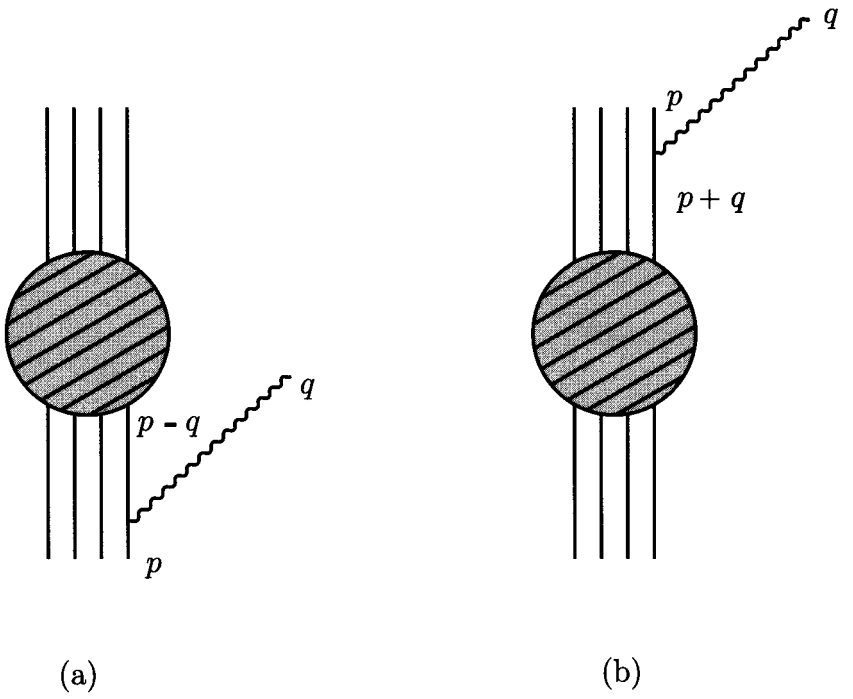

Figure 19.2. Feynman diagrams for pole terms in the matrix element of a symmetry current $J^\mu(x)$ between general states α and β, due to internal lines close to their mass shell. Solid lines indicate 'hard' external particles; wavy lines indicate insertions of the current $J^\mu(x)$; the cross-hatched disk represents the sum of diagrams with the indicated external lines.

$\partial_\mu J^\mu = 0$ for the current J^μ requires (19.2.48) to vanish when contracted with q^μ, and so

$$M_{\beta\alpha} = \frac{i}{F} q_\mu N^\mu_{\beta\alpha} . \tag{19.2.49}$$

One immediate consequence is that unless $N^\mu_{\beta\alpha}$ has a pole at $q \to 0$, the matrix element $M_{\beta\alpha}$ for emitting a Goldstone boson in a transition $\alpha \to \beta$ vanishes as $q \to 0$. This is called an 'Adler zero.'[6]

In fact, it often happens that $N^\mu_{\beta\alpha}$ *does* have a pole at $q = 0$. This is because the vertex for the current $J^\mu(x)$ might be attached to an external line of the process $\alpha \to \beta$. (See Figure 19.2.) For instance, if a four-momentum q^μ is carried away by a current J^μ inserted into an outgoing or incoming particle line of four-momentum p and mass m, then the internal line that connects this vertex to the rest of the diagram will carry a four-momentum $p^\mu + q^\mu$ or $p^\mu - q^\mu$ respectively, and its propagator will

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

therefore contribute to $N^{\mu}_{\beta\alpha}$ a factor

$$\left[(p \pm q)^2 + m^2\right]^{-1} = \left[\pm 2p \cdot q + q^2\right]^{-1} \to \pm \frac{1}{2p \cdot q} . \tag{19.2.50}$$

For a fixed direction of $\mathbf{q}$, the factor $1/|\mathbf{q}|$ from Eq. (19.2.50) cancels the factor $|\mathbf{q}|$ in Eq. (19.2.49), yielding a finite (though $\mathbf{q}$-direction-dependent) result in the limit $|\mathbf{q}| \to 0$. On the other hand, the contribution to $N^{\mu}_{\beta\alpha}$ of diagrams in which the current J^{μ} is attached to an internal line of the process $\alpha \to \beta$ has no singularity for $|\mathbf{q}| \to 0$, and therefore is cancelled by the factor $|\mathbf{q}|$ in Eq. (19.2.49). Thus Eq. (19.2.49) can be interpreted to mean that *the amplitude for emitting a very soft Goldstone boson in a process $\alpha \to \beta$ can be calculated from graphs in which the Goldstone boson is emitted only from the external lines of the process, with the vertex for emitting the Goldstone boson given by applying Eq. (19.2.49) to the transition between single-particle states.*

The effective action formalism can be used to derive interesting results for the interactions of the Goldstone bosons with each other and with other scalars. For this purpose, we note that if we define a set of renormalized Goldstone boson fields π_a for which

$$\langle B_a | \pi_b(x) | \text{VAC} \rangle = \frac{e^{-ip_B \cdot x} \delta_{ab}}{(2\pi)^{3/2} \sqrt{2p^0_B}} , \tag{19.2.51}$$

then Eq. (19.2.39) tells us that

$$\phi_n(x) = \sum_a Z_{an} \pi_a(x) + \dots , \tag{19.2.52}$$

where '...' indicates fields that do not create Goldstone bosons. But Eq. (19.2.40) gives $Z_{an} = \sum_b F^{-1}_{ab}(it_b \bar{\phi})_n$. Thus the amplitude for any reaction among N zero-four-momentum Goldstone bosons $\pi_{a_1}, \dots, \pi_{a_N}$, is the same as would be calculated in the tree approximation from the effective interaction

$$\mathscr{H}_{\text{eff}} = \frac{1}{N!} g_{a_1 \cdots a_N} \pi_{a_1} \cdots \pi_{a_N} , \tag{19.2.53}$$

with

$$g_{a_1 \cdots a_N} = \sum_{b_1 \dots b_N} F^{-1}_{a_1 b_1} \cdots F^{-1}_{a_N b_N} (it_{b_1} \bar{\phi})_{n_1} \cdots (it_{b_N} \bar{\phi})_{n_N} \left. \frac{\partial^N V(\phi)}{\partial \phi_{n_1} \cdots \partial \phi_{n_N}} \right|_{\phi = \bar{\phi}} . \tag{19.2.54}$$

Eq. (19.2.4) at $\phi = \bar{\phi}$ then implies that the sum of all 'tadpole' graphs for a single Goldstone boson line disappearing into the vacuum vanishes, and Eq. (19.2.6) tells us that the amplitudes vanish for Goldstone bosons to make transitions at zero four-momentum into any other scalars. To go beyond these results, we can continue to differentiate with respect to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

the scalar fields. For instance, the derivative of Eq. (19.2.5) gives for any symmetry generator t:

$$\sum_n \frac{\partial^2 V(\phi)}{\partial\phi_n\partial\phi_m}\, t_{n\ell} + \sum_n \frac{\partial^2 V(\phi)}{\partial\phi_n\partial\phi_\ell}\, t_{nm} + \sum_{n,k} \frac{\partial^3 V(\phi)}{\partial\phi_n\partial\phi_m\partial\phi_\ell}\, t_{nk}\phi_k = 0\,. \quad (19.2.55)$$

Taking t as one of the broken symmetry generators t_a, setting $\phi = \bar{\phi}$, contracting with $(t_b\bar{\phi})_m(t_c\bar{\phi})_\ell$, and using Eq. (19.2.6) gives

$$\sum_{nm\ell} \frac{\partial^3 V(\phi)}{\partial\phi_n\partial\phi_m\partial\phi_\ell}\bigg|_{\phi=\bar{\phi}} (t_a\bar{\phi})_n(t_b\bar{\phi})_m(t_c\bar{\phi})_\ell = 0\,, \quad (19.2.56)$$

so the sum of all graphs with three external zero-four-momentum Goldstone boson lines vanishes. In particular, this means that in general processes, to leading order in small Goldstone boson energies, low energy Goldstone bosons are not emitted from external low energy Goldstone boson lines.

19.3 Spontaneously Broken Approximate Symmetries

In the previous section we dealt with exact symmetries of the action that do not leave the vacuum invariant, and are said to be spontaneously broken. We shall now consider the effect of adding small symmetry-breaking terms to the action in such a theory. Such spontaneously broken approximate symmetries are important in the theory of strong interactions, and in some areas of condensed matter physics. As we shall see, the spontaneous breakdown of an approximate symmetry does not lead to the appearance of massless Goldstone bosons, but of low-mass spinless particles, often called *pseudo-Goldstone bosons.*[7]

We continue here to treat translationally invariant theories, in which the effective action is expressed as in Eq. (19.2.3) in terms of an effective potential $V(\phi)$ that depends on a set of spacetime-independent scalar field expectation values ϕ_n. For an action that obeys some set of approximate continuous symmetries with generators t_α, the effective potential may be written

$$V(\phi) = V_0(\phi) + V_1(\phi)\,, \quad (19.3.1)$$

where $V_0(\phi)$ satisfies the invariance condition*

$$\sum_{nm} \frac{\partial V_0(\phi)}{\partial\phi_n}\,(t_\alpha)_{nm}\phi_m = 0 \quad (19.3.2)$$

* We denote general symmetry generators as t_α, t_β, etc., in contrast with the independent broken symmetry generators, for which the subscript α takes values a, b, etc.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

and $V_1(\phi)$ is a small correction due to the symmetry breaking in the action. Suppose that this perturbation shifts the minimum of the potential from ϕ_0, the minimum of $V_0(\phi)$, to $\bar{\phi} = \phi_0 + \phi_1$, where ϕ_1 is small, of first order in the symmetry-breaking perturbation. The equilibrium condition for the vacuum is then

$$\left.\frac{\partial V(\phi)}{\partial \phi_n}\right|_{\phi=\phi_0+\phi_1} = 0 \,. \tag{19.3.3}$$

The zeroth-order term on the left is just $[\partial V_0(\phi)/\partial\phi_n]_{\phi=\phi_0}$, which vanishes because ϕ_0 is defined as the minimum of $V_0(\phi)$. Thus the first-order terms must also vanish, and so:

$$\sum_m \left.\frac{\partial^2 V_0(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=\phi_0} \phi_{1\,m} + \left.\frac{\partial V_1(\phi)}{\partial\phi_n}\right|_{\phi=\phi_0} = 0 \,. \tag{19.3.4}$$

Eq. (19.2.6) holds here with V replaced with the invariant term V_0 and $\bar{\phi}$ replaced with ϕ_0:

$$\sum_{nl} \left.\frac{\partial^2 V_0(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=\phi_0} (t_\alpha)_{nl}\,\phi_{0\,l} = 0 \,. \tag{19.3.5}$$

Hence multiplying Eq. (19.3.4) with $(t_\alpha\phi_0)_n$ and summing over n gives

$$\sum_n (t_\alpha\phi_0)_n \left.\frac{\partial V_1(\phi)}{\partial\phi_n}\right|_{\phi=\phi_0} = 0 \,. \tag{19.3.6}$$

Recalling the interpretation of $V(\phi)$ as the generating function for one-particle-irreducible graphs, and noting that in the absence of the perturbation V_1 the Goldstone components of ϕ_n are those in the direction of $t_\alpha\phi_0$ (see Eq. (19.2.33)), the left-hand side of Eq. (19.3.6) is proportional to the sum of all 'tadpole' graphs, in which a pseudo-Goldstone boson disappears into the vacuum. Eq. (19.3.6) may thus be paraphrased as the condition that, to first order in V_1, *pseudo-Goldstone bosons have no tadpoles.*

The moral of this calculation is that if we do not start with a zeroth-order vacuum expectation value that satisfies Eq. (19.3.6), then even a small perturbation will produce a large change in $\bar{\phi}$, invalidating the expansion of $\bar{\phi}$ around ϕ_0. Fortunately, for compact Lie groups it is always possible to choose ϕ_0 to satisfy Eq. (19.3.6). To see this, note that the invariance of the potential $V_0(\phi)$ under a group of real linear transformations $\phi \to L\phi$ implies that if ϕ_* is one minimum of the potential, then so is $L\phi_*$. For continuous groups of transformations, we can always parameterize the transformations as $L(\theta)$, in such a way that

$$\frac{\partial L(\theta)}{\partial\theta^\alpha} L^{-1}(\theta) = iN_{\alpha\beta}(\theta)t_\beta \,, \tag{19.3.7}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

where $N_{\alpha\beta}$ is a non-singular matrix depending on the group parameters θ_α. (Recall that in a real representation, it is it_α rather than t_α that is real.) Now consider the function $V_1(L(\theta)\phi_*)$. Where the group is compact, $L(\theta)\phi_*$ maps out a compact manifold as θ runs over the group volume, and as long as $V_1(\phi)$ is continuous, it must have a minimum on any such compact surface, say at a point $L(\theta_*)\phi_*$. The derivative of $V_1(L(\theta)\phi_*)$ with respect to θ_α is

$$\frac{\partial V_1(L(\theta)\phi_*)}{\partial\theta_\alpha} = \sum_n \left.\frac{\partial V_1(\phi)}{\partial\phi_n}\right|_{\phi=L(\theta)\phi_*} N_{\alpha\beta}(\theta)\,(it_\beta L(\theta)\phi_*)_n\,. \tag{19.3.8}$$

This must vanish at the minimum θ_*, and since N_{ab} is non-singular, this implies that

$$0 = \sum_n \left.\frac{\partial V_1(\phi)}{\partial\phi_n}\right|_{\phi=L(\theta_*)\phi_*} (t_\beta L(\theta_*)\phi_*)_n\,. \tag{19.3.9}$$

But then Eq. (19.3.6) is satisfied if we make the choice $\phi_0 = L(\theta_*)\phi_*$.

Eq. (19.3.6) is known as a *vacuum alignment* condition,[8] because it generally has the effect of forcing the direction of the symmetry breaking by the vacuum into some sort of alignment with the symmetry-breaking terms in the Hamiltonian. For instance, consider the case of $SO(N)$ spontaneously broken to $SO(N-1)$, introduced in the previous section. In the absence of any symmetry-breaking perturbation, there is no way to tell *which* $SO(N-1)$ subgroup is left unbroken; if the dynamics of the theory leads to a ground state that is invariant under the $SO(N-1)$ subgroup of $SO(N)$ that leaves some N-vector ϕ_{0n} invariant, then by performing an $SO(N)$ rotation we can find a ground state that is invariant under the $SO(N-1)$ subgroup that leaves any other N-vector invariant. If we add a perturbation that transforms under $SO(N)$ like, say, the component $\sum_n u_n\phi_n$ of an N-vector ϕ_n (not necessarily consisting of elementary scalars), then the Hamiltonian is invariant under a specific $SO(N-1)$ subgroup of $SO(N)$, consisting of rotations that leave the vector u invariant. Without a vacuum alignment condition, we might think that the remaining exact symmetry is $SO(N-2)$, consisting of those rotations that leave invariant *both* u and the vector ϕ_0 characterizing the vacuum symmetry. But with $V_1(\phi) = \sum_n u_n\phi_n$, the condition (19.3.6) tells us that at the true vacuum, $\sum_n(t_\alpha\phi_0)_n u_n = 0$ for all $SO(N)$ generators t_α. The $SO(N)$ generators t_α span the space of all antisymmetric $N\times N$ matrices, so this condition requires that ϕ_0 must be in the same direction as u, and so the unbroken symmetry is $SO(N-1)$, not $SO(N-2)$.

According to the general results of Section 16.1, the mass matrix M_{ab}

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

of the pseudo-Goldstone bosons is given to first order by

$$M^2_{ab} = \sum_{mn} Z_{an} Z_{bm} \left. \frac{\partial^2 V(\phi)}{\partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0+\phi_1} , \tag{19.3.10}$$

where Z_{an} is the field renormalization constant defined by Eq. (19.2.39). Since the mass matrix (19.3.10) vanishes in zeroth order, the first-order terms give

$$M^2_{ab} = \sum_{mn} Z_{an} Z_{bm} \left[\sum_{\ell} \left. \frac{\partial^3 V_0(\phi)}{\partial \phi_\ell \partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0} \phi_{1\ell} + \left. \frac{\partial^2 V_1(\phi)}{\partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0} \right] , \tag{19.3.11}$$

where Z_{an} is here given by the zeroth-order approximation to (19.2.40):

$$Z_{an} = \sum_b F^{-1}_{ab} (it_b \phi_0)_n . \tag{19.3.12}$$

To calculate the mass matrix (19.3.11), we take t in Eq. (19.2.55) to be one of the broken symmetry generators t_a, set $\phi = \phi_0$, and contract with $(t_b\phi_0)_m \phi_{1\ell}$:

$$\begin{aligned} 0 = & \sum_{nm\ell} \left. \frac{\partial^3 V_0(\phi)}{\partial \phi_\ell \partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0} \phi_{1\ell} (t_a\phi_0)_n (t_b\phi_0)_m \\ & + \sum_{nm} \left. \frac{\partial^2 V_0(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\phi_0} (t_a\phi_1)_n (t_b\phi_0)_m + \sum_{n\ell} \left. \frac{\partial^2 V_0(\phi)}{\partial \phi_n \partial \phi_\ell} \right|_{\phi=\phi_0} (t_a t_b \phi_0)_n \phi_{1\ell} . \end{aligned}$$

The second term on the right-hand side vanishes according to Eq. (19.2.6), while the third may be rewritten using Eq. (19.3.4), leaving us with

$$\sum_{nm\ell} \left. \frac{\partial^3 V_0(\phi)}{\partial \phi_\ell \partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0} \phi_{1\ell} (t_a\phi_0)_n (t_b\phi_0)_m = \left. \frac{\partial V_1(\phi)}{\partial \phi_n} \right|_{\phi=\phi_0} (t_a t_b \phi_0)_n . \tag{19.3.13}$$

Using this in Eq. (19.3.11) then yields a formula for the pseudo-Goldstone boson mass matrix in terms of V_1:

$$\begin{aligned} M^2_{cd} = - \sum_{ab} F^{-1}_{ca} F^{-1}_{db} \Bigg[& (t_a\phi_0)_n (t_b\phi_0)_m \left. \frac{\partial^2 V_1(\phi)}{\partial \phi_m \partial \phi_n} \right|_{\phi=\phi_0} \\ & + (t_a t_b \phi_0)_n \left. \frac{\partial V_1(\phi)}{\partial \phi_n} \right|_{\phi=\phi_0} \Bigg] . \end{aligned} \tag{19.3.14}$$

For this to be a sensible mass matrix, it had better be positive. To see that it is, it is convenient to rewrite this result in terms of derivatives with respect to the group parameters θ_a. Differentiating Eq. (19.3.8) with respect to θ_β, setting $\theta = \theta_*$, and using Eq. (19.3.9) and $\phi_0 = L(\theta_*)\phi_*$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

gives

$$M_{ab}^2 = \sum_{cd\alpha\beta} N_{a\alpha}^{-1}(\theta_*)\, N_{b\beta}^{-1}(\theta_*) F_{ac}^{-1} F_{bd}^{-1} \left. \frac{\partial^2 V_1(L(\theta)\phi_*)}{\partial\theta_\alpha \partial\theta_\beta} \right|_{\theta=\theta_*} . \tag{19.3.15}$$

The matrix on the right is positive, because θ_* is the *minimum* of the function $V_1(L(\theta)\phi_*)$.

This formula has a somewhat more familiar version, in terms of the vacuum expectation value of a double commutator of symmetry generators with the symmetry-breaking perturbation. Suppose that the symmetry-breaking perturbation H_1 in the Hamiltonian is a linear combination

$$H_1 = \sum_n u_n \Phi_n \tag{19.3.16}$$

of operators Φ_n (not necessarily elementary scalar fields), which furnish a representation of the symmetry group with generators t_α, in the sense that

$$[T_\alpha, \Phi_n] = -(t_\alpha)_{nm} \Phi_m \,, \tag{19.3.17}$$

where T_α are the quantum mechanical generators of the symmetry group. According to the results of Section 16.3, the symmetry-breaking part of the potential is

$$V_1(\phi) = \langle H_1 \rangle_{\langle\Phi\rangle=\phi} = \sum_n u_n \phi_n \,, \tag{19.3.18}$$

the subscript in the middle expression indicating that the expectation value is to be taken in the state of minimum energy in which Φ_n has expectation value ϕ_n. The vacuum alignment condition (19.3.6) then reads

$$0 = \sum_n u_n (t_\alpha \phi_0)_n \,,$$

or using Eq. (19.3.17)

$$0 = \langle [T_\alpha, H_1] \rangle_0 \,, \tag{19.3.19}$$

the subscript 0 now indicating that the expectation value is to be taken in the vacuum state, in which Φ_n has expectation value ϕ_{0n}. Also, Eq. (19.3.14) gives the mass matrix here as

$$M_{cd}^2 = -\sum_{ab} F_{ca}^{-1} F_{db}^{-1} \sum_n u_n (t_a t_b \phi_0)_n \,,$$

and, using Eq. (19.3.17), this is

$$M_{cd}^2 = -\sum_{ab} F_{ca}^{-1} F_{db}^{-1} \langle [T_a, [T_b, H_1]] \rangle_0 \,. \tag{19.3.20}$$

This is symmetric in c and d. To see this, note that the Jacobi identity and group commutation relations may be used to write the difference of Eq. (19.3.20) and the same with c and d interchanged as a linear

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

combination of terms $\langle [T_\alpha, H_1] \rangle_0$, which vanish according to the vacuum alignment condition (19.3.19). The mass matrix (19.3.20) is also positive, because the point $\theta = 0$ is at the *minimum* of the vacuum energy $\langle \exp(-i\theta_a T_a) H_1 \exp(i\theta_a T_a) \rangle_0$ for rotated vacuum states $\exp(i\theta_a T_a)|0\rangle$.

19.4 Pions as Goldstone Bosons

The classic example of a broken symmetry in elementary particle physics is the approximate symmetry of strong interactions known as chiral $SU(2) \times SU(2)$. According to our present understanding, this symmetry arises because there are two quark fields, u and d, that happen to have relatively small masses. (An estimate is given in Section 19.7.) In the approximation that the u and d are massless, the Lagrangian (18.7.5) of quantum chromodynamics is

$$\mathscr{L} = -\bar{u}\gamma^\mu D_\mu u - \bar{d}\gamma^\mu D_\mu d - \cdots , \tag{19.4.1}$$

where D_μ is a color-gauge-covariant derivative (see Eq. (15.1.10)) and '$\cdots$' refers to terms involving only gluon fields and/or other quark flavors, but not u or d. This Lagrangian is invariant under the transformations

$$\begin{pmatrix} u \\ d \end{pmatrix} \to \exp\left(i\vec{\theta}^V \cdot \vec{t} + i\gamma_5 \vec{\theta}^A \cdot \vec{t} \right) \begin{pmatrix} u \\ d \end{pmatrix} , \tag{19.4.2}$$

where $\vec{t}$ is the three-vector* of isospin matrices

$$t_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad t_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad t_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ,$$

and $\vec{\theta}^V$ and $\vec{\theta}^A$ are independent real three-vectors.** This Lie algebra may be written in terms of two commuting $SU(2)$ subalgebras that act respectively only on the left- and right-handed parts of the quark fields, with generators

$$\vec{t}_L = \frac{1}{2}(1+\gamma_5)\vec{t} , \qquad \vec{t}_R = \frac{1}{2}(1-\gamma_5)\vec{t} \tag{19.4.3}$$

* We use arrows for three-vectors in isotopic spin space to distinguish them from ordinary three-vectors, which will continue to be indicated by boldface letters.

** The Lagrangian (19.4.1) has this symmetry because $\overline{\gamma_5\psi}\gamma^\mu = -\bar{\psi}\gamma_5\gamma^\mu = +\bar{\psi}\gamma^\mu\gamma_5$. The Lagrangian also has two other continuous internal symmetries. One is baryon conservation, the invariance under a common phase transformation of the u and d quark fields. This is unbroken and commutes with the other symmetries, so it does not affect our discussion in this section. The other symmetry is invariance under multiplication of the quark doublet with $\exp(i\alpha\gamma_5)$. As discussed in Section 23.5, this $U(1)$ symmetry is strongly intrinsically broken by non-perturbative effects associated with instantons.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

satisfying the commutation relations

$$[t_{Li},\ t_{Lj}] = i\epsilon_{ijk}\, t_{Lk}\ , \tag{19.4.4}$$

$$[t_{Ri},\ t_{Rj}] = i\epsilon_{ijk}\, t_{Rk}\ , \tag{19.4.5}$$

$$[t_{Li},\ t_{Rj}] = 0\ . \tag{19.4.6}$$

The underlying symmetry group is therefore identified as $SU(2) \times SU(2)$. It has another obvious $SU(2)$ subgroup, consisting of ordinary isospin transformations with $\vec{\theta}^A = 0$, and generators

$$\vec{t} = \vec{t}_L + \vec{t}_R\ . \tag{19.4.7}$$

The algebra of $SU(2) \times SU(2)$ may be written in terms of $\vec{t}$ and another triplet of generators:

$$\vec{x} = \vec{t}_L - \vec{t}_R = \gamma_5 \vec{t} \tag{19.4.8}$$

with commutation relations

$$[t_i,\ t_j] = i\epsilon_{ijk}\, t_k\ , \tag{19.4.9}$$

$$[t_i,\ x_j] = i\epsilon_{ijk}\, x_k\ , \tag{19.4.10}$$

$$[x_i,\ x_j] = i\epsilon_{ijk}\, t_k\ . \tag{19.4.11}$$

We will see that the $SU(2) \times SU(2)$ symmetry is spontaneously broken, while its isotopic spin subgroup generated by the $\vec{t}$ is an ordinary unbroken (though approximate) symmetry.

By Noether's method (see, e.g., Section 7.3) we may derive from the Lagrangian (19.4.1) the conserved vector and axial-vector currents

$$\vec{V}^\mu = i\bar{q}\gamma^\mu \vec{t} q\ , \qquad \vec{A}^\mu = i\bar{q}\gamma^\mu\gamma_5 \vec{t} q\ , \tag{19.4.12}$$

$$\partial_\mu \vec{V}^\mu = \partial_\mu \vec{A}^\mu = 0\ , \tag{19.4.13}$$

where q is the quark doublet,

$$q \equiv \begin{pmatrix} u \\ d \end{pmatrix}\ . \tag{19.4.14}$$

Their associated charges are the generators respectively of isospin and of the remaining symmetries

$$\vec{T} = \int d^3x\, \vec{V}^0\ , \tag{19.4.15}$$

$$\vec{X} = \int d^3x\, \vec{A}^0\ . \tag{19.4.16}$$

The currents (19.4.12) are normalized so that the quantum operators $\vec{T}$ and $\vec{X}$ satisfy the same commutation relations as the matrices $\vec{t}$ and $\vec{x}$:

$$[T_i,\ T_j] = i\epsilon_{ijk}\, T_k\ , \tag{19.4.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$$[T_i, X_j] = i\epsilon_{ijk} X_k \,, \tag{19.4.18}$$

$$[X_i, X_j] = i\epsilon_{ijk} T_k \,. \tag{19.4.19}$$

Acting on the quark fields, these operators induce the transformation (19.4.2), in the sense that

$$\left[\vec{T}, q\right] = -\vec{t}\, q \,, \tag{19.4.20}$$

$$\left[\vec{X}, q\right] = -\vec{x}\, q \,. \tag{19.4.21}$$

This symmetry if exact and unbroken would require any one-hadron state $|h\rangle$ to be degenerate with another state $\vec{X}|h\rangle$ of opposite parity and equal spin, baryon number, and strangeness.† No such parity doubling is seen in the hadron spectrum, so we are forced to conclude that if the chiral symmetry $SU(2) \times SU(2)$ is a good approximation at all, then it must be spontaneously broken to its isotopic spin $SU(2)$ subgroup. In this case the operator $\vec{X}$ takes a one-hadron state $|h\rangle$ into a hadron h plus a massless pseudoscalar Goldstone boson, so there is no need for parity doubling of the hadron spectrum.

The question of whether quantum chromodynamics actually exhibits such a pattern of symmetry breaking involves all the complications of strong interaction dynamics. As we shall see in Section 19.9, there are general grounds for believing that the isospin $SU(2)$ is not spontaneously broken in quantum chromodynamics, but it is much more difficult to show that the chiral part of $SU(2) \times SU(2)$ *is* spontaneously broken. (But according to an argument given in Section 22.5, the $SU(3) \times SU(3)$ symmetry of quantum chromodynamics with *three* massless quark flavors must be spontaneously broken.) It was something of a breakthrough in the 1960s to realize that one does not have to have a detailed understanding of the mechanism of the breaking of chiral symmetry; we derive the most interesting consequences of this symmetry breaking by simply *assuming* that $SU(2) \times SU(2)$ is spontaneously broken to $SU(2)$.

The u and d quarks have small but nonzero masses, so the $SU(2) \times SU(2)$ symmetry is not exact. A broken *approximate* chiral symmetry entails the existence of an approximately massless Goldstone boson with the same quantum numbers as the broken symmetry generator $\vec{X}$: it must be a state of negative parity, zero spin, unit isospin, and zero baryon number

† One way to satisfy this condition is for the hadron to have zero mass, with two states $|\pm\rangle$ of helicity $\pm\frac{1}{2}$ and equal spin, baryon number, and strangeness; the states $|+\rangle + |-\rangle$ and $|+\rangle - |-\rangle$ would then have opposite parity. It is not true that an unbroken chiral symmetry necessarily implies a zero nucleon mass, unless we make further assumptions about the matrix elements of the axial-vector current. But as we shall see in Section 22.5, an exact unbroken chiral symmetry would in fact require that *some* baryons be massless.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

and strangeness. In fact the lightest of all hadrons is the pion, which has precisely these quantum numbers, so we are led to identify the pion as the Goldstone boson associated with the spontaneous breaking of approximate chiral symmetry. As we shall see below, it is m_π^2 rather than m_π that is proportional to a linear combination of m_u and m_d, and $m_\pi^2/m_N^2 \simeq 0.022$ is very small, so the consequences we derive from spontaneously broken $SU(2) \times SU(2)$ should be reasonably accurate.

In exploring the consequences of chiral symmetry for pion interactions, it is very useful to note that although the chiral symmetry of the strong interactions does not depend in any way on the existence of weak interactions, the symmetry currents $\vec{V}^\mu$ and $\vec{A}^\mu$ happen to be the currents entering into strangeness-conserving semileptonic weak interactions like nuclear beta decay. As we shall see in Section 21.3, the standard model of electroweak interactions requires that the effective Lagrangian for these interactions at low energy must take the form:

$$\mathscr{L}_{\rm wk} = -\frac{iG_{\rm wk}}{\sqrt{2}}\left(V_+^\lambda + A_+^\lambda\right)\sum_\ell \bar{\ell}\gamma_\lambda(1+\gamma_5)\nu_\ell + \text{H.c.} \quad (19.4.22)$$

where ℓ runs over the renormalized fields of the three charged leptons e, μ and τ; ν_ℓ runs over the renormalized fields of the associated neutrinos; and $V_\pm^\lambda$ and $A_\pm^\lambda$ are the charge changing currents

$$V_\pm^\lambda = V_1^\lambda \pm i V_2^\lambda\,, \qquad A_\pm^\lambda = A_1^\lambda \pm i A_2^\lambda\,. \quad (19.4.23)$$

The constant[††] $G_{\rm wk}$ may be measured from the rates of beta transitions between states of zero spin within the same isotopic multiplet, such as the decays $\pi^+ \to \pi^0 + e^+ + \nu_e$ and $^{14}O \to {}^{14}N^* + e^+ + \nu_e$. The momentum transfer in these transitions is very small, so parity conservation (in strong interactions) and rotational invariance tell us that only the matrix elements of $\int d^3x\, V_-^0 = T_1 - iT_2$ enter in the S-matrix elements for these decays. This operator has matrix elements between states in a given isospin multiplet that are just known Clebsch–Gordan coefficients, so from the rates for these '0 → 0' processes we may calculate a value[9] for the coupling in Eq. (19.4.22): $G_{\rm wk} \simeq 1.14959(38) \times 10^{-5}$ GeV^{-2}. On the other hand, in the process of pion decay, $\pi^+ \to \mu^+ + \nu_\mu$, the only current matrix element we need is the matrix element of A_-^λ between a one-pion state and the vacuum:

$$\langle \text{VAC}|A_i^\mu(x)|\pi_j\rangle = \frac{iF_\pi\,\delta_{ij}\,p_\pi^\mu\, e^{ip_\pi\cdot x}}{2(2\pi)^{3/2}\sqrt{2p_\pi^0}}\,, \quad (19.4.24)$$

[††] $G_{\rm wk}$ is related to the conventional Fermi coupling constant G_F and the Cabibbo angle θ_C by $G_{\rm wk} = G_F \cos\theta_C$; see Section 21.3.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

which is completely known except for the factor F_π. The rate for pion decay turns out to be

$$\Gamma(\pi \to \mu + \nu) = \frac{G_{\rm wk}^2 F_\pi^2 m_\mu^2 (m_\pi^2 - m_\mu^2)^2}{16\pi m_\pi^3} \,. \tag{19.4.25}$$

From the known rate for $\pi^+ \to \mu^+ + \nu_\mu$, $\Gamma = (2.6033(24) \times 10^{-8}\ \text{s})^{-1}$, and the above value of $G_{\rm wk}$, one finds that[‡]

$$F_\pi \simeq 184\ \text{MeV} \,. \tag{19.4.26}$$

Now let us consider the matrix element of $\vec{A}^\mu$ between one-nucleon states. This is of interest in its own right, and as discussed in Section 19.2, it provides information we need to calculate the emission of low energy pions in collisions of nucleons. Following the same reasoning as used for the electromagnetic current in Section 10.6, one finds that Lorentz invariance and parity conservation require this matrix element to take the form[‡‡]

$$\begin{aligned}\langle p|A_+^\mu(x)|n\rangle &= (2\pi)^{-3}\, e^{iq\cdot x} \\ &\times \bar{u}_p\Big[-i\gamma^\mu\gamma_5 f(q^2) + q^\mu\gamma_5 g(q^2) + iq_\nu[\gamma^\mu,\gamma^\nu]\gamma_5 h(q^2)\Big]u_n \,,\end{aligned} \tag{19.4.27}$$

where $q \equiv p_n - p_p$. In the approximation in which the $SU(2) \times SU(2)$ symmetry is exact, the conservation of current requires that

$$q_\mu \langle p|A_+^\mu(x)|n\rangle = 0 \,. \tag{19.4.28}$$

Using the defining equations for the Dirac spinors u_p and u_n,

$$\bar{u}_p(i\not{p}_p + m_N) = (i\not{p}_n + m_N)u_n = 0 \,,$$

we see that

$$q_\mu \bar{u}_p[-i\gamma^\mu\gamma_5]u_n = -2m_N \bar{u}_p\gamma_5 u_n$$

and so Eq. (19.4.28) requires that

$$2m_N f(q^2) = q^2 g(q^2) \,. \tag{19.4.29}$$

If $g(q^2)$ had no singularity at $q^2 = 0$, then (19.4.29) would require that either $m_N = 0$, which is certainly not true, or that $f(0) = 0$, which is not true either. In fact, the quantity $f(0)$ is measured in low energy nuclear

[‡] It is common to encounter a pion decay constant f_π, which in terms of the F_π used here is variously defined as F_π, $F_\pi/\sqrt{2}$, or $F_\pi/2$.

[‡‡] The currents $A_\pm^\mu$ in the standard model have charge conjugation properties that make the coefficient $h(q^2)$ vanish. It is possible that there are 'second-class' terms[10] in the weak currents with opposite charge conjugation properties that give a non-vanishing $h(q^2)$, but there is no evidence for such terms. As we shall see, keeping the $h(q^2)$ term here has no effect on the inferences drawn from chiral symmetry.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

beta decays, like neutron decay, where it is usually called g_A; it is found to have the value

$$f(0) \equiv g_A = 1.2573(28) . \tag{19.4.30}$$

The fact that neither m_N nor $f(0) = g_A$ is small requires that in the limit of exact $SU(2) \times SU(2)$ symmetry, $g(q^2)$ must have a pole as $q^2 \to 0$:

$$g(q^2) \to \frac{2m_N g_A}{q^2} . \tag{19.4.31}$$

Such a pole is naturally provided by the massless pion that would be required by the spontaneous breakdown of an exact $SU(2) \times SU(2)$ symmetry. Suppose that the pion couples to the one-nucleon state as if the interaction Lagrangian were[♮] $-2iG_{\pi N}\vec{\pi}\bar{N}\gamma_5\vec{t}N$. Eq. (19.4.24) tells us that the matrix element of the current between a one-pion state and the vacuum is the same as if there were a term in $\vec{A}^\mu(x)$ of the form $F_\pi \partial^\mu \vec{\pi}/2$. Therefore in the limit $q^2 \to 0$ the matrix element (19.4.27) has the pole

$$\langle p|A_+^\mu(x)|n\rangle \to \left[\frac{e^{iq\cdot x}}{(2\pi)^3}\right]\left[i(2\pi)^4\, 2G_{\pi N}\,\bar{u}_p(i\gamma_5)u_n\right]\left[\frac{-i}{(2\pi)^4 q^2}\right]\left[iq^\mu F_\pi/2\right] .$$

Comparing with Eq. (19.4.27), we see that one-pion exchange gives the function $g(q^2)$ a pole

$$g(q^2) \to \frac{G_{\pi N} F_\pi}{q^2} \tag{19.4.32}$$

for $q^2 \to 0$. Putting together Eqs. (19.4.31) and (19.4.32), we find

$$G_{\pi N} = \frac{2m_N g_A}{F_\pi} . \tag{19.4.33}$$

This is the famous Goldberger–Treiman relation.[11] It works reasonably well; taking $m_N = (m_p + m_n)/2 = 938.9$ MeV, $g_A = 1.257$, and $F_\pi = 184$ MeV gives $G_{\pi N} \simeq 12.7$, in fair agreement with the value[♮♮] $G_{\pi N} = 13.5$ measured in various ways (including the effects of the one-pion pole in nucleon–nuclear scattering and the one-nucleon pole in pion–nucleon scattering.)

In the real world the pion is not massless and the $SU(2) \times SU(2)$ symmetry is not exact (even before being spontaneously broken.) This circumstance may be analyzed using the general formalism presented in

[♮] This is the conventional definition of the pseudoscalar pion–nucleon coupling $G_{\pi N}$. The factor 2 is introduced here to cancel the 1/2 in the isospin matrices.

[♮♮] The textbook[12] value is $G_{\pi N}^2/4\pi = 14.3$, or $G_{\pi N} = 13.4$. More recently, a high precision study[13] of neutron–proton charge-exchange scattering at 162 MeV has given a value $G_{\pi N}^2/4\pi = 14.6 \pm 0.3$, or $G_{\pi N} = 13.5$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

the previous section. The Lagrangian (19.4.1) yields a symmetry-breaking term in the Hamiltonian

$$H_1 = m_u\bar{u}u + m_d\bar{d}d = (m_u + m_d)\Phi_4^+ + (m_u - m_d)\Phi_3^- , \tag{19.4.34}$$

where

$$\Phi_4^+ \equiv \tfrac{1}{2}(\bar{u}u + \bar{d}d) , \qquad \Phi_3^- = \tfrac{1}{2}(\bar{u}u - \bar{d}d) . \tag{19.4.35}$$

The operators Φ_4^+ and Φ_3^- are spatial scalars and, as this notation is meant to suggest, they transform under $SU(2) \times SU(2)$ as components of independent chiral four-vectors $\Phi_\alpha^\pm$:

$$\vec{\Phi}^+ = i\bar{q}\gamma_5\vec{t}q , \qquad \Phi_4^+ = \tfrac{1}{2}\bar{q}q , \tag{19.4.36}$$

$$\vec{\Phi}^- = \bar{q}\vec{t}q , \qquad \Phi_4^- = -\tfrac{1}{2}i\bar{q}\gamma_5 q . \tag{19.4.37}$$

These are chiral four-vectors, in the sense that

$$\left[\vec{T}, \Phi_n^\pm\right] = -\sum_m (\vec{\mathscr{T}})_{nm}\Phi_m^\pm , \tag{19.4.38}$$

$$\left[\vec{X}, \Phi_n^\pm\right] = -\sum_m (\vec{\mathscr{X}})_{nm}\Phi_m^\pm , \tag{19.4.39}$$

where $\vec{\mathscr{T}}$ and $\vec{\mathscr{X}}$ are Hermitian 4×4 matrices that furnish the four-vector representation of the algebra of $SU(2) \times SU(2) \equiv SO(4)$:

$$(\mathscr{T}_a)_{bc} = -i\epsilon_{abc} , \qquad (\mathscr{T}_a)_{b4} = (\mathscr{T}_a)_{4b} = (\mathscr{T}_a)_{44} = 0 , \tag{19.4.40}$$

$$(\mathscr{X}_a)_{b4} = -(\mathscr{X}_a)_{4b} = -i\delta_{ab} , \qquad (\mathscr{X}_c)_{ab} = (\mathscr{X}_c)_{44} = 0 . \tag{19.4.41}$$

This notation makes it easy to see that the vacuum alignment conditions for generators T_1, T_2, X_1, X_2, and X_3 respectively take the form

$$\begin{aligned} 0 &= \langle\Phi_2^-\rangle_0 = \langle\Phi_1^-\rangle_0 = \langle\Phi_1^+\rangle_0 = \langle\Phi_2^+\rangle_0 \\ &= (m_u + m_d)\langle\Phi_3^+\rangle_0 + (m_d - m_u)\langle\Phi_4^-\rangle_0 . \end{aligned} \tag{19.4.42}$$

We have been assuming that in the absence of u and d quark masses, the symmetry $SU(2) \times SU(2)$ is spontaneously broken in such a way as to preserve unbroken the $SU(2)$ symmetry generated by $\vec{T}$ as well as parity, in which case $\langle\Phi_n^+\rangle_0$ points in the four-direction and $\langle\Phi_n^-\rangle_0 = 0$ vanishes, so that the conditions (19.4.42) are satisfied. But with $m_u = m_d = 0$, we can find other symmetry-breaking solutions with other definitions of parity by subjecting this one to an arbitrary $SU(2) \times SU(2)$ transformation. Thus in the absence of u and d quark masses, there would be no way to tell in what direction $\langle\Phi_n^+\rangle_0 = 0$ should point, or which $SU(2)$ subgroup of $SU(2) \times SU(2)$ is left unbroken, though in all cases $\langle\Phi_n^-\rangle_0 = 0$. The vacuum alignment condition (19.4.42) tells us that, with the symmetry broken intrinsically by the perturbation (19.4.34) and spontaneously in such a way that $\langle\Phi_n^-\rangle_0 = 0$, the vacuum must 'line up' in such a way

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

that $\langle \Phi_n^+ \rangle_0 = 0$ points in the four-direction, so that the unbroken $SU(2)$ symmetry is ordinary isospin.

This formalism may be used to calculate the pion mass. From Eqs. (19.4.39)–(19.4.41) we find

$$[X_a, [X_b, \Phi_4^+]] = \delta_{ab}\Phi_4^+ , \qquad [X_a, [X_b, \Phi_3^-]] = \Phi_a^- \delta_{b3} . \tag{19.4.43}$$

Also, isospin invariance tells us that the symmetry-breaking parameter F_{ab} introduced in Section 19.2 is proportional to δ_{ab}, with a proportionality factor that according to (19.4.24) is just $F_\pi/2$, so

$$F_{ab} = \delta_{ab} F_\pi / 2 . \tag{19.4.44}$$

Eq. (19.3.20) thus gives the pion mass matrix as

$$m_{ab}^2 = \delta_{ab} m_\pi^2 , \tag{19.4.45}$$

where

$$m_\pi^2 = -4(m_u + m_d)\langle \Phi_4^+ \rangle_0 / F_\pi^2 . \tag{19.4.46}$$

It is striking that the masses of the charged and neutral pions turn out to be equal, even though we have made no assumption about the ratio of m_u and m_d. We shall see below that this ratio is not near unity; isospin is a good quantum number not because the u and d quark masses are nearly equal, but because they are *small*. Also, as promised, we see that it is the *square* of the pion mass that is proportional to the quark masses, so the quark masses should be quite small. (See Section 19.7 for an estimate.) The observed pion mass difference arises not from the u–d quark mass difference, but from electromagnetism. Indeed, the pion isospin multiplet is the only one whose mass difference has been successfully calculated on the basis of one-photon exchange alone.[14]

Of course, even with quark masses taken into account, Eq. (19.4.24) can still be used to define F_π; its divergence gives

$$\langle \text{VAC} | \partial_\mu A_i^\mu(x) | \pi_j \rangle = \frac{F_\pi \delta_{ij} m_\pi^2 e^{ip_\pi \cdot x}}{2(2\pi)^{3/2}\sqrt{2p_\pi^0}} . \tag{19.4.47}$$

Instead of assuming that $\partial_\mu \vec{A}^\mu$ vanishes, we can now assume that it is *small*, of order m_π^2, except where a pion pole compensates for the smallness of m_π^2. According to Eq. (19.4.47), the behavior of matrix elements of $\partial_\mu \vec{A}^\mu$ near a one-pion pole is the same as if $\partial_\mu \vec{A}^\mu$ were $F_\pi m_\pi^2$ times a properly renormalized pion field. For instance, the one-nucleon matrix element of $\partial_\mu A_+^\mu$ should be

$$\langle p | \partial_\mu A_+^\mu(0) | n \rangle \simeq \frac{F_\pi m_\pi^2}{2} \left[\frac{-i}{(2\pi)^4 (q^2 + m_\pi^2)} \right] \left[i(2\pi)^4 2 G_{\pi N} \bar{u}_p (-i\gamma_5) u_n \right] , \tag{19.4.48}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

so in terms of the form-factors in (19.4.27)

$$q^2 g(q^2) - 2m_N f(q^2) \simeq -\frac{G_{\pi N} F_\pi m_\pi^2}{q^2 + m_\pi^2} \ . \tag{19.4.49}$$

This is expected to be valid when q^2 is of order m_π^2, and not only in the limit $q^2 \to -m_\pi^2$, because the pion pole dominates matrix elements of $\partial_\mu A^\mu$ for all such small q^2. Also, for such q^2, in place of Eq. (19.4.32) we have

$$g(q^2) \simeq \frac{G_{\pi N} F_\pi}{q^2 + m_\pi^2} \ . \tag{19.4.50}$$

From Eqs. (19.4.49) and (19.4.50) we find that for q^2 of order m_π^2,

$$f(q^2) \simeq G_{\pi N} F_\pi \Big/ 2m_N \ . \tag{19.4.51}$$

It should be no surprise that this function is roughly constant over the range of q^2 from zero to of order m_π^2, because it has no one-pion pole, and there is nothing else that could give it a substantial variation in such a small range of q^2. This constant value is approximately $f(0) \equiv g_A$, and Eq. (19.4.51) thus again yields the Goldberger–Treiman relation.

We can now use the results of Section 19.2 to calculate the amplitude for emission of a single low energy pion in an arbitrary process $\alpha \to \beta$. We found that the amplitude is to be calculated from Feynman diagrams in which the pion is emitted only from the external lines of the process, and Eq. (19.2.49) shows that these contributions are to be calculated as if the pion field interaction were $\partial_\mu \vec{\pi} \cdot \vec{A}_N / F_\pi$, in which the subscript N indicates that we are to drop the one-pion pole term in matrix elements of the axial-vector current. From Eq. (19.4.27) (and using isospin invariance) we conclude that for soft pions emitted from a nucleon line this interaction is effectively

$$-\frac{ig_A}{F_\pi} \ \partial_\mu \vec{\pi} \ \bar{N} \gamma^\mu \gamma_5 \vec{t} N \ .$$

Using the free-particle Dirac equation, we can see that for nucleons on the mass shell (that is, at the one-nucleon poles in Figure 19.2) this is equivalent to the pseudoscalar interaction $-2i m_N g_A \vec{\pi} \cdot \bar{N} \gamma_5 \vec{t} N / F_\pi$. This provides yet another demonstration of the Goldberger–Treiman relation (19.4.33).

* * *

Our discussion here has not paralleled the historical development of these ideas. In fact, the course of historical development was chronologically almost exactly opposite to the line of argument presented here. Broken symmetries in particle physics started with the Goldberger–Treiman relation (19.4.33), which was derived[11] in 1957 on the basis of a dynamical

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

calculation of pion decay. In order to explain the surprising success of this very approximate calculation, several theorists[15] introduced the idea of a 'partial conservation of the axial-vector current' (PCAC), the idea that although the axial-vector current is not conserved (as shown by the fact that pions do decay) its divergence $\partial_\mu A^\mu_\pm$ is proportional to the pion field. In itself, this assumption is meaningless — we saw in Chapter 10 that *any* field with a non-vanishing matrix element between the vacuum and one-pion states may be regarded as a pion field. Although it was not clear at the time, what was really being assumed was that the divergence of the axial-vector current is *small*, of order m_π^2, except where a pion pole gives it a large matrix element. The problem was greatly clarified by a 1960 paper of Nambu,[16] who pointed out that the axial-vector current could be regarded as exactly conserved in the limit of zero pion mass, in which case the Goldberger–Treiman relation could be derived as we have done here. In this and a subsequent paper with Jona-Lasinio,[17] Nambu recognized that the appearance of this massless or nearly massless pion was a symptom of a broken exact or approximate symmetry. With other collaborators,[18] Nambu also showed how to calculate the rates of emission of a single soft pion in various processes. Subsequently Goldstone[3] remarked that broken symmetries always entail massless bosons, and this was proved in 1962 by Goldstone, Salam, and myself,[5] using the arguments presented here in Section 19.2.

None of this early work on pions as Goldstone bosons depended on any specific assumptions about the nature of the broken symmetry group; for instance it might have been a direct product of three commuting $U(1)$ groups, whose generators form an isotopic spin vector, or it might have been the non-compact group $SO(3,1)$, for which a minus sign appears on the right-hand side of the commutation relation (19.4.19). The nature of the broken symmetry group became important only with the consideration of processes involving more than one pion, starting with the Adler–Weisberger sum rule in 1965,[19] whose success showed that the broken symmetry is indeed $SU(2) \times SU(2)$. (Such processes are discussed in the following section.) The identification of the symmetry group $SU(2) \times SU(2)$ led to a shift of emphasis,[20] toward the implications of this symmetry within strong interaction physics, and away from an earlier concentration[21] on the currents themselves.

All of this work was done without a specific theory of the strong interactions. One of the reasons for the rapid acceptance of quantum chromodynamics in 1973 as the correct theory of strong interactions was that it explained the $SU(2) \times SU(2)$ symmetry as a simple consequence of the smallness of the u and d quark masses.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

19.5 Effective Field Theories: Pions and Nucleons

In Section 19.2 we learned how to calculate the amplitude for emission of a single low energy Goldstone boson B in a transition $\alpha \to \beta + B$ by applying the condition of current conservation to the matrix element of the symmetry current between the states α and β. In this calculation we never had to use any information about the details of the broken symmetry group; current conservation was all we needed. A new element enters if we wish to calculate the matrix element for the emission and/or absorption of *two* Goldstone bosons, as for instance in a Goldstone boson scattering process. Here we must apply the condition of current conservation to a matrix element like

$$\langle\beta|T\{J_1^{\lambda_1}(x_1), J_2^{\lambda_2}(x_2)\}|\alpha\rangle ,$$

where the states α and β contain the other particles besides the two Goldstone bosons participating in the reaction. But when we let the divergence operator $\partial/\partial x_1^{\lambda_1}$ act on this matrix element we encounter a non-zero contribution from the derivative of the functions $\theta(x_1^0 - x_2^0)$ and $\theta(x_2^0 - x_1^0)$ that appear in the definition of the time-ordered product $T\{\cdots\}$. This contribution is equal to the matrix element of an equal-time commutator $\delta(x_1^0 - x_2^0)[J_1^0(x_1), J_2^{\lambda_2}(x_2)]$, whose value depends on the commutation relations of the group algebra. This makes such multi-Goldstone-boson processes specially interesting, because they can be used to decide experimentally on the nature of the broken symmetry group, in a way that is not possible for processes involving just a single Goldstone boson. Because of the appearance of such current commutators, this approach is known as the method of *current algebra*.[21]

The current algebra method was used in early calculations of multi-Goldstone-boson amplitudes.[22] Unfortunately, such calculations are tedious, especially when three or more Goldstone bosons are involved, and it was also difficult to see how to deal with symmetries like the chiral symmetry of quantum chromodynamics that are not exact. For this reason a different and more physical calculational technique was introduced,[23] based on the use of effective Lagrangians: We simply calculate the Goldstone boson amplitudes by the methods of perturbation theory, using some Lagrangian for the Goldstone bosons and the other particles in the states α and β that obeys the assumed broken symmetry.

Originally the justification for the effective Lagrangian procedure was based on current algebra. By using current algebra one could see that the amplitude for emission of a set of low energy Goldstone bosons in a process $\alpha \to \beta + B_1 + B_2 + \cdots$ is fixed once one knows the equal-time commutation relations of the currents associated with the broken symmetries as well as the matrix element for the process $\alpha \to \beta$ and the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

matrix elements of the currents between various one-particle states. We know that a Lagrangian that respects the broken symmetry will allow the construction of conserved currents with the appropriate equal-time commutators by Noether's technique, so if we simply calculate the low energy Goldstone boson amplitudes with such a Lagrangian and insert the correct values for $M_{\beta\alpha}$ and the one-particle matrix elements of the currents, we must get the same answer as provided by current algebra. In the case of interactions among Goldstone bosons alone, the states α and β may both be taken as the vacuum, and we do not need any extra information beyond the matrix element F between a Goldstone boson state and the vacuum.

In the first example of this sort,[23] the starting point was the Lagrangian of the σ-model.[24] Restricting our attention for the moment to the bosonic sector of this model, its Lagrangian is the $SO(4)$-invariant one used as an example in Section 19.2:

$$\mathscr{L} = -\frac{1}{2}\partial_\mu\phi_n\partial^\mu\phi_n - \frac{\mathscr{M}^2}{2}\phi_n\phi_n - \frac{\lambda}{4}(\phi_n\phi_n)^2 , \tag{19.5.1}$$

where n is understood to be summed over the values 1, 2, 3, 4, with $\vec{\phi}$ an isovector pseudoscalar field and ϕ_4 an isoscalar scalar field.

The immediate problem faced by any sort of effective Lagrangian is that in order to use it to calculate scattering amplitudes, we must either include all Feynman diagrams of all orders of perturbation theory, or else find some rationale for dropping higher-order diagrams. We can find no such rationale with Lagrangians like (19.5.1), in which the broken symmetry is realized through linear transformations on the various fields. Fortunately any such Lagrangian can be recast in a form that allows the use of Feynman diagrams to generate an expansion for scattering amplitudes in powers of Goldstone boson energies. To do this, we perform a symmetry transformation at every point in spacetime that eliminates the fields corresponding to the Goldstone bosons of the theory. The Goldstone boson degrees of freedom reappear in the transformed theory as the parameters of this symmetry transformation. However, since the Lagrangian is invariant under spacetime-independent symmetry transformations, it cannot have any dependence on the new Goldstone boson fields when they are constant, and so every term in the Lagrangian that involves these new Goldstone boson fields must contain at least one spacetime derivative of the field. These derivatives introduce factors of the Goldstone boson energy when we calculate S-matrix elements for Goldstone boson reactions, and as we shall see we can use the Lagrangian to construct a series for the S-matrix elements in powers of these energies.

For instance, to recast the Lagrangian (19.5.1) in a useful form, we write the four-vector ϕ_n as a chiral rotation R acting on a four-vector $\{0, 0, 0, \sigma\}$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

whose first three components (the Goldstone part of ϕ_n) vanish:

$$\phi_n(x) = R_{n4}(x)\sigma(x) \tag{19.5.2}$$

with $R_{nm}(x)$ an orthogonal matrix

$$R^T(x)\,R(x) = 1 \tag{19.5.3}$$

and therefore

$$\sigma(x) = \sqrt{\sum_n \phi_n(x)^2}\,. \tag{19.5.4}$$

The Lagrangian (19.5.1) then becomes

$$\mathscr{L} = -\frac{1}{2}\sum_{n=1}^{4}\left(R_{n4}\partial_\mu\sigma + \sigma\partial_\mu R_{n4}\right)^2 - \frac{1}{2}\,\mathscr{M}^2\sigma^2 - \frac{\lambda}{4}\,\sigma^4\,. \tag{19.5.5}$$

Because R is an orthogonal matrix, the $\partial_\mu\sigma\partial^\mu\sigma$ term is R-independent, and the cross-term vanishes

$$\sum_n R_{n4}^2 = 1\,, \qquad \sum_n R_{n4}\partial_\mu R_{n4} = \frac{1}{2}\,\partial_\mu\sum_n R_{n4}^2 = 0\,,$$

so $\mathscr{L}$ becomes

$$\mathscr{L} = -\frac{1}{2}\,\partial_\mu\sigma\partial^\mu\sigma - \frac{1}{2}\,\sigma^2\sum_{n=1}^{4}\partial^\mu R_{n4}\,\partial_\mu R_{n4} - \frac{1}{2}\mathscr{M}^2\sigma^2 - \frac{\lambda}{4}\sigma^4\,. \tag{19.5.6}$$

If $\mathscr{M}^2$ is negative then σ has a non-vanishing vacuum expectation value, given in lowest order by the position of the minimum of the sum of the last two terms, at $\sigma = |\mathscr{M}|/\sqrt{\lambda}$.

In place of the field variables ϕ_n, our variables now are $\sigma - |\mathscr{M}|/\sqrt{\lambda}$ and whatever other variables are needed to parameterize the rotation R. For instance, these parameters could be simply chosen as the R_{a4} themselves (where $a, b, \cdots$ from now are isovector indices running over the values 1,2,3), with R_{44} given by the condition that R is orthogonal. A different parameterization will give simpler final results, and was historically the first to be used for these purposes. It is to define

$$\zeta_a \equiv \frac{\phi_a}{\phi_4 + \sigma} \tag{19.5.7}$$

and take

$$R_{a4} = \frac{2\zeta_a}{1+\vec{\zeta}^2} = -R_{4a}\,, \quad R_{44} = \frac{1-\vec{\zeta}^2}{1+\vec{\zeta}^2}\,, \quad R_{ab} = \delta_{ab} - \frac{2\zeta_a\zeta_b}{1+\vec{\zeta}^2}\,, \tag{19.5.8}$$

so that

$$\phi_a/\sigma = R_{a4} = \frac{2\zeta_a}{1+\vec{\zeta}^{\,2}}\,, \qquad \phi_4/\sigma = R_{44} = \frac{1-\vec{\zeta}^{\,2}}{1+\vec{\zeta}^{\,2}}\,. \tag{19.5.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Then the Lagrangian (19.5.6) becomes

$$\mathscr{L} = -\frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - 2\sigma^2\vec{D}_\mu\cdot\vec{D}^\mu - \frac{1}{2}\mathscr{M}^2\sigma^2 - \frac{\lambda}{4}\sigma^4\,, \tag{19.5.10}$$

where

$$\vec{D}_\mu \equiv \frac{\partial_\mu\vec{\zeta}}{1+\vec{\zeta}^{\,2}}\,. \tag{19.5.11}$$

Whatever parameterization we use, it is clear that the fields $\vec{\zeta}$ describe particles of zero mass, whose interactions all involve field derivatives. These are (up to a normalization) our new pion fields.

Despite appearances, this Lagrangian is still invariant under $SO(4)$, but with $SO(4)$ now realized non-linearly. Under an isospin transformation with infinitesimal parameters $\vec{\theta}$, the field $\vec{\zeta}$ simply rotates like an ordinary isovector, and σ is an isoscalar:

$$\delta\vec{\zeta} = \vec{\theta}\times\vec{\zeta}\,, \qquad \delta\sigma = 0\,, \tag{19.5.12}$$

so the Lagrangian (19.5.10) is manifestly isospin-invariant. On the other hand, under the broken symmetry transformations parameterized by an infinitesimal vector $\vec{\epsilon}$, the original fields transform according to

$$\delta\vec{\phi} = 2\vec{\epsilon}\phi_4\,, \qquad \delta\phi_4 = -2\vec{\epsilon}\cdot\vec{\phi}\,. \tag{19.5.13}$$

From Eq. (19.5.7), we then find

$$\delta\vec{\zeta} = \vec{\epsilon}(1-\vec{\zeta}^{\,2}) + 2\vec{\zeta}(\vec{\epsilon}\cdot\vec{\zeta})\,, \qquad \delta\sigma = 0\,. \tag{19.5.14}$$

The Lagrangian (19.5.10) is invariant under the broken symmetry transformation (19.5.14) because $\vec{D}_\mu$ undergoes a linear (though field-dependent) isospin rotation:[25]

$$\delta\vec{D}_\mu = 2(\vec{\zeta}\times\vec{\epsilon})\times\vec{D}_\mu \tag{19.5.15}$$

and Eq. (19.5.10) is isospin-invariant. Because of the transformation rule (19.5.15), $\vec{D}_\mu$ is often called the *covariant derivative* of the pion field.

The transformation rules (19.5.12) and (19.5.14) specify what is called a *non-linear realization* of the group $SU(2)\times SU(2)$.[25] The general theory of non-linear realizations of Lie groups is given in the next section; we shall show there that, up to field redefinitions, the transformation rules (19.5.12) and (19.5.14) provide the most general realization of $SU(2)\times SU(2)$ in which the isospin $SU(2)$ subgroup is realized linearly on $\vec{\zeta}$.

We see that each interaction of these new pion fields is accompanied with a spacetime derivative, so that the effective coupling is weak for low pion energies. (This remark will be made more precise below.) Therefore for sufficiently small pion energies we may use this Lagrangian in the tree approximation to reproduce the soft-pion theorems of current algebra. For

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

this purpose, it is only necessary that the Lagrangian be $SO(4)$-invariant. But since the σ field is an $SO(4)$ scalar, it plays no role in maintaining the $SO(4)$ invariance of the Lagrangian, and may be simply discarded.* Of course, this procedure changes the physical content of the theory, but it does not change the amplitudes given by soft-pion theorems. The Lagrangian (19.5.10) then simplifies to

$$\mathscr{L} = -\frac{F^2}{2}\vec{D}_\mu \vec{D}^\mu = -\frac{F^2}{2}\frac{\partial_\mu \vec{\zeta}\cdot\partial^\mu\vec{\zeta}}{(1+\vec{\zeta}^{\,2})^2}\,, \tag{19.5.16}$$

where $F = 2\langle\sigma\rangle = 2|\mathscr{M}|/\sqrt{\lambda}$. (As we shall soon see, this F is the same as the constant F_π discussed in the previous section.) For many purposes it is more convenient to work with a conventionally normalized pion field

$$\vec{\pi} \equiv F\vec{\zeta} \tag{19.5.17}$$

for which the Lagrangian (19.5.16) reads

$$\mathscr{L} = -\frac{1}{2}\frac{\partial_\mu\vec{\pi}\cdot\partial^\mu\vec{\pi}}{(1+\vec{\pi}^2/F^2)^2}\,. \tag{19.5.18}$$

The factor $1/F$ acts as a coupling parameter that accompanies the interaction of each additional pion. Eq. (19.5.18) describes what is often called the 'non-linear σ-model,' for the special case of $SU(2)\times SU(2)$ spontaneously broken to $SU(2)$.

An important point: to derive Eq. (19.5.18) it was not really necessary to start with the 'linear σ-model' Lagrangian (19.5.1). Indeed, we did not need to start with *any* specific theory. Eq. (19.5.18) can be used simply because it is invariant under the $SO(4)$ transformation (19.5.12), (19.5.14), and current algebra tells us that this is all we need to get the right results for low energy pion reaction amplitudes.

Some years after the introduction of effective Lagrangians for soft pions, there emerged a different justification for the effective field theory technique,[26] one that does not rely on current algebra and allows calculations that are not limited to the limit of vanishingly small Goldstone boson energies. It is based on the realization (not yet formally embodied in a theorem) that when we calculate a physical amplitude from Feynman diagrams using the most general Lagrangian that involves the relevant degrees of freedom and satisfies the assumed symmetries of the theory, we are simply constructing the most general amplitude that is consistent with general principles of relativity, quantum mechanics, and the assumed symmetries. This was the point of view underlying Volume I of this book. In the present context, the 'relevant' degrees of freedom are the Goldstone

* Alternatively, we can pass to the limit where $\mathscr{M}$ and λ go to infinity together, keeping the expectation value of σ constant.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

bosons themselves, together with the particles in the states α and β and any other particle states that can be produced from them by interactions with low energy Goldstone bosons. By invoking this justification for effective field theories, we are freed from any need to wrestle with the complications of current algebra. More importantly, the modern effective field theory approach yields results that take us beyond the extreme low energy limit and allows a systematic treatment of any intrinsic symmetry breaking.

According to this approach, in order to calculate pion interaction amplitudes to any desired order in pion energies, we must use the most general Lagrangian involving a pion field $\vec{\zeta}$ that transforms according to the rules (19.5.12) and (19.5.14):

$$\mathscr{L}_{\text{eff}} = -\frac{F^2}{2}\vec{D}_\mu\vec{D}^\mu - \frac{c_4}{4}(\vec{D}_\mu\vec{D}^\mu)^2 - \frac{c_4'}{4}(\vec{D}_\mu\cdot\vec{D}_\nu)(\vec{D}^\mu\cdot\vec{D}^\nu) - \ldots . \quad (19.5.19)$$

The terms indicated by ... will contain higher powers of the covariant derivative $\vec{D}_\mu$, or higher covariant derivatives, whose general structure is described in the next section. The coefficients c_4 and c_4' are dimensionless, and all higher terms have coefficients with the dimensionality of negative powers of mass.

Consider a general process involving arbitrary numbers of incoming and outgoing pions. We suppose that their energies and momenta are all at most of some order Q, which is small compared with a typical quantum chromodynamics energy scale (say, the nucleon or ρ mass). Even though Lagrangians like (19.5.19) are not renormalizable in the usual sense, we saw in Section 12.3 that such Lagrangians can yield finite results as long as they contain all possible terms allowed by symmetries, for then there will be a counterterm available to cancel every infinity. If we define the renormalized values of the constants F^2, c_4, c_4', ... by specifying the values of various Goldstone boson scattering amplitudes at energies of order Q, then the integrals in momentum-space Feynman diagrams will be dominated by contributions from virtual momenta which are also of order Q (because renormalization makes them finite, and there is no other possible effective cut-off in the theory.) We can then develop perturbation theory as a power series expansion in Q.

Each derivative and hence each $\vec{D}_\mu$ in each interaction vertex contributes one factor of Q to the order of magnitude of the diagram; each internal pion propagator contributes a factor Q^{-2}; and each integration volume d^4q associated with the loops of the diagram contributes a factor Q^4; so a general connected diagram makes a contribution of order Q^ν, where

$$\nu = \sum_i V_i d_i - 2I + 4L . \quad (19.5.20)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Here d_i is the number of derivatives in an interaction of type i, V_i is the number of interaction vertices of type i in the diagram, I is the number of internal pion lines, and L is the number of loops. These quantities are related by the familiar topological identity (see Eq. (4.4.7)):

$$L = I - \sum_i V_i + 1 \,, \tag{19.5.21}$$

so we can eliminate I and write

$$\nu = \sum_i V_i(d_i - 2) + 2L + 2 \,. \tag{19.5.22}$$

The point of this is that each term here is positive; every interaction in Eq. (19.5.19) has at least two derivatives, and of course $L \geq 0$. Therefore the leading term of each process is of order Q^2, and arises solely from *tree* graphs (that is, $L = 0$) constructed solely from the term in Eq. (19.5.19) with only two derivatives (that is, $V_i = 0$ for $d_i \neq 2$):

$$\mathscr{L}_2 = -\frac{F^2}{2}\,\vec{D}_\mu \vec{D}^\mu = -\frac{1}{2}\,\frac{\partial_\mu \vec{\pi}\partial^\mu \vec{\pi}}{(1+\vec{\pi}^2/F^2)^2} \,. \tag{19.5.23}$$

For instance, the invariant amplitude M that appears in the pion–pion scattering S-matrix element

$$S = i(2\pi)^4\delta^4(p_A + p_B - p_C - p_D)M(2\pi)^{-6}(16E_AE_BE_CE_D)^{-1/2} \tag{19.5.24}$$

is given to this order by

$$\begin{aligned} M^{(\nu=2)}_{abcd} = 4F^{-2}\Big[&\delta_{ab}\delta_{cd}(-p_A\cdot p_B - p_C\cdot p_D) + \delta_{ac}\delta_{bd}(p_A\cdot p_C + p_B\cdot p_D) \\ &+ \delta_{ad}\delta_{bc}(p_A\cdot p_D + p_B\cdot p_C)\Big] \,, \end{aligned} \tag{19.5.25}$$

where a, b, c, d are the isovector indices of the pions A, B, C, D, respectively. (The effect of the finite pion mass will be taken up a little later in this section.) Using Eq. (19.5.25) as the leading term does not depend on any assumption of the smallness of the coupling constant λ in the original Lagrangian (19.5.1), or even on the validity of this formula for the Lagrangian, but only on the assumed smallness of the typical pion energy Q.

The next term in the amplitude for any Goldstone boson reaction will be of order Q^4, and arises both from one-loop graphs involving only the Lagrangian (19.5.16), and also from tree graphs constructed solely from the interaction (19.5.16) plus a single vertex arising from the $d = 4$ terms

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

in Eq. (19.5.19):

$$M^{(\nu=4)}_{abcd} = \frac{\delta_{ab}\delta_{cd}}{F^4}\left[-\frac{1}{2\pi^2}\, s^2 \ln(-s) - \frac{1}{12\pi^2}\,(u^2 - s^2 + 3t^2)\ln(-t) \right.$$
$$- \frac{1}{12\pi^2}\,(t^2 - s^2 + 3u^2)\ln(-u) + \frac{1}{3\pi^2}(s^2 + t^2 + u^2)\ln\Lambda^2$$
$$\left. - \frac{1}{2}\, c_4 s^2 - \frac{1}{4}\, c'_4(t^2 + u^2)\right] + \text{crossed terms}\,, \qquad (19.5.26)$$

where 'crossed terms' denotes terms given by the interchanges $B \leftrightarrow C$ and $B \leftrightarrow D$, and s, t, and u are the Mandelstam variables

$$s = -(p_A + p_B)^2,\ t = -(p_A - p_C)^2,\ u = -(p_A - p_D)^2\,.$$

The dependence of this result on the cut-off Λ can be eliminated by a redefinition of the constants c_4 and c'_4. The renormalized couplings are

$$c_{4R} = c_4 - \frac{2}{3\pi^2}\,\ln\left(\frac{\Lambda^2}{\mu^2}\right)\,, \qquad (19.5.27)$$

$$c'_{4R} = c'_4 - \frac{4}{3\pi^2}\,\ln\left(\frac{\Lambda^2}{\mu^2}\right)\,, \qquad (19.5.28)$$

where μ is an arbitrary renormalization scale of order Q inserted in order to make the logarithms well-defined. In terms of these renormalized couplings, the amplitude (19.5.26) takes the form

$$M^{(\nu=4)}_{abcd} = \frac{\delta_{ab}\delta_{cd}}{F^4}\left[-\frac{1}{2\pi^2}\, s^2 \ln\left(\frac{-s}{\mu^2}\right) - \frac{1}{12\pi^2}\,(u^2 - s^2 + 3t^2)\ln\left(\frac{-t}{\mu^2}\right)\right.$$
$$\left. - \frac{1}{12\pi^2}\,(t^2 - s^2 + 3u^2)\ln\left(\frac{-u}{\mu^2}\right) - \frac{1}{2}\, c_{4R} s^2 - \frac{1}{4}\, c'_{4R}(t^2 + u^2)\right]$$
$$+ \text{crossed terms}\,. \qquad (19.5.29)$$

This sort of calculation can be carried to arbitrary orders in Q, always with the result that in each order we encounter a finite number of new couplings whose renormalization serves to eliminate the cut-off dependence of physical amplitudes. Notice that the ratio of the leading $\nu = 2$ terms and the $\nu = 4$ corrections is of order $Q^2/8\pi^2F^2$, so this is likely to be a useful expansion as long as the pion energies are all much less than an amount of order $2\pi F$.

This perturbative expansion may also be used to relate the ubiquitous parameter F to the measured pion decay amplitude F_π. Recalling the transformation rule (19.5.14) for $\vec{\zeta}$, the axial-vector current is given by

$$\vec{\epsilon}\cdot\vec{A}^\mu = -\frac{\partial\mathscr{L}}{\partial(\partial_\mu\vec{\zeta})}\cdot\delta\vec{\zeta}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

and so

$$\vec{A}^\mu = -(1-\vec{\zeta}^{\,2})\frac{\partial \mathscr{L}}{\partial(\partial_\mu \vec{\zeta})} - 2\vec{\zeta}\,\vec{\zeta}\cdot\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \vec{\zeta})}\right). \tag{19.5.30}$$

(This $\vec{A}^\mu$ is the Noether current of the symmetry generated by $2\vec{x}$, which on the nucleon doublet is represented by $2\gamma_5\vec{t} = \gamma_5\vec{\tau}$.) After integrating out all other fields, and using Eq. (19.5.17) to express $\vec{\zeta}$ in terms of the canonically normalized pion field $\vec{\pi}$, we find

$$\vec{A}^\mu = F\left[\partial^\mu\vec{\pi}\frac{(1-\vec{\pi}^2/F^2)}{(1+\vec{\pi}^2/F^2)} + \frac{2\vec{\pi}(\vec{\pi}\cdot\partial^\mu\vec{\pi})}{F^2(1+\vec{\pi}^2/F^2)^2}\right] + \cdots . \tag{19.5.31}$$

Using our perturbative expansion in powers of the pion energy to calculate $\langle \mathrm{VAC}|\vec{A}^\mu|\pi\rangle$, we see that in lowest order the pion decay amplitude here is

$$F_\pi = F . \tag{19.5.32}$$

Furthermore Lorentz invariance and Eq. (19.5.22) tell us that the higher-order corrections must be proportional to powers of p_π^2/F_π^2, which vanishes for a massless pion, so Eq. (19.5.32) is actually exact in the limit $m_\pi \to 0$. We may therefore guess that our perturbation expansion will be useful for pion energies that are less than an amount of order $2\pi F_\pi = 1200$ MeV.

In order to make contact with experiment, we must deal with the fact that the pion mass is not zero. On the mass shell it is not possible for the time component of a pion four-momentum to be less than m_π, so in counting powers of the typical pion energy and momentum Q, we must regard m_π as being of order Q. But we saw in the previous section that m_π^2 is proportional to a linear combination of quark masses, so our formula (19.5.22) for the order Q^ν of a given Feynman diagram should be rewritten to read

$$\nu = \sum_i V_i(d_i + 2m_i - 2) + 2L + 2 , \tag{19.5.33}$$

where m_i is the number of factors of quark masses in the interaction of type i.

The interactions involving quark masses may be distinguished by their transformation properties under $SU(2) \times SU(2)$, or equivalently, under $SO(4)$. Eq. (19.4.34) shows that the terms of first order in quark masses transform as a linear combination of two scalars, the fourth component of a chiral four-vector Φ_n^+, with coefficient $m_u + m_d$, and the third component of a different chiral four-vector Φ_n^-, with coefficient $m_u - m_d$. Thus the terms $\Delta\mathscr{L}^+$ and $\Delta\mathscr{L}^-$ in the effective Lagrangian that are of first order in $m_u + m_d$ and $m_u - m_d$ must have the chiral transformation properties of the fourth and third components of chiral four-vectors, respectively, and of course be Lorentz scalars. One obvious candidate for a chiral

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

four-vector whose fourth component is a scalar is the ϕ_n with which we started in this section. According to Eq. (19.5.9), its fourth component is just $\sigma(1-\vec{\zeta}^2)/(1+\vec{\zeta}^2)$. The factor σ may be dropped, as it is a chiral scalar, and therefore has no effect on the chiral transformation properties of this quantity. We can fix the normalization of this term by requiring that the coefficient of $\vec{\pi}^2 = F_\pi^2\vec{\zeta}^2$ be $-m_\pi^2/2$, so that, apart from an additive constant,

$$\Delta\mathscr{L}^+ = -\frac{m_\pi^2 F_\pi^2}{2}\frac{\vec{\zeta}^2}{1+\vec{\zeta}^2} = -\frac{m_\pi^2}{2}\frac{\vec{\pi}^2}{1+\vec{\pi}^2/F_\pi^2}\,. \tag{19.5.34}$$

We shall see in the next section that this is the unique scalar function of the pion field without derivatives that transforms as the fourth component of a chiral four-vector. On the other hand, there is *no* scalar function of the pion field without derivatives that transforms as the third component of a chiral four-vector because such a function, if the third component of a chiral four-vector, would have to be odd in the pion field, and therefore pseudoscalar rather than scalar. Thus (19.5.34) is the only interaction with $d_i = 0$ and $m_i = 1$. It is striking that the isospin violating difference in the u and d quark masses has not only no effect on the pion masses, as we saw in the previous section, but also has no effect on any non-derivative multipion interaction.

We are now in a position to do a realistic calculation of the leading $\nu = 2$ terms in the pion–pion scattering amplitude. According to Eq. (19.5.33), these terms arise only from tree graphs constructed from the $d = 2$, $m = 0$ pion interaction in Eq. (19.5.23) or from the $d = 0$, $m = 1$ pion interaction in Eq. (19.5.34). To this order, the invariant amplitude M defined by Eq. (19.5.24) is now

$$M_{abcd}^{(\nu=2)} = 4F_\pi^{-2}\left[\delta_{ab}\delta_{cd}(s-m_\pi^2)+\delta_{ac}\delta_{bd}(t-m_\pi^2)+\delta_{ad}\delta_{bc}(u-m_\pi^2)\right]. \tag{19.5.35}$$

In particular, at threshold $s = 4m_\pi^2$, $t = u = 0$, so

$$\begin{aligned} M_{abcd}^{(\nu=2)}(\text{threshold}) &= 4m_\pi^2 F_\pi^{-2}[3\delta_{ab}\delta_{cd} - \delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}] \\ &= 4m_\pi^2 F_\pi^{-2}\left[7\,M_{ab,cd}^{(0)} - 2\,M_{ab,cd}^{(2)}\right], \end{aligned} \tag{19.5.36}$$

where $M^{(0)}$ and $M^{(2)}$ are the appropriate tensors representing two-pion states with isospin $T = 0$ or $T = 2$:

$$M_{ab,cd}^{(0)} = \tfrac{1}{3}\delta_{ab}\delta_{cd}\,, \tag{19.5.37}$$

$$M_{ab,cd}^{(2)} = \tfrac{1}{2}\left(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc} - \tfrac{2}{3}\delta_{ab}\delta_{cd}\right), \tag{19.5.38}$$

normalized so that $\mathrm{Tr}M^{(T)} = 2T + 1$. This result is usually expressed in terms of the scattering lengths. According to Sections 3.6 and 3.7, the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

scattering length a_T for two pions in a state of isospin T is given[27] by $1/32\pi m_\pi$ times the coefficient of $M^{(T)}$ in Eq. (19.5.36):

$$a_0 = \frac{7m_\pi}{8\pi F_\pi^2} = 0.16\, m_\pi^{-1}\,, \qquad a_2 = -\frac{m_\pi}{4\pi F_\pi^2} = -0.046\, m_\pi^{-1}\,.$$

The pion scattering lengths are difficult to measure, but careful study of processes like $\pi+N \to \pi+\pi+N$ and $K \to \pi+\pi+e+\nu$ has given the results[28] $(0.26\pm0.05)\, m_\pi^{-1}$ and $(-0.028\pm0.012)\, m_\pi^{-1}$ for a_0 and a_2, respectively, which are consistent with these theoretical values. Corrections of higher order in $m_\pi/2\pi F_\pi$ seem to improve the agreement.

This formalism can be extended to describe the interactions of pions with nucleons or other particles. The easiest approach is to suppose that we add a term involving the nucleon doublet field N to the Lagrangian with which we started:

$$\mathscr{L}_N = -\bar{N}\left(\not{\partial} + g[\phi_4 + 2i\vec{t}\cdot\vec{\phi}\gamma_5]\right) N\,, \tag{19.5.39}$$

where $\vec{t}$ is the isospin matrix vector for isospin 1/2 (that is, half the Pauli spinor $\vec{\tau}$.) This is invariant under chiral $SU(2)\times SU(2)$ transformations with

$$\delta\vec{\phi} = 2\vec{\epsilon}\,\phi_4\,, \qquad \delta\phi_4 = -2\vec{\epsilon}\cdot\vec{\phi}\,, \tag{19.5.40}$$

$$\delta N = -2i\gamma_5\vec{\epsilon}\cdot\vec{t}\,N\,. \tag{19.5.41}$$

Now in eliminating the non-derivative pion couplings we must express the nucleon field N as the $SO(4)$ rotation R in the representation (19.5.41) acting upon a new nucleon field $\tilde{N}$

$$N \equiv \frac{(1-2i\gamma_5\vec{t}\cdot\vec{\zeta})}{\sqrt{1+\vec{\zeta}^{\,2}}}\,\tilde{N} \tag{19.5.42}$$

with $\vec{\zeta}$ given again by Eq. (19.5.7). With this transformation the non-derivative term in Eq. (19.5.39) now depends only on $\tilde{N}$ and σ:

$$\bar{N}[\phi_4 + 2i\vec{t}\cdot\vec{\phi}\gamma_5]N = \sigma\bar{\tilde{N}}\tilde{N}\,.$$

On the other hand, the derivative term involves derivatives of the matrix in Eq. (19.5.41), yielding a nucleonic Lagrangian:

$$\mathscr{L}_N = -\bar{\tilde{N}}\left[\not{\partial} + g\sigma + 2i\,\frac{\vec{t}\cdot(\vec{\zeta}\times\not{\partial}\vec{\zeta})}{1+\vec{\zeta}^{\,2}} + 2i\gamma_5\vec{t}\cdot\not{\partial}\right]\tilde{N} \tag{19.5.43}$$

or in terms of the canonically normalized pion field (19.5.17):

$$\mathscr{L}_N = -\bar{\tilde{N}}\left[\not{\partial} + g\sigma + \frac{2i\vec{t}\cdot(\vec{\pi}\times\not{\partial}\vec{\pi})}{F_\pi^2[1+\vec{\pi}^{\,2}/F_\pi^2]} + \frac{2i\gamma_5\vec{t}\cdot\not{\partial}\vec{\pi}}{F_\pi[1+\vec{\pi}^{\,2}/F_\pi^2]}\right]\tilde{N}\,. \tag{19.5.44}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Since σ has a non-vanishing vacuum expectation value, we see that the nucleon here has a non-zero rest mass, a result that would be prohibited if the symmetry under the transformation (19.5.40), (19.5.41) were unbroken.

By its construction, the Lagrangian (19.5.44) is chiral-invariant. This can also be seen directly, using the previously obtained chiral transformation properties of σ and $\vec{\pi}$, and now also noting that under the chiral transformation (19.5.40), (19.5.41), the new nucleon field defined by Eq. (19.5.42) transforms as

$$\delta\tilde{N} = 2i\vec{t}\cdot[\vec{\zeta}\times\vec{\epsilon}]\,\tilde{N}\,. \tag{19.5.45}$$

That is, under a chiral transformation $\tilde{N}$ simply undergoes the same isospin rotation (19.5.15) as the quantity $\vec{D}_\mu$, but of course in the $T = 1/2$ representation. The isospin rotation parameter $\vec{\zeta}\times\vec{\epsilon}$ is spacetime-dependent, so derivatives of $\tilde{N}$ do not have the same chiral transformation property, but it is straightforward to check that the combination of the first and third terms in Eq. (19.5.43) behaves as a chiral-covariant derivative: that is,

$$\delta\mathscr{D}_\mu\tilde{N} = 2i\vec{t}\cdot[\vec{\zeta}\times\vec{\epsilon}]\mathscr{D}_\mu\tilde{N}\,, \tag{19.5.46}$$

where

$$\mathscr{D}_\mu\tilde{N} \equiv \left[\partial_\mu + 2i\,\frac{\vec{t}\cdot(\vec{\zeta}\times\partial_\mu\vec{\zeta})}{1+\vec{\zeta}^{\,2}}\right]\tilde{N}\,. \tag{19.5.47}$$

Thus the Lagrangian (19.5.43) (and hence (19.5.44)) is obviously chiral-invariant, because it is isospin-invariant, and constructed solely from the ingredients $\tilde{N}$, $\mathscr{D}_\mu\tilde{N}$, σ, and $\vec{D}_\mu$, all of which transform under chiral transformations with the same isospin rotation.

Again, the specific Lagrangian with which we started here is not important. As in the case of the pure pion theory considered earlier, the important thing is the chiral invariance of the Lagrangian. For chiral invariance, the Lagrangian must conserve isospin, and be constructed only from the ingredients $\tilde{N}$, $\mathscr{D}_\mu\tilde{N}$, and $\vec{D}_\mu$ (together with higher covariant derivatives of these objects). The most general chiral-invariant Lagrangian that is bilinear in the new nucleon field and involves no more than one derivative therefore takes the form

$$\mathscr{L}_{N,0} = -\bar{\tilde{N}}\left[\not{\mathscr{D}} + m_N + 2ig_A\gamma_5\vec{t}\cdot\not{\vec{D}}\right]\tilde{N} \tag{19.5.48}$$

or in terms of the pion field (19.5.17):

$$\mathscr{L}_{N,0} = -\bar{\tilde{N}}\left[\not{\partial} + m_N + \frac{2i\vec{t}\cdot(\vec{\pi}\times\not{\partial}\vec{\pi})}{F_\pi^2[1+\vec{\pi}^{\,2}/F_\pi^2]} + \frac{2ig_A\gamma_5\vec{t}\cdot\not{\partial}\vec{\pi}}{F_\pi[1+\vec{\pi}^{\,2}/F_\pi^2]}\right]\tilde{N}\,. \tag{19.5.49}$$

Note that we have inserted an arbitrary constant g_A in the last term in Eq. (19.5.48) because this term is chiral symmetric by itself, and so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

chiral symmetry cannot dictate its coefficient. (This is in contrast with the third term in Eq. (19.5.43), whose coefficient is fixed by the condition that it and the first term together give a chiral invariant.) We can check that the constant g_A inserted here is indeed the axial-vector coupling of beta decay by constructing the extra term in the axial-vector current that arises from the nucleonic Lagrangian (19.5.48). Alternatively, integrating by parts and using the Dirac equation, we find that the pion–nucleon interaction $-2ig_A\bar{\tilde{N}}\gamma_5\vec{t}\cdot(\not{\partial}\vec{\pi})\tilde{N}/F_\pi$ is equivalent on the nucleon mass shell to an interaction $-4im_N g_A\bar{\tilde{N}}\gamma_5\vec{t}\cdot\vec{\pi}\tilde{N}/F_\pi$, corresponding to a pion–nucleon coupling constant

$$G_{\pi N} = 2m_N g_A/F_\pi \ ,$$

which is just the Goldberger–Treiman relation (19.4.33).

Incidentally, if we had used the Lagrangian (19.5.44) in this calculation, we would have obtained the Goldberger–Treiman relation with $g_A = 1$. However, this result is in fact an artifact of the particular form (19.5.39) of the interaction with which we started. We might have included a non-renormalizable derivative coupling term**

$$\mathscr{L}'_N = ig'\bar{N}\left[(\vec{t}\cdot\vec{\phi}\not{\partial}\phi_4 - \phi_4\vec{t}\cdot\not{\partial}\vec{\phi})\gamma_5 + \vec{t}\cdot(\vec{\phi}\times\not{\partial}\vec{\phi})\right]N \ . \qquad (19.5.50)$$

It is straightforward to show that in terms of the transformed fields $\sigma, \tilde{N}$, and $\vec{\zeta}$, this takes the form

$$\mathscr{L}'_N = -8i\sigma^2 g'\bar{\tilde{N}}\vec{t}\cdot\not{\partial}\vec{\zeta}\gamma_5\tilde{N} \ .$$

Since σ has a non-vanishing vacuum expectation value, this makes a contribution to the pion–nucleon coupling constant and to g_A proportional to g'. Hence the value $g_A = 1$ is *not* dictated by the broken $SU(2)\times SU(2)$ symmetry alone; by including the interaction (19.5.50) and adjusting g', we can make g_A anything we like.

Now let us consider how to use this Lagrangian to calculate amplitudes for reactions involving both pions and nucleons. We will have to give special attention to the nucleon propagators, since a nucleon can never be a 'soft' particle like a pion. A nucleon line that enters a diagram with an on-shell four-momentum p of order m_N, and then from interactions with

** The left- and right-handed parts of the nucleon doublet transform according to the representations $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ of $SU(2)\times SU(2)$, respectively. The bilinear $\bar{N}\gamma_\mu\gamma_5 N$ is thus a sum of terms quadratic in $(\frac{1}{2},0)$ or $(0,\frac{1}{2})$ terms, so it transforms like a direct sum of the (1,0), (0,0), and (0,1) representations. In Eq. (19.5.50) we have coupled the $(1,0)+(0,1)$ terms in $\bar{N}\gamma_\mu\gamma_5 N$ to the antisymmetric tensor formed from the $SU(2)\times SU(2)$ four-vectors ϕ_n and $\partial_\mu\phi_n$. Its invariance under the transformation (19.5.40), (19.5.41) can of course be checked directly.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

soft pions absorbs a net four-momentum q with components much less than m_N, will have a propagator:

$$\frac{-i(\not{p}+\not{q})+m_N}{(p+q)^2+m_N^2} \xrightarrow[q\to 0]{} \frac{-i\not{p}+m_N}{2p\cdot q}\,. \tag{19.5.51}$$

(The neglected terms may be taken into account by including higher derivative terms in the nucleon Lagrangian.) Suppose again that all external pion four-momenta have components at most of order Q and define all renormalized couplings at renormalization points of order Q, so that integrals converge in such a way that internal pion lines also have four-momenta Q. Then Eq. (19.5.51) shows that internal nucleon lines make a contribution of order $1/Q$. A general Feynman diagram for such a process will make a contribution to the invariant amplitude of order Q^ν, where now

$$\nu = \sum_i V_i(d_i + 2m_i) - 2I_\pi - I_N + 4L\,. \tag{19.5.52}$$

Here V_i is the number of vertices associated with interactions of type i, d_i is the number of derivatives in each such interaction, m_i is the number of quark mass factors in each interaction, I_π and I_N are the numbers of internal pion and nucleon lines, respectively, and L is the number of loops in the graph. We use the familiar topological relations for connected graphs:

$$L = I_\pi + I_N - \sum_i V_i + 1 \tag{19.5.53}$$

and

$$2I_N + E_N = \sum_i V_i n_i\,, \tag{19.5.54}$$

where n_i is the number of nucleon fields in interactions of type i, and E_N is the number of external nucleon lines. Eliminating the quantities I_N and I_π, this gives

$$\nu = \sum_i V_i\left(d_i + 2m_i + \frac{n_i}{2} - 2\right) + 2L - \frac{E_N}{2} + 2\,. \tag{19.5.55}$$

The important point here is that the coefficient $d_i + 2m_i + \frac{1}{2}n_i - 2$ in the first term is always positive or zero. We have already seen that $d_i + 2m_i \geq 2$ for the purely pionic interactions with $n_i = 0$, and inspection of (19.5.49) shows that $d_i \geq 1$ for the pion–nucleon interactions with $n_i = 2$ and $m_i = 0$. Any interaction with $n_i = 2$ and $m_i \geq 1$ or $n_i \geq 4$ clearly also has $d_i \geq 2$. Hence the leading terms for $Q \ll 2\pi F_\pi$ are tree graphs (that is,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

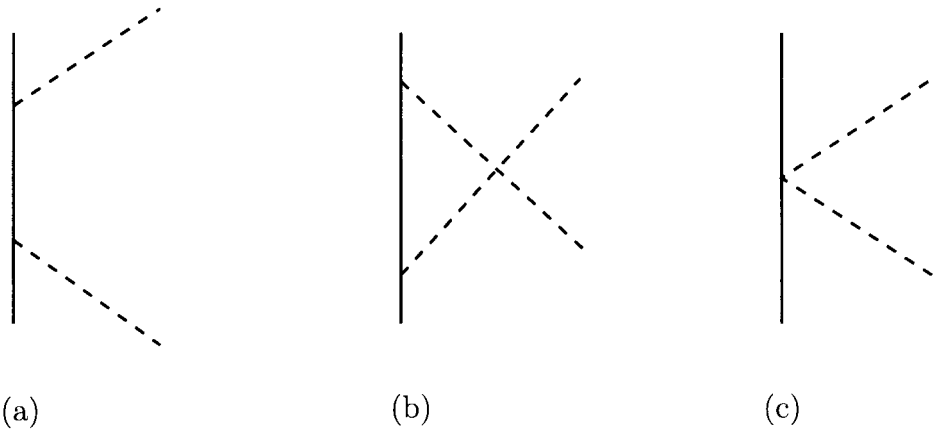

Figure 19.3. Feynman diagrams used with an effective chiral Langrangian to calculate the scattering of a soft pion from a nucleon. Dashed lines are pions; solid lines are nucleons.

$L = 0$) for which all interactions have

$$d_i + 2m_i + \frac{n_i}{2} - 2 = 0 \,.$$

The interactions that satisfy this condition are just those shown explicitly in Eqs. (19.5.23), (19.5.34), and (19.5.49), plus possible interactions with $d_i = 0$ and $n_i = 4$:

$$\left(\bar{\tilde{N}}\,\Gamma_\alpha \tilde{N}\right)\left(\bar{\tilde{N}}\,\Gamma^\alpha \tilde{N}\right), \tag{19.5.56}$$

where Γ_α and Γ^α are any matrices in spin and isospin space that yield Lorentz, space inversion, and isospin-invariant four-fermion interactions. These last interactions are important for multinucleon processes,[29] which will not be considered here.

Let us now apply this method to pion–nucleon scattering. For pion energies roughly of order m_π, the pion–nucleon scattering amplitude is given by the Feynman graphs of Figure 19.3, all of which make contributions of order m_π to the invariant amplitude. However, at threshold the leading contribution is from Figure 19.3(c), the others being suppressed by an extra factor of m_π/m_N. This is because at threshold in the rest frame the incoming and outgoing pion four-momenta q, q' and nucleon four-momenta p, p' are given by

$$q = q' = (0,0,0,m_\pi) = \left(\frac{m_\pi}{m_N}\right) p = \left(\frac{m_\pi}{m_N}\right) p' \,. \tag{19.5.57}$$

Thus the invariant amplitude at threshold from either Figure 19.3(a) or

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

19.3(b) is proportional to

$$\begin{aligned}
&\bar{u}\gamma_5 \not{q} \frac{(-i(\not{p} \pm \not{q}) + m_N)}{(p \pm q)^2 + m_N^2} \gamma_5 \not{q} u \\
&= \frac{m_\pi^2 \bar{u}\gamma_5(-i\not{p}(1 \pm m_\pi/m_N) + m_N)\gamma_5 u}{(\mp 2m_\pi m_N - m_\pi^2)} \\
&= \frac{\mp m_\pi}{2m_N \pm m_\pi} \bar{u}(i\not{p}(1 \pm m_\pi/m_N) + m_N)u = \frac{m_\pi^2}{2m_N \pm m_\pi}
\end{aligned}$$

where u is a Dirac spinor with $\bar{u}u = 1$. In contrast, the diagram of Figure 19.3(c) gives a contribution to M of order m_π. We may write this amplitude as a 2×2 matrix in the nucleon isospin indices:

$$M_{ba} = \frac{-2i}{F_\pi^2} t_c \epsilon_{abc} \bar{u}(-i\not{q} - i\not{q}')u ,$$

where q' and q are the final and initial pion four-momenta and b and a are the corresponding isovector indices. Using Eq. (19.5.57) and the momentum-space Dirac equation $(i\not{p} + m_N)u = 0$, this becomes

$$M_{ba} = \frac{-4im_\pi}{F_\pi^2} t_c \epsilon_{abc} = \frac{-4m_\pi}{F_\pi^2} \vec{t} \cdot [\vec{t}^{(\pi)}]_{ba} , \tag{19.5.58}$$

where $[t_c^{(\pi)}]_{ba} \equiv -i\epsilon_{bac}$ is the pion isovector matrix. The matrix $\vec{t} \cdot \vec{t}^{(\pi)}$ has eigenvalues in states of total isospin T equal to $\frac{1}{2}[T(T+1) - 2 - \frac{3}{4}]$, so in the two isospin states, $T = 1/2$ and $T = 3/2$, the invariant amplitudes are[30]

$$M_{1/2} = 4m_\pi/F_\pi^2 , \qquad M_{3/2} = -2m_\pi/F_\pi^2 ,$$

These results are usually expressed in terms of the scattering lengths, which are defined as the invariant amplitudes divided by $4\pi(1 + m_\pi/m_N)$:

$$a_{1/2} = \frac{m_\pi}{\pi F_\pi^2(1 + m_\pi/m_N)} = 0.15\, m_\pi^{-1} , \tag{19.5.59}$$

$$a_{3/2} = -\frac{m_\pi}{2\pi F_\pi^2(1 + m_\pi/m_N)} = -0.075\, m_\pi^{-1} . \tag{19.5.60}$$

These are in reasonable agreement with the experimental values[12] $a_{1/2} = (0.173 \pm 0.003)\, m_\pi^{-1}$ and $a_{3/2} = (-0.101 \pm 0.004)\, m_\pi^{-1}$. (The results (19.5.59) and (19.5.60) are only supposed to be valid to lowest order in m_π/m_N, but we retain the factor $1 + m_\pi/m_N$, because it arises just from the definition of the scattering lengths.)

The corrections to the scattering lengths of next order in m_π arise from several sources. There are the Born diagram graphs of Figures 19.3(a) and 19.3(b), which are nominally of leading order but as we have

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

seen at threshold are suppressed by an extra factor of m_π/m_N. There are additional tree graphs containing a vertex with two derivatives. Of special interest are tree graphs containing a vertex with no derivatives that arises from a symmetry-breaking interaction proportional to m_π^2. These interactions must have the chiral and spacetime transformation properties of the operators in Eq. (19.4.34): the fourth and third components of two different chiral four-vectors Φ_n^+ and Φ_n^-. There are two obvious candidates for such operators that are bilinear in the nucleon fields: a Φ_4^+ term,

$$\bar{N}N = \left(\frac{1-\vec{\zeta}^2}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\tilde{N} - 4i\left(\frac{\vec{\zeta}}{1+\vec{\zeta}^2}\right)\cdot\bar{\tilde{N}}\gamma_5\vec{t}\tilde{N}\,,$$

and a Φ_3^- term,

$$\bar{N}t_3N = \bar{\tilde{N}}t_3\tilde{N} - 2\left(\frac{\zeta_3}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\vec{t}\cdot\vec{\zeta}\tilde{N} - i\left(\frac{\zeta_3}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\gamma_5\tilde{N}\,.$$

Chiral symmetries act on $\tilde{N}$ like ordinary isospin rotations, so they cannot mix up scalar and pseudoscalar nucleon bilinears. Thus there are really two independent Φ_4^+ operators:

$$\left(\frac{1-\vec{\zeta}^2}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\tilde{N} \tag{19.5.61}$$

and

$$i\left(\frac{\vec{\zeta}}{1+\vec{\zeta}^2}\right)\cdot\bar{\tilde{N}}\gamma_5\vec{t}\tilde{N}\,, \tag{19.5.62}$$

and two independent Φ_3^- operators:

$$\bar{\tilde{N}}t_3\tilde{N} - 2\left(\frac{\zeta_3}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\vec{t}\cdot\vec{\zeta}\tilde{N} \tag{19.5.63}$$

and

$$i\left(\frac{\zeta_3}{1+\vec{\zeta}^2}\right)\bar{\tilde{N}}\gamma_5\tilde{N}\,. \tag{19.5.64}$$

We shall show in the next section that these are the only operators with the transformation properties of the terms in Eq. (19.4.34) that are bilinear in the nucleon fields. The operators (19.5.62) and (19.5.64) evidently provide isospin-conserving and isospin-violating corrections to the Goldberger–Treiman formula for the pion-nucleon couplings. The other two operators, (19.5.61) and (19.5.63), contribute directly to both the nucleon mass and to low energy pion–nucleon scattering. From their contribution to the nucleon mass, we see that these latter terms

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

enter into the effective Lagrangian in the form (now replacing $\vec{\zeta}$ with the conventionally normalized pion field):

$$\delta\mathscr{L}_{eff} = -\frac{\delta m_p + \delta m_n}{2}\left(\frac{1-\vec{\pi}^2/F_\pi^2}{1+\vec{\pi}^2/F_\pi^2}\right)\bar{\widetilde{N}}\widetilde{N} - (\delta m_p - \delta m_n)\left[\bar{\widetilde{N}}t_3\widetilde{N} - \frac{2}{F_\pi^2}\left(\frac{\pi_3}{1+\vec{\pi}^2/F_\pi^2}\right)\bar{\widetilde{N}}\vec{t}\cdot\vec{\pi}\widetilde{N}\right], \quad (19.5.65)$$

where δm_p and δm_n are the contributions of the quark mass terms (19.4.34) to the proton and neutron masses. This makes a contribution to the pion–nucleon scattering amplitude (again written as a matrix in the isospin space of the nucleon):

$$\delta M_{ba} = \frac{2[\delta m_p + \delta m_n]}{F_\pi^2}\delta_{ab} + \frac{2[\delta m_p - \delta m_n]}{F_\pi^2}(t_a\delta_{3b} + t_b\delta_{3a}). \quad (19.5.66)$$

The first term in Eq. (19.5.65) is often known as the 'σ-term'. The second term is an isospin violating correction to the σ-term, which shows up only in processes involving neutral pions, such as the charge-exchange processes $\pi^+ + n \to \pi^0 + p$ and $\pi^- + p \to \pi^0 + n$. Even though it is not possible to measure the nucleon mass shifts $\delta m_{p,n}$ directly, we shall see in Section 19.7 that an $SU(3)$ symmetry allows their difference to be calculated from hyperon mass differences. This yields the result $\delta m_p - \delta m_n = -2.5$ MeV. Unfortunately, we still do not have a firm theoretical estimate of the coefficient $\delta m_p + \delta m_n$ of the first term in Eq. (19.5.65).

Eq. (19.4.34) shows that in quantum chromodynamics the coefficients $\delta m_p + \delta m_n$ and $\delta m_p - \delta m_n$ in Eq. (19.5.66) are respectively proportional to $m_u + m_d$ and $m_u - m_d$. We shall see in Section 19.7 that m_u and m_d are not at all degenerate, so these coefficients are roughly of the same order of magnitude; the isospin-violating corrections to the σ-term are not much smaller than the σ-term itself. We see again that the reason that isospin conservation is such a good approximation in hadronic physics is not that the u and d quark masses are nearly equal, but just that they are small.

The use here of Lagrangians like (19.5.19) or (19.5.49) is one example of the method of *effective field theories*, already introduced in Section 12.3. Similar techniques have been employed to deal with meson and baryon interactions including strange particles (see Section 19.7), with quark and lepton interactions at energies below the scale of electroweak symmetry breaking (see Section 21.3), and even with superconductivity (see Section 21.6). In all these cases effective field theories provide the most convenient method for working out the consequences of symmetries and the general principles underlying quantum field theory.

* * *

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

The low energy theorems provided by broken symmetries when combined with dispersion relations yield useful sum rules. Let's see how this works for chiral symmetry, ignoring the small u and d quark masses and the pion mass. Consider the forward scattering of a massless pion with isovector index a and four-momentum q on a nucleon of four-momentum p, yielding a pion with isovector index b. The scattering amplitude M, defined by $S_{fi} = -2\pi i\delta^4(P_f - P_i)M_{fi}$, is given to first order in q by the Feynman diagrams of Figure 19.3 as

$$\begin{aligned}
-2\pi i M = &\frac{1}{(2\pi)^6(2q^0)} \\
&\times \bar{u}\left[\left(-i\frac{(2\pi)^4 2g_A\gamma_5 t_b \not{q}}{F_\pi}\right)\left(\frac{-i}{(2\pi)^4}\frac{-i(\not{p}+\not{q})+m_N}{(p+q)^2+M^2}\right)\left(i\frac{(2\pi)^4 2g_A\gamma_5 t_a \not{q}}{F_\pi}\right)\right. \\
&+\left(i\frac{(2\pi)^4 2g_A\gamma_5 t_a \not{q}}{F_\pi}\right)\left(\frac{-i}{(2\pi)^4}\frac{-i(\not{p}-\not{q})+m_N}{(p-q)^2+M^2}\right)\left(-i\frac{(2\pi)^4 2g_A\gamma_5 t_b \not{q}}{F_\pi}\right) \\
&\left.-4i\frac{t_c\epsilon_{abc}\not{q}}{F_\pi^2}\right] u\,. \qquad (19.5.67)
\end{aligned}$$

To this order in q, the nucleon propagator in the first two terms inside the square brackets may by approximated by

$$\frac{-i(\not{p}\pm\not{q})+m_N}{(p\pm q)^2+M^2} \approx \frac{-i\not{p}+m_N}{\pm 2p\cdot q}\,.$$

Eq. (19.5.67) can then be further simplified, using the relations $\not{q}\not{q} = q^2 = 0$, $\bar{u}\gamma^\mu u = -ip\cdot q/m_N$, and $[t_a, t_b] = i\epsilon_{abc}t_c$. Also, in the laboratory frame (which for small q is the same as the center-of-mass frame) we have $p\cdot q = -m_N\omega$, where $\omega \equiv q^0$ is the pion energy in the nucleon rest frame. Finally, the conventional forward scattering amplitude (whose absolute square is the differential cross section for forward scattering) is given by Eq. (3.6.9) as $f \equiv -4\pi^2\omega M$. Putting this all together, we find the forward scattering amplitude at pion energy ω is given for $\omega \ll m_N$ by

$$f_{ba}(\omega) \to -i\frac{\omega}{\pi F_\pi^2}(1-g_A^2)\,\epsilon_{abc}t_c\,. \qquad (19.5.68)$$

In particular, for π^+–proton scattering, we must contract Eq. (19.5.68) with $v_b^* v_a$, where $\vec{v}$ is the normalized isovector $(1,i,0)/\sqrt{2}$, and set $t_3 = +\frac{1}{2}$. The forward low energy π^+–proton scattering amplitude is then

$$f_{\pi^+ p}(\omega) \to -\frac{\omega}{2\pi F_\pi^2}(1-g_A^2)\,. \qquad (19.5.69)$$

Now, the dispersion relation for the forward scattering of a massless π^+ on a proton is given by Eq. (10.8.24) (with subtraction polynomial

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$P(E) \propto E$) as

$$f_{\pi^+p}(\omega) = R + \frac{i\omega}{4\pi}\sigma_{\pi^+p}(\omega) + \frac{\omega}{4\pi^2}\int_0^\infty \left[\frac{\sigma_{\pi^+p}(E)}{E-\omega} - \frac{\sigma_{\pi^-p}(E)}{E+\omega}\right] dE \,, \tag{19.5.70}$$

where R is a possible subtraction constant, and ω and E are pion energies in the proton rest frame. Comparison of this dispersion relation with the low energy limit (19.5.69) shows that $R = 0$, and that

$$g_A^2 = 1 + \frac{F_\pi^2}{2\pi}\int_0^\infty \left[\sigma_{\pi^+p}(E) - \sigma_{\pi^-p}(E)\right] dE/E \,. \tag{19.5.71}$$

This is the celebrated Adler–Weisberger sum rule,[19] for the case of exact chiral symmetry, with $m_\pi = 0$. A similar sum rule may be derived for the scattering of pions on any baryon or meson, including the pion itself.[30a]

19.6 Effective Field Theories: General Broken Symmetries

The techniques described in the previous section for constructing effective Lagrangians for the case of $SU(2) \times SU(2)$ broken to $SU(2)$ were soon generalized[31] to the case of a general group G broken to an arbitrary subgroup H. Consider a quantum field theory whose Lagrangian is invariant under an arbitrary compact Lie group G of ordinary linear spacetime-independent transformations g of the fields $\psi_n(x)$:

$$\psi_n(x) \to \sum_m g_{nm}\psi_m(x) \,. \tag{19.6.1}$$

We suppose that this symmetry group is spontaneously broken to some subgroup $H \subset G$ of symmetry transformations that leave all vacuum expectation values invariant: for $h \in H$,

$$\sum_m h_{nm}\langle\psi_m(x)\rangle_{\text{VAC}} = \langle\psi_n(x)\rangle_{\text{VAC}} \,. \tag{19.6.2}$$

(Not all of the ψs need be scalars, but of course only the scalars will have non-vanishing vacuum expectation values.) In the example at the beginning of the previous section, G was the group $SO(4)$, the fields ψ_n furnished a four-vector representation of this group, and their vacuum expectation values broke $SO(4)$ down to the subgroup $H = SO(3)$ of rotations that leave the four-axis invariant. In the general case the fields ψ_n may furnish a reducible rather than an irreducible representation of G, as for instance became the case when we introduced the nucleon fields in the previous section.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

We next express a general 'point' ψ_n in field space as a G transformation acting on fields $\tilde{\psi}_n$ from which the massless Goldstone mode has been eliminated:

$$\psi_n(x) = \sum_m \gamma_{nm}(x)\tilde{\psi}_m(x) \,. \tag{19.6.3}$$

For instance, in the previous section γ was the $SO(4)$ rotation we called R, and the condition satisfied by $\tilde{\psi}_n$ was that the first three components of the four-vector should vanish.

In order to formulate this condition more generally, let us suppose that we are working with a real representation of G. (There is no loss of generality here, because we are not insisting on an irreducible representation, so if we start with a set of fields forming a complex representation of G we can always take the ψ_n to be the real and imaginary parts of these complex fields.) As shown in Section 19.2, the massless eigenvectors of the mass matrix are just the independent linear combinations of the vectors $\sum_m [t^\alpha]_{nm}\langle\psi_m(0)\rangle_{\rm VAC}$, where t^α are the generators of G. (The reality of the representation means that it^α is real, and the fact that G is compact means that we can choose the representation to be unitary and hence orthogonal, so that the t^α are imaginary and antisymmetric.) The condition that $\tilde{\psi}_n$ does not contain Goldstone modes may thus be formulated as

$$\sum_{nm} \tilde{\psi}_n(x)[t^\alpha]_{nm}\langle\psi_m(0)\rangle_{\rm VAC} = 0 \,. \tag{19.6.4}$$

The number of independent conditions here is the dimensionality of the group G minus the dimensionality of the subgroup H whose generators annihilate $\langle\psi_m(0)\rangle_{\rm VAC}$. As in the previous section, the Goldstone bosons whose fields are eliminated in this way will reappear in the $\dim(G) - \dim(H)$ fields needed to parameterize the transformation $\gamma_{nm}(x)$. In general the $\tilde{\psi}_n$ include all the heavy fields of the theory, including those (like nucleons in the previous section) that have different spacetime symmetry properties from the Goldstone bosons.

It is necessary to show that we may always choose the transformation $\gamma_{nm}(x)$ so as to satisfy Eq. (19.6.4). For this purpose, consider the quantity

$$V_\psi(g) \equiv \sum_{nm} \psi_n \, g_{nm}\langle\psi_m(0)\rangle_{\rm VAC} \,, \tag{19.6.5}$$

where g runs over the whole group G in the real orthogonal representation furnished by the ψ_n. This is obviously a continuous real function of g, and since the group is compact $V_\psi(g)$ is also a bounded function of g. At each spacetime point x, $V_{\psi(x)}(g)$ therefore reaches a maximum value for some group element, which we shall call $\gamma(x)$. At $g = \gamma(x)$, $V_{\psi(x)}(g)$ must be stationary with respect to arbitrary variations in g. But any infinitesimal

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

shift in a group element g may always be written as a linear combination

$$\delta g = i \sum_\alpha \epsilon_\alpha g t^\alpha ,$$

with real infinitesimal coefficients ϵ_α that may depend on g. Hence the condition that $V_{\psi(x)}(g)$ be stationary with respect to the variation δg at $g = \gamma(x)$ reads

$$\begin{aligned} 0 = \delta V_{\psi(x)}\Big(\gamma(x)\Big) &= i \sum_\alpha \epsilon_\alpha \sum_{nml} \psi_n(x) \gamma_{nl}(x)\, t^\alpha_{lm} \langle \psi_m(0) \rangle_{\text{VAC}} \\ &= i \sum_\alpha \epsilon_\alpha \sum_{nml} [\gamma^{-1}(x)]_{ln} \psi_n(x)\, t^\alpha_{lm} \langle \psi_m(0) \rangle_{\text{VAC}} . \end{aligned}$$

This must be satisfied for all variations, and thus for all ϵ_α, so we see that Eq. (19.6.4) is satisfied for $\tilde{\psi}(x) = \gamma^{-1}(x)\psi(x)$, as was to be shown. In the purely bosonic theory with Lagrangian (19.4.1), $\tilde{\psi}_n$ was the four-vector field $(0, 0, 0, \sigma)$. This points in the direction of the vacuum expectation value $\langle \psi_n \rangle$, but this is not always the case.

Incidentally, it is not necessary in dealing with complex fields to formulate the condition on $\tilde{\psi}(x)$ explicitly in terms of its real and imaginary parts. If a set of fields $\chi(x)$ transform according to some complex representation of G with Hermitian generators T^α, then in the real representation

$$\psi(x) = \begin{pmatrix} \text{Re}\, \chi(x) \\ \text{Im}\, \chi(x) \end{pmatrix}$$

the generators are

$$i t^\alpha = \begin{pmatrix} -\text{Im}\, T^\alpha & -\text{Re}\, T^\alpha \\ \text{Re}\, T^\alpha & -\text{Im}\, T^\alpha \end{pmatrix} .$$

The condition (19.6.4) may then be expressed directly in terms of T^α and $\chi(x)$ as

$$\text{Im}\left(\tilde{\chi}(x)^\dagger\, T^\alpha\, \langle \chi(0) \rangle_{\text{VAC}}\right) = 0 .$$

Because the Lagrangian is assumed to be only invariant under spacetime-independent G transformations, it will be found after the transformation (19.6.3) to depend on $\gamma(x)$ as well as on $\tilde{\psi}(x)$, though always with at least one derivative in each γ-dependent term. As already remarked, the spacetime-dependent parameters needed to specify $\gamma(x)$ will play the role of Goldstone boson fields. We now have to consider how to parameterize $\gamma(x)$.

It must be recognized from the outset that the choice of γ in Eq. (19.6.3) is generally not unique. Because $\langle \psi_m(0) \rangle_{\text{VAC}}$ is H-invariant in the sense of Eq. (19.6.2), the quantity $V_\psi(g)$ defined by Eq. (19.6.5) is invariant under

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

right multiplication of g by any element h of the unbroken subgroup H:

$$V_\psi(g) = V_\psi(gh) \qquad \text{for} \qquad h \in H \,. \tag{19.6.6}$$

It follows that if γ maximizes $V_\psi(g)$, then so does γh, so that condition (19.6.4) is satisfied for $\tilde{\psi} = h^{-1}\gamma^{-1}\psi$ as well as $\tilde{\psi} = \gamma^{-1}\psi$. Hence γ is only defined so far up to right multiplication by an element of H. (For instance in the example of the previous section, the four-dimensional rotation R could have been multiplied on the right with any rotation acting only on the first three components of four-vectors.) We may think of two group elements γ_1 and γ_2 as equivalent if $\gamma_1 = \gamma_2 h$ with $h \in H$. This is an equivalence relation because it is reflexive (if γ_1 is equivalent to γ_2 then γ_2 is equivalent to γ_1), symmetric (γ is equivalent to itself), and transitive (if γ_1 is equivalent to γ_2 and γ_2 is equivalent to γ_3 then γ_1 is equivalent to γ_3). The elements of the group G can therefore be sorted into disjoint 'equivalence classes', each consisting of elements γ that differ only by right multiplication by an element of H. These are known as the *right cosets* of G with respect to H. What we need is a parameterization of the space (known as G/H) of right cosets.

For this purpose we need only choose one representative group element from each right coset. For the $SO(4)$ symmetry broken to $SO(3)$ of the previous section, it was convenient to choose these representative elements as the rotations R parameterized by the three-vector $\vec{\zeta}$. There is another choice that is available for any compact group G broken to any subgroup H. Let us first adapt our notation for the generators of G to the pattern of symmetry breaking. The complete set of independent generators of H will be called t_i. According to Eq. (19.6.2) these satisfy

$$\sum_m (t_i)_{nm} \langle \psi_m \rangle_{\text{VAC}} = 0 \,. \tag{19.6.7}$$

Since H is a subgroup, the t_i form a subalgebra

$$[t_i, t_j] = i \sum_k C_{ijk} t_k \,. \tag{19.6.8}$$

We will take the x_a to be the *other* independent generators of G, in any basis with totally antisymmetric structure constants. (Such a basis always exists for compact groups; see Section 15.2.) Since x_a does not appear on the right-hand side of Eq. (19.6.8), the structure constants C_{ija} all vanish, and since they are totally antisymmetric, it follows that $C_{iaj} = 0$, so

$$[t_i, x_a] = i \sum_b C_{iab} x_b \,. \tag{19.6.9}$$

However, it is *not* necessarily the case that commutators of xs with each other are linear combinations of ts; this depends on the nature of G and H, and also on how H is embedded in G. (Where C_{abc} vanishes, the coset

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

space G/H is known as a *symmetric space*). In general we may write

$$[x_a, x_b] = i \sum_i C_{abi} t_i + i \sum_c C_{abc} x_c \,. \tag{19.6.10}$$

Any set of generators with commutation relations of the form (19.6.8)–(19.6.10) is called a *Cartan decomposition* of the Lie algebra. An example is provided by the chiral $SU(2) \times SU(2)$ Lie algebra (19.4.9)–(19.4.11), for which the C_{abc} did happen to vanish.

Because t_i and x_a span the Lie algebra of G, any finite element of G may be expressed in the form

$$g = \exp\left[i \sum_a \xi_a x_a\right] \exp\left[i \sum_i \theta_i t_i\right] \,, \tag{19.6.11}$$

where ξ_a and θ_i are a set of real parameters. But the transformation $\gamma(x)$ in Eq. (19.6.3) is only defined so far up to right-multiplication with an element of H, so we may standardize our definition of γ by taking it in the form

$$\gamma(x) = \exp\left[i \sum_a \xi_a(x) x_a\right] \,. \tag{19.6.12}$$

The $\xi_a(x)$ may be identified (apart from normalization) with the Goldstone boson fields.

In what follows we will simply assume that representative elements have been chosen from each right coset, and expressed as continuous functions $\gamma(\xi)$ of some parameters ξ_a. Eq. (19.6.12) in general provides one explicit example, but we will not limit ourselves to this parameterization.

Now, suppose we use Eq. (19.6.3) to replace all fields $\psi(x)$ in the Lagrangian with $\gamma(x)\tilde{\psi}(x)$. The derivatives of the fields are given by

$$\partial_\mu \psi(x) = \gamma(x) \left[\partial_\mu \tilde{\psi}(x) + (\gamma^{-1}(x)\partial_\mu \gamma(x))\tilde{\psi}(x)\right] \,. \tag{19.6.13}$$

Therefore when we express the Lagrangian in terms of $\tilde{\psi}$ rather than ψ, the Goldstone boson fields appear through the dependence of $\gamma^{-1}(x)\partial_\mu\gamma(x)$ on $\xi_a(x)$ and its derivatives.

Any variation of a group element like $\gamma(x)$ may be written as the group element times a linear combination of the generators of the group. In our case, we may write this as

$$\gamma^{-1}(x)\partial_\mu \gamma(x) = i \sum_a x_a D_{a\mu}(x) + i \sum_i t_i E_{i\mu}(x) \,, \tag{19.6.14}$$

where $D_{a\mu}$ and $E_{i\mu}$ take the forms

$$D_{a\mu}(x) = \sum_b D_{ab}\Big(\xi(x)\Big) \partial_\mu \xi_b(x) \,, \tag{19.6.15}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$$E_{i\mu}(x) = \sum_b E_{ib}\Big(\xi(x)\Big)\partial_\mu \xi_b(x)\,. \tag{19.6.16}$$

The Goldstone boson fields will thus enter the Lagrangian through the appearance of the quantities $D_{a\mu}(x)$ and $E_{i\mu}(x)$ (and their derivatives). Note that for exact broken symmetries every interaction of the ξs must involve at least one derivative, so mass terms $m^2_{ab}\xi_a\xi_b$ cannot appear in the Lagrangian, and all interactions vanish when any Goldstone boson four-momentum vanishes.

Even where we do not know the details of the underlying Lagrangian, we can learn a great deal about the way that the quantities $D_{a\mu}$ and $E_{i\mu}$ appear in the transformed Lagrangian from their transformation properties. Under an arbitrary element g of the group G, the original field ψ transforms according to Eq. (19.6.1):

$$\psi(x) \to \psi'(x) = g\psi(x) = g\gamma\Big(\xi(x)\Big)\tilde{\psi}(x)\,. \tag{19.6.17}$$

Now, $g\gamma(\xi)$ is an element of G, so it must be in the same right coset as some $\gamma(\xi')$, and may therefore be written in the form

$$g\gamma\Big(\xi(x)\Big) = \gamma\Big(\xi'(x)\Big)h\Big(\xi(x), g\Big)\,, \tag{19.6.18}$$

where h is some element of the unbroken subgroup H. Using this in Eq. (19.6.17), we find that $\psi'(x)$ is of the form (19.6.3)

$$\psi'(x) = \gamma\Big(\xi'(x)\Big)\tilde{\psi}'(x)\,, \tag{19.6.19}$$

with $\xi'(x)$ defined by Eq. (19.6.18), and

$$\tilde{\psi}'(x) \equiv h\Big(\xi(x), g\Big)\tilde{\psi}(x)\,. \tag{19.6.20}$$

In the example at the beginning of the previous section, $\tilde{\psi}$ consisted of a single isoscalar σ, and since this was invariant under the unbroken isospin subgroup, it was also chiral invariant. More generally, we find here that $\tilde{\psi}(x)$ is not necessarily invariant under general G transformations, but that its G transformation depends only on its transformation under the unbroken subgroup H. We saw this in the previous section when we introduced the nucleon fields; under a general infinitesimal broken chiral symmetry, the fields $\tilde{N}$ transformed according to the field-dependent isospin rotation (19.5.45).

The transformation rules $\xi \to \xi'$ and $\tilde{\psi} \to \tilde{\psi}'$ specified by Eqs. (19.6.18) and (19.6.20) make no reference to the particular linear G transformation properties of the original fields ψ_n. Indeed, we did not need to start with a Lagrangian that was invariant under linear G transformations in order to deduce these transformation rules. Given any set of fields on which a group G acts linearly or non-linearly, with a subgroup H that leaves one special set of field values (their vacuum expectation values) invariant,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

we can always express these fields (in at least a finite neighborhood of the special field values) in terms of a set ξ_a and $\tilde{\psi}_n$, which transform under G according to the standard realization $\xi_a \to \xi'_a$, $\tilde{\psi} \to \tilde{\psi}'$ defined by Eqs. (19.6.18) and (19.6.20). The essential uniqueness of this realization was first proved[25] for $SU(2) \times SU(2)$ broken to $SU(2)$, and then for general Lie groups broken to any subgroup.[31] For an easier argument, we may note that for any set of fields that transforms according to a non-linear realization of an arbitrary compact Lie group G, it is always possible to define functions of these fields (and perhaps additional fields*) that transform linearly under G.[32] Starting from this linear realization, the arguments of this section allow us to construct fields ξ and $\tilde{\psi}$ with the transformation rules (19.6.18) and (19.6.20).

Although the transformation of $\xi(x)$ and $\tilde{\psi}(x)$ is complicated for general $g \in G$, it is much simpler when g is itself a member h of the unbroken subgroup H. It is usual and always possible to choose the Goldstone boson fields $\xi_a(x)$ to transform according to some *linear* representation $\mathscr{D}_{ab}(h)$ of the unbroken subgroup H:

$$h\gamma(\xi)h^{-1} = \gamma\Big(\mathscr{D}(h)\xi\Big)\,. \tag{19.6.21}$$

For instance, in the parameterization of the coset space $SO(4)/SO(3)$ used in the previous section, the Goldstone boson fields ζ_a formed an isospin three-vector. For the more generally applicable parameterization based on Eq. (19.6.12), the commutation relations (19.6.9) show that the x_a transform linearly under H:

$$hx_b h^{-1} = \sum_a \mathscr{D}_{ba}(h)x_a\,, \tag{19.6.22}$$

from which Eq. (19.6.21) follows immediately. Comparing Eq. (19.6.21) with Eq. (19.6.18), we see that for $g = h \in H$,

$$\xi'_a(x) = \sum_b \mathscr{D}_{ab}(h)\xi_b(x)\,, \tag{19.6.23}$$

$$\tilde{\psi}'_n(x) = \sum_m h_{nm}\tilde{\psi}_m(x)\,. \tag{19.6.24}$$

In other words, ξ_a and $\tilde{\psi}_n$ transform under H according to the representations $\mathscr{D}_{ab}(h)$ and h_{nm} itself, respectively.

We must now consider how to construct the most general Lagrangian that is invariant under the transformations (19.6.18) and (19.6.20). The transformation $\xi \to \xi'$ for general $g \in G$ is non-linear and rather complicated, and rules out the appearance of $\xi_a(x)$ in the Lagrangian except in

* For instance, the polar and azimuthal angles θ and φ furnish a non-linear realization of $SO(3)$; by adding an additional variable r we can construct quantities x_1, x_2, and x_3 that transform linearly under $SO(3)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

quantities like $D_{a\mu}(x)$ and $E_{i\mu}(x)$. Fortunately, these quantities obey very simple transformation rules. Differentiate Eq. (19.6.18) with respect to x^μ and multiply on the left with its inverse. This gives

$$\begin{aligned}\gamma^{-1}\Big(\xi(x)\Big)\partial_\mu\gamma\Big(\xi(x)\Big) &= \Big[g\gamma\Big(\xi(x)\Big)\Big]^{-1}\partial_\mu\Big[g\gamma\Big(\xi(x)\Big)\Big] \\ &= h^{-1}\Big(\xi(x),g\Big)\gamma^{-1}\Big(\xi'(x)\Big)\partial_\mu\Big[\gamma\Big(\xi'(x)\Big)h\Big(\xi(x),g\Big)\Big] \\ &= h^{-1}\Big(\xi(x),g\Big)\Big[\gamma^{-1}\Big(\xi'(x)\Big)\partial_\mu\gamma\Big(\xi'(x)\Big)\Big]\ h\Big(\xi(x),g\Big) \\ &\quad + h^{-1}\Big(\xi(x),g\Big)\partial_\mu h\Big(\xi(x),g\Big)\end{aligned}$$

and so

$$\begin{aligned}\gamma^{-1}\Big(\xi'(x)\Big)\partial_\mu\gamma\Big(\xi'(x)\Big) &= h\Big(\xi(x),g\Big)\Big[\gamma^{-1}\Big(\xi(x)\Big)\partial_\mu\gamma\Big(\xi(x)\Big)\Big]\,h^{-1}\Big(\xi(x),g\Big) \\ &\quad - \Big[\partial_\mu h\Big(\xi(x),g\Big)\Big]\ h^{-1}\Big(\xi(x),g\Big)\,. \qquad (19.6.25)\end{aligned}$$

Eq. (19.6.14) gives the quantity $\gamma^{-1}\partial_\mu\gamma$ as a linear combinations of xs and ts. Also, the second term on the right-hand side of Eq. (19.6.25) is a linear combination of ts alone. Since the x_a and t_i are all independent, the coefficients of each generator on both sides of Eq. (19.6.25) must all be equal:

$$\sum_a x_a D'_{a\mu} = h(\xi,g)\left(\sum_a x_a D_{a\mu}\right)\ h^{-1}(\xi,g)\,, \qquad (19.6.26)$$

$$\begin{aligned}\sum_i t_i E'_{i\mu} &= h(\xi,g)\left(\sum_i t_i E_{i\mu}\right)\ h^{-1}(\xi,g) \\ &\quad + i\left[\partial_\mu h(\xi,g)\right]\ h^{-1}(\xi,g)\,, \qquad (19.6.27)\end{aligned}$$

or in more detail

$$\begin{aligned}D'_{a\mu}(x) &= \sum_b \mathscr{D}_{ab}\left(h\Big(\xi(x),g\Big)\right)\ D_{b\mu}(x)\,, \\ E'_{i\mu}(x) &= \sum_j \mathscr{E}_{ij}\left(h\Big(\xi(x),g\Big)\right)E_{j\mu}(x) \\ &\quad - \sum_b \mathscr{H}_{ib}\left(h\Big(\xi(x),g\Big)\right)\partial_\mu\xi_b(x)\,, \qquad (19.6.28)\end{aligned}$$

where $D_{a\mu}$ and $E_{i\mu}$ are defined by Eqs. (19.6.15) and (19.6.16), and

$$\Big[\partial_\mu h\Big(\xi(x),g\Big)\Big]\,h^{-1}\Big(\xi(x),g\Big) = i\sum_{ib}\mathscr{H}_{ib}\Big(\xi(x)\Big)t_i\partial_\mu\xi_b(x)\,,$$

$$h\,t_j\,h^{-1} = \sum_i \mathscr{E}_{ij}(h)t_i\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

while $\mathscr{D}_{ab}$ is defined by Eq. (19.6.22). We see that the quantities $D_{a\mu}(x)$ are 'covariant derivatives' of the Goldstone boson fields, transforming under a general element $g \in G$ much like the fields $\tilde{\psi}$: both are subjected to the H transformation $h(\xi(x), g)$, though in different representations. For instance, in the $SO(4)$ theory of the previous section, the covariant derivative of the pion field $\vec{\zeta}$ was the quantity $\partial_\mu \vec{\zeta}/(1+\vec{\zeta}^2)$, which is transformed under an infinitesimal chiral transformation by an isospin rotation (19.5.15).

On the other hand, $E_{j\mu}(x)$ transforms inhomogeneously, much like a gauge field. The extra term in Eq. (19.6.28) allows a cancellation of the inhomogeneous term in derivatives of $\tilde{\psi}$. Differentiating Eq. (19.6.20), we have

$$\partial_\mu \tilde{\psi}'(x) = h\Big(\xi(x), g\Big)\Big[\partial_\mu \tilde{\psi}(x) + h^{-1}\Big(\xi(x), g\Big)\partial_\mu h\Big(\xi(x), g\Big)\tilde{\psi}(x)\Big] \ .$$

Combining this with Eq. (19.6.27) gives

$$\Big(\mathscr{D}_\mu \tilde{\psi}(x)\Big)' = h\Big(\xi(x), g\Big)\mathscr{D}_\mu \tilde{\psi}(x) \ , \tag{19.6.29}$$

where $\mathscr{D}_\mu \tilde{\psi}$ is a covariant derivative of the heavy particle fields:

$$\mathscr{D}_\mu \tilde{\psi}(x) \equiv \partial_\mu \tilde{\psi}(x) + i\sum_i t_i E_{i\mu}(x)\tilde{\psi}(x) \ . \tag{19.6.30}$$

(Eq. (19.5.47) provides an example of such a covariant derivative in the case of $SO(4)$ broken to $SO(3)$.) *Any Lagrangian that is invariant under the unbroken subgroup H and is constructed from $\tilde{\psi}$, $\mathscr{D}_\mu \tilde{\psi}$, and $D_{a\mu}$ will thus also be invariant under the full group G.*

Renormalizability is not an issue here, so we can also consider quantities involving more than one spacetime derivative. One such is $\mathscr{D}_\nu \mathscr{D}_\mu \tilde{\psi}$, obtained by just repeating the operation (19.6.30):

$$\mathscr{D}_\nu \mathscr{D}_\mu \tilde{\psi} = \left[\partial_\nu + i\sum_j t_j E_{j\nu}\right]\left[\partial_\mu + i\sum_i t_i E_{i\mu}\right]\tilde{\psi} \ . \tag{19.6.31}$$

This transforms just like $\mathscr{D}_\mu \tilde{\psi}$

$$(\mathscr{D}_\nu \mathscr{D}_\mu \tilde{\psi})' = h(\xi, g)\mathscr{D}_\nu \mathscr{D}_\mu \tilde{\psi} \ . \tag{19.6.32}$$

Another covariant quantity can be constructed from the covariant derivative of the covariant Goldstone boson derivative (19.6.15), which is defined just like (19.6.30), but with t_i replaced with the corresponding matrix in the representation of H furnished by the x_a:

$$\mathscr{D}_\nu D_{a\mu} = \partial_\nu D_{a\mu} + \sum_{ib} C_{iab}\, E_{i\nu} D_{b\mu} \ , \tag{19.6.33}$$

which transforms just like $D_{a\mu}$:

$$(\mathscr{D}_\nu D_{a\mu})' = \sum_b \mathscr{D}_{ab}(h, g)\mathscr{D}_\nu D_{b\mu} \ . \tag{19.6.34}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Also, by antisymmetrizing the derivative of $E_{i\mu}$ we can eliminate the inhomogeneous term in (19.6.27). This yields a 'curvature'

$$R_{i\mu\nu} \equiv \partial_\nu E_{i\mu} - \partial_\mu E_{i\nu} - i \sum_{j,k} C_{ijk} E_{j\mu} E_{k\nu} \,. \tag{19.6.35}$$

This is not really new; it is easy to see that the antisymmetric covariant derivative of the 'matter' field $\tilde{\psi}$ is proportional to this curvature

$$[\mathscr{D}_\nu, \mathscr{D}_\mu]\tilde{\psi} = \sum_i t_i R_{i\mu\nu} \tilde{\psi} \,. \tag{19.6.36}$$

Of course, we can continue this process and construct yet more covariants by taking higher and higher covariant derivatives of $\mathscr{D}_\mu \tilde{\psi}$ and $D_{a\mu}$.

It is useful to note that the quantities $D_{a\mu}$ and $E_{i\mu}$ may be calculated once and for all for given groups G and H and a given parameterization of the coset space G/H, and do not depend on the assumed transformation properties of the fields ψ_n or the specific matrices x_a, t_i used to represent the Lie algebra of the broken symmetry group. For instance, in the exponential parameterization (19.6.12), the first few terms in the power series for these covariants are easily calculated from Eqs. (19.6.12) and (19.6.14) as

$$\begin{aligned} D_{a\mu} = {} & \partial_\mu \xi_a + \tfrac{1}{2} \sum_{bc} C_{abc} \xi_b \partial_\mu \xi_c \,, \\ & + \tfrac{1}{6} \sum_{bcd} \left[\sum_e C_{cde} C_{bea} + \sum_i C_{cdi} C_{bia} \right] \xi_b \xi_c \partial_\mu \xi_d \\ & + O(\xi^3 \partial_\mu \xi) \,, \end{aligned} \tag{19.6.37}$$

$$\begin{aligned} E_{i\mu} = {} & \tfrac{1}{2} \sum_{ab} C_{abi} \xi_a \partial_\mu \xi_b + \tfrac{1}{6} \sum_{abcd} C_{acd} C_{bdi} \xi_a \xi_b \partial_\mu \xi_c \\ & + O(\xi^3 \partial_\mu \xi) \,. \end{aligned} \tag{19.6.38}$$

All we need to know in order to construct the most general Lagrangian involving the Goldstone boson fields ξ_a and a set of heavy particle fields is the pattern of spontaneous symmetry breaking, of G to H, and the transformation of the heavy particle fields under H, not G: the Lagrangian is taken to be the most general H-invariant function of $\tilde{\psi}$, $\mathscr{D}_\mu \tilde{\psi}$, $D_{a\mu}$, and higher covariant derivatives.

Now consider the case where the symmetry under G is not only spontaneously broken, but not even exact to begin with. Suppose that there is a term $\Delta\mathscr{L}$ in the initial Lagrangian that is *not* invariant under the group G of linear transformations $\psi \to g\psi$, but transforms as a linear combination of the components of some representation (reducible or irreducible) of G.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

That is, we take

$$\Delta\mathscr{L} = \sum_A c_A \mathcal{O}_A \,, \tag{19.6.39}$$

where $\mathcal{O}_A$ transforms under G according to some representation $D[g]_{AB}$:

$$\mathcal{O}_A \to \sum_B D[g]_{AB}\mathcal{O}_B \,. \tag{19.6.40}$$

If we now replace the fields ψ with the set $\xi_a, \tilde{\psi}$, this term in the Lagrangian will still take the form (19.6.39), but Eq. (19.6.40) now reads

$$\mathcal{O}_A\Big[f_a(\xi,g),\ h(\xi,g)\tilde{\psi}\Big] = \sum_B D[g]_{AB}\mathcal{O}_B\Big[\xi,\tilde{\psi}\Big] \,, \tag{19.6.41}$$

where $f_a(\xi,g)$ is the result of applying the transformation g to ξ_a, defined by Eq. (19.6.18):

$$g\gamma\Big(\xi\Big) = \gamma\Big(f(\xi,g)\Big)h\Big(\xi,g\Big) \,. \tag{19.6.42}$$

As we shall see, it is easy to find sets of operators satisfying (19.6.41), and the solution is unique up to a specification of certain H-invariant functions of the $\tilde{\psi}$.

First, consider the case $\xi_a = 0$, with $g = \gamma(\xi')$. In this case $g\gamma(\xi) = \gamma(\xi')$, so

$$f_a\Big(0,\gamma(\xi')\Big) = \xi'_a \,, \qquad h\Big(0,\gamma(\xi')\Big) = 1 \,.$$

Applying this in Eq. (19.6.41), we find (now dropping the prime):

$$\mathcal{O}_A\Big[\xi_a,\tilde{\psi}\Big] = \sum_B D[\gamma(\xi)]_{AB}\,\mathcal{O}_B\Big[0,\tilde{\psi}\Big] \,. \tag{19.6.43}$$

This gives the operators $\mathcal{O}_A[\xi_a,\tilde{\psi}]$ for all ξ_a if we know them for $\xi_a = 0$.

Now take $\xi_a = 0$, with g an element h of the unbroken subgroup H. Here we have $g\gamma(\xi) = h$, so

$$f_a(0,h) = 0 \,, \qquad h(0,h) = h \,.$$

Applying this in Eq. (19.6.41), we find here

$$\mathcal{O}_A\Big[0,h\tilde{\psi}\Big] = \sum_B D[h]_{AB}\,\mathcal{O}_B\Big[0,\tilde{\psi}\Big] \,. \tag{19.6.44}$$

In other words, the $\mathcal{O}_A$ for $\xi = 0$ must have the same transformation under linear H transformations as are found in the representation (19.6.40) of G. (However, there is nothing in this that fixes the normalization of the different irreducible H representations found in the G representation (19.6.40); some of them may even be absent altogether.)

Finally, we note that any set of $\mathcal{O}$ operators satisfying (19.6.44) allows the construction using Eq. (19.6.43) of operators satisfying the general G-transformation rules (19.6.40). Using Eqs. (19.6.43) and (19.6.44), the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

left-hand side of Eq. (19.6.41) becomes

$$\begin{aligned}\mathcal{O}_A\left[f_a(\xi,g),\ h(\xi,g)\tilde{\psi}\right] &= \sum_B D[\gamma\left(f(\xi,g)\right)]_{AB}\mathcal{O}_B\left[0,h(\xi,g)\tilde{\psi}\right]\\ &= \sum_{BC} D[\gamma\left(f(\xi,g)\right)]_{AB}D[h(\xi,g)]_{BC}\ \mathcal{O}_C\left[0,\tilde{\psi}\right]\\ &= \sum_B D[\gamma\left(f(\xi,g)\right)h(\xi,g)]_{AB}\ \mathcal{O}_B\left[0,\tilde{\psi}\right]\\ &= \sum_B D[g\gamma(\xi)]_{AB}\ \mathcal{O}_B\left[0,\tilde{\psi}\right]\\ &= \sum_B D[g]_{AB}\ \mathcal{O}_B\left[\xi,\tilde{\psi}\right],\end{aligned}$$

as was to be shown.

As an example, consider the group $SO(4)$ spontaneously broken to $SO(3)$ as in the previous section, and suppose we want to construct operators that transform according to the four-vector representation of $SO(4)$ out of the Goldstone boson field $\vec{\zeta}(x)$ and the other fields $\tilde{\psi}(x)$. According to (19.6.43), these must take the form

$$\mathcal{O}_n\left[\vec{\zeta},\tilde{\psi}\right] = R_{nm}(\vec{\zeta})\,\mathcal{O}_m\left[0,\tilde{\psi}\right],$$

where R is the $SO(4)$ rotation defined by Eqs. (19.5.8) and (19.5.9):

$$R_{a4} = \frac{2\zeta_a}{1+\vec{\zeta}^{\,2}}, \qquad R_{44} = \frac{1-\vec{\zeta}^{\,2}}{1+\vec{\zeta}^{\,2}}.$$

The condition that R is an orthogonal matrix is satisfied if we choose its other components as

$$R_{ab} = \delta_{ab} - \frac{2\zeta_a\zeta_b}{1+\vec{\zeta}^{\,2}}, \qquad R_{4a} = -\frac{2\zeta_a}{1+\vec{\zeta}^{\,2}}.$$

Thus any scalar (as opposed to pseudoscalar) operators that transform as the fourth and third components of chiral four-vectors must appear in the effective Lagrangian in the forms

$$\Phi_4^+\left[\vec{\zeta},\tilde{\psi}\right] = -\left(\frac{2\vec{\zeta}}{1+\vec{\zeta}^{\,2}}\right)\cdot\vec{\Phi}^+\left[0,\tilde{\psi}\right] + \left(\frac{1-\vec{\zeta}^{\,2}}{1+\vec{\zeta}^{\,2}}\right)\Phi_4^+\left[0,\tilde{\psi}\right]$$

and

$$\Phi_3^-\left[\vec{\zeta},\tilde{\psi}\right] = \left[\Phi_3^-\left[0,\tilde{\psi}\right] - \left(\frac{2\zeta_3}{1+\vec{\zeta}^{\,2}}\right)\vec{\zeta}\cdot\overrightarrow{\Phi^-}\left[0,\tilde{\psi}\right]\right] + \left(\frac{2\zeta_3}{1+\vec{\zeta}^{\,2}}\right)\Phi_4^-\left[0,\tilde{\psi}\right].$$

The only condition on the operators $\vec{\Phi}^+[0,\tilde{\psi}]$, $\Phi_4^+[0,\tilde{\psi}]$, $\vec{\Phi}^-[0,\tilde{\psi}]$, and $\Phi_4^-[0,\tilde{\psi}]$ is that they should transform under Lorentz and isospin transformations respectively as a pseudoscalar isovector, a scalar isoscalar, a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

scalar isovector, and a pseudoscalar isoscalar, but they do not have to be related in any way. For processes involving only pions we use an effective Lagrangian that involves no fields except the pion field $\vec{\zeta}$, so the operators $\vec{\Phi}^+[0,\tilde{\psi}]$, $\vec{\Phi}^-[0,\tilde{\psi}]$, and $\Phi_4^-[0,\tilde{\psi}]$ must all vanish, while $\Phi_4^+[0,\tilde{\psi}]$ is just a numerical constant. We see as promised in the previous section that the only non-derivative chiral symmetry breaking operator in the effective Lagrangian is proportional to $(1-\vec{\zeta}^2)/(1+\vec{\zeta}^2)$, and hence to the operator (19.5.34). In order to include symmetry-breaking operators bilinear in the nucleon field, parity and isospin conservation require that we take

$$\vec{\Phi}^+\left[0,\tilde{N}\right] \propto \bar{\tilde{N}}\gamma_5\vec{t}\tilde{N}\,, \qquad \Phi_4^+\left[0,\tilde{N}\right] \propto \bar{\tilde{N}}\tilde{N}\,,$$
$$\vec{\Phi}^-\left[0,\tilde{N}\right] \propto \bar{\tilde{N}}\vec{t}\tilde{N}\,, \qquad \Phi_4^-\left[0,\tilde{N}\right] \propto \bar{\tilde{N}}\gamma_5\tilde{N}\,.$$

Thus as claimed in the previous section the only non-derivative symmetry-breaking terms involving nucleon bilinears are of the form (19.5.61)–(19.5.64).

The phenomenological Lagrangian may be used to calculate the amplitudes for processes involving particles of small three-momentum in much the same way as was done for the chiral theory of pions and nucleons in the previous section. In counting powers of a characteristic small momentum Q, Goldstone boson internal lines always contribute factors of order Q^{-2}. Also, an internal line for one of the heavy particles (of any spin) described by the fields $\tilde{\psi}$ that has absorbed a net four-momentum q from the Goldstone boson field will have a propagator

$$\frac{N(p+q)}{(p+q)^2+M^2} \to \frac{N(p)}{2p\cdot q}$$

(where N is some polynomial, depending on the particle spin and field type, and M is the particle mass) so it will contribute a factor of order Q^{-1}. The same argument that led to (19.5.55) applies here, and gives the number of powers of Q in a given Feynman diagram as

$$\nu = \sum_i V_i\,(d_i + h_i/2 - 2) + 2L - E_h/2 + 2\,, \tag{19.6.45}$$

where V_i is the number of vertices of type i, d_i is the number of derivatives (or for approximate broken symmetries, derivatives or Goldstone boson masses) at a vertex of type i, h_i is the number of heavy particle lines at a vertex of type i, L is the number of loops, and E_h is the number of external lines of heavy particles. In general the coefficient $d_i + h_i/2 - 2$ is non-negative for all interactions allowed by chiral symmetries. The interactions among Goldstone bosons alone always have $d_i \geq 2$, because they must either be at least bilinear in the covariant derivative (19.6.15), or proportional to symmetry-breaking parameters (like quark masses in the previous section) which are of order of the squares of the Goldstone boson

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

masses. Interactions between Goldstone bosons and heavy particles must involve the covariant derivatives (19.6.15) or (19.6.30), and so must have $d_i \geq 1$ as well as $h_i \geq 2$. The only interactions for which $d_i + h_i/2 < 2$ would be trilinear interactions among the heavy particles, which we exclude because we are only considering graphs in which all heavy hadrons remain non-relativistic. With $d_i + h_i/2 \geq 2$ for all interactions, the leading graphs are those constructed solely from interactions with $d_i + h_i/2 = 2$ and no loops. Corrections are again provided by interactions with more derivatives and/or more heavy-particle fields, and/or one or more loops.

In the absence of intrinsic symmetry breaking, the part of the Lagrangian that involves only the Goldstone boson fields has a unique $d_i = 2$ term

$$\mathscr{L}_{GB} = -\frac{1}{2}\sum_{ab} F_{ab}^2 D_{a\mu} D_b^\mu \,, \tag{19.6.46}$$

with F_{ab}^2 some real positive-definite matrix. For (19.6.12) and a large class of other parameterizations of the cosets, $\gamma(\xi)$ approaches $1 + i\xi_a x_a$ for $\xi_a \to 0$, so Eq. (19.6.14) shows that the linear term in $D_{a\mu}$ is just $\partial_\mu \xi_a$. The canonically normalized Goldstone boson fields π_a with a kinematic Lagrangian $-\frac{1}{2}\sum_a \partial_\mu \pi_a \partial^\mu \pi_a$ are then

$$\pi_a = \sum_b F_{ab} \xi_b \,. \tag{19.6.47}$$

In the $SO(4)$ theory of the previous section, the generators x_a and the fields ξ_a transform according to an irreducible representation of the unbroken subgroup $H = SO(3)$, so in this case $F_{ab}^2 = F_\pi^2 \delta_{ab}$, with F_π a single constant that characterizes the energy scale of the spontaneous symmetry breakdown. In the general case we can choose the generators (without changing the structure constants) so that F_{ab}^2 is diagonal, with diagonal elements all equal within each irreducible representation of H, but otherwise independent.

* * *

We have seen how to construct theories of Goldstone boson fields by starting with fields belonging to linear representations of the broken symmetry group, such as the $SU(2) \times SU(2)$ four-vector $\phi_n(x)$ with which we started in Section 19.5. At very low energy the only significant degrees of freedom are the Goldstone bosons, so we generally throw away the non-Goldstone parts of the field, such as the field $\sigma(x)$ given by Eq. (19.5.4). But there are circumstances in which it is necessary to return to full linear representations of the broken symmetry group, in which the Goldstone boson fields are just one part.

This happens when, by varying the temperature or turning on an

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

external field of some sort, a system is brought close to a second-order transition, in which it smoothly goes from broken to unbroken symmetry. On one side of this transition the symmetry is broken, so we have massless Goldstone bosons, and various other massive excitations, not generally forming complete multiplets that would furnish linear representations of the broken symmetry group. On the other side of the transition the symmetry is unbroken, so here we have complete linear multiplets, not generally massless. If this transition is continuous, then very near the phase transition the Goldstone boson must form part of a complete linear multiplet of nearly massless excitations. Barring accidents or other symmetries, this multiplet will form an *irreducible* representation of the broken symmetry group. This irreducible multiplet of fields, which become massless only just at the phase transition, is known as the *order parameter*.

This is a more precise definition of order parameters than usual. Often one speaks of an order parameter as any set of fields whose expectation values break the symmetry, but this is too vague — there is no one set of fields whose expectation values can be blamed for a broken symmetry. For instance, the $SU(2) \times SU(2)$ symmetry of quantum chromodynamics with massless u and d quarks is broken by the expectation value of $\bar{u}u+\bar{d}d$, which is the fourth component of a chiral four-vector, but quartic and higher powers of quark fields also have non-vanishing vacuum expectation values, and these belong to other representations of $SU(2) \times SU(2)$. In contrast, at a smooth phase transition at which $SU(2) \times SU(2)$ becomes unbroken the Goldstone boson becomes a member of just one massless representation of $SU(2) \times SU(2)$, and this provides an unambiguous definition of the order parameter.

It is important to identify the correct order parameters, both in order to calculate the critical exponents discussed in Section 18.5, and to deal with configurations like vortex lines or magnetic monopoles, where, as we shall see in Sections 21.6 and 23.3, there is a singularity at which the broken symmetry becomes unbroken. It is often assumed that the order parameter for the chiral $SU(2) \times SU(2)$ symmetry of quantum chromodynamics is a chiral four-vector, but this is not known to be the case.[32a]

19.7 Effective Field Theories: $SU(3) \times SU(3)$

The approximate $SU(2)$ isospin symmetry of nuclear physics was extended in the early 1960s by Gell-Mann[33] and Ne'eman[34] to an even less exact $SU(3)$ symmetry, which grouped the known baryons and mesons in various irreducible representations: an octet of $1/2^+$ baryons p, n, Λ^0, $\Sigma^{\pm,0}$, $\Xi^{-,0}$, an octet of 0^- mesons $K^{+,0}$, $\pi^{\pm,0}$, η^0, $\overline{K}^{-,0}$, an octet of 1^- mesons $K^{*+,0}$, $\rho^{\pm,0}$, ω, $\overline{K^*}^{-,0}$, and a decuplet of $3/2^+$ baryons $\Delta^{++,+,0,-}$, $\Sigma^{*+,0,-}$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$\Xi^{*0,-}$, Ω^-. (The η and Ω were not discovered until later.) After the successes of the chiral $SU(2)\times SU(2)$ symmetry in the mid-1960s, it became natural to suppose that the strong interactions also respect an approximate $SU(3) \times SU(3)$ symmetry, which like $SU(2) \times SU(2)$ is spontaneously broken to its diagonal subgroup, the $SU(3)$ of Gell-Mann and Ne'eman. Then after the advent of quantum chromodynamics it became clear that this symmetry arises because there are not two but three fairly light quarks; the u and d, and a third quark s which like d has charge $-1/3$. In this case the $SU(3) \times SU(3)$ symmetry consists of independent $SU(3)$ transformations (analogous to Eq. (19.4.2)) on the left- and right-handed parts of the u, d, s quark fields:

$$\begin{pmatrix} u \\ d \\ s \end{pmatrix} \to \exp\left[i \sum_a (\theta_a^V \lambda_a + \theta_a^A \lambda_a \gamma_5) \right] \begin{pmatrix} u \\ d \\ s \end{pmatrix} , \tag{19.7.1}$$

where λ_a are a complete set of traceless Hermitian matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} , \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} ,$$
$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} , \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} , \tag{19.7.2}$$

normalized so that $\mathrm{Tr}(\lambda_a \lambda_b) = 2\delta_{ab}$. The generators of $SU(3) \times SU(3)$ are thus represented on the quark fields by the generators $t_a = \lambda_a$ of the unbroken $SU(3)$ symmetry and the broken symmetry generators $x_a = \lambda_a \gamma_5$.

To define the Goldstone boson fields and work out their transformation properties, we note that in the representation furnished by the quark fields, any $SU(3) \times SU(3)$ transformation may be written as $\exp(-i\gamma_5 \sum_a \xi_a \lambda_a)$ times a transformation $\exp(i \sum_a \theta_a \lambda_a)$ belonging to the unbroken $SU(3)$ subgroup of $SU(3) \times SU(3)$. (The minus sign accompanying ξ is inserted for convenience in later comparing components of this field with the pion fields introduced in Section 19.5.) Hence in this representation, each right coset of $SU(3) \times SU(3)/SU(3)$ is represented by the matrix of the form $\gamma(\xi) = \exp(-i\gamma_5 \sum_a \xi_a \lambda_a)$ that it contains, and aside from normalization, the parameters ξ_a of these right cosets may be taken as our Goldstone boson fields. According to Eq. (19.6.18), the transformation rule of these

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

fields is dictated by

$$\exp\left[i\sum_a(\theta_a^V\lambda_a+\theta_a^A\lambda_a\gamma_5)\right]\exp\left(-i\gamma_5\sum_a\xi_a(x)\lambda_a\right)$$
$$=\exp\left(-i\gamma_5\sum_a\xi'_a(x)\lambda_a\right)\exp\left(i\sum_a\theta_a(x)\lambda_a\right) \quad (19.7.3)$$

with $\theta_a(x)$ some function of θ^V, θ^A, and $\xi(x)$. Also, according to Eq. (19.6.3), the Goldstone-free quark fields $\tilde{q}(x)$ are here defined by

$$q(x)\equiv\begin{pmatrix}u(x)\\d(x)\\s(x)\end{pmatrix}=\exp\left(-i\gamma_5\sum_a\xi_a(x)\lambda_a\right)\tilde{q}(x) \quad (19.7.4)$$

and have the transformation rule given by Eq. (19.6.24):

$$\tilde{q}'(x)=\exp\left(i\sum_a\theta_a(x)\lambda_a\right)\tilde{q}(x)\,. \quad (19.7.5)$$

We could proceed to introduce covariant derivatives (19.6.15) and (19.6.30), and use them to construct chiral-invariant Lagrangian densities, but for chiral symmetries there is a simpler approach available.

Note that the parts of Eq. (19.7.3) that are proportional to $(1+\gamma_5)$ and $(1-\gamma_5)$ read

$$\exp\left(i\sum_a\theta_a^L\lambda_a\right)\exp\left(-i\sum_a\xi_a(x)\lambda_a\right)$$
$$=\exp\left(-i\sum_a\xi'_a(x)\lambda_a\right)\exp\left(i\sum_a\theta_a(x)\lambda_a\right) \quad (19.7.6)$$

and

$$\exp\left(i\sum_a\theta_a^R\lambda_a\right)\exp\left(i\sum_a\xi_a(x)\lambda_a\right)$$
$$=\exp\left(i\sum_a\xi'_a(x)\lambda_a\right)\exp\left(i\sum_a\theta_a(x)\lambda_a\right)\,, \quad (19.7.7)$$

where

$$\theta_a^L\equiv\theta_a^V+\theta_a^A\,,\qquad\qquad\theta_a^R\equiv\theta_a^V-\theta_a^A\,.$$

By multiplying Eq. (19.7.7) on the right by the inverse of Eq. (19.7.6), we find the simple transformation rule

$$U'(x)=\exp\left(i\sum_a\lambda_a\theta_a^R\right)U(x)\exp\left(-i\sum_a\lambda_a\theta_a^L\right)\,, \quad (19.7.8)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

where $U(x)$ is the unitary unimodular matrix

$$U(x) \equiv \exp\left(2i\sum_a \xi_a(x)\lambda_a\right) . \tag{19.7.9}$$

In other words, $U(x)$ transforms according to the $(\bar{\mathbf{3}},\mathbf{3})$ representation of $SU(3) \times SU(3)$. This is still a non-linear realization of $SU(3) \times SU(3)$ because the components of $U(x)$ are not independent; they are subject to the nonlinear constraints $U^\dagger U = 1$ and $\text{Det}\, U = 1$.

The unique $(SU(3) \times SU(3))$-invariant term in the Goldstone boson Lagrangian that is of second order in spacetime derivatives is

$$\mathscr{L}_{2\,\text{deriv}} = -\tfrac{1}{16}F^2\,\text{Tr}\{\partial_\mu U\,\partial^\mu U^\dagger\}\,, \tag{19.7.10}$$

with F^2 a positive constant to be determined. We can express the ξ_a in terms of conventionally normalized pseudoscalar meson fields by writing

$$\sum_a \lambda_a\xi_a = \frac{\sqrt{2}}{F}\begin{bmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & K^0 \\ \bar{K}^- & \bar{K}^0 & -\sqrt{\frac{2}{3}}\eta^0 \end{bmatrix} \equiv \frac{\sqrt{2}B}{F}\,, \tag{19.7.11}$$

so that the kinematic term in Eq. (19.7.10) is of the usual form

$$\mathscr{L}_{\text{kin}} = -\tfrac{1}{2}\partial_\mu\pi^0\partial^\mu\pi^0 - \partial_\mu\pi^+\partial^\mu\pi^- - \partial_\mu K^+\partial^\mu\bar{K}^- - \partial_\mu K^0\partial^\mu\bar{K}^0 - \tfrac{1}{2}\partial_\mu\eta^0\partial^\mu\eta^0\,.$$

To determine the constant F, we note by comparing Eq. (19.7.4) with Eq. (19.5.42), that for ξ infinitesimal, the components ξ_1, ξ_2, and ξ_3 are the same as the Goldstone boson fields ζ_1, ζ_2, and ζ_3 introduced in Section 19.5. Then, comparing (19.7.11) with (19.5.17) and (19.5.32), we see that the constant F defined by Eq. (19.7.10) is the same as the $F_\pi = 184$ MeV introduced in Section 19.4.

The $SU(3) \times SU(3)$ symmetry of quantum chromodynamics is broken by the quark mass term, which in terms of the quark field $\tilde{q}(x)$ defined by Eq. (19.7.4) is

$$\mathscr{L}_{\text{mass}} = -\bar{q}\,M_q\,q = -\bar{\tilde{q}}\,e^{-i\sqrt{2}\gamma_5 B/F_\pi}\,M_q\,e^{-i\sqrt{2}\gamma_5 B/F_\pi}\,\tilde{q} \tag{19.7.12}$$

where

$$M_q = \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{bmatrix} . \tag{19.7.13}$$

Eq. (19.7.12) contains a purely bosonic part, obtained by replacing the quark bilinear with its vacuum expectation value, given by the unbroken

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$SU(3)$ and parity symmetries as*

$$< \bar{\tilde{q}}_n \gamma_5 \tilde{q}_m >_0 = 0 , \qquad < \bar{\tilde{q}}_n \tilde{q}_m >_0 = -v\delta_{nm} . \tag{19.7.14}$$

The Goldstone boson mass term in the Lagrangian is then

$$\begin{aligned} -\frac{v}{F_\pi^2} \operatorname{Tr}\{B,\{B,M_q\}\} = -\frac{v}{F_\pi^2}\Bigg[& 4m_u\left(\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0\right)^2 + 4(m_u+m_d)\pi^+\pi^- \\ & + 4(m_u+m_s)K^+\bar{K}^- + 4m_d\left(-\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0\right)^2 \\ & + 4(m_d+m_s)K^0\bar{K}^0 + \frac{8}{3}m_s(\eta^0)^2\Bigg] . \end{aligned} \tag{19.7.15}$$

From this, we can read off that[35]

$$\begin{aligned} m_{\pi^+}^2 &= m_{\pi^0}^2 = \frac{4v}{F_\pi^2}[m_u + m_d] , \\ m_{K^+}^2 &= \frac{4v}{F_\pi^2}[m_u + m_s] , \\ m_{K^0}^2 &= \frac{4v}{F_\pi^2}[m_d + m_s] , \\ m_{\eta^0}^2 &= \frac{4v}{F_\pi^2}\left[\frac{4m_s + m_d + m_u}{3}\right] , \end{aligned} \tag{19.7.16}$$

and there is also a term that mixes the π^0 and η^0:

$$m_{\pi\eta}^2 = \frac{4v}{\sqrt{3}F_\pi^2}[m_u - m_d] . \tag{19.7.17}$$

The same results can also be obtained by operator methods. Noether's theorem allows us to construct generators T_a and X_a, for which

$$[T_a, q] = -\lambda_a q , \qquad [X_a, q] = -\gamma_5 \lambda_a q . \tag{19.7.18}$$

We can write the Goldstone boson fields in a real basis as π_a, with $\pi^\pm = (\pi_1 \pm i\pi_2)/\sqrt{2}$, $\pi^0 = \pi_3$, $K^\pm = (\pi_4 \pm i\pi_5)/\sqrt{2}$, $K^0 = (\pi_6 + i\pi_7)/\sqrt{2}$, $\bar{K}^0 = (\pi_6 - i\pi_7)/\sqrt{2}$, and $\eta^0 = \pi_8$. The $SU(3)$ symmetry that survives the spontaneous symmetry breaking tells us that the F-matrix

* The fact that the conservation laws of parity, charge, and strangeness respected by the vacuum are the same as those respected by the quark mass matrix is a result of the vacuum alignment condition discussed in Section 19.3. Likewise, if the quark fields are defined so that the quark masses are all positive, then with $v > 0$ the minus sign in Eq. (19.7.3) is required by the condition that the true vacuum be at a minimum rather than a maximum of the vacuum energy for vacuum states rotated by $SU(3) \times SU(3)$ transformations, for as we shall see, it is this sign that will yield positive masses for the pseudo-Goldstone boson octet.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

for these bosons takes the form $F_{ab} = F_\pi \delta_{ab}$. The $SU(3) \times SU(3)$ symmetry is also intrinsically broken by the mass term in the Hamiltonian $H_1 = m_u \bar{u}u + m_d \bar{d}d + m_s \bar{s}s = \bar{q} M_q q$, where M_q is the quark mass matrix (19.7.13). The mass matrix of the pseudo-Goldstone bosons is given here by (19.3.20) as

$$M^2_{ab} = -F_\pi^{-2} \langle \bar{q} \{\lambda_a, \{\lambda_b, M_q\}\} q \rangle_0 \,. \tag{19.7.19}$$

Since this is already of first order in quark masses, to this order we may use the unbroken $SU(3)$ relations $\langle \bar{u}u \rangle_0 = \langle \bar{d}d \rangle_0 = \langle \bar{s}s \rangle_0 \equiv -v$, and find the results (19.7.16) and (19.7.17), as before.

The masses $m_{K^+} = 493.65$ MeV and $m_{K^0} = 497.7$ MeV of the kaon doublet are quite close, and considerably larger than the pion mass, so we can see from (19.7.16) that the m_u and m_d must be considerably smaller than m_s. In calculating quantities that are sensitive to m_u and m_d, like the kaon mass difference or the pion masses, we should also take into account another small correction, the effect of electromagnetism. The electromagnetic current is $J^0 = ie\bar{q}\gamma^\mu Q q$, where Q is a diagonal matrix with elements $2/3$, $-1/3$, and $-1/3$. Its commutators with the $SU(3) \times SU(3)$ generators are

$$[T_a, J^\mu] = -ie\, \tfrac{1}{2} \bar{q} \gamma^\mu [Q, \lambda_a] q \,, \qquad [X_a, J^\mu] = -ie\, \tfrac{1}{2} \bar{q} \gamma^\mu \gamma_5 [Q, \lambda_a] q \,.$$

We see that J^μ commutes with X_3, X_6, X_7, and X_8 as well as T_3, T_6, T_7, and T_8, so the electromagnetic part of the Hamiltonian is invariant under the $SU(2) \times SU(2) \times U(1) \times U(1)$ subgroup with these generators. Therefore in the limit of zero quark masses, *electromagnetic effects give no mass to the neutral pseudo-Goldstone bosons*[36] π^0, K^0, $\bar{K}^0$, and η^0, the Goldstone bosons associated with the spontaneous breakdown of the symmetries with generators X_3, X_6, X_7 and X_8. Also, in the zero-quark-mass limit there is an unbroken $SU(2)$ symmetry generated by T_6, T_7, and $\sqrt{3}T_8 - T_3$, under which the K^+ and π^+ transform as a doublet, so that for zero quark masses *the electromagnetic corrections to the K^+ and π^+ masses are equal.*[36] Since the quark mass terms and electromagnetic corrections are all small, it is reasonable to treat these effects as additive corrections to the effective quark Hamiltonian. We see then that with electromagnetism taken into account, the mass formulas (19.7.16) should be corrected to read[37]

$$\begin{aligned}
m^2_{K^\pm} &= 4v(m_u + m_s)/F_\pi^2 + \Delta \,, \\
m^2_{K^0} &= m^2_{\bar{K}^0} = 4v(m_d + m_s)/F_\pi^2 \,, \\
m^2_{\pi^\pm} &= 4v(m_d + m_u)/F_\pi^2 + \Delta \,, \\
m^2_{\pi^0} &= 4v(m_d + m_u)/F_\pi^2 \,, \\
m^2_\eta &= 4v(m_u + m_d + 4m_s)/3F_\pi^2 \,,
\end{aligned} \tag{19.7.20}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

where Δ is the common electromagnetic correction to the K^+ and π^+ squared masses.

These formulas impose one linear relation on the five pseudo-Goldstone masses, a version of the Gell-Mann–Okubo relation:**

$$3m_\eta^2 + 2m_{\pi^+}^2 - m_{\pi^0}^2 = 2m_{K^+}^2 + 2m_{K^0}^2 \,. \tag{19.7.21}$$

Taking the kaon and pion masses from experiment, this relation yields an η mass of 566 MeV, in comparison with the experimental value of 547 MeV. The small discrepancy is usually attributed to the mixing of the η state with a heavier pseudoscalar particle, the η' at 958 MeV.

From the mass formulas (19.7.20) we may derive formulas for the quark mass ratios in terms of the pion and kaon masses:[37]

$$\frac{m_d}{m_s} = \frac{m_{K^0}^2 + m_{\pi^+}^2 - m_{K^+}^2}{m_{K^0}^2 + m_{K^+}^2 - m_{\pi^+}^2} \,, \qquad \frac{m_u}{m_s} = \frac{2m_{\pi^0}^2 - m_{K^0}^2 - m_{\pi^+}^2 + m_{K^+}^2}{m_{K^0}^2 + m_{K^+}^2 - m_{\pi^+}^2} \,. \tag{19.7.22}$$

Using the mass values $m_{\pi^+} = 139.57$ MeV, $m_{\pi^0} = 134.974$ MeV, $m_{K^+} = 493.65$ MeV and $m_{K^0} = 497.7$ MeV gives the ratios $m_d/m_s = 0.050$ and $m_u/m_s = 0.027$. Thus the ratio of the d and u quark masses is closer to 2 than 1. (A 1996 calculation[38a] using various other pieces of information as well as pseudoscalar meson masses has given values $m_d/m_s = 0.053 \pm 0.002$ and $m_u/m_s = 0.029 \pm 0.003$.)

None of this gives definite values for the individual quark masses. Indeed, these masses are not well defined until we define a renormalization prescription for the quark bilinears. Often this prescription is fashioned so that m_s equals the mass difference between isomultiplets differing by one unit of strangeness in some $SU(3)$ multiplet. For instance, the lightest vector meson octet consists of an isotopic doublet K^* at 892 MeV, interpreted as a bound state of an $\bar{s}$ antiquark and a u or d; its antidoublet, consisting of an s and a $\bar{u}$ or $\bar{d}$; and a $T = 1$ ρ at 770

** This relation was originally derived[38] on the basis of the approximate $SU(3)$ symmetry of Gell-Mann and Ne'eman generated by the T_a, ignoring mass differences within isotopic multiplets. From this derivation, one cannot tell whether this relation should be applied to the pseudoscalar meson masses themselves, or as here to their squares. Applying this relation to the pseudoscalar masses themselves would not in fact work very well; it gives an η mass of 613 MeV. The fact that the Gell-Mann–Okubo relation works well for the squares of the meson masses, but not for their first powers, supports the interpretation of these particles as Goldstone bosons of a spontaneously broken approximate $SU(3) \times SU(3)$ symmetry. Gell-Mann and Okubo also used the approximate $SU(3)$ symmetry generated by the T_a to derive relations among the masses of other particles, such as (ignoring isospin violating effects) the relation $2m_N + 2m_\Xi = 3m_\Lambda + m_\Sigma$ among the masses of the lightest baryon octet. For such multiplets, the average mass is so much larger than the mass differences within the multiplet that it makes little difference whether one applies the relation to the masses or their squares.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

MeV and a $T = 0$ ω at 783 MeV, both interpreted as bound states of a $\bar{u}$ or $\bar{d}$ and a u or d. If we wish to attribute the difference between the K^* mass and the average of the ρ and ω masses to the relatively large s quark mass, then we must renormalize the quark bilinear so that $m_s - \frac{1}{2}(m_u + m_d) = m_{K^*} - \frac{1}{2}(m_\rho + m_\omega) = 120$ MeV, yielding $m_s = 125$ MeV. With the above quark mass ratios, this gives $m_d = 6.0$ MeV and $m_u = 3.3$ MeV. However, these estimates of mass values are much less reliable than the values of the mass ratios given in Eq. (19.7.22). Often m_s is estimated[38b] as 180 MeV rather than 125 MeV.

The mass term (19.7.12) includes meson–meson interactions. Using Eqs. (19.7.14), (19.7.11), and (19.7.9), we may write the purely bosonic part of this term in the Lagrangian as

$$\mathscr{L}_{\text{mass, bosonic}} = \tfrac{1}{2}v \operatorname{Tr}\left\{M_q(U^\dagger + U)\right\} . \tag{19.7.23}$$

The form of this term could have been deduced from general symmetry considerations, which also allow us to find the allowed terms of higher order in M_q. Suppose we invent an external 3×3 field χ, and replace the mass term (19.7.12) in the underlying quantum chromodynamics Lagrangian with the χ–quark coupling term

$$\mathscr{L}_\chi = -\bar{q}\left[\tfrac{1}{2}(1+\gamma_5)\chi + \tfrac{1}{2}(1-\gamma_5)\chi^\dagger\right] q , \tag{19.7.24}$$

which becomes the same as Eq. (19.7.12) when we make the replacements $\chi = \chi^\dagger = M_q$. The point of this procedure is that the Lagrangian (19.7.24) becomes formally invariant under $SU(3) \times SU(3)$ if we give χ the formal transformation rule

$$\chi \to \exp\left(i\sum_a \lambda_a \theta_a^R\right)\ \chi\ \exp\left(-i\sum_a \lambda_a \theta_a^L\right) . \tag{19.7.25}$$

Thus we can work out the allowed bosonic terms involving quark masses by writing the most general $SU(3) \times SU(3)$ Lagrangian (up to some given order in derivatives and M_q) involving U and χ, requiring also invariance under the parity transformation

$$U(\mathbf{x}, t) \leftrightarrow U^\dagger(-\mathbf{x}, t) , \qquad \chi \leftrightarrow \chi^\dagger , \tag{19.7.26}$$

and then making the replacements

$$\chi = \chi^\dagger = M_q . \tag{19.7.27}$$

For instance, the interaction $\operatorname{Tr}(U^\dagger\chi + U\chi^\dagger)$ is invariant under $SU(3) \times SU(3)$ and parity, and becomes of the same form as Eq. (19.7.23) when we give χ the values (19.7.27).

Using this technique, Gasser and Leutwyler[39] have given the complete effective Lagrangian for the pseudoscalar octet of fourth order in momenta

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

or meson masses (with quark masses counted as being of second order in meson masses) as

$$\begin{aligned}\mathscr{L}_4 &= L_1 \operatorname{Tr}\left\{\partial_\mu U^\dagger \partial^\mu U\right\}^2 + L_2 \operatorname{Tr}\left\{\partial_\mu U \partial_\nu U^\dagger\right\} \operatorname{Tr}\left\{\partial^\mu U \partial^\nu U^\dagger\right\} \\ &+ L_3 \operatorname{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger \partial_\nu U \partial^\nu U^\dagger\right\} + L_4 \operatorname{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger\right\} \operatorname{Tr}\left\{M_q(U + U^\dagger)\right\} \\ &+ L_5 \operatorname{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger (M_q U + U^\dagger M_q)\right\} + L_6 \left[\operatorname{Tr}\left\{M_q(U + U^\dagger)\right\}\right]^2 \\ &+ L_7 \left[\operatorname{Tr}\left\{(U^\dagger - U)M_q\right\}\right]^2 + L_8 \operatorname{Tr}\left\{((UM_q)^2 + (U^\dagger M_q)^2)\right\}, \end{aligned} \tag{19.7.28}$$

where $L_1, \ldots, L_8$ are constants to be determined by comparison with experiment. The complete effective Lagrangian up to fourth order in meson masses and momenta is

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_2 + \mathscr{L}_4 , \tag{19.7.29}$$

where $\mathscr{L}_2$ is the sum of the terms (19.7.10) and (19.7.23):

$$\mathscr{L}_2 = -\tfrac{1}{16} F^2 \operatorname{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger\right\} + \tfrac{1}{2} v \operatorname{Tr}\left\{M_q(U^\dagger + U)\right\} . \tag{19.7.30}$$

Electroweak interactions may be included by replacing the derivatives ∂_μ with suitable gauge-covariant derivatives D_μ, and adding a few additional terms to $\mathscr{L}_{\text{eff}}$.

Following the same power-counting arguments as in Sections 19.5 and 19.6, to calculate S-matrix elements to fourth order in meson masses and momenta we must include both tree graphs to all orders in $\mathscr{L}_2$ and up to first order in $\mathscr{L}_4$, and also one-loop graphs constructed from $\mathscr{L}_2$ alone. Gasser and Leutwyler and others have carried out a comprehensive study of meson dynamics (and associated electroweak interactions) using this effective Lagrangian.[40]

* * *

The quark mass term (19.7.12) naturally affects other $SU(3)$ multiplets. For any multiplet other than the pseudoscalar octet this can be treated as a first-order perturbation, and therefore gives a shift in the mass matrix of a generic multiplet $|i\rangle$ equal to

$$\delta m_{ij} = \langle i|\bar{\tilde{q}} M_q \tilde{q}|j\rangle \tag{19.7.31}$$

and $SU(3)$ may be used[41] to relate the various matrix elements $\langle i|\bar{\tilde{q}}_r \tilde{q}_s|j\rangle$. In this way, it is straightforward to show that

$$\frac{\delta m_p - \delta m_n}{m_u - m_d} = \frac{m_\Xi - m_\Sigma}{m_s} . \tag{19.7.32}$$

Using this together with Eq. (19.7.22) gives the quark mass contribution to the nucleon mass splitting $\delta m_p - \delta m_n \approx -2.5$ MeV. This result may be used in Eqs. (19.5.65) and (19.5.66) to calculate the leading violations of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

isospin symmetry in low energy pion–nucleon interactions. Of course, as defined here $\delta m_p - \delta m_n$ is not the full proton–neutron mass difference, which receives an important contribution also from photon emission and absorption. Because the neutron is electrically neutral, this electromagnetic term is almost certainly positive, in agreement with the fact that the observed proton–neutron mass difference is -1.3 MeV, leaving $+1.2$ MeV to be accounted for by electromagnetism. Unfortunately, accurate calculation of this electromagnetic mass difference has proved difficult.

19.8 Anomalous Terms in Effective Field Theories*

In implementing the effective field theory program described in the previous three sections, it is necessary that the action should contain all possible terms allowed by the assumed symmetries of the theory. The methods described in these sections allow us to identify all *manifestly* invariant terms, either by using the general formalism of covariant derivatives described in Section 19.6, or for chiral symmetries by using a linearly transforming field subject to non-linear constraints, like the $U(x)$ of Section 19.7. But it is possible that there may be other terms in the action of the effective field theory that are anomalous in the sense of being given by integrals of four-dimensional Lagrangian densities that are *not* invariant, but whose variation under the broken symmetry is a spacetime derivative, preserving the invariance of the corresponding term in the action.

As we shall see in Section 22.7, such a term was discovered for $SU(3) \times SU(3)$ by Wess and Zumino,[42] in a study of 'anomalies' due to quark loop graphs. However, the structure of this term can be understood without knowing anything about the underlying theory of quarks and gluons.

The easiest way to describe the Wess–Zumino term is by an extension of spacetime to five dimensions, introduced for this purpose by Witten.[43] As long as we require the fields in an effective field theory to approach a common limit as $x^\mu \to \infty$ in any direction, we can think of spacetime as having the topology of a sphere S_4, with the point at infinity included as an ordinary point. As remarked in Section 19.6, when a group G is broken to a subgroup H, the possible values of the Goldstone boson fields ξ_a at any spacetime point may be regarded as defining a point in the coset space G/H (the space of elements of G, with two elements identified if they differ by multiplication on the right by elements of H), so a set of functions $\xi_a(x)$ represents a mapping of the spacetime S_4 sphere into G/H. Depending on the topology of G/H, it may be possible smoothly to

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

deform any four-sphere in G/H to a point; that is, it may be possible to extend any $\xi_a(x)$ to a continuous function $\xi_a(x;s)$ defined for $0 \geq s \geq 1$, for which $\xi_a(x;0) = \xi_a(x)$ and $\xi_a(x;1)$ is any fixed point (say, $\xi_a = 0$) on the original sphere. Where this is true, it is expressed mathematically as the statement that the homotopy group $\pi_4(G/H)$ is trivial. (For a discussion of homotopy groups, see Section 23.2.) It is known that this is the case for $SU(N) \times SU(N)$ spontaneously broken to $SU(N)$, where the coset space $SU(N) \times SU(N)/SU(N)$ has the topology of $SU(N)$ itself.** Hence in the case of physical interest, where $G = SU(3) \times SU(3)$ and $H = SU(3)$, we may extend the Goldstone boson fields, or equivalently $U(x)$, (see Eq. (19.7.9)) to a unitary unimodular matrix $U(y)$ defined in a five-dimensional ball B_5 with coordinates x^μ and s, whose surface is the four-dimensional sphere of spacetime.

Now consider the following function formed from $U(y)$:

$$\omega(y) \equiv -\frac{i}{240\pi^2}\epsilon^{ijklm}\mathrm{Tr}\left\{U^{-1}\frac{\partial U}{\partial y^i}U^{-1}\frac{\partial U}{\partial y^j}U^{-1}\frac{\partial U}{\partial y^k}U^{-1}\frac{\partial U}{\partial y^l}U^{-1}\frac{\partial U}{\partial y^m}\right\}, \tag{19.8.1}$$

where indices i, j, etc. run over the five coordinate directions for the coordinates x^μ and s. (The phase and numerical coefficient are chosen for later convenience.) This is manifestly invariant under the chiral transformations (19.7.8). Also, because ϵ^{ijklm} is a tensor density, the integral of $\omega(y)$ over the five-ball is manifestly independent (up to a sign) of the choice of five-dimensional coordinates y^i. Furthermore, this integral depends only on the values taken by $U(y)$ on the ball's surface; that is, in spacetime. To check this last point, note that when we make an infinitesimal variation $\delta U(y)$ in $U(y)$ in the interior of the ball, $\omega(y)$ changes by a derivative:

$$\delta\omega(y) = -\frac{i}{48\pi^2}\epsilon^{ijklm}\frac{\partial}{\partial y^m}\mathrm{Tr}\left\{U^{-1}\frac{\partial U}{\partial y^i}U^{-1}\frac{\partial U}{\partial y^j}U^{-1}\frac{\partial U}{\partial y^k}U^{-1}\frac{\partial U}{\partial y^l}U^{-1}\delta U\right\}, \tag{19.8.2}$$

(for this calculation, see Section 23.4), so a change in $U(y)$ that does not affect its value in spacetime also does not affect the integral of (19.8.1) over the five-ball B_5 whose surface is spacetime. We can therefore include this integral as a term in the action:

$$I_{\mathrm{WZW}}[U] = n\int_{B_5} d^5y\ \omega(y) \tag{19.8.3}$$

with n a coefficient that so far is arbitrary.

** This is shown by the fact that we can use $SU(N)$ elements $U(x)$ to represent the right cosets of $SU(N)$ in $SU(N) \times SU(N)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

This term may be written as the four-dimensional integral of a Lagrangian density, but not of an $(SU(3) \times SU(3))$-invariant Lagrangian density. Using Eq. (19.7.9), the leading term in $\omega(x)$ in the limit of small meson fields is

$$\omega(x) \to \frac{8\sqrt{2}}{15\pi^2 F_\pi^5} \epsilon^{ijklm} \operatorname{Tr}\left\{ \frac{\partial B}{\partial y^i} \frac{\partial B}{\partial y^j} \frac{\partial B}{\partial y^k} \frac{\partial B}{\partial y^l} \frac{\partial B}{\partial y^m} \right\} , \tag{19.8.4}$$

where B is the matrix (19.7.11) of Goldstone boson fields. It follows then from Gauss's theorem that

$$I_{\rm WZW}[U] = \frac{8\sqrt{2}n}{15\pi^2 F_\pi^5} \epsilon^{\mu\nu\rho\sigma} \int_{S_4} d^4x \operatorname{Tr}\left\{ B \frac{\partial B}{\partial x^\mu} \frac{\partial B}{\partial x^\nu} \frac{\partial B}{\partial x^\rho} \frac{\partial B}{\partial x^\sigma} \right\} + O\left(\frac{B^6}{F_\pi^6}\right) . \tag{19.8.5}$$

Although chiral invariant, this cannot be written as the integral of a chiral-invariant density over spacetime, because any chiral invariant density would have to be constructed out of the first and higher covariant derivatives of the Goldstone boson fields, and so, when expanded in powers of the Goldstone boson fields, such an invariant density would begin with a term involving only derivatives of B, not B itself.

As Witten noted, the inclusion of this term in the effective action resolved what otherwise would have been a conflict between the $SU(3) \times SU(3)$ effective field theory and experiment. Because no $\epsilon^{\mu\nu\rho\sigma}$ terms appear in the effective Lagrangians (19.7.28) and (19.7.30) (or in higher-order terms of this sort) parity conservation imposes the requirement that these terms are *even* in the Goldstone boson fields, ruling out processes like $K + \bar{K} \to 3\pi$. Not only is there no symmetry in the underlying theory of quantum chromodynamics that would account for such a selection rule — there is even experimental evidence against it, for as Witten pointed out, the ϕ meson is observed to decay both into $K + \bar{K}$ and 3π final states. Eq. (19.8.5) shows that this unwanted selection rule is removed by the Wess–Zumino–Witten term in the action.

Remarkably, the coefficient of the Wess–Zumino–Witten term is not a freely adjustable parameter. As Witten showed,[43] this is because, although this term is not changed by smooth deformations of the function $U(y)$ in B_5 that do not affect $U(x)$ on the spacetime boundary S_4, the Wess–Zumino–Witten term *can* be changed by a discontinuous change in $U(y)$ that leaves $U(x)$ unchanged on the boundary. We may think of the five-ball B_5 as half of a five-sphere S_5, with the spacetime S_4 as the border between B_5 and the other half B_5'. (Think of S_5 as analogous to the surface of the Earth, with the spacetime S_4 as the equator and B_5 and B_5' as the northern and southern hemispheres.) Because S_4 is also the boundary of the other half of S_5, we could just as well have written a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

Wess–Zumino–Witten term as

$$I'_{\rm WZW}[U] = -n \int_{B'_5} d^5y\ \omega(y) \tag{19.8.6}$$

with a minus sign inserted because the boundary of B'_5 is the four-sphere with opposite orientation. It is not possible to require that the terms (19.8.3) and (19.8.6) should be equal for arbitrary Goldstone boson fields without setting $n = 0$, but in order that the weighting factor $\exp(iI)$ in path integrals should be unaffected by the difference between the terms (19.8.3) and (19.8.6), it is only necessary to require that this difference should be 2π times an integer. That is,

$$I_{\rm WZW}[U] - I'_{\rm WZW}[U] = n \int_{S_5} d^5y\ \omega(y) = 2\pi \times \text{ integer}\,. \tag{19.8.7}$$

With the normalization factors we have inserted in the definition (19.8.1) of $\omega(y)$, the integral of $\omega(y)$ over any five-sphere turns out to have the value[44] 2π. It follows that *the coefficient n must be an integer.*

The example of the Wess–Zumino–Witten term raises the question whether there may be other anomalous terms in the action, not necessarily related to quark loops, that also are invariant under $SU(3) \times SU(3)$, despite not being four-dimensional integrals of $(SU(3) \times SU(3))$-invariant Lagrangian densities. Fortunately, the answer is no. It has been shown[45] that for a general group G broken to an arbitrary subgroup H (with $\pi_4(G/H) = 0$), any term $F[\xi]$ in the action of the Goldstone boson fields $\xi_a(x)$ may always be written as the integral of a G-invariant five-form Ω over a five-ball B_5 whose boundary is the spacetime four-sphere S_4:

$$F[\xi] = \int_{B_5} d^5y\ \epsilon^{ijklm} \frac{\partial \xi_a}{\partial y^i} \frac{\partial \xi_b}{\partial y^j} \frac{\partial \xi_c}{\partial y^k} \frac{\partial \xi_d}{\partial y^l} \frac{\partial \xi_e}{\partial y^m}\ \Omega_{abcde}(\xi(y))\,. \tag{19.8.8}$$

In order that this should be independent of the particular way that $\xi_a(x)$ is extended into the interior of the five-ball, it is necessary that Ω should be exact, in the sense that there it is the exterior derivative of a four-form

$$\Omega_{abcde}(\xi) = (\partial/\partial \xi_{[a}) \mathscr{L}_{bcde]}(\xi) \tag{19.8.9}$$

(with square brackets as usual indicating antisymmetrization with respect to the enclosed indices), so that

$$F[\xi] = \int_{S_4} d^4x\ \epsilon^{\mu\nu\rho\sigma} \frac{\partial \xi^a}{\partial x^\mu} \frac{\partial \xi^b}{\partial x^\nu} \frac{\partial \xi^c}{\partial x^\rho} \frac{\partial \xi^d}{\partial x^\sigma} \mathscr{L}_{abcd}(\xi)\,. \tag{19.8.10}$$

It follows that Ω is also closed; that is, it has a vanishing exterior derivative

$$(\partial/\partial \xi_{[f}) \Omega_{abcde]}(\xi) = 0\,. \tag{19.8.11}$$

Where the four-form $\mathscr{L}_{abcd}(\xi)$ is also G-invariant, the functional (19.8.10) is just one of the ordinary manifestly G-invariant terms in the action,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

discussed in the previous three sections. The anomalous terms in the action arise from the possibility that, although every term in $\Omega(y)$ is G-invariant and the exterior derivative of a four-form, some terms may not be the exterior derivatives of G-invariant four-forms. Thus the new terms in the action may be identified with the closed G-invariant five-forms that are independent, in the sense that no real linear combination of them is the exterior derivative of a G-invariant four-form. These are known in mathematics as the generators of the *de Rham cohomology group* $H^5(G/H;\mathbf{R})$. (The group multiplication rule here is just simple addition.) The de Rham cohomology groups have been calculated for manifolds of various topologies.[46] In particular, $H^5(SU(N)\times SU(N)/SU(N);\mathbf{R})$ has a single generator, given by Eq. (19.8.1). Thus without knowing anything about the underlying theory of quarks and gluons, we can learn everything about the anomalous terms in the Goldstone boson action, with the single exception of the value of the integer n. We will see in Section 22.7 that in $SU(N_c)$ gauge theories this integer equals the number N_c of colors, which in quantum chromodynamics is $n = 3$.

19.9 Unbroken Symmetries

We have seen how the properties of Goldstone bosons and their low energy interactions may be deduced from an assumption that the theory is invariant (or approximately invariant) under a group G spontaneously broken to a subgroup H. But in applying these methods to the cases where G is $SU(2)\times SU(2)$ or $SU(3)\times SU(3)$, we had to take the pattern of symmetry breaking, of $SU(2)\times SU(2)$ or $SU(3)\times SU(3)$ to their non-chiral $SU(2)$ or $SU(3)$ subgroups, from experiment. In Section 22.5 we shall show that the $SU(3)\times SU(3)$ symmetry for massless u, d, and s quarks must in fact be spontaneously broken in quantum chromodynamics, but it is more difficult to show on the basis of quantum chromodynamics that the $SU(2)\times SU(2)$ symmetry with only u and d massless is also spontaneously broken.* On the other hand, there is an intuitive argument that their non-chiral $SU(2)$ or $SU(3)$ subgroups are *not* broken, based on a conjecture known as the *persistent mass condition*[48], which states that composite particles will not be massless if the particles of which they are

* Weingarten[47] has used lattice methods to show that whether or not chiral symmetry is broken, the lightest particle in quantum chromodynamics with massless u and d quarks must have the quantum numbers of the pion. As we will see in Section 22.5, the existence of anomalies due to fermion loops in quantum chromodynamics together with the assumption of quark trapping requires that some hadron be massless, so it follows that the pion is massless, which strongly suggests that chiral symmetry is spontaneously broken.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

composed are massive. Non-chiral symmetries like isospin conservation are not violated if we give the quarks equal masses, and then if they were spontaneouly broken we would have the massless Goldstone bosons formed as composites of massive quarks, in contradiction with the persistent mass condition. In what follows we shall present a proof by Vafa and Witten[49] that in gauge theories like quantum chromodynamics those non-chiral symmetries that are not violated by quark masses can not be spontaneously broken. This result is of more than academic interest; as we shall see in Section 21.4, it is possible that the spontaneous breakdown of electroweak gauge symmetries is described by a 'technicolor' theory similar to quantum chromodynamics, and in testing this idea it is important to know what symmetries are left unbroken in this theory.

Consider a gauge theory like quantum chromodynamics, with a number of fermion 'flavors' in identical representations of the gauge group. If all fermions have masses, then the theory will be invariant under all unitary global non-chiral transformations on the fermion flavors that commute with the fermion mass matrix. For instance, if n_1 fermions are degenerate with a common mass m_1, n_2 fermions are degenerate with some other common mass m_2, and so on, then this global symmetry group is $U(n_1)\times U(n_2)\times\ldots$. (As a special case, if there are no degeneracies we have a global symmetry under $U(1)\times U(1)\times\ldots$; an example is the conservation of baryon number, strangeness, etc. in quantum chromodynamics.) *These symmetries cannot be spontaneously broken.*

To prove this, let us consider a general Greens function for r fermion and antifermion fields.** In the path-integral formalism this is†

$$\begin{aligned}&\langle T\left\{\Psi_{u_1k_1}(x_1)\cdots\Psi_{u_rk_r}(x_r)\,\Psi^\dagger_{v_1l_1}(y_1)\cdots\Psi^\dagger_{v_rl_r}(y_r)\right\}\rangle_{\rm VAC}\\ &=\frac{1}{Z}\int[dA][d\psi][d\psi^\dagger]\,\psi_{u_1k_1}(x_1)\cdots\psi_{u_rk_r}(x_r)\,\psi^\dagger_{v_1l_1}(y_1)\cdots\psi^\dagger_{v_rl_r}(y_r)\\ &\quad\times\exp\left(iI_{\rm gauge}[A]+iI_{\rm Dirac}[\psi,\psi^\dagger;A]\right),\end{aligned}\tag{19.9.1}$$

where $k_1\ldots k_r$ and $l_1\ldots l_r$ are Dirac spin indices, $u_1\ldots u_r$ and $v_1\ldots v_r$ are fla-

** In the original work of Vafa and Witten,[49] they first proved the absence of symmetry breaking in the case $r=1$ with $x=y$, then observed that the absence of symmetry breaking in this vacuum expectation value did not rule out the possibility of a spontaneous symmetry breakdown occurring in other Greens functions, and so went on to different methods of proof.

† As we saw in Section 15.5, both numerator and denominator are proportional to the infinite volume of the gauge group, which cancels in the ratio (19.9.1). The presence of this infinite factor detracts from the rigor of the following arguments, but if we were to remove it by introducing ghosts then some of the steps below that depend on the positivity of the action would raise difficulties. One way to deal with this problem is to replace the spacetime continuum with a finite lattice of points, in which case the gauge group has a finite volume, and no gauge fixing or ghosts are needed.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

vor indices, $I_{\text{gauge}}[A]$ is the action for a pure gauge theory, $I_{\text{Dirac}}[\psi, \psi^\dagger; A]$ is the action for the Dirac fields in the presence of a gauge field $A^\mu_\alpha(x)$, and Z is the vacuum–vacuum amplitude

$$Z \equiv \int [dA][d\psi][d\psi^\dagger] \exp\left(iI_{\text{gauge}}[A] + iI_{\text{Dirac}}[\psi, \psi^\dagger; A]\right) . \tag{19.9.2}$$

It will be necessary here to work in a Euclidean spacetime, with $x^4 = x_4 = ix^0$, $y^4 = y_4 = iy^0$, and $A^4_\alpha = A_{4\alpha} = iA_{0\alpha}$ all real. (See Appendix A of Chapter 23.) In this case, the Dirac action is

$$I_{\text{Dirac}}[\psi, \psi^\dagger; A] = i \int d^3x \int dx^4 \, \psi^\dagger [\not{D} + M] \psi , \tag{19.9.3}$$

where M is the fermion mass matrix, and $\not{D}$ is the Euclidean gauge-covariant derivative contracted with the Euclidean Dirac matrices

$$\not{D} = \sum_{i=1}^{4} (\partial_i - it^\alpha A_{i\alpha}) \gamma_i , \tag{19.9.4}$$

where as usual $\gamma_4 = i\gamma^0$. Because the action is quadratic in fermion fields, we can explicitly perform the integral over these fields

$$\begin{aligned} &\langle T \left\{ \Psi_{u_1 k_1}(x_1) \cdots \Psi_{u_r k_r}(x_r) \Psi^\dagger_{v_1 l_1}(y_1) \cdots \Psi^\dagger_{v_r l_r}(y_r) \right\} \rangle_{\text{VAC}} \\ &\quad = \frac{1}{Z} \int [dA] \, \text{Det} \, (\not{D} + M) \, \exp\left(iI_{\text{gauge}}[A]\right) \\ &\quad \times \left[(\not{D} + M)^{-1}_{x_1 u_1 k_1, \, y_1 v_1 l_1} \cdots (\not{D} + M)^{-1}_{x_r u_r k_r, \, y_r v_r l_r} \pm \text{permutations} \right] , \end{aligned} \tag{19.9.5}$$

where '$\pm$ permutations' indicates that we must sum over all $r!$ permutations of the ψ fields, with a minus sign for odd permutations, and

$$Z = \int [dA] \, \text{Det} \, (\not{D} + M) \exp\left(iI_{\text{gauge}}[A]\right) . \tag{19.9.6}$$

The expression (19.9.5) is manifestly invariant under any unitary transformation on flavor indices that commutes with the mass matrix M. It does not matter what non-perturbative effects are produced by the functional integral over gauge fields and ghosts; these fields are inert under the symmetries in question here, so that the remaining functional integral in Eq. (19.9.5) cannot break these symmetries.[††] But for this argument to be convincing, we must show that the expression (19.9.5) is well defined.

[††] It is more difficult to show that P, C, and T are not spontaneously broken in quantum chromodynamics with massive quarks, because these symmetries act non-trivially on the gauge fields. Vafa and Witten[50] have applied the methods of Ref. 49 to show that P is not spontaneously broken in quantum chromodynamics.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

This is not merely an academic question of mathematical rigor. As we saw in Section 19.1, the sign that a symmetry is not spontaneously broken is not just that the ground state is invariant under the symmetry transformations, but also that the symmetry of the ground state is stable under small perturbations. If we break the symmetries of M by adding a small perturbation δM, and if the expression (19.9.5) becomes singular as $\delta M \to 0$, then factors of δM in symmetry-breaking matrix elements may be cancelled by the singularities in the matrix elements for $\delta M = 0$. Precisely this happens for chiral symmetries, which arise in a symmetry limit where some of the eigenvalues of M vanish. In this case, as we approach the symmetry limit of zero mass, the factors of mass in the numerators of symmetry-breaking expectation values are cancelled by factors of mass in the denominators of the propagators $(\not{D}+M)^{-1}$ wherever $\not{D}$ has zero eigenvalues. For this reason, the arguments here will apply only to the non-chiral symmetries of theories with mass matrices M whose eigenvalues are all non-zero.

To set a bound on the matrix element (19.9.5) in this case, note first that the differential operator (19.9.4) is here antihermitian, so as long as the hermitian matrix M has no null eigenvalues, $\not{D}+M$ has a well-defined inverse. It is still necessary to show that the remaining integration over the gauge fields in Eq. (19.9.5), including that in Z, does not make this expression singular when M satisfies the conditions for a symmetry, for instance that n of its eigenvalues are equal for a $U(n)$ symmetry. As we shall see, this will insure that when a fermion mass matrix M that is invariant under some global symmetry transformation is perturbed by adding a small term δM that breaks this symmetry, the change in the expectation value (19.9.5) under this symmetry transformation vanishes in the limit $\delta M \to 0$.

It is difficult to show that the coordinate-space Greens functions Eq. (19.9.5) are non-singular, so let us consider instead the matrix element for smeared fields

$$\Psi_u[f] \equiv \int d^4x\, f^{k\dagger}(x)\, \Psi_{uk}(x)\,, \tag{19.9.7}$$

where $f^k(x)$ are arbitrary smooth square-integrable functions. In these terms, Eq. (19.9.5) reads

$$\begin{aligned}\left\langle T\left\{\Psi_{u_1}[f_1]\cdots\Psi_{u_r}[f_r]\;\Psi^\dagger_{v_1}[g_1]\cdots\Psi^\dagger_{v_r}[g_r]\right\}\right\rangle_{\text{VAC}} \\ = \frac{1}{Z}\int [dA]\ \text{Det}\,(\not{D}+M)\ \exp\left(iI_{\text{gauge}}[A]\right) \\ \times\ \left[(\not{D}+M)^{-1}_{f_1 u_1,\, g_1 v_1}\cdots(\not{D}+M)^{-1}_{f_r u_r,\, g_r v_r} \pm \text{permutations}\right]\,,\end{aligned} \tag{19.9.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

where

$$(\not{D} + M)^{-1}_{f\,u,\,g\,v} \equiv \int d^4x \int d^4y\, f^{k\dagger}(x)\, (\not{D} + M)^{-1}_{x\,u\,k,\,y\,v\,l}\, g^l(y)\,. \qquad (19.9.9)$$

In this basis, the fermion propagators for a given gauge field are not only well defined, but bounded uniformly in $A^\mu_\alpha(x)$:‡

$$\left|(\not{D} + M)^{-1}_{f\,u,\,g\,v}\right| \leq \sum_a \frac{1}{|m_a|}\,, \qquad (19.9.10)$$

where m_a are the eigenvalues of M, and for convenience we are now normalizing the functions $f_i(x)$ and $g_j(y)$ so that

$$\int f_i^\dagger(x) f_i(x)\, d^4x = \int g_j^\dagger(y) g_j(y)\, d^4y = 1\,. \qquad (19.9.11)$$

(The summation convention is suspended here.) Furthermore, the weight function for the average over gauge fields is positive:‡‡

$$\begin{aligned}\text{Det}\,(\not{D} + M) \exp\left(iI_{\text{gauge}}[A]\right) = \exp\left(-\tfrac{1}{4}\int d^3x \int dx^4 \sum_{i=1}^{4}\sum_{j=1}^{4} F_{ij}{}^2\right) \\ \times \sqrt{\text{Det}\left[(iD)^2 + M^2\right]}\,. \qquad (19.9.12)\end{aligned}$$

With a positive weight function, the average of any function is bounded by the bound of that function. Using the bound (19.9.10) in Eq. (19.9.8),

‡ To see this, we may expand in the eigenvectors of M

$$M_{uv} c_v^a = m_a c_u^a\,, \qquad c_u^{a*} c_u^b = \delta_{ab}$$

and, adapting a trick of Vafa and Witten,[49] write Eq. (19.9.9) as

$$\begin{aligned}(\not{D} + M)^{-1}_{f\,u,\,g\,v} &= \sum_a c_u^{a*} c_v^a \int d^4x \int d^4y\, f^{k\dagger}(x)\, (\not{D} + m_a)^{-1}_{x\,k,\,y\,l}\, g^l(y) \\ &= \sum_a \pm c_u^{a*} c_v^a \int_0^\infty dt \exp(-|m_a|\tau) \int d^4x \int d^4y\, f^{k\dagger}(x)\, (\exp(\mp\tau\, \not{D}))_{x\,k,\,y\,l}\, g^l(y)\,,\end{aligned}$$

where $\pm$ is the sign of m_a. The matrix $c_u^{a*} c_v^a$ and the operator $(\exp(\mp\tau\ \not{D}))_{x\,k,\,y\,l}$ are both unitary, so

$$\left|c_u^{a*} c_v^a \int d^4x \int d^4y\, f^{k\dagger}(x)\, (\exp(\mp\tau\, \not{D}))_{x\,k,\,y\,l}\, g^l(y)\right|^2 \leq \int d^4x\, |f(x)|^2 \int d^4y\, |g(y)|^2 = 1\,.$$

Using this inside the integral yields Eq. (19.9.10).

‡‡ In the fermion determinant, we use the fact that $\text{Det}\,(\not{D} + M) = \text{Det}\,\gamma_5 (\not{D} + M)\gamma_5 = \text{Det}\,(-\not{D} + M)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

the matrix element is bounded by

$$\left|\langle T\left\{\Psi_{u_1}[f_1]\cdots\Psi_{u_r}[f_r]\;\Psi^\dagger_{v_1}[g_1]\cdots\Psi^\dagger_{v_r}[g_r]\right\}\rangle_{\rm VAC}\right| \le r!\left[\sum_a \frac{1}{|m_a|}\right]^r . \tag{19.9.13}$$

Thus there are no singularities as the symmetry-breaking terms δM in M go to zero that could cancel factors of δM. To see this more explicitly, we can use the same methods to show that if we perturb M by a small symmetry-breaking term δM, then the term in Eq. (19.9.8) of first order in δM is bounded by

$$\begin{aligned}&\left|\delta\langle T\left\{\Psi_{u_1}[f_1]\cdots\Psi_{u_r}[f_r]\;\Psi^\dagger_{v_1}[g_1]\cdots\Psi^\dagger_{v_r}[g_r]\right\}\rangle_{\rm VAC}\right| \\ &\quad\le r\,r!\sum_{a,b}\frac{|(\delta M)_{ab}|}{|m_a m_b|}\left[\sum_a\frac{1}{|m_a|}\right]^{r-1},\end{aligned} \tag{19.9.14}$$

so this vanishes when $\delta M \to 0$.

In the real world none of the quark masses are zero or degenerate, so it follows immediately from the above arguments that the $U(1)$ symmetries like conservation of baryon number, strangeness, etc., are neither spontaneously nor intrinsically broken. The other strong-interaction symmetries like isospin or $SU(3)$ are more problematic. These symmetries would be completely unbroken if the u and d quarks, or u, d, and s quarks, had equal non-zero masses, but as we saw in Section 19.7, these masses are not at all degenerate; the isospin and $SU(3)$ symmetries arise because the quark masses are small, not equal. These symmetries do remain unbroken if we give two or three quarks equal masses and let the masses become arbitrarily small, so isospin or $SU(3)$ will be good symmetries in any process that is not sensitive to the small quark masses. But not all processes are insensitive to these masses, because for two or three massless quarks, the pion or the pseudoscalar meson octet to which it belongs becomes massless. This is not a problem for isospin, because as we saw in Section 19.5, the pion triplet remains degenerate to first order in quark masses even though $m_u \neq m_d$. For $SU(3)$ the quark mass differences produce first-order mass differences among the pion, kaon, and eta, so that processes dominated by one-meson poles *can* show large departures from $SU(3)$.

19.10 The $U(1)$ Problem

The success of quantum chromodynamics in explaining the pattern of strong-interaction symmetries seemed at first to be marred by one failure. As we saw in Section 19.5, the broken $SU(2)\times SU(2)$ symmetry of the strong interactions is a natural consequence of the smallness of the u and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

d quark masses. But the Lagrangian of quantum chromodynamics with small u and d quark masses also has another chiral symmetry,[51] the $U(1)_A$ symmetry under transformations

$$u \to \exp(i\gamma_5\theta)u\,, \qquad d \to \exp(i\gamma_5\theta)d\,. \tag{19.10.1}$$

Such a symmetry if unbroken would, like $SU(2) \times SU(2)$, impose a parity doubling on the hadron spectrum, but no such parity doubling is observed. On the other hand, a broken $U(1)_A$ symmetry would imply the existence of an isoscalar 0^- Goldstone boson with a mass comparable to that of the pion, and this Goldstone boson is also not observed. It is true that the η meson is an isoscalar 0^- boson, but it is considerably heavier than the pion, and as we saw in Section 19.7, it is well understood as one of the Goldstone bosons of $SU(3) \times SU(3)$. With the s as well as the u and d quarks regarded as relatively light, the Lagrangian of quantum chromodynamics would have a $U(1)$ chiral symmetry in addition to $SU(3) \times SU(3)$, under the transformations

$$u \to \exp(i\gamma_5\theta)u\,, \qquad d \to \exp(i\gamma_5\theta)d\,, \qquad s \to \exp(i\gamma_5\theta)s\,. \tag{19.10.2}$$

The spontaneous breakdown of such a symmetry would require the existence of *two* isoscalar 0^- mesons: one the η, and the other with a mass comparable to that of the pion.

This prediction can be made more explicit.[52] If we include the transformation (19.10.2) among the spontaneously broken symmetries of massless quantum chromodynamics, then in the presence of quark masses we encounter a term (19.7.12):

$$\mathscr{L}_{\text{mass}} = -\bar{q}\, M_q\, q = -\bar{\tilde{q}}\, e^{-i\sqrt{2}\gamma_5 B/F_\pi}\, M_q\, e^{-i\sqrt{2}\gamma_5 B/F_\pi}\, \tilde{q}\,, \tag{19.10.3}$$

where again

$$M_q = \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{bmatrix}, \tag{19.10.4}$$

but now B includes a Goldstone boson field ζ for the broken symmetry (19.10.2):

$$B = \begin{bmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & K^0 \\ \bar{K}^- & \bar{K}^0 & -\sqrt{\frac{2}{3}}\pi^0 \end{bmatrix} + \frac{F_\pi}{\sqrt{3}F_\zeta} \begin{bmatrix} \zeta & 0 & 0 \\ 0 & \zeta & 0 \\ 0 & 0 & \zeta \end{bmatrix}, \tag{19.10.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

with F_ζ the unknown coupling of the $U(1)_A$ Goldstone boson to the corresponding current (and the factor $\sqrt{3}$ inserted for future convenience.) We can again use the unbroken $SU(3)$ symmetry to write the vacuum expectation value of the quark bilinear in the form (19.7.14), and expanding in powers of the boson fields, we find a Goldstone boson mass term in the Lagrangian of the form

$$\begin{aligned}-\frac{v}{F_\pi^2}\,\mathrm{Tr}\,\{B,\{B,M_q\}\} = -\frac{v}{F_\pi^2}\Bigg[&4m_u\left(\frac{1}{\sqrt{2}}\pi^0+\frac{1}{\sqrt{6}}\eta^0+\frac{F_\pi}{\sqrt{3}F_\zeta}\zeta\right)^2\\&+4(m_u+m_d)\pi^+\pi^-+4(m_u+m_s)K^+\bar{K}^-\\&+4m_d\left(-\frac{1}{\sqrt{2}}\pi^0+\frac{1}{\sqrt{6}}\eta^0+\frac{F_\pi}{\sqrt{3}F_\zeta}\zeta\right)^2\\&+4(m_d+m_s)K^0\bar{K}^0+4m_s\left(-\sqrt{\frac{2}{3}}\eta^0+\frac{F_\pi}{\sqrt{3}F_\zeta}\zeta\right)^2\Bigg].\end{aligned} \tag{19.10.6}$$

The charged and strange meson masses are the same as before, but now the neutral non-strange mesons have a mass matrix

$$M_0^2 = 8v\begin{bmatrix} \frac{m_u+m_d}{2F_\pi^2} & \frac{m_u-m_d}{2\sqrt{3}F_\pi^2} & \frac{m_u-m_d}{\sqrt{6}F_\pi F_\zeta} \\ \frac{m_u-m_d}{2\sqrt{3}F_\pi^2} & \frac{m_u+m_d+4m_s}{6F_\pi^2} & \frac{m_u+m_d-2m_s}{3\sqrt{2}F_\pi F_\zeta} \\ \frac{m_u-m_d}{\sqrt{6}F_\pi F_\zeta} & \frac{m_u+m_d-2m_s}{3\sqrt{2}F_\pi F_\zeta} & \frac{m_u+m_d+m_s}{3F_\zeta^2} \end{bmatrix} \tag{19.10.7}$$

with rows and columns listed in the order π^0, η^0, and ζ.

In the limit where m_u and m_d vanish, this has two eigenvectors with eigenvalue zero:

$$u_a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad u_b = \frac{1}{\sqrt{F_\pi^2+2F_\zeta^2}}\begin{pmatrix} 0 \\ F_\pi \\ \sqrt{2}F_\zeta \end{pmatrix}. \tag{19.10.8}$$

In this orthonormal basis, the mass-squared matrix to first order in m_u and m_d is

$$m_{aa}^2 \simeq u_a^{\mathrm{T}} M_0^2 u_a = \frac{4v\,(m_u+m_d)}{F_\pi^2}, \tag{19.10.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

$$m_{bb}^2 \simeq u_b^{\mathrm{T}} M_0^2 u_b = \frac{12v\,(m_u + m_d)}{F_\pi^2 + 2F_\zeta^2}\,, \tag{19.10.10}$$

$$m_{ab}^2 = m_{ba}^2 \simeq \frac{\sqrt{3}v(m_u - m_d)}{2F_\pi\sqrt{F_\pi^2 + 2F_\zeta^2}}\,. \tag{19.10.11}$$

The effect of the off-diagonal term (19.10.11) is to decrease the product of eigenvalues by a negligible fractional amount $(m_u - m_d)^2/64(m_u + m_d)^2$, while of course leaving the sum of the eigenvalues unchanged, so the eigenvalues are given to a good approximation by the diagonal elements (19.10.9) and (19.10.10). Comparing Eq. (19.10.9) with Eq. (19.7.16), we see that the particle corresponding to the eigenvector u_a is the π^0. The other particle, corresponding to u_b, has a mass

$$m_b \simeq \sqrt{m_{bb}^2} \simeq \frac{\sqrt{3}m_\pi F_\pi}{\sqrt{F_\pi^2 + 2F_\zeta^2}} \leq \sqrt{3}m_\pi\,. \tag{19.10.12}$$

Thus a broken $U(1)_A$ symmetry would require a neutral pseudoscalar Goldstone boson with mass less than $\sqrt{3}m_\pi$, in addition to the pion itself. It hardly needs to be said that no such strongly-interacting particle exists.* We shall see in Section 23.5 that this problem was eventually solved by the discovery of non-perturbative effects that violate the extra $U(1)_A$ symmetry.

Problems

1. Apply the general theory of broken global symmetries to the case where an $SO(3)$ symmetry group (with generators t_1, t_2, t_3) is spontaneously broken to its $SO(2)$ subgroup (with generator t_3). How do the Goldstone boson fields transform under infinitesimal $SO(3)$ transformations? (Use the exponential parameterization of the coset space $SO(3)/SO(2)$.) Evaluate the covariant derivative $D_{a\mu}$ of the Goldstone boson field, and of a general field with a non-vanishing value q for the unbroken symmetry generator t_3. What is the most general $SO(3)$-invariant Lagrangian involving Goldstone bosons alone with no more than two derivatives? Use this Lagrangian to calculate the terms in the invariant amplitude for elastic scattering of

* Note that taking $F_\zeta \gg F_\pi$ would give the extra neutral scalar a very light mass, but its interactions would be so weak that it might have escaped detection. It seems unlikely that a very large ratio of F_π to F_ζ could arise in quantum chromodynamics, but something like this happens in theories with extra field variables that have been proposed to avoid parity violation by instantons; see Section 23.6.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

the Goldstone bosons to lowest order in their energy. What is the most general $SO(3)$-invariant Lagrangian with two non-Goldstone field factors and at most one derivative? What is the most general function of the Goldstone boson field alone (with no derivatives) that transforms like the three-component of an $SO(3)$ three-vector? Add this term to the Lagrangian, with a coefficient chosen to give the Goldstone boson a mass m, and recalculate the lowest order amplitude for Goldstone boson scattering.

2. Consider a theory with an $SU(N)$ global symmetry, spontaneously broken to $SU(N-1)$. Suppose we add a small symmetry-breaking perturbation, belonging to the defining representation N of $SU(N)$. Taking vacuum alignment into account, what symmetry group is completely unbroken? What about the case where the symmetry-breaking perturbation belongs to the adjoint representation of $SU(N)$?

3. Calculate the pion–pion scattering amplitude to one-loop order, taking account of a finite pion mass.

4. Derive the 'Adler sum rule,' the analog for the case of pion–pion scattering of the Adler–Weisberger sum rule.

5. Calculate the $SU(2) \times SU(2)$ transformation properties of the pion field ξ_a in the case where the cosets of $SU(2) \times SU(2)/SU(2)$ are parameterized as $\exp(i\xi_a x_a)$.

6. Use Eq. (19.7.31) and $SU(3)$ symmetry to derive the relation (19.7.32).

References

1. S. Coleman, 'Secret symmetry: an introduction to spontaneous symmetry breakdown and gauge fields,' in *Aspects of Symmetry: Selected Erice Lectures of Sidney Coleman* (Cambridge University Press, Cambridge, 1985).

2. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, Oxford, 1937): Section 1.4.

3. J. Goldstone. *Nuovo Cimento* **9**, 154 (1961).

4. Y. Nambu, *Phys. Rev. Lett.* **4**, 380 (1960).

5. J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127**, 965 (1962).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

6. S. Adler, *Phys. Rev.* **137**, B1022 (1965).

7. S. Weinberg, *Phys. Rev. Lett.*, **29**, 1698 (1972).

8. R. Dashen, *Phys. Rev.* **183**, 1245 (1969). For corrections to this relation, see J. F. Donoghue, B. R. Holstein, and D. Wyler, *Phys. Rev.* **D47**, 2089 (1993); J. Bijnens, *Phys. Lett.* **B306**, 343 (1993); K. Maltman and D. Kotchan, *Mod. Phys. Lett.*, **A5**, 2457 (1990); R. Urech, *Nucl. Phys.* **B433**, 234 (1995); R. Baur and R. Urech, Zurich–Karlsruhe preprint ZU-TH 22/95, TTP95-31, hep-ph/9508393 (1995).

9. For a review, see I. S. Towner *et al.*, Chalk River preprint nucl-th/9507005 (1995).

10. S. Weinberg, *Phys. Rev.* **112**, 1375 (1958).

11. M. L. Goldberger and S. Treiman, *Phys. Rev.* **111**, 354 (1966).

12. T. Ericson and W. Weise, *Pions and Nuclei* (Clarendon Press, Oxford, 1988).

13. T. Ericson *et al.*, CERN preprint CERN-TH/95-50 (1995).

14. T. Das, G. S. Guralnik, V. S. Mathur, F. E. Low, and J. E. Young, *Phys. Rev. Lett.* **18**, 759 (1967).

15. J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, *Nuovo Cimento* **17**, 757 (1960); M. Gell-Mann and M. Lévy, Ref. 24; K-C. Chou, *Soviet Physics JETP* **12**, 492 (1961).

16. Y. Nambu, *Phys. Rev. Lett.* **4**, 380 (1960).

17. Y. Nambu and G. Jona-Lasinio, *Phys. Rev.* **122**, 345 (1961).

18. Y. Nambu and D. Lurie, *Phys. Rev.* **125**, 1429 (1962); Y. Nambu and E. Shrauner, *Phys. Rev.* **128**, 862 (1962).

19. S. L. Adler, *Phys. Rev. Lett.* **14**, 1051 (1965); *Phys. Rev.* **140**, B736 (1965); W. I. Weisberger, *Phys. Rev. Lett.* **14**, 1047 (1965); *Phys. Rev.* **143**, 1302 (1965). The effect of a finite pion mass was included in a different way in Ref. 27, by combining the current algebra results for the pion–nucleon scattering lengths with the sum rule for the scattering lengths derived from the forward scattering dispersion relations by M. L. Goldberger, H. Miyazawa, and R. Oehme, *Phys. Rev.* **99**, 986 (1955); M. L. Goldberger, in *Dispersion Relations and Elementary Particles* (Wiley, New York, 1960): p. 146.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

20. S. Weinberg, 'Current Algebra – Rapporteur's Report' in *Proceedings of the International Conference on High-Energy Physics, Vienna, 1968* (CERN, Geneva, 1968): p. 253.

21. M. Gell-Mann, *Physics* **1**, 63 (1964).

22. S. Weinberg, *Phys. Rev. Lett.* **16**, 879 (1966).

23. S. Weinberg, *Phys. Rev. Lett.* **18**, 188 (1967).

24. M. Gell-Mann and M. Lévy, *Nuovo Cimento* **16**, 705 (1960).

25. S. Weinberg, *Phys. Rev.* **166**, 1568 (1968).

26. S. Weinberg, *Physica* **96A**, 327 (1979).

27. These scattering lengths were first calculated, using current algebra techniques, by S. Weinberg, *Phys. Rev. Lett.* **17**, 616 (1966).

28. For a review, see J. F. Donoghue, E. Golowich, and B. R. Holstein, *Dynamics of the Standard Model* (Cambridge University Press, Cambridge, 1992): Section VI-4. Measurements of the strong-interaction energy shift and width of the 1*s* state in $\pi^- p$ atoms have yielded results $a(\pi^- p \to \pi^- p) = \frac{2}{3}a_{1/2} + \frac{1}{3}a_{3/2} = 0.0885(9)m_\pi^{-1}$ and $a(\pi^- p \to \pi^0 n) = \frac{1}{3}\sqrt{2}(a_3 - a_1) = -0.136(10)m_\pi^{-1}$; see D. Sigg *et al.*, preprint ETHZ-IPP PR-95-4, July 1995, to be published in *Phys. Rev. Lett.*

29. S. Weinberg, *Phys. Lett.* **B251**, 288 (1990); *Nucl. Phys.* **B363**, 3 (1991); *Phys. Lett.* **B295**, 114 (1992). C. Ordóñez and U. van Kolck, *Phys. Lett.* **B291**, 459 (1992); C. Ordóñez, L. Ray, and U. van Kolck, *Phys. Rev. Lett.* **72**, 1982 (1994); U. van Kolck, *Phys. Rev.*, **C49**, 2932 (1994); U. van Kolck, J. Friar, and T. Goldman, to appear in *Phys. Lett. B*. This approach to nuclear forces is summarized in C. Ordóñez, L. Ray, and U. van Kolck, nucl-th/9511380, submitted to *Phys. Rev. C*; J. Friar, *Few-Body Systems Suppl.* **99**, 1 (1996). For application of these techniques to related nuclear processes, see T.-S. Park, D.-P. Min, and M. Rho, *Phys. Rep.* **233**, 341 (1993); nucl-th/9505017; S. R. Beane, C. Y. Lee, and U. van Kolck, *Phys. Rev.*, **C52**, 2915 (1995); T. Cohen, J. Friar, G. Miller, and U. van Kolck, nucl-th/9512036; D. B. Kaplan, M. Savage, and M. Wise, nucl-th/9605002.

30. S. Weinberg, Ref. 27. The pion–nucleon scattering lengths were independently calculated by Y. Tomozawa, *Nuovo Cimento* **46A**, 707 (1966).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

30*a*. S. L. Adler, *Phys. Rev.* **140**, B736 (1965).

31. S. Coleman, J. Wess, and B. Zumino, *Phys. Rev.* **177**, 2239 (1969) ; C. G. Callan, S. Coleman, J. Wess, and B. Zumino, *Phys. Rev.* **177**, 2247 (1969).

32. R. S. Palais, *J. Math. Mech.* **6**, 673 (1957); G. D. Mostow, *Annals of Math.* **65**, 432 (1957).

32*a*. For a discussion of other possibilities, see S. Weinberg, *Physica Scripta* **21**, 773 (1980).

33. M. Gell-Mann, Cal. Tech. Synchotron Laboratory Report CTSL–20 (1961), unpublished. This is reproduced along with other articles on $SU(3)$ symmetry in M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).

34. Y. Ne'eman, *Nucl. Phys.* **26**, 222 (1961).

35. M. Gell-Mann, R. J. Oakes, and B. Renner, *Phys. Rev.* **175**, 2195 (1968); S. Glashow and S. Weinberg, *Phys. Rev. Lett.* **20**, 224 (1968).

36. R. Dashen, *Phys. Rev.* **183**, 1245 (1969). Corrections were considered by P. Langacker and H. Pagels, *Phys. Rev.* **D8**, 4620 (1973). The absence of electromagnetic corrections to the π^0 mass had been proved in the model of ref. 17 by G. S. Guralnik, *Nuovo Cimento* **36**, 1002 (1965).

37. S. Weinberg, contribution to a *festschrift* for I. I. Rabi, *Trans. N. Y. Acad. Sci.* **38**, 185 (1977).

38. M. Gell-Mann, Ref. 33; S. Okubo, *Prog. Theor. Phys.*, **27**, 949 (1962).

38*a*. H. Leutwyler, Bern–CERN preprint CERN-TH/96-44, hep-ph/-9602366, to be published (1996).

38*b*. J. F. Donoghue, E. Golowich, and B. R. Holstein, Ref. 28.

39. J. Gasser and H. Leutwyler, *Nucl. Phys.* **B250**, 465 (1985). Also see J. Gasser and H. Leutwyler, *Ann. Phys.* **158**, 142 (1984).

40. For surveys, see H. Leutwyler, in *Proceedings of the XXVI International Conference on High Energy Nuclear Physics, Dallas, 1992*, ed. J. Sanford (American Institute of Physics, New York, 1993): 185; U. G. Meissner, *Rep. Prog. Phys.* **56**, 903 (1993); A. Pich, Valencia preprint FTUV/95-4, February 1995, to be published in *Reports on Progress in Physics*; J. Bijnens, G. Ecker, and J. Gasser, in *The Daphne Physics Handbook*, Vol. 1, eds. L. Maiani, G. Pancheri, and N. Paver (INFN, Frascati, 1995): Chapters 3 and 3.1; G. Ecker,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

preprint hep-ph/9501357, to be published in *Progress in Particle and Nuclear Physics*, Vol. 35 (Pergamon Press, Oxford).

41. S. Coleman and S. Glashow, *Phys. Rev. Lett.* **6**, 423 (1961); S. Okubo, *Phys. Lett.* **4**, 14 (1963).
42. J. Wess and B. Zumino, *Phys. Lett.* **37B**, 95 (1971).
43. E. Witten, *Nucl. Phys.* **B223**, 422 (1983).
44. R. Bott and R. Seely, *Comm. Math. Phys.* **62**, 235 (1978).
45. E. D'Hoker and S. Weinberg, *Phys. Rev.* **D50**, R6050 (1994).
46. *Encyclopedic Dictionary of Mathematics*, eds. S. Iyanaga and Y. Kawada (MIT Press, Cambridge, 1980); W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature and Cohomology*, Vol III, (Academic Press, New York, 1976); A. Borel, *Ann. Math.* (2) **57**, 115 (1953); also in A. Borel, *Collected Papers*, Vol I (Springer Verlag, Berlin, 1983).
47. D. Weingarten, *Phys. Rev. Lett.* **51**, 1830 (1983).
48. J. Preskill and S. Weinberg, *Phys. Rev.* **D24**, 1059 (1981). This was a modified version of an earlier argument by 't Hooft, who had considered the limit of large quark masses; see G. 't Hooft, lecture given at the Cargèse Summer Institute, 1979, in *Recent Developments in Gauge Theories*, eds. G 't Hooft *et al.* (Plenum, New York, 1980), reprinted in *Dynamical Gauge Symmetry Breaking*, eds. E. Farhi and R. Jackiw (World Scientific, Singapore, 1982), and in G. 't Hooft, *Under the Spell of the Gauge Principle* (World Scientific, Singapore, 1994). Also see Section 22.5.
49. C. Vafa and E. Witten, *Nucl. Phys.* **B234**, 173 (1984). Also see C. Vafa and E. Witten, *Commun. Math. Phys.* **95**, 257 (1984).
50. C. Vafa and E. Witten, *Phys. Rev. Lett.* **53**, 535 (1984).
51. S. L. Glashow, in *Hadrons and Their Interactions*, ed. A. Zichichi (Academic Press, New York, 1968); S. L. Glashow, R. Jackiw, and S-S. Shei, *Phys. Rev.* **187**, 1916 (1969); M. Gell-Mann, in *Proc. Third Topical Conf. on Part. Phys.*, eds. W. A. Simonds and S. F. Tuan (Western Periodicals, Los Angeles, 1970); in *Elementary Particle Physics* (Springer-Verlag, Bonn, 1972); *Acta Phys. Austriaca Suppl.* **IX**, 1972 (1972); H. Fritzsch and M. Gell-Mann, in *Proceedings of the XVI International Conference on High Energy Physics*, eds. J. D. Jackson and A. Roberts (Fermi National Accelerator Laboratory, Batavia, IL, 1972); *Phys. Lett.* **47B**, 365 (1973).
52. S. Weinberg, *Phys. Rev.* **D11**, 3583 (1975).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.007

20

Operator Product Expansions

We often find ourselves needing to know how an amplitude behaves when the four-momentum brought in by one operator and out by another goes to infinity, with all other external lines held at fixed four-momenta. For instance, we will see in Section 20.6 that the total cross section for the scattering of an electron by an initial hadron H, with arbitrary hadrons in the final state, is given by unitarity as a linear combination (with known coefficients) of the components of the amplitude

$$\int d^4x\, e^{-ik\cdot x} \langle H|J^\mu(x)J^\nu(0)|H\rangle ,$$

where k is the four-momentum transferred from the electron to the hadrons, and $J^\mu(x)$ is the electromagnetic current. In the case of deep inelastic electric scattering the momentum k that is carried in by one current operator and out by the other is allowed to go to infinity. Similarly, in studying the high momentum limit of various propagators and deriving corresponding spectral function sum rules in Section 20.5, we shall encounter the high momentum limit of similar Fourier transforms, but with the one-hadron state $|H\rangle$ replaced with the vacuum.

If an operator product such as $J^\mu(x)J^\nu(0)$ were analytic in x^μ, then its Fourier transform would decrease exponentially as the Fourier variable k goes to infinity. The leading terms in the high momentum limit of the Fourier transform arise from the singularities of the operator product as the spacetime arguments approach one another. The study of such operator products was initiated in 1969 by Wilson,[1] originally as an attempt to formulate a substitute for conventional quantum field theory. As has happened earlier (for instance with dispersion relations and Feynman's diagrammatic rules) the effort to bypass quantum field theory led to valuable general results, but results that can best be understood as general properties of quantum field theory.

The operator product expansion will be stated in Section 20.1. The standard proof of this expansion was given in perturbation theory in 1970 by Zimmerman.[2] In Section 20.1 we shall offer a non-perturbative and simpler though less rigorous derivation based on the path-integral

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

formulation of field theory. Section 20.2 will present a different perspective on the operator product expansion, in terms of the flow of large momenta through Feynman graphs, which will lead us to a perturbative proof.

There are several aspects of the operator product expansion that make it particularly useful for drawing consequences from theories like quantum chromodynamics. One such property, discussed in Section 20.3, is that the functions describing the singularities in this expansion have a momentum dependence governed by renormalization group equations, so that in asymptotically free theories they can be calculated at large momenta using perturbation theory. Another aspect, shown in Section 20.4, is that these functions exhibit the full symmetry of the underlying theory, unaffected by possible spontaneous symmetry breaking. Applications will be considered in Sections 20.5 and 20.6.

20.1 The Expansion: Description and Derivation

Wilson[1] hypothesized that the singular part as $x \to y$ of the product $A(x)B(y)$ of two operators is given by a sum over other local operators

$$A(x)B(y) \to \sum_C F_C^{AB}(x-y)\, C(y)\,, \tag{20.1.1}$$

where $F_C^{AB}(x-y)$ are singular c-number functions. Dimensional analysis suggests that $F_C^{AB}(x-y)$ behaves for $x \to y$ like the power $d_C - d_A - d_B$ of $x-y$, where d_O is the dimensionality of the operator O in powers of mass or momentum. Since d_O increases as we add more fields or derivatives to an operator O, the strength of the singularity of $F_C^{AB}(x-y)$ decreases for operators C of increasing complexity. The remarkable thing about the operator product expansion is that it is an *operator* relation; that is, in applying it to any matrix element $\langle\beta|A(x)B(y)|\alpha\rangle$, we get the same functions $F_C^{AB}(x-y)$ for all states $|\alpha\rangle$ and $|\beta\rangle$.

It is the decrease of the singularity in Eq. (20.1.1) with operators $C(y)$ of increasing complexity that makes this expansion useful in drawing conclusions about the behavior of the product $A(x)B(y)$ for $x \to y$. The simple power-counting argument above is modified by renormalization effects; the expansion (20.1.1) must be formulated in terms of operators renormalized at some scale μ, and then μ appears along with $x-y$ in the coefficient function $F_C^{AB}(x-y)$. We shall see in Section 20.3 that in asymptotically free theories $F_C^{AB}(x-y)$ does behave like the power $d_C - d_A - d_B$ of $x-y$ suggested by dimensional analysis only up to a power of $\ln(x-y)^2$. Even in more general theories, it is plausible that the singularities associated with various operators $C(y)$ will decrease with the complexity of the operators.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

The corresponding statement in momentum space is that for $k \to \infty$,

$$\int d^4x\, e^{-ik\cdot x} A(x)B(0) \to \sum_C V_C^{AB}(k)\, C(0) \tag{20.1.2}$$

and correspondingly

$$\int d^4x\, e^{-ik\cdot x} T\{A(x)B(0)\} \to \sum_C U_C^{AB}(k)\, C(0)\,, \tag{20.1.3}$$

where $V_C^{AB}(k)$ and $U_C^{AB}(k)$ are functions of k^μ that for large k decrease increasingly rapidly for more and more complicated terms in the series.

We are going to derive a generalized version of the Wilson expansion, in which the momenta carried by any number of operators go to infinity together. For this purpose, let us consider a Greens function for local operators $A_1(x_1)$, $A_2(x_2)$, etc. whose arguments approach a point x, as well as other local operators $B_1(y_1)$, $B_2(y_2)$, etc. with fixed arguments:

$$\begin{aligned}&\langle T\{A_1(x_1), A_2(x_2), \ldots B_1(y_1), B_2(y_2), \ldots\}\rangle_0 \\ &\quad = \int \left[\prod_{\ell, z} d\phi_\ell(z)\right] a_1(x_1)\, a_2(x_2) \cdots b_1(y_1)\, b_2(y_2) \cdots \exp(iI[\phi])\,,\end{aligned} \tag{20.1.4}$$

where the lower-case letters a and b indicate replacement of the field operators in the As and Bs with the c-number fields ϕ. Now surround the point x with a ball $B(R)$ of radius R which is much larger than the separations among the x_1, x_2, etc. but much smaller than the separations between x, y_1, y_2, etc. Since the action is local, it may be written as

$$I = \int_{z\in B(R)} d^4z\, \mathscr{L}(z) + \int_{z\notin B(R)} d^4z\, \mathscr{L}(z)\,. \tag{20.1.5}$$

Eq. (20.1.4) may then be put in the form

$$\begin{aligned}&\langle T\{A_1(x_1), A_2(x_2), \ldots B_1(y_1), B_2(y_2), \ldots\}\rangle_0 \\ &= \int \left[\prod_{z\notin B(R),\,\ell} d\phi_\ell(z)\right] b_1(y_1)\, b_2(y_2) \ldots \exp\left(i\int_{z\notin B(R)} d^4z\, \mathscr{L}(z)\right) \\ &\quad \times \int \left[\prod_{z\in B(R),\,\ell} d\phi_\ell(z)\right] a_1(x_1)\, a_2(x_2) \ldots \exp\left(i\int_{z\in B(R)} d^4z\, \mathscr{L}(z)\right)\,,\end{aligned} \tag{20.1.6}$$

in which the path integral over the fields inside the ball is constrained by the boundary condition that the fields merge smoothly at the ball's surface with the fields outside the ball. Aside from this boundary condition, the path integral over the fields inside the ball is completely unaffected by the behavior of the field outside the ball, so the integral over the fields inside

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

the ball may be expressed in terms of the values and derivatives of the fields on the surface of the ball, which in turn may be expressed in terms of the fields and their derivatives extrapolated from outside the ball to the interior point x. If we express this integral as a series in products* $o(x)$ of the c-number fields and their derivatives extrapolated to x, then the coefficients can only be functions $U_O^{A_1,A_2,\ldots}(x_1 - x, x_2 - x, \ldots)$ of the coordinate differences. Since the points y_1, y_2, etc. are all far outside the ball $B(R)$, the exclusion of this ball in the action for the fields outside the ball has no effect in the limit $R \to 0$, so in this limit Eq. (20.1.6) becomes

$$\begin{aligned}\langle T\{A_1(x_1), A_2(x_2), \ldots B_1(y_1), B_2(y_2), \ldots\}\rangle_0 \to \int \left[\prod_{\ell, z} d\phi_\ell(z)\right] \\ \times b_1(y_1)\, b_2(y_2) \ldots \exp\left(i \int d^4z \mathscr{L}(z)\right) \\ \times \sum_O U_O^{A_1,A_2,\ldots}(x_1 - x, x_2 - x, \ldots)\ o(x) \\ = \sum_O U_O^{A_1,A_2,\ldots}(x_1 - x, x_2 - x, \ldots)\, \langle T\{O(x), B_1(y_1), B_2(y_2), \ldots\}\rangle_0\end{aligned} \tag{20.1.7}$$

for x_1, x_2, etc. all approaching x, where $O(x)$ is the quantum-mechanical Heisenberg-picture operator corresponding to $o(x)$. In particular, by Fourier transforming with respect to the y variables and multiplying with appropriate coefficient functions, this yields

$$\langle\beta|T\{A_1(x_1), A_2(x_2), \ldots\}|\alpha\rangle \to \sum_O U_O^{A_1,A_2,\ldots}(x_1 - x, x_2 - x, \ldots)\, \langle\beta|O(x)|\alpha\rangle \tag{20.1.8}$$

for arbitrary states $|\alpha\rangle$ and $\langle\beta|$. Because this applies for arbitrary states, it is the operator product expansion in a generalized version:

$$T\{A_1(x_1), A_2(x_2), \ldots\} \to \sum_O U_O^{A_1,A_2,\ldots}(x_1 - x, x_2 - x, \ldots)\, O(x)\,. \tag{20.1.9}$$

20.2 Momentum Flow*

We shall now consider the simplest example of the operator product expansion: the asymptotic behavior of an $(n+2)$-point Feynman amplitude

* In order for the coefficients in this series to be finite, these products must be renormalized by multiplication with suitable infinite factors. This will be made clearer in the derivation presented in the next section.

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

in the theory of a single real scalar field $\phi(x)$ with mass m and interaction Lagrangian density $-\frac{1}{24}g\phi^4$, when a large four-momentum k flows in one line and out another, with all other external lines held at fixed momentum. This will lead us to a perturbative proof of the leading term in the operator product expansion in this case, but our real purpose here will be to gain some insight into the way that the flow of large momenta through Feynman diagrams leads to this expansion. An appendix to this chapter will discuss the extension of these results to the general case.

Let us define $\Gamma(k;p_1\cdots p_n)$ as the sum of all connected graphs for the n-particle scattering amplitude, whose external lines carry incoming momenta k, $p-k$, and outgoing momenta $p_1\cdots p_n$, where $p\equiv p_1+p_2+\cdots+p_n$. (It will be convenient to specify further that $\Gamma(k;p_1\cdots p_n)$ includes propagators for the external lines with momenta k and $p-k$, but not for those with the fixed momenta $p_1\cdots p_n$.) We wish to show that in any finite order of perturbation theory, for $k\to\infty$,

$$\Gamma(k;p_1\cdots p_n)\to U_{\phi^2}(k)F_{\phi^2}(p_1\cdots p_n)+O(k^{-5})\,, \tag{20.2.1}$$

where $U_{\phi^2}(k)$ is a sum of terms of order** k^{-4}, that is independent of $p_1\cdots p_n$ and of n, and $F_{\phi^2}(p_1\cdots p_n)$ is the amplitude for n ϕ lines with insertion of a single ϕ^2 vertex, times a suitable renormalization constant Z_{ϕ^2} to make it finite. Because $F_{\phi^2}(p_1\cdots p_n)$ is a matrix element of the renormalized operator $(\phi^2)_R\equiv Z_{\phi^2}\phi^2(0)$, Eq. (20.2.1) corresponds to the statement that the leading term in the operator product expansion for $k\to\infty$ is†

$$\int d^4x\, e^{-ik\cdot x}\, T\{\phi_R(x)\phi_R(0)\}_C\to U_{\phi^2}(k)\,(\phi^2(0))_R\,. \tag{20.2.2}$$

In assessing the asymptotic behavior of Feynman amplitudes, we must take account of the fact that there are parts of the range of integration in momentum space where some of the internal lines carry momenta of the same order as the external line momenta that are going to infinity, while other internal lines do not. The contribution to $\Gamma(k;p_1\cdots p_n)$ from the part of the region of integration where the lines in some subgraph $\mathcal{S}$ have momenta of order k has an asymptotic behavior of order $k^{D_{\mathcal{S}}}$, where $D_{\mathcal{S}}$

** Throughout this chapter, whenever it is said that an amplitude is of order k^A, it should be understood that for $k^\rho=\kappa n^\rho$ where $\kappa\to\infty$ with n^ρ a fixed generic four-vector, the amplitude approaches κ^A times a sum of powers of $\ln\kappa$.

† The subscript C on the time-ordered product here indicates that we are including only connected graphs. We have dropped the superscript $\phi\phi$ on V_{ϕ^2}, because in this section we are concerned only with the operator product expansion for two ϕ fields. These fields are themselves renormalized, because we implicitly include counterterms along with radiative corrections for the propagators of the lines carrying momenta k and $p-k$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

is the dimensionality of the subgraph $\mathscr{S}$, calculated according to the rules of Section 12.1. If $\mathscr{S}$ has m external lines that connect it to the rest of the graph as well as the two external lines with momenta k and $p-k$, then from Eq. (12.1.8) we have $D_{\mathscr{S}} = 4-2-m-4 = -2-m$. (The term -4 arises from the two propagators of the lines with momenta k and $p-k$, which we have specified are to be included in $\Gamma(k;p_1\cdots p_n)$.) *Thus the asymptotic behavior of* Γ *is dominated by the part of the momentum-space integral where the large momentum k flows either through the whole graph or through some subgraph, whichever has the smallest number of external lines.*[3]

For $n=0$ this is always the whole graph; that is, the dominant part of the integral comes from the part of the region of integration where every line carries a momentum of order k, giving an asymptotic behavior of order k^{-2}. In this case the only operator in the operator product expansion contributing to this matrix element is the unit operator, $C=1$. This term is excluded here for $n>0$ because for the present we are limiting ourselves to connected graphs.

For $n=2$ the dominant contribution comes both from the whole graph, *and* from subgraphs in which the two external lines with momenta k and $p-k$ are connected to the other two external lines by a bridge consisting of two internal lines,[††] giving an asymptotic behavior of order k^{-4}. For $n\geq 4$ the dominant contribution comes *only* from subgraphs in which the two external lines with momenta k and $p-k$ are connected to the $n>2$ other external lines by a bridge consisting of two internal lines, again giving an asymptotic behavior of order k^{-4}.

The analysis of the cases $n=2$ and $n\geq 4$ is complicated by the fact that a general graph may contain several of these two-particle bridges. Let us first consider the case $n=2$. We define $I(k,k',p)$ as the sum of all graphs contributing to $\Gamma(k;p_1,p_2)$ (with $p_1=k'$, $p_2=p-k'$) that are two-particle-irreducible, in the sense that the two external lines with incoming momenta k and $p-k$ cannot be disconnected from the two external lines with momenta k' and $p-k'$ by cutting through any pair of internal lines. Then $\Gamma(k;k',p-k')-I(k,k',p)$ consists of graphs that can be disconnected in this way, and may therefore be written (see Figure 20.1):

$$\Gamma(k;k',p-k')-I(k,k',p)=\int d^4k''\, I(k,k'',p)\Gamma(k'';k',p-k')\,. \qquad (20.2.3)$$

(Like $\Gamma(k;k',p-k')$, the kernel $I(k,k',p)$ includes propagators for the lines

[††] The possibility $m=1$ is excluded because the symmetry of this theory under $\phi\to-\phi$ rules out graphs or subgraphs with odd numbers of external lines. The possibility $m=0$ is excluded because Γ is defined to arise only from connected graphs.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

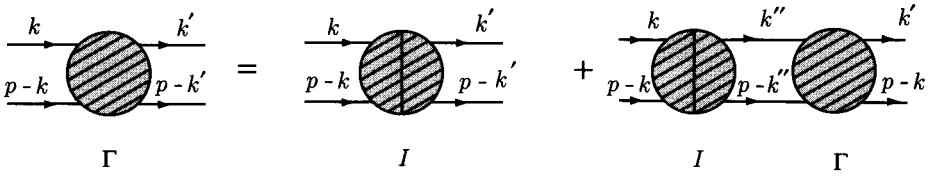

Figure 20.1. Diagrammatic representation of the integral equation (20.2.3). The cross-hatched disks marked Γ represent the sum of all connected Feynman diagrams with the indicated external lines, while the cross-hatched disks marked I, which are divided by a vertical line, represent the sum of all connected diagrams for which the lines on the left cannot be separated from those on the right by cutting through any pair of external lines.

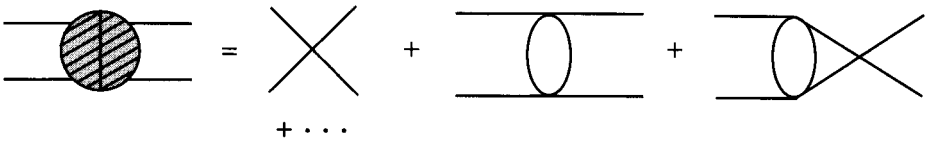

Figure 20.2. Tree and one-loop graphs for the kernel $I(k,\ k',\ p)$ in the theory of a scalar field with interaction Φ^4.

with momenta k and $p-k$, but, to avoid double counting, not for the lines with momenta k' and $p-k'$.) The Feynman diagrams for $I(k,k',p)$ to order g^2 are shown in Figure 20.2.

To evaluate the behavior of the right-hand side of Eq. (20.2.3), let's first consider the asymptotic behavior of the kernel $I(k,k',p)$ when $k \to \infty$ with k' and p fixed. This is dominated by the region of momentum space in which all internal lines carry momenta of order k, which makes a contribution of order k^{-4}, because in any other region the subgraph consisting of internal lines that carry momenta of order k would have more than four external lines, and hence would make a contribution that decreases faster than k^{-4}. It follows that differentiating $I(k,k',p)$ with respect to k' or p would reduce the asymptotic behavior of this kernel by a factor of k^{-1}. Hence for $k \to \infty$ with k' and p fixed, we have

$$I(k,k',p) \to I_\infty(k)\,, \tag{20.2.4}$$

where $I_\infty(k)$ is a function only of k, of order k^{-4}.

Unfortunately, we cannot simply replace $I(k,k',p)$ in Eq. (20.2.3) with this asymptotic limit, because however large k may become, the integral will receive a large contribution from values of k' of order k. To deal with this complication, we shall now employ a trick, based on mathematical induction. In lowest order $\Gamma(k;k',p-k')$ is given by a single vertex with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

two attached bare propagators

$$\Gamma(k;k',p-k') = \frac{ig}{(2\pi)^4(k^2+m^2)((p-k)^2+m^2)} \qquad \text{(lowest order)}\,,$$

for which it is easy to verify an asymptotic behavior of form Eq. (20.2.1). Let us therefore assume that Eq. (20.2.1) for $n = 2$ holds up to some given order N in g — that is, that up to this order, the asymptotic behavior for $k \to \infty$ takes the form

$$\Gamma(k;k',p-k) \to U_{\phi^2}(k)F_{\phi^2}(k',p-k') + O(k^{-5})\,, \tag{20.2.5}$$

and try to verify this behavior to the next order. In order to eliminate the contribution to the integral in Eq. (20.2.3) from values of k' of order k, we rewrite Eq. (20.2.3) as

$$\begin{aligned}\Gamma(k;k',p-k) &= I(k,k',p)\\ &+\int d^4k''\, I(k,k'',p)\left[\Gamma(k'';k',p-k') - U_{\phi^2}(k'')F_{\phi^2}(k',p-k')\right]\\ &+F_{\phi^2}(k',p-k')\int d^4k''\, I(k,k'',p)U_{\phi^2}(k'')\,.\end{aligned} \tag{20.2.6}$$

Since $I(k,k',p)$ is at least of first order, we may use Eq. (20.2.5) in the right-hand side of Eq. (20.2.6). Hence in the second term on the right-hand side, the part of the region of integration where k'' is of order k gives a contribution that vanishes like k^{-4+4-5}, and may therefore be neglected compared with the part where k'' remains finite, which yields the convergent integral

$$I_\infty(k)\int d^4k''\left[\Gamma(k'';k',p-k') - U_{\phi^2}(k'')F_{\phi^2}(k',p-k')\right]\,.$$

Further, since the dominant part of $\int d^4k''\, I(k,k'',p)U_{\phi^2}(k'')$ comes from the region of integration where every internal line of the graphs for $I(k,k'',p)$ carries a momentum of order k, differentiating this integral with respect to p would lower its asymptotic behavior by a factor of order k^{-1}, so asymptotically $I(k,k',p)$ may be replaced in this integral with $I(k,k') \equiv I(k,k',0)$. Therefore for $k \to \infty$, Eq. (20.2.6) becomes

$$\begin{aligned}\Gamma(k;k',p-k') &\to F_{\phi^2}(k',p-k')\int d^4k''\, I(k,k'')U_{\phi^2}(k'')\\ &+I_\infty(k)\left\{1+\int d^4k''\,\left[\Gamma(k'';k',p-k') - U_{\phi^2}(k'')F_{\phi^2}(k',p-k')\right]\right\}.\end{aligned} \tag{20.2.7}$$

Let us therefore define $U_{\phi^2}(k)$ and $F_{\phi^2}(k',p-k')$ to order $N+1$ in g in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

terms of these functions in lower orders of perturbation theory, by

$$U_{\phi^2}(k) = CI_\infty(k) + \int d^4k'\, I(k,k')U_{\phi^2}(k')\,, \tag{20.2.8}$$

$$F_{\phi^2}(k', p-k') = C^{-1} \times \left\{1 + \int d^4k'' \left[\Gamma(k'';k',p-k') - U_{\phi^2}(k'')F_{\phi^2}(k',p-k')\right]\right\}\,, \tag{20.2.9}$$

where C is a constant that can be chosen as we like. With these definitions, Eq. (20.2.5) follows from Eq. (20.2.7).

It will be convenient to choose the constant C so that $F^2_\phi(k', p-k')$ has the value unity at some renormalization point $k' = k(\mu)$ and $p = p(\mu)$, where $k(\mu)$ and $p(\mu)$ are standard four-momenta of order μ. Then

$$C = 1 + \int d^4k''\Gamma(k'';k(\mu),p(\mu)-k(\mu)) - \int d^4k''\, U_{\phi^2}(k'')\,. \tag{20.2.10}$$

Using Eq. (20.2.5), we see that the divergences in the two integrals in Eq. (20.2.10) cancel.

Eq. (20.2.9) may now be written

$$F_{\phi^2}(k', p-k') = Z_{\phi^2}\left\{1 + \int d^4k''\,\Gamma(k'';k',p-k')\right\}\,, \tag{20.2.11}$$

where

$$Z_{\phi^2} = \left[1 + \int d^4k''\,\Gamma(k'';k(\mu),p(\mu)-k(\mu))\right]^{-1}\,. \tag{20.2.12}$$

We can think of Z_{ϕ^2} as the renormalization constant for the composite operator ϕ^2, defined in such a way that the operator $Z_{\phi^2}\phi^2$ has the finite two-particle matrix element $F_{\phi^2}(k, p-k)$, with the value unity for $k = k(\mu)$ and $p = p(\mu)$.

It is not particularly convenient to calculate $U_{\phi^2}(k)$ or $F_{\phi^2}(k, p-k)$ by using Eqs. (20.2.8) and (20.2.11). Rather, it is simpler to calculate $\Gamma(k;k',p-k')$, and read off $U_{\phi^2}(k)$ or $F_{\phi^2}(k,p-k)$ by comparison with Eq. (20.2.5). By multiplying the function (12.2.26) with the product of propagators for the lines with momenta k and $p-k$, we see that, to one-loop order

$$\Gamma(k;k',p-k') = \left[\frac{-i}{(2\pi)^4(k^2+m^2)}\right]\left[\frac{-i}{(2\pi)^4((p-k)^2+m^2)}\right]\left[-i(2\pi)^4\, g\right] \times\left\{1 - \frac{g}{32\pi^2}\int_0^1 dx\left\{\ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2-sx(1-x)}\right) + \ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2-tx(1-x)}\right) + \ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2-ux(1-x)}\right) + \cdots\right\}\right., \tag{20.2.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

where s, t, and u are the Mandelstam variables

$$s = -p^2\,, \qquad t = -(k-k')^2\,, \qquad u = -(p-k-k')^2\,,$$

μ is the renormalization scale and g the corresponding renormalized coupling, defined as the Feynman amplitude at $s = t = u = -4\mu^2/3$. This has the asymptotic behavior

$$\Gamma(k;k',p-k') \to \frac{ig}{(2\pi)^4(k^2)^2} \times\left\{1 - \frac{g}{32\pi^2}\int_0^1 dx\left\{\ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2+p^2x(1-x)}\right) +2\ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2+k^2x(1-x)}\right)+\cdots\right\}\right. . \tag{20.2.14}$$

To order g^2, this agrees with Eq. (20.2.5) if we take

$$U_{\phi^2}(k) = \frac{ig}{(2\pi)^4(k^2)^2} \times\left\{1 - \frac{g}{16\pi^2}\int_0^1 dx\ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2+k^2x(1-x)}\right)+\cdots\right\}, \tag{20.2.15}$$

and

$$F_{\phi^2}(k,p-k) = 1-\frac{g}{32\pi^2}\int_0^1 dx\,\ln\left(\frac{m^2+4x(1-x)\mu^2/3}{m^2+p^2x(1-x)}\right)+\cdots\,. \tag{20.2.16}$$

We have here chosen the renormalization point $k(\mu), p(\mu)$ for the operator ϕ^2 to be related to that for the coupling constant g in such a way that $p(\mu)^2 = 4\mu^2/3$, so that $F_{\phi^2}(k, p-k) = 1$ when $p^2 = 4\mu^2/3$.

Now consider the case where the number n of external lines with fixed momenta is greater than two. In accordance with our earlier discussion, in the limit $k \to \infty$, the leading graphs for $\Gamma(k; p_1 \cdots p_n)$ are those in which the two external lines carrying momenta k and $p-k$ can be disconnected from the n external lines with fixed momenta by cutting through a pair of internal lines:

$$\Gamma(k;p_1\cdots p_n) \to \int d^4k'\, I(k,k',p)\,\Gamma(k';p_1\cdots p_n)\,. \tag{20.2.17}$$

As before, we cannot simply use the asymptotic limit of the kernel $I(k,k',p)$ for $k \to \infty$ on the right-hand side, because this integral receives important contributions from k' of order k. We deal with this complication by

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

rewriting Eq. (20.2.17) in the form

$$\Gamma(k;p_1\cdots p_n)\to\int d^4k'\,I(k,k',p)\left[\Gamma(k';p_1\cdots p_n)-U_{\phi^2}(k')F_{\phi^2}(p_1\cdots p_n)\right]$$
$$+F_{\phi^2}(p_1\cdots p_n)\int d^4k'\,I(k',k,p)\,U_{\phi^2}(k')\,, \tag{20.2.18}$$

where by mathematical induction, we suppose that up to some given order N

$$\Gamma(k;p_1\cdots p_n)\to U_{\phi^2}(k)F_{\phi^2}(p_1\cdots p_n)\,, \tag{20.2.19}$$

with correction terms of order $1/k^5$. We can now use the limit (20.2.4) together with Eq. (20.2.8) to rewrite Eq. (20.2.18) as

$$\Gamma(k;p_1\cdots p_n)\to I_\infty(k)\int d^4k'\left[\Gamma(k';p_s\cdots p_n)-U_{\phi^2}(k')F_{\phi^2}(p_1\cdots p_n)\right]$$
$$+F_{\phi^2}(p_1\cdots p_n)\left[U_{\phi^2}(k)-CI_\infty(k)\right]\,.$$

This agrees with Eq. (20.2.19) to order $N+1$, provided we take

$$CF_{\phi^2}(p_1\cdots p_n)=\int d^4k'\,\left[\Gamma(k';p_s\cdots p_n)-U_{\phi^2}(k')F_{\phi^2}(p_1\cdots p_n)\right]$$

or, using Eqs. (20.2.10) and (20.2.12),

$$F_{\phi^2}(p_1\cdots p_n)=Z_{\phi^2}\int d^4k'\,\Gamma(k';p_1\cdots p_n)\,. \tag{20.2.20}$$

This just says that $F_{\phi^2}(p_1\cdots p_n)$ is the matrix element of the renormalized operator $Z_{\phi^2}\phi^2$, so that Eq. (20.2.1) corresponds to the operator product formula (20.2.2). Note in particular that $U_{\phi^2}(k)$ is the same coefficient function whatever value n or the momenta $p_1\cdots p_n$ may take, as was to be shown.

Strictly speaking, the $(\phi^2)_R$ operator is not the leading term in the expansion of the product of two ϕs. There is also the operator unity, which has lower dimensionality than $(\phi^2)_R$, but was excluded from Eq. (20.2.2) because (as indicated by the subscript C) we are excluding disconnected graphs. As mentioned earlier, the graphs for $\Gamma(k)$ with $n=0$ (and $p^\mu=0$) are dominated for $k\to\infty$ by the region of integration where all internal lines carry momenta of order k, and $U_1(k)$ is the contribution of this region.

Higher-dimensional operators and more general theories will be considered in an appendix to this chapter.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

20.3 Renormalization Group Equations for Coefficient Functions

As mentioned earlier, one of the aspects of the operator product expansion that makes it so useful is that the momentum dependence of the coefficient functions is governed by renormalization group equations. This is because they arise from limiting values of sums of Feynman graphs (such as $I_\infty(k)$ in Section 20.2) in which all relevant momenta are going to infinity together, so that masses may be set equal to zero without introducing singularities. Nevertheless the operator product coefficient functions do not obey simple scaling laws, because of renormalization effects: the functions are multiplied with scale-dependent renormalization constants, and depend on scale-dependent renormalized couplings.

Consider the operator product expansion for a Greens function $\Gamma_{\ell\ell'}(k,k',p)$ in which the incoming momenta (collectively labelled k, with sum p) of a set ℓ of lines all go to infinity together, with the set ℓ' of remaining lines having fixed outgoing momenta (collectively labelled k', with sum p):

$$\Gamma_{\ell\ell'}(k,k',p) \to \sum_{\mathcal{O}} U^{\ell}_{\mathcal{O}}(k) F_{\mathcal{O},\ell'}(k',p)\,. \tag{20.3.1}$$

The function $F_{\mathcal{O},\ell'}(k',p)$ is the matrix element of a renormalized operator $\mathcal{O}_R = \sum_{\mathcal{O}'} Z_{\mathcal{O},\mathcal{O}'}\mathcal{O}'$, so its coefficient $U^{\ell}_{\mathcal{O}}(k)$ in the product of fields corresponding to the lines ℓ is proportional to $Z^{-1}_{\mathcal{O},\mathcal{O}'}$, as well as to $Z_{\ell',\ell}$, the direct product of all the renormalization matrix factors for the field (or composite) operators in the set ℓ. Hence

$$\mu\frac{d}{d\mu}U^{\ell}_{\mathcal{O}} = \sum_{\ell'}\gamma_{\ell\ell'}U^{\ell'}_{\mathcal{O}} - \sum_{\mathcal{O}'}U^{\ell}_{\mathcal{O}'}\gamma_{\mathcal{O}',\mathcal{O}} + \beta(g)\frac{\partial}{\partial g}U^{\ell}_{\mathcal{O}}\,, \tag{20.3.2}$$

where

$$\mu\frac{\partial}{\partial\mu}Z_{\ell\ell'} = \sum_{\ell''}\gamma_{\ell\ell''}Z_{\ell''\ell'}\,, \qquad \mu\frac{\partial}{\partial\mu}Z_{\mathcal{O}\mathcal{O}'} = \sum_{\mathcal{O}''}\gamma_{\mathcal{O}\mathcal{O}''}Z_{\mathcal{O}''\mathcal{O}'}\,, \tag{20.3.3}$$

and for simplicity we assume a single renormalizable coupling g_μ, defined as the value of some Feynman amplitude at a renormalization point with momenta of order μ, with $\mu\, dg_\mu/d\mu = \beta(g_\mu)$. In order to be able to use dimensional analysis, we multiply all operators by powers of μ in such a way that they become dimensionless. This has the effect that the components of the Z and γ matrices are also dimensionless, with the values in the limit of zero coupling given by

$$\gamma_{\ell\ell'} \to \delta_{\ell\ell'}N(\ell)\,, \qquad \gamma_{\mathcal{O}\mathcal{O}'} \to \delta_{\mathcal{O}\mathcal{O}'}N(\mathcal{O})\,, \tag{20.3.4}$$

where $N(\mathcal{O})$ is the dimensionality of the operator $\mathcal{O}$, and $N(\ell)$ is the total dimensionality of the set ℓ of fields (the sum of $s+1$ for each field, where

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

$s = 0$ for scalar and massless gauge fields, $s = 1/2$ for Dirac fields, etc.) Also, dimensional analysis tells us that for $k^\mu = \kappa n^\mu$ with n^μ fixed, the amplitude can depend on κ only through the ratio κ/μ, aside from a factor $\kappa^{4-4n(\ell)}$ arising from the integrals used in defining the Fourier transform. The solution of Eq. (20.3.2) is then of the form

$$U^\ell_{\mathcal{O}}(\kappa n) = \kappa^{4-4n(\ell)} \sum_{\ell',\mathcal{O}'} \left[M\left\{ \exp\left(\int^\kappa \frac{d\mu}{\mu} \gamma(g_\mu) \right) \right\} \right]_{\ell,\ell'} \mathcal{U}^{\ell'}_{\mathcal{O}'}(g_\kappa, n) \times \left[M\left\{ \exp\left(\int^\kappa \frac{d\mu}{\mu} \gamma(g_\mu) \right) \right\} \right]^{-1}_{\mathcal{O}',\mathcal{O}}, \tag{20.3.5}$$

where M denotes the 'μ-ordered' product, that is, each term in the expansion of the exponentials rearranged so that the factors are in the order of decreasing μ from left to right.

This gives especially simple results where g_μ approaches a fixed point g_* for $\mu \to \infty$. The contribution of large μs to $M\{\exp(\int^\kappa \gamma(g_\mu) d\mu/\mu)\}$ is then a matrix factor $\kappa^{\gamma(g_*)}$ that because of the μ-ordering appears on the *left*. Thus Eq. (20.3.5) becomes

$$U^\ell_{\mathcal{O}}(\kappa n) = \kappa^{4-4n(\ell)} \sum_{\ell',\mathcal{O}'} \left[\kappa^{\gamma(g_*)} \right]_{\ell,\ell'} \mathcal{C}_{\ell',\mathcal{O}'} \left[\kappa^{-\gamma(g_*)} \right]_{\mathcal{O}',\mathcal{O}}, \tag{20.3.6}$$

where $\mathcal{C}$ is either a constant or a sum of powers of $\ln \kappa$, depending on the rate at which g_κ approaches g_*.

Asymptotically free theories like quantum chromodynamics are a special case of particular physical interest. Here the fixed point is at $g_* = 0$, and, according to Eq. (20.3.4), near this fixed point the γ matrices go as

$$\gamma(g)_{\ell\ell'} \to N(\ell)\delta_{\ell\ell'} + g^2 c_{\ell\ell'}, \quad \gamma(g)_{\mathcal{O}\mathcal{O}'} \to N(\mathcal{O})\delta_{\mathcal{O}\mathcal{O}'} + g^2 c_{\mathcal{O}\mathcal{O}'}. \tag{20.3.7}$$

Also, if we write the renormalization group equation for the coupling in the form

$$\mu \frac{d}{d\mu} g^2_\mu = -\frac{b}{8\pi^2} g^4_\mu, \tag{20.3.8}$$

then

$$\int^\kappa \frac{d\mu}{\mu} g^2_\mu \to -\frac{8\pi^2}{b} \ln g^2_\kappa + \text{constant}.$$

Using this in Eq. (20.3.5) yields the asymptotic behavior

$$U^\ell_{\mathcal{O}}(\kappa n) \to \kappa^{4-4n(\ell)+N(\ell)-N(\mathcal{O})} \sum_{\ell',\mathcal{O}'} \left[\left(g^2_\kappa\right)^{-8\pi^2 c/b} \right]_{\ell,\ell'} \mathcal{C}^{\ell'}_{\mathcal{O}'} \left[\left(g^2_\kappa\right)^{8\pi^2 c/b} \right]_{\mathcal{O}',\mathcal{O}}, \tag{20.3.9}$$

where $\mathcal{C}^{\ell'}_{\mathcal{O}'}$ is a constant matrix, equal to $\mathcal{U}^{\ell'}_{\mathcal{O}'}(0, n)$ times constant factors that are not calculable in perturbation theory because they come from the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

parts of the integrals in Eq. (20.3.5) where g_μ is not small. For $\kappa \to \infty$, the behavior of the coupling constant is $g_\kappa^2 \to 8\pi^2/b \ln \kappa$, so Eq. (20.3.9) may also be written

$$U_{\mathcal{O}}^{\ell}(\kappa n) \to \kappa^{4-4n(\ell)+N(\ell)-N(\mathcal{O})} \sum_{\ell',\mathcal{O}'} \left[\left(\ln \kappa \right)^{8\pi^2 c/b} \right]_{\ell,\ell'} \mathscr{B}_{\mathcal{O}'}^{\ell'} \left[\left(\ln \kappa \right)^{-8\pi^2 c/b} \right]_{\mathcal{O}',\mathcal{O}} , \tag{20.3.10}$$

where $\mathscr{B}_{\mathcal{O}'}^{\ell'}$ is another constant matrix. The condition for Eqs. (20.3.9) and (20.3.10) to be valid is that $g_\kappa^2/8\pi^2 \ll 1$, but it is not necessary for $\ln \kappa$ to be so large that only one eigenvector of the c matrix contributes to the asymptotic behavior. Eq. (20.3.10) will be used to study deep inelastic scattering in Section 20.6.

20.4 Symmetry Properties of Coefficient Functions

The usefulness of the operator product expansion is greatly enhanced by the fact that the coefficient functions exhibit the full symmetry of the underlying theory, even where part or all of that symmetry is spontaneously broken.[4] To prove this, we consider the operator product expansion for a product of renormalized operators $\mathcal{O}_i(x)$ that transform linearly under some symmetry with conserved current $J^\mu(x)$, in the sense that

$$[J^0(\mathbf{x}, t), \mathcal{O}_i(\mathbf{y}, t)] = -\delta^3(\mathbf{x} - \mathbf{y}) \sum_j t_{ij} \mathcal{O}_j(\mathbf{y}, t) , \tag{20.4.1}$$

where t_{ij} is a constant matrix. We can write the operator product expansion as the statement that, as $x_1, \cdots x_n$ approach x together (with $x_1 - x, \cdots x_n - x$ all having fixed ratios)

$$\langle \beta | T\{\mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_n}(x_n)\} | \alpha \rangle \to \sum_i U_i^{i_1 \cdots i_n}(x_1 - x, \cdots x_n - x) \langle \beta | \mathcal{O}_i(x) | \alpha \rangle . \tag{20.4.2}$$

Now suppose that the symmetry with current J^μ is spontaneously broken, with the corresponding Goldstone boson π satisfying

$$\langle \mathrm{VAC} | J^\mu(0) | \pi \rangle = \frac{F p_\pi^\mu}{(2\pi)^{3/2} \sqrt{2p_\pi^0}} . \tag{20.4.3}$$

Then, as we saw in Section 19.2, the matrix element of the operator product between states with an additional low energy Goldstone boson is

$$\langle \beta | T\{\mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_n}(x_n)\} | \pi\alpha \rangle = \frac{1}{(2\pi)^{3/2} \sqrt{2p_\pi^0} F} \times \int d^4x \, \frac{\partial}{\partial x^\mu} \langle \beta | T\{\mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_n}(x_n) J^\mu(x)\} | \alpha \rangle , \tag{20.4.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

because only the Goldstone pole term survives in the integral on the right-hand side of Eq. (20.4.4). Using Eq. (20.4.1) and the conservation of the current, this can be put in the form

$$\langle\beta|T\{\mathcal{O}_{i_1}(x_1)\cdots\mathcal{O}_{i_n}(x_n)\}|\pi\alpha\rangle = -\frac{1}{(2\pi)^{3/2}\sqrt{2p_\pi^0}F} \times \sum_{r=1}^{n}\sum_{j_r} t_{i_r j_r}\langle\beta|T\{\mathcal{O}_{i_1}(x_1)\cdots\mathcal{O}_{j_r}(x_r)\cdots\mathcal{O}_{i_n}(x_n)\}|\alpha\rangle\ . \quad (20.4.5)$$

Now apply the operator product expansion to both sides of this formula. In the limit as $x_1\cdots x_n$ all approach x together, we find

$$\sum_i U_i^{i_1\cdots i_n}(x_1-x,\cdots x_n-x)\langle\beta|\mathcal{O}_i(x)|\pi\alpha\rangle = -\frac{1}{(2\pi)^{3/2}\sqrt{2p_\pi^0}F} \times \sum_{r=1}^{n}\sum_{j_r}\sum_i t_{i_r j_r}\, U_i^{i_1\cdots j_r\cdots i_n}(x_1-x,\cdots x_n-x)\langle\beta|\mathcal{O}_i(x)|\alpha\rangle\ . \quad (20.4.6)$$

But as a special case of Eq. (20.4.5), we have

$$\langle\beta|\mathcal{O}_i(x)|\pi\alpha\rangle = -\frac{1}{(2\pi)^{3/2}\sqrt{2p_\pi^0}F}\sum_j t_{ij}\langle\beta|\mathcal{O}_j(x)|\alpha\rangle\ . \quad (20.4.7)$$

Since all this holds for arbitrary states $\langle\beta|$ and $|\alpha\rangle$, the coefficients of $\langle\beta|\mathcal{O}_j(x)|\alpha\rangle$ on both sides of Eq. (20.4.6) must be equal, so

$$0 = -\sum_i t_{ij}U_i^{i_1\cdots i_n}(x_1-x,\cdots x_n-x)+\sum_{r=1}^{n}\sum_{j_r} t_{i_r j_r}\, U_j^{i_1\cdots j_r\cdots i_n}(x_1-x,\cdots x_n-x)\ . \quad (20.4.8)$$

This can be restated as the condition that $U_i^{i_1\cdots i_n}(x_1-x,\cdots x_n-x)$ is invariant under the symmetry generated by t, with the action of this symmetry on the lower index contragredient to its action on the upper indices, in the sense that the matrix t is replaced with $-t^{\mathrm{T}}$. This is the same relation that would be expected if the symmetry generated by J^μ were not spontaneously broken.

20.5 Spectral Function Sum Rules

Spectral function sum rules are constraints on the spectral functions of various currents.[5] We will start here with a set of currents J_a^μ that are arbitrary, except for being Lorentz four-vectors, and then later consider more special examples. To define their spectral functions, we use Lorentz

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

invariance to write

$$\sum_N \delta^4(p-p_N)\langle \mathrm{VAC}|J_\alpha^\mu(0)|N\rangle\langle \mathrm{VAC}|J_\beta^\nu(0)|N\rangle^* = (2\pi)^{-3}\theta(p^0)$$
$$\times \left[\left(\eta^{\mu\nu}-p^\mu p^\nu/p^2\right)\rho_{\alpha\beta}^{(1)}(-p^2)+p^\mu p^\nu \rho_{\alpha\beta}^{(0)}(-p^2)\right], \tag{20.5.1}$$

in analogy with Eqs. (19.2.19) and (19.2.20) or Eq. (10.7.4). Taking the Fourier transform and using the completeness of the states $|N\rangle$, this can be written

$$\langle J_\alpha^\mu(x)\, J_\beta^\nu(0)\rangle_{\mathrm{VAC}} = \int d\mu^2$$
$$\times \left[\eta^{\mu\nu}\rho_{\alpha\beta}^{(1)}(\mu^2) - \left(\rho_{\alpha\beta}^{(0)}(\mu^2)+\rho_{\alpha\beta}^{(1)}(\mu^2)/\mu^2\right)\partial^\mu\partial^\nu\right]\Delta_+(x;\mu^2), \tag{20.5.2}$$

where $\Delta_+(x;\mu^2)$ is the function defined in Eq. (5.2.7):

$$\Delta_+(x;\mu^2) = \frac{1}{(2\pi)^3}\int d^4p\,\theta(p^0)\,\delta(p^2+\mu^2)e^{ip\cdot x}. \tag{20.5.3}$$

Assuming the currents to have been chosen as Hermitian operators, it is immediately apparent from Eq. (20.5.1) that $\rho_{\alpha\beta}^{(1)}(\mu^2)$ and $\rho_{\alpha\beta}^{(0)}(\mu^2)$ are positive Hermitian matrices. Also, taking x^μ to be spacelike (for which $\Delta_+(x)$ is even) and using translation invariance and causality in Eq. (20.5.2), we see that $\rho_{\alpha\beta}^{(1)}(\mu^2)$ and $\rho_{\alpha\beta}^{(0)}(\mu^2)$ are also symmetric.

Now, for $x\to 0$ with $x^2>0$, the function $\Delta_+(x;\mu^2)$ goes as

$$\Delta_+(x;\mu^2) \to \frac{1}{4\pi^2x^2}+\frac{\mu^2}{8\pi^2}\left[\ln\left(\frac{\gamma\mu\sqrt{x^2}}{2}\right)-\frac{1}{2}\right]+O(x^2), \tag{20.5.4}$$

where γ is the Euler constant. The first few terms in the vacuum expectation value of the expansion of $J_\alpha^\mu(x)\,J_\beta^\nu(0)$ are thus

$$\langle J_\alpha^\mu(x)\,J_\beta^\nu(0)\rangle_{\mathrm{VAC}} \to -\frac{1}{2\pi^2}\left[\frac{\eta_{\mu\nu}}{(x^2)^2}-\frac{4x_\mu x_\nu}{(x^2)^3}\right]\int d\mu^2\left(\rho_{\alpha\beta}^{(0)}(\mu^2)+\rho_{\alpha\beta}^{(1)}(\mu^2)/\mu^2\right)$$
$$-\frac{\eta_{\mu\nu}}{4\pi^2x^2}\int d\mu^2\,\rho_{\alpha\beta}^{(0)}(\mu^2)\mu^2+\frac{x_\mu x_\nu}{2\pi^2(x^2)^2}\int d\mu^2\left(\rho_{\alpha\beta}^{(0)}(\mu^2)\mu^2+\rho_{\alpha\beta}^{(1)}(\mu^2)\right)$$
$$+O(\ln x^2). \tag{20.5.5}$$

Hence if some linear combination $\sum_{\alpha\beta}c_{\alpha\beta}\langle J_\alpha^\mu(x)\,J_\beta^\nu(0)\rangle_{\mathrm{VAC}}$ of the two-point functions has a singularity as $x\to 0$ which can be shown to be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

weaker than of order $1/x^4$, we have

$$\sum_{\alpha\beta} c_{\alpha\beta} \int d\mu^2 \left(\rho^{(0)}_{\alpha\beta}(\mu^2) + \rho^{(1)}_{\alpha\beta}(\mu^2)/\mu^2 \right) = 0 \,, \tag{20.5.6}$$

while if its singularity is also weaker than $1/x^2$, then

$$\sum_{\alpha\beta} c_{\alpha\beta} \int d\mu^2 \, \rho^{(1)}_{\alpha\beta}(\mu^2) = 0 \tag{20.5.7}$$

and

$$\sum_{\alpha\beta} c_{\alpha\beta} \int d\mu^2 \, \rho^{(0)}_{\alpha\beta}(\mu^2)\mu^2 = 0 \,. \tag{20.5.8}$$

Eqs. (20.5.6), (20.5.7), and (20.5.8) are known respectively as the first, second, and third spectral function sum rules.

Let us see how this works out in the case of greatest interest, where the $J^\mu_\alpha(x)$ are conserved currents in a theory like quantum chromodynamics. The conservation of current tells us that (20.5.1) vanishes when contracted with p_μ, so $\rho^{(0)}_{\alpha\beta}(-p^2)$ must be proportional to $\delta(-p^2)$, and therefore automatically satisfies the third spectral function sum rule (20.5.8) for any $c_{\alpha\beta}$. With $\rho^{(0)}_{\alpha\beta}(-p^2) \propto \delta(-p^2)$, it can receive contributions only from the terms $|B_a\rangle$ in the sum over states in (20.5.1) consisting of a single massless particle B_a of zero spin, which in practice means a Goldstone boson. For such one-particle states, Lorentz invariance gives

$$\langle \text{VAC}|J^\mu_\alpha(0)|B_a\rangle = \frac{iF_{\alpha a}\, p^\mu_B}{(2\pi)^{3/2}\sqrt{2p^0_B}} \,. \tag{20.5.9}$$

Using the relation $\delta(p^0 - |\mathbf{p}|)/2p^0 = \theta(p^0)\delta(-p^2)$, we see that

$$\rho^{(0)}_{\alpha\beta}(-p^2) = \delta(-p^2) \sum_a F_{\alpha a} F^*_{\beta a} \,. \tag{20.5.10}$$

In contrast, $\rho^{(1)}_{\alpha\beta}(-p^2)$ is non-vanishing only for $-p^2 > 0$.

To be more specific, consider a renormalizable asymptotically free gauge theory with a number N of massless (or nearly massless) spin 1/2 fermions belonging to the same representation of the gauge group. Quantum chromodynamics fits this description, with $N = 3$ if we neglect the masses of the u, d, and s quarks, and with $N = 2$ if only the u and d are taken to be massless. As we have seen in Chapter 19, such a theory has an $SU(N) \times SU(N)$ global symmetry* under which the left- and right-handed

* There is also a $U(1)$ symmetry which is vectorial, in the sense that it acts the same way on the left- and right-handed parts of the light fermion fields. This is just the conservation of light quark number, and will not concern us here. The axial $U(1)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

parts of the light fermion fields transform under the representations $(N, 1)$ and $(1, N)$, respectively, where 'N' and '1' denote the defining and identity representations of $SU(N)$, respectively. The currents of the left- and right-handed $SU(N)$ symmetries are

$$J^{\mu}_{La}(x) = -i\bar{\psi}(x)\gamma^{\mu}(1+\gamma_5)\lambda_a\psi(x)\,, \quad J^{\mu}_{Ra}(x) = -i\bar{\psi}(x)\gamma^{\mu}(1-\gamma_5)\lambda_a\psi(x)\,, \tag{20.5.11}$$

where the λ_a form a complete set of Hermitian traceless matrices acting on the 'flavor' index that distinguishes the N light quarks, as in Eq. (19.7.2) for $N = 3$. These currents have dimensionality (in powers of mass) $+3$, so the coefficient of an operator $\mathcal{O}$ of dimensionality $d(\mathcal{O})$ in the expansion (20.3.6) of the product of two currents is expected (apart from logarithms) to have a singularity of order $x^{-6+d(\mathcal{O})}$ as the separation x of their arguments goes to zero. Thus, if the expansion of a linear combination of products of currents contains the operator unity, then the corresponding linear combination of spectral functions goes as x^{-6}, and therefore in general satisfies neither the first nor the second spectral function sum rule; if the lowest-dimensional operator in the vacuum expectation value of this expansion is a fermion bilinear with zero or one derivatives, then the corresponding linear combination of spectral functions goes as x^{-3} or x^{-2}, and therefore satisfies the first spectral function sum rule but generally not the second; and if the lowest-dimensional operator in the vacuum expectation value of this expansion is a fermion bilinear with two or more derivatives or a fermion quadrilinear then the corresponding linear combination of spectral functions is less singular than x^{-2}, and therefore satisfies both the first and second spectral function sum rules.

In order to tell which operators appear in the expansion of the product of two currents, we need to classify the $SU(N) \times SU(N)$ representations contained in the product, and to tell whether these operators have non-vanishing vacuum expectation values, we have to ask which of them are invariant under the subgroup of $SU(N) \times SU(N)$ that is not spontaneously broken. To answer these questions, we note that the currents $J^{\mu}_{La}(x)$ and $J^{\mu}_{Ra}(x)$ transform under $SU(N) \times SU(N)$ respectively according to the $(A, 1)$ and $(1, A)$ representations of $SU(N) \times SU(N)$, where A and 1 are the adjoint and identity representations of $SU(N)$.

Also, we shall assume that the subgroup of $SU(N) \times SU(N)$ that is not spontaneously broken is the vectorial $SU(N)_V$ whose currents are $J^{\mu}_{La}(x) + J^{\mu}_{Ra}(x)$, as is the case in quantum chromodynamics and (as we saw in Section 19.9) a wide range of other theories. We also assume that parity conservation is not spontaneously broken. These unbroken

symmetry of the Lagrangian, which acts differently on the left- and right-handed parts of the light fermion fields, is broken by the quantum effects discussed in Section 23.5.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

symmetries have the consequence that

$$\rho^{(1)}_{La,Lb}(\mu^2) = \rho^{(1)}_{Ra,Rb}(\mu^2) = \delta_{ab}\left[\rho^{(1)}_V(\mu^2) + \rho^{(1)}_A(\mu^2)\right] \tag{20.5.12}$$

and

$$\rho^{(1)}_{La,Rb}(\mu^2) = \rho^{(1)}_{Ra,Lb}(\mu^2) = \delta_{ab}\left[\rho^{(1)}_V(\mu^2) - \rho^{(1)}_A(\mu^2)\right] , \tag{20.5.13}$$

where $\delta_{ab}\rho^{(1)}_V(\mu^2)$ and $\delta_{ab}\rho^{(1)}_A(\mu^2)$ are the spectral functions defined by (20.5.2) for the currents

$$J^\mu_{Va} = -i\bar\psi\gamma^\mu\lambda_a\psi , \qquad J^\mu_{Aa} = -i\bar\psi\gamma^\mu\gamma_5\lambda_a\psi , \tag{20.5.14}$$

with generators λ_a and $\lambda_a\gamma_5$, respectively. Also

$$F_{La\,b} = -F_{Ra\,b} = \delta_{ab}F , \tag{20.5.15}$$

where Eq. (20.5.9) here reads

$$\langle \mathrm{VAC}|J^\mu_{Aa}(0)|B_b\rangle = \frac{iF\delta_{ab}\,p^\mu_B}{(2\pi)^{3/2}\sqrt{2p^0_B}} \tag{20.5.16}$$

so that (20.5.10) reads

$$\rho^{(0)}_{La\,Lb}(\mu^2) = \rho^{(0)}_{Ra\,Rb}(\mu^2) = -\rho^{(0)}_{La\,Rb}(\mu^2) = -\rho^{(0)}_{Ra\,Lb}(\mu^2) = F^2\delta(\mu^2)\delta_{ab} . \tag{20.5.17}$$

It will be convenient to consider separately the products of currents of like and unlike chirality.

Like Chirality: The products $J^\mu_{La}(x)J^\nu_{Lb}(0)$ and $J^\mu_{Ra}(x)J^\nu_{Rb}(0)$ transform respectively as the $(A\times A, 1)$ and $(1, A\times A)$ representations of $SU(N)\times SU(N)$, where A and 1 are the adjoint and identity representations, respectively. For any group, $A\times A$ contains the identity representation, so the unit operator appears in the expansions of these products, with equal coefficients proportional to δ_{ab}. Hence only the traceless parts of the operator product can satisfy spectral function sum rules. But as we have seen, the spectral functions have no traceless part, so the like-chirality spectral functions cannot satisfy any sum rules.

Unlike Chirality: The products $J^\mu_{La}(x)J^\nu_{Rb}(0)$ and $J^\mu_{Ra}(x)J^\nu_{Lb}(0)$ both transform as the (A, A) representation of $SU(N)\times SU(N)$. The unit operator (and operators like $F_{\alpha\mu\nu}F^{\mu\nu}_\alpha$) are of course $SU(N)\times SU(N)$ singlets, and therefore cannot appear in the expansion of these products. Non-derivative fermion bilinears like $\bar\psi\psi$ transform according to the $(\bar N, N)$ and $(N, \bar N)$ representations of $SU(N)\times SU(N)$, so they also cannot appear in the expansions of the products of currents with unlike chirality. The only gauge- and Lorentz-invariant fermion bilinears with a single derivative involve the gauge-covariant derivative operator $\gamma^\mu D_\mu$ acting on ψ, which the field equations tell us vanishes. Thus these spectral functions satisfy

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

both the first and second spectral function sum rules, which here read

$$\int d\mu^2 \left[\rho_V^{(1)}(\mu^2) - \rho_A^{(1)}(\mu^2)\right]/\mu^2 = F^2 \tag{20.5.18}$$

and

$$\int d\mu^2 \left[\rho_V^{(1)}(\mu^2) - \rho_A^{(1)}(\mu^2)\right] = 0\,. \tag{20.5.19}$$

In the original work on spectral function sum rules,[5] the $SU(2) \times SU(2)$ spectral functions were assumed to be sharply peaked at values of μ that for $\rho_V^{(1)}(\mu^2)$ could be assumed to be $m_\rho = 770$ MeV, and for $\rho_A^{(1)}(\mu^2)$ was taken to be at an unknown mass m_A. That is,

$$\rho_V^{(1)}(\mu^2) \simeq g_\rho^2 \delta(\mu^2 - m_\rho^2)\,, \qquad \rho_A^{(1)}(\mu^2) \simeq g_A^2 \delta(\mu^2 - m_A^2)\,.$$

Eqs. (20.5.18) and (20.5.19) then read**

$$\frac{g_\rho^2}{m_\rho^2} - \frac{g_A^2}{m_A^2} = F_\pi^2$$

and

$$g_\rho^2 = g_A^2\,.$$

Eliminating the unknown g_A, this gives a formula

$$g_\rho^2 = F_\pi^2 \left(\frac{1}{m_\rho^2} - \frac{1}{m_A^2}\right)^{-1}\,.$$

Originally in 1967 this result was used together with a formula[6] $g_\rho^2 = 2F_\pi^2 m_\rho^2$ (whose justification was unclear but which agreed with experimental measurements of the rate of the decay $\rho \to e^+ + e^-$) to derive the result

$$m_A = \sqrt{2} m_\rho\,.$$

The status of a possible a_1 resonance with the right quantum numbers to couple to the axial-vector current (that is, 1^+ with $T = 1$ and $\mathrm{C}(a_1^0) = +1$) at around a mass $\sqrt{2}m_\rho$ remained unclear for many years, but a resonance with these quantum numbers is now reasonably well established at a mass 1230 MeV $= 1.6\, m_\rho$. It seems preferable today to take the ratio m_A/m_ρ as an input, using either the value $\sqrt{2}$ suggested by certain models[7] or the experimental value 1.6, and use it to predict g_ρ.

** Taking the $SU(2)$ generators λ_a to be the Pauli matrices, as in Eq. (19.7.2), we have $F = F_\pi = 184$ MeV.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

Since 1967 not only g_ρ but the whole vector spectral function $\rho_V^{(1)}(\mu^2)$ of the $SU(3) \times SU(3)$ current in quantum chromodynamics has been calculated with precision for a wide range of energies from the measured cross section for the process $e^+ + e^- \to \gamma \to$ hadrons, using the fact that the electromagnetic current is a linear combination of $SU(3)$ currents. The axial-vector spectral function $\rho_A^{(1)}(\mu^2)$ of the $SU(3) \times SU(3)$ currents can in principle be measured in the process $\bar{\nu} + e \to$hadrons, since the charged components of the currents (20.5.14) are the same as the hadron currents to which the leptons are coupled. However, although antineutrino–electron scattering has been studied experimentally, the low rate of these reactions precludes the use of colliding beams, so that the electron target is essentially at rest. In order to reach typical hadronic energies of, say, 3 GeV in the center-of-mass frame, it would be necessary to have neutrino energies of $(3\,\text{GeV})^2/2m_e \simeq 10$ TeV in the laboratory frame. Intense beams of neutrinos with energies this high will not be available for many years, if ever. Fortunately, it has become possible to study the spectral functions in the process $\tau \to \nu +$ hadrons, but the hadron energies here are strictly bounded by $m_\tau = 1.7$ GeV. It is also possible to use effective chiral Lagrangians to calculate the spectral functions at small μ^2, and to use quantum chromodynamics to calculate $\rho_V^{(1)} - \rho_A^{(1)}$ at large μ^2, where it is quite small. A careful 1993 analysis by Donoghue and Golowich[8] of all these inputs shows that the spectral function integrals are indeed dominated by the ρ and a_1 resonances, and gives results that are consistent with the first and second spectral function sum rules.

20.6 Deep Inelastic Scattering

The renormalization group coupled with the operator product expansion found one of its most important applications in the analysis of deep inelastic lepton–nucleon scattering. We will first review the early phenomenological models for these reactions, and then see how the operator product expansion justifies these models and provides corrections to them.

Consider a process in which an electron of four-momentum k collides with a nucleon N of four-momentum p, yielding a electron of four-momentum k' and a general unobserved hadron state H, over which we sum. To calculate the spin-averaged inclusive cross section, we shall need to know the quantity

$$(m_N/p_N^0)W^{\mu\nu}(q,p) \equiv \frac{1}{2}\sum_{\sigma_N}\sum_H \delta^4(p_H - p - q)\langle H|J^\mu(0)|N\rangle\langle H|J^\nu(0)|N\rangle^* , \tag{20.6.1}$$

where J^μ is the electromagnetic current (divided here by a factor e) and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

$q = k - k'$ is the momentum transfer from the electrons to the hadrons. Lorentz invariance tells us that $W^{\mu\nu}(q,p)$ must be a linear combination of $p^\mu p^\nu$, $p^\mu q^\nu$, $q^\mu p^\nu$, $q^\mu q^\nu$, and $\eta^{\mu\nu}$, with coefficients that can only depend on the two independent scalar functions of q and p: q^2 and $\nu \equiv -q\cdot p/m_N$. Current conservation requires that $q_\mu W^{\mu\nu} = q_\nu W^{\mu\nu} = 0$, so W must take the form*

$$W^{\mu\nu}(q,p) = -\left(\frac{q^\mu q^\nu}{q^2} - \eta^{\mu\nu}\right) W_1(\nu, q^2) + \frac{1}{m_N^2}\left(p^\mu - \frac{p\cdot q}{q^2}q^\mu\right)\left(p^\nu - \frac{p\cdot q}{q^2}q^\nu\right) W_2(\nu, q^2)\,. \qquad (20.6.2)$$

Also, Eq. (20.6.1) shows that $W^{\mu\nu *} = W^{\nu\mu}$, so W_1 and W_2 are real, and $W^{\mu\nu}$ is a positive matrix, so W_1 and W_2 are positive. The differential cross section in the nucleon rest frame is

$$\frac{d^2\sigma}{d\Omega d\nu} = \left(\frac{d\sigma}{d\Omega}\right)_{\rm MOTT}\left(W_2 + 2W_1\tan^2(\theta/2)\right)\,, \qquad (20.6.3)$$

where $d\Omega = \sin\theta\, d\theta\, d\phi$ is the solid angle into which the electron is scattered, and $(d\sigma/d\Omega)_{\rm MOTT}$ is the differential cross section for relativistic elastic scattering by a point spinless particle

$$\left(\frac{d\sigma}{d\Omega}\right)_{\rm MOTT} = \frac{e^4}{4E_e^2}\,\frac{\cos^2(\theta/2)}{\sin^4(\theta/2)}\,, \qquad (20.6.4)$$

with $E_e = -k\cdot p/m_N$ the incident electron energy in the nucleon rest frame.

One might expect that for fixed values of $-p_H^2 = -q^2 + 2m_N\nu + m_N^2$, the differential cross section should fall off very rapidly as $q^2 \to \infty$, because it should be proportional to the square of the form-factor for the transition from the nucleon to whatever particles or resonance have mass near $-p_H^2$. It was therefore somewhat of a surprise that, two years after the 1966 opening of the Stanford Linear Accelerator Center, a SLAC–MIT collaboration headed by Friedman, Kendall, and Taylor[9] discovered that in fact $\nu W_2(q^2,\nu)$ is roughly constant in q^2 for fixed values of $\omega \equiv 2m_N\nu/q^2 > 1$. (To be specific, $W_2(q^2,\nu)$ for the proton was fitted to the curve $\nu W_2(q^2,\nu) \simeq 0.35 - 0.004\omega$ for $E_e = 10$, 13.5, and 16 GeV, and $\theta = 6°$ and $10°$. These experiments were insensitive to W_1, because $\tan^2(10°/2) = 7.6\times 10^{-3}$.) Note that in this limit $-p_H^2 \to (\omega - 1)q^2 \to \infty$, which is why this is called 'deep-inelastic' scattering.

At about the same time Bjorken[10] was using current algebra to argue that $W_2(q^2,\nu)$ and $W_1(q^2,\nu)$ satisfy scaling laws: for q^2 and ν going to

* The same formalism can be used for other deep-inelastic lepton scattering processes, such as $\nu_\mu + p \to \mu^- + H$, except that since parity is not conserved in these processes there is an additional term in $W^{\mu\nu}(q,p)$ proportional to $\epsilon^{\mu\nu\rho\sigma}p_\rho q_\sigma$. For simplicity, we will limit ourselves here to deep-inelastic electron scattering.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

infinity together,

$$\nu W_2(\nu, q^2) \to F_2(\omega)\,, \qquad W_1(\nu, q^2) \to F_1(\omega)\,, \tag{20.6.5}$$

where again $\omega \equiv 2m_N\nu/q^2$. A more intuitive explanation was given a little later by Feynman.[11] He supposed that in deep-inelastic scattering on a highly relativistic nucleon of momentum $\mathbf{p}$, the nucleon behaves as if it consists of 'partons' of various types labelled i, each with a probability $\mathscr{F}_i(x)dx$ of having momentum between $x\mathbf{p}$ and $(x+dx)\mathbf{p}$. Then for each i,

$$\int dx\, \mathscr{F}_i(x) = 1\,. \tag{20.6.6}$$

The condition that the nucleon has total momentum $\mathbf{p}$ yields the additional sum rule

$$\int_0^1 \sum_i \mathscr{F}_i(x)\, x\, dx = 1\,. \tag{20.6.7}$$

For elastic scattering of an electron (with m_e neglected) from a parton of four-momentum xp, we have $x^2m_N^2 = -(q+xp)^2 = -q^2 - 2\nu m_N x + x^2 m_N^2$, so $\nu = q^2/2m_N x$. The inelastic cross section in this model is thus**

$$\frac{d^2\sigma}{d\Omega d\nu} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{MOTT}} \sum_i Q_i^2 \int_0^1 dx \left(1 + \frac{q^2}{2m_N^2x^2}\tan^2\left(\frac{\theta}{2}\right)\right)\delta\left(\nu - \frac{q^2}{2m_N x}\right) \tag{20.6.8}$$

where Q_i is the charge of the ith type of parton in units of e. (The $\tan^2(\theta/2)$ term here is appropriate for a 'Dirac' parton, with magnetic moment $eQ_i/2m_N x$.) Comparing this with Eq. (20.6.3) gives

$$W_2(\nu, q^2) = \sum_i Q_i^2 \int_0^1 dx\, \mathscr{F}_i(x)\delta\left(\nu - \frac{q^2}{2m_N x}\right) = \frac{1}{\nu\omega}\sum_i Q_i^2 \mathscr{F}_i\left(\frac{1}{\omega}\right)\,, \tag{20.6.9}$$

$$W_1(\nu, q^2) = \frac{1}{2}\sum_i Q_i^2 \int_0^1 dx\, \mathscr{F}_i(x)\frac{q^2}{2m_N^2x^2}\delta\left(\nu - \frac{q^2}{2m_N x}\right) = \frac{\omega\nu}{2m_N} W_2(\nu, q^2)\,. \tag{20.6.10}$$

** This may be derived from the formulas (8.7.7) and (8.7.38) for Compton scattering on a spin 1/2 particle of four-momentum p and mass m in a general Lorentz frame. For this purpose, it is useful to note that in the gauge used to derive these formulas, for which the initial and final real polarization vectors e_μ and e'_μ satisfy $e^2 = e'^2 = 1$, $e\cdot p = e'\cdot p = e\cdot k = e'\cdot k' = 0$, the polarization sum gives

$$\sum_{e,e'} (e\cdot e')^2 = 2 + \frac{m^4(k\cdot k')^2}{(k\cdot p)^2(k'\cdot p)^2} + \frac{2m^2(k\cdot k')}{(k\cdot p)(k'\cdot p)}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

This agrees with the Bjorken scaling rules (20.6.5), with

$$F_2(\omega) = \frac{1}{\omega}\sum_i Q_i^2 \mathscr{F}_i\left(\frac{1}{\omega}\right) \tag{20.6.11}$$

and

$$F_1(\omega) = (\omega/2m_N)F_2(\omega)\,. \tag{20.6.12}$$

Eq. (20.6.12) was originally derived by Callan and Gross.[12] It agrees with experiment within about 10–15%.

On the assumption that the proton and neutron respectively consist of two u and one d or one u and two d quarks, plus any number of neutral partons, Eqs. (20.6.11) and (20.6.6) yield the sum rule for W_2

$$\int_1^\infty F_2(\omega)\frac{d\omega}{\omega} = \sum_i Q_i^2 = \begin{cases} 1 & p \\ 2/3 & n \end{cases}\,. \tag{20.6.13}$$

The integral receives a large contribution from large ω, where F_2 is difficult to measure. If we also assume that the total momentum of the nucleon is equally shared among three quarks (and no neutral partons) then in place of Eq. (20.6.7) we have the stronger relation $\int \mathscr{F}_i(x)x\,dx = 1/3$ for each quark, which with Eq. (20.6.11) yields

$$\int_1^\infty F_2(\omega)\frac{d\omega}{\omega^2} = \tfrac{1}{3}\sum_i Q_i^2 = \begin{cases} 1/3 & p \\ 2/9 & n \end{cases}\,.$$

This integral is easier to measure, and turns out to disagree badly with the sum rule, showing that a large fraction of the nucleon momentum is carried by the neutral partons.

None of the above phenomenology depends on any specific field theory. It was the operator product expansion that finally provided a way of applying an underlying field theory to deep-inelastic scattering. In particular, the operator product expansion made it clear that one needed an asymptotically free field theory to explain the scaling assumptions (20.6.5) (and to calculate the corrections to scaling). This asymptotically free theory was ultimately provided by quantum chromodynamics.

To apply the operator product expansion to deep-inelastic scattering, first Fourier transform Eq. (20.6.1). Using translation invariance and the completeness of the hadron states $|H\rangle$, this gives

$$(m_N/p_N^0)W^{\mu\nu}(q,p) \equiv \frac{1}{2(2\pi)^4}\sum_{\sigma_N}\int d^4z\; e^{-iq\cdot z}\langle N|J^\nu(z)\,J^\mu(0)|N\rangle\,. \tag{20.6.14}$$

The asymptotic behavior of $W^{\mu\nu}(q,p)$ as $q \to \infty$ is therefore related to the singularity of the operator product at $z \to 0$.

Feynman diagram calculations of the coefficient functions in operator product expansions refer directly not to the expansion for matrix elements

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

like those appearing in Eq. (20.6.14) for $W^{\mu\nu}$, but rather to matrix elements of the two-point Greens function

$$(m_N/p_N^0)T^{\mu\nu}(q,p) \equiv \frac{1}{2(2\pi)^4}\sum_{\sigma_N}\int d^4z\, e^{-iq\cdot z}\langle N|T\{J^\nu(z), J^\mu(0)\}|N\rangle\,. \tag{20.6.15}$$

We can express this in terms of the structure functions for $T^{\mu\nu}$, defined by the analog to Eq. (20.6.2):

$$T^{\mu\nu}(q,p) = -\left(\frac{q^\mu q^\nu}{q^2} - \eta^{\mu\nu}\right)T_1(\nu,q^2) + \frac{1}{m_N^2}\left(p^\mu - \frac{p\cdot q}{q^2}q^\mu\right)\left(p^\nu - \frac{p\cdot q}{q^2}q^\nu\right)T_2(\nu,q^2)\,. \tag{20.6.16}$$

The connection between the $T_r(\nu,q^2)$ and $W_r(\nu,q^2)$ (with $r = 1,2$) is provided[13] by the dispersion relations† for fixed q^2:

$$T_r(\nu,q^2) = \frac{1}{2}W_r(-\nu,q^2) + \frac{1}{2}W_r(\nu,q^2) + \frac{1}{2\pi i}\int_{-\infty}^{\infty}d\nu'\,\frac{W_r(-\nu',q^2) - W_r(\nu',q^2)}{\nu'-\nu}\,, \tag{20.6.17}$$

with the denominator in the integrand interpreted as a principal value function. The functions $W_r(\nu,q^2)$ vanish except for $\nu > q^2/2m_N$, so the dispersion relations may be rewritten

$$T_r(\nu,q^2) = \frac{1}{2}W_r(-\nu,q^2) + \frac{1}{2}W_r(\nu,q^2) - \frac{1}{2\pi i}\int_{q^2/2m_N}^{\infty}d\nu'\,W_r(\nu',q^2)\left(\frac{1}{\nu'+\nu} + \frac{1}{\nu'-\nu}\right)\,. \tag{20.6.18}$$

We can categorize the operators that contribute to the operator product expansion for T_r according to the irreducible representation of the Lorentz group to which they belong. The only Lorentz-covariant functions of a single four-vector p^μ with $p^2 = -m_N^2$ fixed are proportional to the symmetric tensors $p^{\mu_1}\cdots p^{\mu_s}$, so the only operators that can contribute to the spin-averaged nucleon expectation value are symmetric traceless tensors $\mathcal{O}_{si}^{\mu_1\cdots\mu_s}$, the subscript i distinguishing any different operators with this tensor structure. (For reasons that will become clear, we are using the same label i to distinguish operators that we used in the parton model to

† For $q^2 = 0$, these may be derived exactly as in the derivation of the dispersion relation (10.8.16) for forward photon scattering. The derivation for fixed $q^2 \neq 0$ is more difficult.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

label types of partons.) These operators have matrix elements of the form

$$\frac{1}{2}\sum_{\sigma_N}\langle N|\mathcal{O}_{si}^{\mu_1\cdots\mu_s}|N\rangle = (m_N/p_N^0)\Big[\, p^{\mu_1}\cdots p^{\mu_s} - \text{traces}\,\Big]\,\langle\mathcal{O}_{si}\rangle\,, \qquad (20.6.19)$$

where the $\langle\mathcal{O}_{si}\rangle$ are constant coefficients. Such an operator makes a contribution to $T^{\mu\nu}(q,p)$ proportional to s factors of the four-vector p, and hence makes a contribution to $T_1(\nu,q^2)$ and $T_2(\nu,q^2)$ proportional to ν^s and ν^{s-2}, respectively. (We are dropping terms involving p^2, because such terms will be suppressed by factors of m_N^2/q^2 or $m_N^2/p\cdot q$.) If we ignore logarithmic corrections, then in an asymptotically free theory the q^2 dependence of the coefficients is of the forms $(q^2)^{(-4+6-d(s,i)-s)/2}$ and $(q^2)^{(-4+6-d(s,i)-s+2)/2}$, respectively, where $d(s,i)$ is the dimensionality of the operator $\mathcal{O}_{si}$.[††] Using $\nu \propto q^2\omega$, we see that the contributions of an operator $\mathcal{O}_{si}$ to the structure functions are asymptotically of the forms

$$T_{1,si} \propto \nu^s(q^2)^{(2-d(s,i)-s)/2} \propto \omega^s(q^2)^{(2-\tau(s,i))/2} \qquad (20.6.20)$$

and

$$\nu T_{2,si} \propto \nu^{s-1}(q^2)^{(4-d(s,i)-s)/2} \propto \omega^{s-1}(q^2)^{(2-\tau(s,i))/2}\,, \qquad (20.6.21)$$

where $\tau(s,i)$ is the 'twist' of the operator $\mathcal{O}_{si}$, defined as[14]

$$\tau(s,i) \equiv d(s,i) - s\,. \qquad (20.6.22)$$

We see that the dominant terms in T_1 and νT_2 for $q^2\to\infty$ with fixed ω are contributed by operators of minimum twist. Also, Eq. (20.6.18) shows there are no terms in $T_r(\nu,q^2)$ of odd order in ν, so the only operators $\mathcal{O}_{si}$ that can contribute here are those with even s.

The symmetric traceless tensors of rank s with minimum dimensionality are the operators

$$(\mathcal{O}_{sf})_{\mu_1\cdots\mu_s} \equiv (i^{s-2}/s!)\,\bar\psi_f\gamma_{\{\mu_1}D^{\leftrightarrow}_{\mu_2}\cdots D^{\leftrightarrow}_{\mu_s\}}\psi_f \qquad (20.6.23)$$

and

$$(\mathcal{O}_{s0})_{\mu_1\cdots\mu_s} \equiv (i^{s-2}/2s!)\,F_{\alpha\,\nu\{\mu_1}D^{\leftrightarrow}_{\mu_3}\cdots D^{\leftrightarrow}_{\mu_s}F^{\nu}_{\alpha\,\mu_2\}}\,, \qquad (20.6.24)$$

where f labels quark flavors; D_μ is the gauge-covariant derivative; and the brackets indicate a sum over permutations and subtraction of trace terms for the enclosed spacetime indices. (The symbol $\leftrightarrow$ indicates half the difference of the derivative acting to the right and left. We take this difference of derivatives, because their sum would vanish in any

[††] The -4 in the exponents arises from the integral over z in Eq. (20.6.14), and the $+6$ in these exponents is the dimensionality of the two electric current operators. The terms $-s/2$ and $-(s-2)/2$ serve to compensate for the powers of q^μ already in ν^s and ν^{s-2}, respectively.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

matrix element between states of equal four-momenta.) These operators have dimensionality $3 + (s-1) = 4 + (s-2) = 2 + s$, so they all have twist $\tau = 2$. Thus an infinite number of operators contribute for $q^2 \to \infty$, all of which give contributions to T_1 and νT_2 that depend on ω but only logarithmically on q^2. Asymptotic freedom thus confirms Bjorken scaling, but only up to logarithmic corrections. Keeping only operators of twist two, we see that (as anticipated in our notation) for each s there is indeed one operator for each parton type, with i running over quark types f (lumping quarks and antiquarks together) and also a value $i = 0$ for the gluon.

Now let us consider the logarithmic corrections. The asymptotic behavior of the coefficient functions in an asymptotically free theory is governed by Eq. (20.3.9). As discussed in Section 10.4, when we ignore electromagnetic radiative corrections no renormalization factors are needed for the electromagnetic current, because these currents are conserved. Thus the matrix $c_{\ell\ell'}$ in Eq. (20.3.9) vanishes. The matrix $c_{\mathcal{O},\mathcal{O}'}$ has no elements connecting operators of different Lorentz transformation type, so that in the notation we are now using, $c_{si,s'i'} = \delta_{ss'} c_{ii'}(s)$. Thus Eq. (20.3.9) takes the form

$$T_1(\nu, q^2) \to \sum_{sij} \omega^s \mathscr{A}_{si} \left[\left(g_q^2 / g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,, \tag{20.6.25}$$

$$\nu T_2(\nu, q^2) \to \sum_{sij} \omega^{s-1} \mathscr{B}_{si} \left[\left(g_q^2 / g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,, \tag{20.6.26}$$

where $\mathscr{A}$ and $\mathscr{B}$ are constants appearing in the coefficient function for the operator $\mathcal{O}_{si}$ in the operator product expansion of the product of two electric currents, ℓ^2 is a particular value of q^2 at which we choose to define the coefficient functions, b is the constant in the one-loop renormalization group equation (20.3.8) for the strong coupling constant g_q, and $\langle \mathcal{O}_{si} \rangle$ is the constant coefficient in the matrix elements (20.6.19).

The operator product expansion coefficients are independent of the particular process in question and unaffected by quark trapping, so we can calculate the coefficients $\mathscr{A}_{si}$ and $\mathscr{B}_{si}$ by considering a fictitious simpler process, the scattering of an electron by a free quark of flavor f. The renormalized operators $\mathcal{O}_{si}$ may conveniently be defined so that the one-quark matrix elements of the operators (20.6.23) and (20.6.24) are given by the tree approximation:

$$\langle f', \sigma' | \mathcal{O}_{sf} | f'', \sigma'' \rangle = \frac{(-1)^s i}{s!} \left(\bar{u}' \gamma^{\{\mu_1} u \right) p^{\mu_2} \cdots p^{\mu_s\}} \delta_{ff'} \delta_{ff''} \,, \tag{20.6.27}$$

$$\langle f', \sigma' | \mathcal{O}_{s0} | f'', \sigma'' \rangle = 0 \,. \tag{20.6.28}$$

Averaging Eq. (20.6.27) over $\sigma' = \sigma''$ and comparing the result with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

(20.6.19) yields

$$\langle \mathcal{O}_{sf} \rangle_{f'} \frac{m_f}{p^0}\, p^{\mu_1} p^{\mu_2} \cdots p^{\mu_s} = \delta_{ff'} \frac{(-1)^s i}{2s!} \mathrm{Tr} \left\{ \left(\frac{-i\, \not{p} + m_f}{2p^0} \right) \gamma^{\{\mu_1} p^{\mu_2} \cdots p^{\mu_s\}} \right\} ,$$

so that (for even s)

$$\langle \mathcal{O}_{sf} \rangle_{f'} = \delta_{ff'}/m_f \tag{20.6.29}$$

and

$$\langle \mathcal{O}_{s0} \rangle_{f'} = 0 . \tag{20.6.30}$$

With this definition of the operators $\mathcal{O}_{si}$, at a renormalization momentum scale ℓ^2 which is large enough for the operator product expansion functions to be calculated using the tree approximation, these functions can be derived from the tree-approximation formulas for the electron-quark scattering cross section. Retracing the parton-model derivations of Eqs. (20.6.9) and (20.6.10), we see that W_r for an electron scattering on a quark of flavor f is given in the tree approximation by‡

$$\nu W_{2,f} = (m_N/m_f) Q_f^2 \delta(\omega - 1) , \tag{20.6.31}$$

$$W_{1,f} = Q_f^2 \delta(\omega - 1)/2m_f . \tag{20.6.32}$$

Inserting these results in the dispersion relations (20.6.18) yields, for $\omega \neq 1$,

$$T_{1,f} = \frac{Q_f^2}{2\pi i m_f} \frac{m_N}{m_f} \frac{1}{\omega^2 - 1} , \tag{20.6.33}$$

$$\nu T_{2,f} = \frac{2Q_f^2}{2\pi i} \frac{\omega^2}{\omega^2 - 1} , \tag{20.6.34}$$

results that may also be obtained directly from the Feynman graphs for electron–quark scattering. Comparing the coefficients of ω^s in Eqs. (20.6.33) and (20.6.34) with those in Eqs. (20.6.25) and (20.6.26) at the renormalization point $q^2 = \ell^2$ and using Eqs. (20.6.29) and (20.6.30), we find that

$$\mathscr{A}_{si} = \frac{iQ_i^2}{2\pi} , \qquad \mathscr{B}_{si} = \frac{im_N Q_i^2}{\pi} , \tag{20.6.35}$$

with the gluon charge Q_0 of course taken as zero. To repeat, even though these values have been derived for electron–quark scattering, $\mathscr{A}_{si}$ and $\mathscr{B}_{si}$ are factors in the operator product expansion of two electric currents, and therefore do not depend on the state in which we calculate the expectation

‡ The factor (m_N/m_f) is inserted in Eq. (20.6.31) because ν is defined as $-q \cdot p/m_N$ instead of $-q \cdot p/m_f$, while it is W_2/m_N^2 rather than W_2/m_f^2 that appears in Eq. (20.6.2). Eq. (20.6.32) then follows from Eq. (20.6.10). The quantity ω may be written in a mass-independent way, as $\omega = -2q \cdot p/q^2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

value of this operator product. So we can use Eq. (20.6.35) in Eqs. (20.6.25) and (20.6.26), and find

$$T_1(\nu, q^2) \to \frac{i}{2\pi} \sum_{sij} \omega^s Q_i^2 \left[\left(g_q^2/g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,, \tag{20.6.36}$$

$$\nu T_2(\nu, q^2) \to \frac{i m_N}{\pi} \sum_{sij} \omega^{s-1} Q_i^2 \left[\left(g_q^2/g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,. \tag{20.6.37}$$

Now let us at last return to the structure functions $W_r(\nu, q^2)$. We note that the coefficient of $\omega^s = (2m_N \nu/q^2)^s$ in $T_r(\nu, q^2)$ is given by Eq. (20.6.18) as

$$\frac{-2}{2\pi i} \left(\frac{q^2}{2m_N} \right)^s \int_{q^2/2m_N}^{\infty} d\nu' \, \nu'^{-1-s} W_r(\nu', q^2) = \frac{i}{\pi} \int_1^{\infty} d\omega \, \omega^{-1-s} W_r \left(\frac{\omega q^2}{2m_N}, q^2 \right) \,.$$

Comparing with Eqs. (20.6.36) and (20.6.37), we find

$$\int_1^{\infty} d\omega \, \omega^{-1-s} W_1 \left(\frac{\omega q^2}{2m_N}, q^2 \right) \to \frac{1}{2} \sum_{ij} Q_i^2 \left[\left(g_q^2/g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,, \tag{20.6.38}$$

$$\int_1^{\infty} d\omega \, \omega^{-s} \nu W_2 \left(\frac{\omega q^2}{2m_N}, q^2 \right) \to m_N \sum_{ij} Q_i^2 \left[\left(g_q^2/g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,. \tag{20.6.39}$$

The W_r satisfying these equations may evidently be expressed in a manner similar to the equations (20.6.9) and (20.6.10) of the parton model

$$W_1 \left(\frac{\omega q^2}{2m_N}, q^2 \right) \to \sum_i Q_i^2 \mathcal{F}_i(1/\omega, q^2) \,, \tag{20.6.40}$$

$$\nu W_2 \left(\frac{\omega q^2}{2m_N}, q^2 \right) \to \frac{2m_N}{\omega} \sum_i Q_i^2 \mathcal{F}_i(1/\omega, q^2) \,, \tag{20.6.41}$$

where $\mathcal{F}_i$ is a parton distribution function, now defined by the moment equations

$$\int_0^1 dx \, x^{s-1} \mathcal{F}_i(x, q^2) = \frac{1}{2} \sum_j \left[\left(g_q^2/g_\ell^2 \right)^{8\pi^2 c(s)/b} \right]_{ij} \langle \mathcal{O}_{sj} \rangle \,. \tag{20.6.42}$$

In particular, we see that asymptotic freedom implies not only a corrected version of Bjorken scaling, but also the Callan–Gross relation (20.6.12) between W_1 and W_2.

There is an elegant reformulation of the moment equations (20.6.42) due to Altarelli and Parisi[15] that has become widely used in studies of deep-inelastic scattering. Note that Eqs. (20.6.42) together with the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

renormalization group equation (20.3.8) imply the differential equations

$$q^2\frac{d}{dq^2}\int_0^1 dx\, x^{s-1}\mathscr{F}_i(x,q^2) = -g_q^2\sum_j c_{ij}(s)\int_0^1 dx\, x^{s-1}\mathscr{F}_j(x,q^2)\,. \tag{20.6.43}$$

These equations, together with the 'initial' condition at the renormalization point ℓ^2

$$\int_0^1 dx\, x^{s-1}\mathscr{F}_i(x,\ell^2) = \tfrac{1}{2}\langle\mathcal{O}_{si}\rangle\,, \tag{20.6.44}$$

have a unique solution, so they can be used instead of the moment equations. Now, Eq. (20.6.43) is satisfied by the solution of the differential equation for $\mathscr{F}_i$:

$$q^2\frac{d}{dq^2}\mathscr{F}_i(x,q^2) = \frac{g_q^2}{4\pi^2}\sum_j\int_x^1\frac{dy}{y}P_{ij}\left(\frac{x}{y}\right)\mathscr{F}_j(y,q^2)\,, \tag{20.6.45}$$

where the matrix function $P_{ij}(z)$ is subject to the conditions

$$\int_0^1 z^{s-1}P_{ij}(z)\,dz = -4\pi^2 c_{ij}(s)\,. \tag{20.6.46}$$

(The factor $4\pi^2$ is conventional.)

The matrix $c_{ij}(s)$ was calculated in quantum chromodynamics by Georgi and Politzer[16] and Gross and Wilczek.[17] They assumed that there are N flavors of quarks that are sufficiently light to be treated as massless, while quarks of all other flavors are treated as very heavy and integrated out, so that they can be ignored except for their effect on the strong coupling constant. Their results in the basis provided by the operators (20.6.23) and (20.6.24) are

$$c_{00}(s) = \frac{1}{2\pi^2}\left\{C_1\left[\frac{1}{12} - \frac{1}{s(s-1)} - \frac{1}{(s+1)(s+2)} + \sum_{t=2}^{s}\frac{1}{t}\right] + \frac{N}{3}C_2\right\}\,, \tag{20.6.47}$$

$$c_{f0}(s) = \frac{1}{\pi^2}C_2\left[\frac{1}{s+2} + \frac{2}{s(s+1)(s+2)}\right]\,, \tag{20.6.48}$$

$$c_{0f}(s) = \frac{1}{8\pi^2}C_3\left[\frac{1}{s+1} + \frac{2}{s(s-1)}\right]\,, \tag{20.6.49}$$

$$c_{ff'}(s) = \frac{1}{8\pi^2}C_3\left[1 - \frac{2}{s(s+1)} + 4\sum_{t=2}^{s}\frac{1}{t}\right]\delta_{ff'}\,, \tag{20.6.50}$$

where 0 and f denote the operators (20.6.24) and (20.6.23), respectively; the constants C_1 and C_2 are defined by Eqs. (17.5.33) and (17.5.34) (with only a single quark flavor included in the trace in Eq. (17.5.34)); N is the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

number of quark flavors; and C_3 is defined (using the notation of Section 17.4) by

$$t_\alpha t_\alpha = C_3 g^2 1 . \tag{20.6.51}$$

In the realistic case of an $SU(3)$ gauge group with quarks in its defining representation **3**, these numbers are

$$C_1 = 3 , \qquad C_2 = \tfrac{1}{2} , \qquad C_3 = \tfrac{4}{3} . \tag{20.6.52}$$

These rather complicated results become simpler when expressed in term of the Altarelli–Parisi functions. It is straightforward to check that Eq. (20.6.46) is satisfied if

$$P_{ff'} = \delta_{ff'} \left[\frac{4}{3} \left(\frac{1+x^2}{(1-x)_+} \right) + 2\delta(1-x) \right] , \tag{20.6.53}$$

$$P_{f0} = x^2 - x + \frac{1}{2} , \tag{20.6.54}$$

$$P_{0f} = \frac{4}{3} \left[\frac{2}{x} - 2 + x \right] , \tag{20.6.55}$$

$$P_{00} = 6 \left[\frac{1-x}{x} + x(1-x) + \frac{x}{(1-x)_+} + \frac{11}{12}\delta(1-x) \right] - \frac{N}{3}\delta(1-x) , \tag{20.6.56}$$

where in an integral over x up to $x = 1$, $1/(1-x)_+$ is defined by

$$\frac{f(x)}{(1-x)_+} \equiv \frac{f(x) - f(1)}{1-x} . \tag{20.6.57}$$

For each s there is an obvious $(N-1)$-fold degenerate eigenvalue of the matrix $c(s)$, equal to the coefficient of $\delta_{ff'}$ in Eq. (20.6.50):

$$c(s, \text{adjoint}) = \frac{1}{8\pi^2} C_3 \left[1 - \frac{2}{s(s+1)} + 4 \sum_{t=2}^{s} \frac{1}{t} \right] , \tag{20.6.58}$$

with eigenoperators given by all independent linear combinations of the operators (20.6.23) with coefficients a_f satisfying $\sum_f a_f = 0$, which belong to the adjoint representation of the unbroken global $SU(N)$ symmetry group of quantum chromodynamics with N quark flavors. In addition for each s there are two eigenoperators belonging to the singlet representation of $SU(N)$, given by linear combinations of the operator (20.6.24) and the sum over f of the operators (20.6.23). These eigenoperators and the corresponding eigenvalues can be found by diagonalizing the 2×2 matrix:

$$c(s)_{\text{singlet}} = \begin{bmatrix} c_{00}(s) & c_{0f}(s)N \\ c_{f0}(s) & c(s, \text{adjoint}) \end{bmatrix} . \tag{20.6.59}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

For $s = 2$ this matrix takes the form

$$c(2)_{\text{singlet}} = \begin{bmatrix} NC_2/6\pi^2 & NC_3/6\pi^2 \\ C_2/3\pi^2 & C_3/3\pi^2 \end{bmatrix} . \tag{20.6.60}$$

This has one zero eigenvalue, corresponding to the linear combination of $\mathcal{O}_{20}^{\mu\nu}$ and $\sum_f \mathcal{O}_{2f}^{\mu\nu}$ that is equal to the energy-momentum tensor, which like J^μ is not renormalized. The other eigenvalue of the matrix (20.6.60) is given by its trace, $NC_2/6\pi^2 + C_3/3\pi^2$. Now, the minimum of the eigenvalues of $c_{ij}(s)$ for a given s must be at least as large as the minimum of the eigenvalues of $c_{ij}(s')$ for any $s' < s$, because otherwise for sufficiently large q^2 the integral (20.6.42) would eventually become larger for s than for s', in contradiction with the fact that this integral is a strictly decreasing function of s. Because the minimum eigenvalue for $s = 2$ is zero, we can conclude that all of the other eigenvalues for $s > 2$ are positive. In fact, they are all positive-definite, because there are no unrenormalized operators for $s > 2$. Hence strict Bjorken scaling occurs only in the extreme limit where $g_q^2 \to 0$, where only the contribution of the energy-momentum tensor survives. The prediction of violations of strict Bjorken scaling have been confirmed experimentally by exhaustive studies of deep-inelastic electron–nucleon and muon–nucleon scattering.

20.7 Renormalons*

Since the beginning of quantum field theory theorists have wondered whether the perturbation series for physical matrix elements converges, and if it does not, then what can be done about it? Very early in the modern period Dyson[18] observed that the number of diagrams of nth order typically grows as $n!$, which suggested that the perturbation series has zero radius of convergence.

There is a well-known technique known as a Borel transformation[19] for improving the convergence of a power series whose nth order term grows as $n!$, either to make the series converge, or at least to improve the behavior of the series so that it can be used as an asymptotic expansion for a larger range of coupling constants. For a given series

$$F(g) = \sum_n f_n g^n \tag{20.7.1}$$

we consider the related series

$$B(z) \equiv \sum_n f_n z^n / n! \, . \tag{20.7.2}$$

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

If f_n grows no faster than $n!$ then $B(z)$ will generally have at least a finite radius of convergence. The question is, can we recover the original series (20.7.1) from the resummed series (20.7.2)? Using the familiar formula

$$\int_0^\infty \exp(-z/g)\, z^n\, dz = n!\, g^{n+1}$$

we see that, at least formally,

$$gF(g) = \int_0^\infty \exp(-z/g)\, B(z)\, dz\ . \tag{20.7.3}$$

Singularities of $B(z)$ anywhere in the complex plane limit the radius of convergence of the series (20.7.2), but this is not an insuperable problem if these singularities are off the positive real axis. To calculate $F(g)$ using Eq. (20.7.3) we need $B(z)$ only for real positive values of z less than or of order g, which can be obtained from the power series (20.7.1) if the singularities of $B(z)$ in the complex plane are all at distances from the origin much greater than g. Even if a few poles z_1, z_2, etc. have moduli of order g or less, we can calculate $B(z)$ out to values of z of order g by using the power series for $(z - z_1)(z - z_2)\cdots B(z)$, but for this purpose we have to know where the poles are.

Singularities of $B(z)$ on the positive real axis are much worse, for they invalidate Eq. (20.7.3). The contour in this integral may be distorted to avoid singularities on the positive real axis, but then we have an ambiguity: do we distort the contour above or below the singularity?

This section will show that some of the singularities of the Borel transform $B(z)$ are associated with solutions of the classical field equations known as *instantons*, while other singularities, known as *renormalons*, are associated with terms in the operator product expansion. In quantum chromodynamics it is the renormalons that obstruct the use of the Borel transformation to sum the perturbation series.

It was Lipatov[20] who in 1976 showed that some of the singularities of the Borel transform $B(z)$ are associated with the existence of classical solutions of the field equations. Consider a function $F(g)$ defined by a Euclidean path integral:

$$F(g) \equiv \int [d\phi] \exp\left(I[\phi, g]\right)\ . \tag{20.7.4}$$

(The use of Euclidean path integration is discussed in Appendix A of Chapter 23.) The coefficients in the power series (20.7.1) are given by

$$\begin{aligned} f_n &= \frac{1}{2\pi i} \int [d\phi] \oint dg\ g^{-n-1} \exp\left(I[\phi, g]\right) \\ &= \frac{1}{2\pi i} \int [d\phi] \oint dg\ \exp\Big(I[\phi, g] - (n+1)\ln g\Big)\ , \end{aligned} \tag{20.7.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

where $\oint$ denotes the integral counterclockwise over a closed curve in the complex g plane surrounding the point $g = 0$. For very large n, it is reasonable to suppose that the integral is dominated by the point ϕ_n, g_n where the argument of the exponential in the last line of Eq. (20.7.5) is stationary in both ϕ and g:

$$\left.\frac{\delta I[\phi, g_n]}{\delta \phi(x)}\right|_{\phi=\phi_n} = 0 , \tag{20.7.6}$$

$$\left.\frac{\partial I[\phi_n, g]}{\partial g}\right|_{g=g_n} = \frac{n+1}{g_n} . \tag{20.7.7}$$

For instance, suppose that $I[\phi, g]$ is the action for a massless scalar field

$$I[\phi, g] = -\frac{1}{2}\int \partial_i\phi\partial_i\phi \, d^4x - \frac{g}{24}\int \phi^4 \, d^4x , \tag{20.7.8}$$

the sum running over the Euclidean coordinate directions 1, 2, 3, 4. Then the field equation (20.7.6) reads

$$\Box\phi_n = \tfrac{1}{6} g_n \phi_n^3 . \tag{20.7.9}$$

We will see that g_n is negative, so the solution has

$$\phi_n(x) = (-g_n)^{-1/2}\chi(x) , \tag{20.7.10}$$

with $\chi(x)$ the g-independent solution of the equation

$$\Box\chi = -\tfrac{1}{6}\chi^3 . \tag{20.7.11}$$

The condition (20.7.7) tells us that

$$-\frac{1}{24}\int d^4x \, \phi_n^4 = \frac{n+1}{g_n}$$

or in terms of the rescaled field (20.7.10)

$$g_n = -\frac{1}{24(n+1)}\int d^4x \, \chi^4 . \tag{20.7.12}$$

At this stationary point the action (20.7.8) becomes

$$\begin{aligned} I[\phi_n, g_n] &= -\frac{1}{2}\int \partial_i\phi_n\partial_i\phi_n \, d^4x - \frac{g_n}{24}\int \phi_n^4 \, d^4x = \frac{g_n}{24}\int \phi_n^4 \, d^4x \\ &= -n - 1 . \end{aligned} \tag{20.7.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

Evaluating Eq. (20.7.5) at the stationary point then gives for $n \to \infty$**

$$f_n \approx g_n^{-n-1} \exp\Big(I(\phi_n, g_n)\Big) = (n+1)^{n+1} \left(-\frac{e}{24}\int \chi^4\, d^4x\right)^{-n-1}$$
$$\approx n! \left(-\frac{1}{24}\int \chi^4\, d^4x\right)^{-n} . \tag{20.7.14}$$

The leading singularity in $B(z)$ is therefore a pole at $z = z_1$, where

$$z_1 = -\frac{1}{24}\int \chi^4\, d^4x\,. \tag{20.7.15}$$

Because this is negative, it does not prevent us from carrying out the integration in Eq. (20.7.3). To calculate the pole position (20.7.15), we note that the field equation (20.7.11) has the solution

$$\chi = \frac{4\sqrt{3}a}{r^2 + a^2}\,, \tag{20.7.16}$$

where $r = (x_i x_i)^{1/2}$, and a is an arbitrary parameter. This solution of the field equation is an elementary example of the 'instanton' solutions to be discussed in Section 23.5. (These are called instantons because instead of being concentrated along a worldline they are concentrated near a point in spacetime — in this case, the origin.) Fortunately the pole position does not depend on a:

$$z_1 = -96\pi^2 a^4 \int_0^\infty \frac{r^2\, dr^2}{(r^2 + a^2)^4} = -16\pi^2\,. \tag{20.7.17}$$

We see that the perturbation series for $B(z)$ can be used in Eq. (20.7.3) if $g \ll 16\pi^2$. If $g/16\pi^2$ is of the order or greater than unity, we may still be able to calculate $B(z)$ from the perturbation series for $(z + 16\pi^2)B(z)$.

We will see in Section 23.5 that there are instanton solutions in non-Abelian gauge theories like quantum chromodynamics, but these also yield relatively harmless singularities of $B(z)$ on the negative real axis. The real problem in quantum chromodynamics is with a different class of singularities, known as renormalons.[21] These were first discovered through the realization[22] that single $2n$-th order diagrams, like that shown in Figure 20.3, can make individual contributions that grow like $n!$, and hence according to Eq. (20.7.2) may lead to additional singularities in $B(z)$. In this particular case the singularity is known as an *infrared* renormalon,

** Here the symbol '$\approx$' should be interpreted to mean 'asymptotically equal up to factors of constants and powers of n.' These factors arise from the factor $\sqrt{12\pi n}$ in Stirling's formula for $(n+1)!$, from the ratio of $(n+1)!$ and $n!$, and from the integral over the fluctuations of g and $\phi(x)$ around the stationary point. Since we are not going to calculate the factors from the last source, there is no point in keeping track of the factors from the first two sources either.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

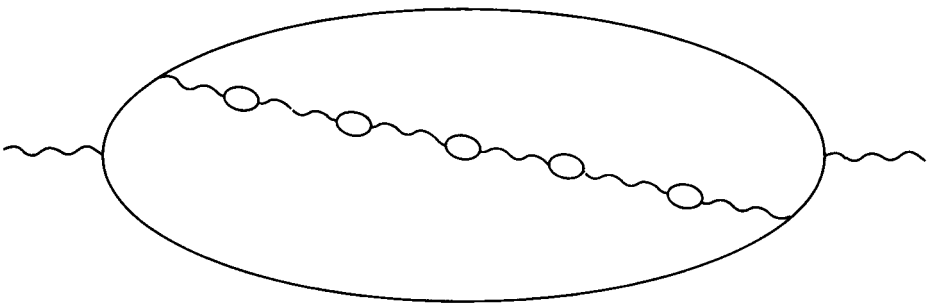

Figure 20.3. One of a class of N-loop Feynman diagrams that grow as $N!$. Solid lines are fermions; wavy lines are gauge bosons.

as it arises from virtual momenta that are much smaller than those used to define the running quantum chromodynamics coupling g_μ). Fortunately, it is possible to use the operator product expansion to locate the infrared renormalons, without having to look at individual Feynman diagrams.

For a simple but important example, consider the sum $\Pi^{\mu\nu}_{\alpha\beta}(q)$ of all vacuum diagrams in quantum chromodynamics with insertions of four-vector currents J^μ_α and J^ν_β carrying a four-momentum q into and out of the diagram. ($\Pi^{\mu\nu}(q)$ gives the hadronic contribution to the electroweak vacuum polarization, and its imaginary part yields the cross section for e^+–e^- and electron–antineutrino annihilation into hadrons.) As we saw in Section 20.5, this receives contributions that go as q^2 (the Fourier transform of x^{-6}) from the operator 1, contributions that go as q^{-2} from the operator $F^{\mu\nu}_\alpha F_{\alpha\mu\nu}$, contributions that go as q^{-4} from four-fermion operators, and so on. (For instance, the contribution of the operator $F^{\mu\nu}_\alpha F_{\alpha\mu\nu}$ arises from diagrams like Figure 20.3, with a momentum much less than q flowing through the chain of bubbles.) Dimensional analysis tells us that these momentum-dependent factors must be accompanied with vacuum matrix elements proportional to Λ^0, Λ^4, Λ^6, etc. But if we calculate Feynman diagrams using a running coupling defined at a scale $\mu \gg \Lambda$, where the coupling constants are small, then according to Eq. (18.7.7)

$$\Lambda^2 = \mu^2 \exp\left(\frac{12\pi}{(33-2n_f)\alpha_s(\mu)}\right) , \tag{20.7.18}$$

where n_f is the number of quark flavors with mass much less than μ. Operators of dimension $d > 0$ would make a contribution in the operator product expansion with the coupling-constant dependence

$$\Lambda^d \propto \exp\left(\frac{6\pi d}{(33-2n_f)\alpha_s(\mu)}\right) . \tag{20.7.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

In quantum chromodynamics, perturbation theory yields a series in powers of $\alpha_s \equiv g^2/4\pi$ rather than g, so we would write Eqs. (20.7.1) and (20.7.3) in terms of α_s:

$$F(\alpha_s) = \sum_n f_n \alpha_s^n \,, \tag{20.7.20}$$

$$\alpha_s F(\alpha_s) = \int_0^\infty \exp(-z/\alpha_s)\, B(z)\, dz \,. \tag{20.7.21}$$

The presence of terms in $\Pi^{\mu\nu}$ with coupling-constant dependence (20.7.19) indicates that $B(z)$ must have singularities (not necessarily poles) at

$$z_1 = \frac{6\pi d}{(33 - 2n_f)} \,. \tag{20.7.22}$$

These are on the positive real axis, and make the integral (20.7.21) ambiguous. Thus Borel summation cannot be used to deal with low-energy effects in quantum chromodynamics.

The fact that diagrams with small virtual momenta impede the use of perturbation theory is of course nothing new. As we saw in Section 20.2, the whole point of the operator product expansion is to separate the parts of Feynman diagrams where every line carries a large momentum, which in asymptotically free theories can be calculated using perturbation theory, from the contribution of the parts of Feynman diagrams through which small momenta flow, which cannot be calculated perturbatively.

Appendix Momentum Flow: The General Case

In this appendix we shall consider the asymptotic behavior of an amplitude in a general renormalizable quantum field theory when the momenta of any set of two or more external lines become large, taking into account operators containing arbitrary numbers of field factors and derivatives, with dimensionality up to some limit N. To deal with this problem we shall have to introduce a more compact notation than that of Sections 20.1 or 20.2. A letter ℓ, ℓ', etc. will denote a set of external lines i of specified types entering or leaving a Feynman diagram or part of a Feynman diagram. A letter k, k', etc. will denote the set of four-momenta k_i of such lines, subject to the condition that their sum has some fixed value p. The amplitude $\Gamma_{\ell\ell'}(k,k',p)$ is the sum of all graphs with a set ℓ of incoming lines carrying momenta k and a set ℓ' of outgoing lines carrying momenta k', including the final bare propagators for the set ℓ but not ℓ'.

As shown by the power-counting theorem[3] quoted in Section 12.1, the part of the region of integration where the momenta of order k flow only through a subgraph with external lines ℓ and ℓ'' makes a contribution to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

$\Gamma_{\ell\ell'}(k,k',p)$ of order $k^{d(\ell,\ell'')}$, where $d(\ell,\ell'')$ is the dimensionality (in powers of mass or momentum) of this subgraph:

$$d(\ell,\ell'') = 4 - \sum_{i\in\ell,\ell''}(1+s_i) - \sum_{i\in\ell}(2-2s_i)\,. \tag{20.A.1}$$

(The last term in Eq. (20.A.1) arises from the propagators for the lines in the set ℓ.) It will be convenient to rewrite Eq. (20.A.1) as

$$d(\ell,\ell') = 4 - 4n(\ell) - N(\ell') + N(\ell)\,, \tag{20.A.2}$$

where $n(\ell)$ is the number of lines in the set ℓ, and $N(\ell)$ is the total dimensionality of the fields for these lines:

$$n(\ell) \equiv \sum_{i\in\ell} 1\,, \qquad N(\ell) \equiv \sum_{i\in\ell}(s_i+1)\,. \tag{20.A.3}$$

Here s_i is the 'spin' of a line of type i, in the sense used in Section 12.3: the dimensionality of the field of type i is $1+s_i$, and the bare propagator of such a field goes as k^{-2+2s_i}. (For scalars and gauge bosons, $s_i = 0$; for spin 1/2, $s_i = 1/2$.).

By taking account of the asymptotic behavior associated with these subgraphs, we wish to show that as $k \to \infty$ (all components going to infinity together, in generic directions), with k' and p fixed, the asymptotic behavior of $\Gamma_{\ell\ell'}(k,k',p)$ in each order of perturbation theory is of the form

$$\Gamma_{\ell\ell'}(k,k',p) \to \sum_{\mathcal{O}}^{(N)} U^{\ell}_{\mathcal{O}}(k)\, F_{\mathcal{O},\ell'}(k',p) + o\left(k^{4-4n(\ell)+N(\ell)-N}\right)\,, \tag{20.A.4}$$

where the sum runs over operators $\mathcal{O}$ with dimensionality $N(\mathcal{O}) \le N$; the function $U^{\ell}_{\mathcal{O}}(k)$ is of order $k^{4+N(\ell)-4n(\ell)-N(\mathcal{O})}$; and $o(k^A)$ denotes terms that vanish faster (by at least one factor $1/k$) than k^A.

To isolate the contribution of operators of dimensionality $\le N$, we shall define an 'N-irreducible' amplitude $I^N_{\ell\ell'}(k,k',p)$ as the sum of all graphs for $\Gamma_{\ell\ell'}(k,k',p)$ in which the lines in the set ℓ cannot be disconnected from the lines in the set ℓ' by cutting through any set ℓ'' of internal lines with $N(\ell'') \le N$. Since the difference $\Gamma - I^N$ consists of graphs that *can* be disconnected in this way, it may be written as

$$\Gamma_{\ell\ell'}(k,k',p) - I^N_{\ell\ell'}(k,k',p) = \sum_{\ell''}^{(N)} \int dk''\, I^N_{\ell\ell''}(k,k'',p)\,\Gamma_{\ell''\ell'}(k'',k',p)\,, \tag{20.A.5}$$

where $\sum^{(N)}_{\ell''}$ denotes the sum over sets ℓ'' of particle lines with $N(\ell'') \le N$, and $\int dk''$ is the integral over the components of the four-momenta in the set ℓ'', subject to the constraint that the sum of these momenta is p.

The asymptotic behavior of the kernel $I^N_{\ell,\ell''}(k,k'',p)$ is much simpler than that of $\Gamma_{\ell,\ell''}(k,k'',p)$. For $k \to \infty$ with k'' and p fixed it is dominated

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

by the part of the region of integration where every internal line carries a momentum of order k, which gives an asymptotic behavior $k^{d(\ell,\ell'')}$, because the terms where only some subregion carries such large momenta would have to be connected to the rest of the graph by a bridge ℓ''' of lines with $N(\ell''') > N \geq N(\ell'')$, and this would be smaller as $k \to \infty$ by at least a factor $k^{N(\ell''')-N-1}$. Thus differentiating $I^N_{\ell,\ell''}(k,k'',p)$ with respect to any component of k'' or p reduces its asymptotic behavior to order $k^{d(\ell,\ell'')-1}$. But differentiating d times lowers the asymptotic behavior to $k^{d(\ell,\ell'')-d}$ only if $d \leq N - N(\ell'') + 1$, because for higher derivatives we get a larger contribution from the region of integration where the subgraph carrying momenta of order k is connected to the rest of the graph by a bridge of lines ℓ''' with total dimensionality greater than N, with the derivatives all acting on lines carrying momenta of order k'' or p. Therefore in order to take into account the contribution of operators containing derivatives of fields, we may write the asymptotic behavior of I^N as

$$I^N_{\ell\ell''}(k,k'',p) = \sum_{\nu : d_\nu + N(\ell'') \leq N} I^N_{\ell\ell''\nu}(k)\, P_{\ell''\nu}(k'',p) + o\left(k^{d(\ell,\ell'')-N+N(\ell'')}\right), \quad (20.A.6)$$

where $P_{\ell''\nu}(k'',p)$ are a complete set of homogeneous polynomials of order d_ν in the $n(\ell'')$ momenta k'' and p, and $I^N_{\ell\ell''\nu}(k)$ are functions only of k, of order $k^{d(\ell,\ell'')-d}$ as $k \to \infty$.

We cannot immediately use Eq. (20.A.6) in Eq. (20.A.5), because the integral over k'' receives important contributions from the region where some of the k'' are of order k. To deal with this, we use mathematical induction: we assume that Eq. (20.A.4) holds up to some given order in perturbation theory, and use it on the right-hand side of Eq. (20.A.5) to calculate the asymptotic behavior of Γ to the next order of perturbation theory. We rewrite Eq. (20.A.5) in the form

$$\begin{aligned}\Gamma_{\ell\ell'}(k,k',p) &= I^N_{\ell\ell'}(k,k',p) \\ &+ \sum_{\ell''}^{(N)} \int dk''\, I^N_{\ell\ell''}(k,k'',p) \left[\Gamma_{\ell''\ell'}(k'',k',p) - \sum_{\mathcal{O}}^{(N)} U^{\ell''}_{\mathcal{O}}(k'')\, F_{\mathcal{O},\ell'}(k',p)\right] \\ &+ \sum_{\mathcal{O}}^{(N)} F_{\mathcal{O},\ell'}(k',p) \sum_{\ell'' :\, N(\ell'') \leq N} \int dk''\, I^N_{\ell\ell''}(k,k'',p) U^{\ell''}_{\mathcal{O}}(k'')\,. \end{aligned} \quad (20.A.7)$$

According to Eq. (20.A.4), the quantity in square brackets in the second term on the right-hand side of Eq. (20.A.7) vanishes for $k'' \to \infty$ faster than $(k'')^{4-4n(\ell'')+N(\ell'')-N}$, so the product of this factor with a polynomial $P_{\ell''\nu}(k'')$ of order $\nu \leq N - N(\ell'')$ vanishes faster than $(k'')^{4-4n(\ell'')}$, and therefore has a finite integral over the $4(n(\ell'')-1)$ independent components

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

of k''.* Hence we may use Eq. (20.A.6) in this term, and find

$$\Gamma_{\ell\ell'}(k,k',p) \to \sum_{v\,:\,d_v+N(\ell')\le N} I^N_{\ell\ell' v}(k)\,P_{\ell' v}(k',p) + \sum_{\ell''}^{(N)} \sum_{v\,:\,d_v+N(\ell'')\le N} I^N_{\ell\ell'' v}(k) \int dk''\,P_{\ell'' v}(k'',p) \times \left[\Gamma_{\ell''\ell'}(k'',k',p) - \sum_{\mathcal{O}}^{(N)} U^{\ell''}_{\mathcal{O}}(k'')\,F_{\mathcal{O},\ell'}(k',p)\right] + \sum_{\mathcal{O}}^{(N)} F_{\mathcal{O},\ell'}(k',p) \sum_{\ell''}^{(N)} \int dk''\,I^N_{\ell\ell''}(k,k'',p) U^{\ell''}_{\mathcal{O}}(k'')\,, \quad (20.A.8)$$

the correction being smaller than the terms shown by a factor $1/k$. (Of course, the first term on the right-hand side of Eq. (20.A.8) is present only if $N(\ell') \le N$.)

Now, for each value of ℓ and v with $d_v + N(\ell) \le N$, there is an operator $\mathcal{O}$ with field factors corresponding to the lines in ℓ and with d_v derivatives, such that in zeroth-order perturbation theory the vertex function, with incoming momentum p carried by the operator $\mathcal{O}$ and outgoing momenta k carried by external lines ℓ, is the polynomial $P_{\ell v}(k,p)$. Then the corresponding complete vertex function for a renormalized operator $\mathcal{O}_R = \sum_{\mathcal{O}'} Z_{\mathcal{O},\mathcal{O}'}\mathcal{O}'$ is

$$F_{\mathcal{O},\ell}(k,p) = \sum_{\mathcal{O}'} Z_{\mathcal{O},\mathcal{O}'} \left\{ \delta_{\ell_{\mathcal{O}'},\ell}\,P_{\ell v_{\mathcal{O}'}}(k,p) + \int dk'\,P_{\ell_{\mathcal{O}'} v_{\mathcal{O}'}}(k',p)\,\Gamma_{\ell_{\mathcal{O}'}\ell}(k',k,p) \right\}, \quad (20.A.9)$$

where $\ell_{\mathcal{O}}$ and $v_{\mathcal{O}}$ label the types of fields and spacetime derivatives in the operator $\mathcal{O}$. We see then that Eq. (20.A.4) is satisfied, with

$$U^{\ell}_{\mathcal{O}}(k) = \sum_{\mathcal{O}'}^{(N)} I^N_{\ell\mathcal{O}'}(k) \left[Z^{-1}_{\mathcal{O}',\mathcal{O}} - \int dk''\,U^{\ell_{\mathcal{O}'}}_{\mathcal{O}}(k')\,P_{\mathcal{O}'}(k'') \right] + \sum_{\ell''}^{(N)} \int dk''\,I^N_{\ell\ell''}(k,k'')\,U^{\ell}_{\mathcal{O}''}(k'')\,, \quad (20.A.10)$$

* This overlooks the possibility that even though power-counting indicates the convergence of the integral in the second term on the right-hand side of Eq. (20.A.7) over the region where all k'' go to infinity together, for $n(\ell'') \ge 3$ subintegrations may diverge. For this reason the arguments given in this appendix do not constitute a proof of the operator product expansion except for the simple case treated in Section 20.2, where we consider the momenta of just two external lines to go to infinity, and we look for the terms in the power series expansion associated with operators that are quadratic in the fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

in which we now use the abbreviations

$$I_{\ell\mathcal{O}}(k) \equiv I_{\ell\,\ell_{\mathcal{O}}\,\nu_{\mathcal{O}}}(k)\,, \qquad P_{\mathcal{O}}(k) \equiv P_{\ell_{\mathcal{O}}\,\nu_{\mathcal{O}}}(k)\,. \tag{20.A.11}$$

(The p dependence of $\Gamma_{\ell_{\mathcal{O}''}\ell'}$ is dropped for the same reason as in Section 20.2: whether or not k'' is comparable with k, p is negligible here compared with k.) We will define the normalization constant $Z_{\mathcal{O}',\mathcal{O}}$ so that at a renormalization point $k(\mu)$, $p(\mu)$ the function $F_{\mathcal{O},\ell}(k(\mu),p(\mu))$ has the same value $\delta_{\ell\ell_{\mathcal{O}}}P_{\mathcal{O}}(k(\mu),p(\mu))$ that it would have in the absence of interactions:

$$\delta_{\ell\ell_{\mathcal{O}}}P_{\mathcal{O}}(k(\mu),p(\mu)) = F_{\mathcal{O},\ell}(k(\mu),p(\mu)) = \sum_{\mathcal{O}'} Z_{\mathcal{O},\mathcal{O}'}$$
$$\times\left\{\delta_{\ell_{\mathcal{O}'},\ell}\,P_{\ell\nu_{\mathcal{O}'}}(k(\mu),p(\mu)) + \int dk'\,P_{\ell_{\mathcal{O}'}\nu_{\mathcal{O}'}}(k',p(\mu))\,\Gamma_{\ell_{\mathcal{O}'}\ell}(k',k(\mu),p(\mu))\right\}. \tag{20.A.12}$$

For $\Gamma = 0$ this has a solution $Z = 1$ which is unique (because the polynomials for a given set of lines are supposed to be linearly independent), so by continuity Eq. (20.A.12) will have a unique solution for coupling constants in some finite range. Eq. (20.A.10) therefore provides a recursive definition of the coefficient functions $U^{\ell}_{\mathcal{O}}(k)$ appearing in the general operator product expansion (20.A.4).

Problems

1. Consider a theory of a fermion field ψ interacting with a scalar field ϕ with interactions of the form $\bar{\psi}\psi$ and ϕ^4. List the operators that appear in the operator product expansion of $\bar{\psi}(x)\psi(0)$ with coefficient functions that (judging from perturbation theory) are singular and non-vanishing for $x \to 0$. Describe how you would calculate these coefficient functions to one-loop order.

2. Consider quantum chromodynamics with N massless quarks, and define the spectral functions of the scalar and pseudoscalar quark bilinears by

$$\sum_N \langle \mathrm{VAC}|\bar{\psi}(0)\lambda_\alpha\psi(0)|\mathrm{VAC}\rangle\langle \mathrm{VAC}|\bar{\psi}(0)\lambda_\beta\psi(0)|\mathrm{VAC}\rangle^*$$
$$= (2\pi)^{-3}\theta(p^0)\rho^S_{\alpha\beta}(-p^2)\,,$$
$$\sum_N \langle \mathrm{VAC}|\bar{\psi}(0)\gamma_5\lambda_\alpha\psi(0)|\mathrm{VAC}\rangle\langle \mathrm{VAC}|\bar{\psi}(0)\gamma_5\lambda_\beta\psi(0)|\mathrm{VAC}\rangle^*$$
$$= (2\pi)^{-3}\theta(p^0)\rho^P_{\alpha\beta}(-p^2)\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

where λ_α are a complete set of traceless Hermitian $N \times N$ matrices, normalized so that $\mathrm{Tr}(\lambda_\alpha \lambda_\beta) = 2\delta_{\alpha\beta}$. What spectral function sum rules are satisfied by linear combinations of $\rho^S_{\alpha\beta}(\mu^2)$ and $\rho^P_{\alpha\beta}(\mu^2)$?

3. Derive Eq. (20.6.8) in the parton model from the formulas (8.7.7) and (8.7.38) for Compton scattering.

4. List the gauge-invariant symmetric traceless tensors of twist four in quantum chromodynamics.

5. In the massless scalar field theory with interaction $-g\phi^4/24$ (with $g > 0$), where in the complex plane would you expect the function (20.7.3) to have renormalon singularities?

References

1. K. Wilson, *Phys. Rev.* **179**, 1499 (1969)

2. W. Zimmerman, in *Lectures on Elementary Particles and Quantum Field Theory — 1970 Brandeis University Summer Institute in Theoretical Physics*, eds. S. Deser, H. Pendleton, and M. Grisaru, (MIT Press, Cambridge, 1970).

3. S. Weinberg, *Phys. Rev.* **118**, 838 (1960).

4. C. Bernard, A. Duncan, J. LoSecco, and S. Weinberg, *Phys. Rev.* **D12**, 792 (1975).

5. S. Weinberg, *Phys. Rev. Lett.* **18**, 507 (1967). For the general case, see Ref. 4.

6. K. Kawarabayashi and M. Suzuki, *Phys. Rev. Lett.* **16**, 255 (1966); Riazuddin and Fayyazuddin, *Phys. Rev.* **147**, 1071 (1966).

7. M. Ademollo, G. Veneziano, and S. Weinberg, *Phys. Rev. Lett.* **22**, 83 (1969).

8. J. F. Donoghue and E. Golowich, *Phys. Rev.* **D49**, 1513 (1994).

9. This result was reported at the 1968 'Rochester' conference at Vienna, and published in E. D. Bloom *et al.*, *Phys. Rev. Lett.* **23**, 930 (1969); M. Breidenbach *et al.*, *Phys. Rev. Lett.* **23**, 935 (1969).

10. J. D. Bjorken, *Phys. Rev.* **179**, 1547 (1969).

11. R. P. Feynman, *Phys. Rev. Lett.* **23**, 1415 (1969); *Photon-Hadron Interactions* (Benjamin, New York, 1972).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

12. C. G. Callan and D. J. Gross, *Phys. Rev. Lett.* **22**, 156 (1969). The reader should be warned that the symbol ω as used by Callan and Gross is $2/\omega$ in the notation used here.

13. N. Christ, B. Hasslacher, and A. H. Mueller, *Phys. Rev.*, **D6**, 3543 (1972).

14. D. Gross and S. Treiman, *Phys. Rev.* **D4**, 1059 (1971).

15. G. Altarelli and G. Parisi, *Nucl. Phys.* **B126**, 298 (1972).

16. H. Georgi and H. D. Politzer, *Phys. Rev.* **D9**, 416 (1974).

17. D. J. Gross and F. Wilczek, *Phys. Rev.* **D9**, 980 (1974).

18. F. J. Dyson, *Phys. Rev.* **85**, 631 (1952).

19. See, e.g., G. N. Hardy, *Divergent Series* (Oxford University Press, Oxford, 1949).

20. L. N. Lipatov, Leningrad Nuclear Physics Institute report, 1976 (unpublished); *Proceedings of the XVIII International Conference on High Energy Physics at Tbilisi*, 1976.

21. G. 't Hooft, in *The Whys of Subnuclear Physics – Proceedings of the 1977 Erice Summer School*, ed. A. Zichichi (Plenum, New York, 1978). For a collection of articles on renormalons and high orders of perturbation theory, see *Large Order Behavior of Perturbation Theory*, eds. J. C. Le Guillou and J. Zinn-Justin (North-Holland, Amsterdam, 1990). For further work on infrared renormalons, see A. H. Mueller, *Phys. Lett. B* **308**, 355 (1993); *Nucl. Phys.* **B250**, 327 (1995); A. Duncan and S. Pernice, *Phys. Rev. D* **51**, 1956 (1995), and articles quoted therein.

22. B. Lautrup, *Phys. Lett.* **76B**, 109 (1977).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.008

21

Spontaneously Broken Gauge Symmetries

The 1961 theorem[1] that broken symmetry implies massless spin zero Goldstone bosons was seen at first as a serious obstacle to the search for broken symmetries in nature. A few years later several authors noted an exception to this theorem, already mentioned in Section 19.2: the Goldstone bosons are absent where the broken symmetry is local, rather than global.[2] Instead, these degrees of freedom show up as the helicity zero states of the vector particles associated with the broken local symmetries, which thereby acquire a mass. This phenomenon, now generally known as the Higgs mechanism, was not at first applied in any sort of realistic theory, perhaps because by the mid-1960s it had become clear that the pion is a Goldstone boson of a spontaneously broken approximate symmetry, and attention therefore shifted away from the effort to avoid Goldstone bosons. But soon after, spontaneously broken local symmetries turned out to provide the natural framework for understanding the weak and electromagnetic interactions of the elementary particles.[3]

21.1 Unitarity Gauge

We saw in Chapter 19 that in a theory with a global symmetry group G that is spontaneously broken to a subgroup H, there is a massless 'Goldstone' boson for every independent broken symmetry, in the sense that the mass matrix M^2_{nm} of real spinless fields $\phi_n(x)$ has a zero eigenvalue with eigenvector $\sum_m (t_\alpha)_{nm} v_m$ for each independent broken symmetry generator t_α of G. (We are here considering the case where these Goldstone bosons are included among the elementary spinless particles represented by scalar or pseudoscalar fields ϕ_n appearing in the Lagrangian; the more general case will be taken up in Section 21.4.) We also saw that we could rotate away these Goldstone modes, by subjecting the fields to a G transformation $\gamma^{-1}(x)$

$$\tilde{\phi}_n(x) = \sum_m \gamma^{-1}_{nm}(x)\phi_m(x)\,, \tag{21.1.1}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

such that the new fields are orthogonal to the Goldstone directions*

$$0 = \sum_{nm} \tilde{\phi}_n(x)(t_\alpha)_{nm} v_m \,, \tag{21.1.2}$$

where v_n is the vacuum expectation value, $v_n \equiv \langle \phi_n(0) \rangle_{\text{VAC}}$. After rotating the fields so that they satisfy Eq. (21.1.2), the Goldstone boson fields then reemerged as the spacetime-dependent parameters in $\gamma(x)$. The point of this procedure was that since the Lagrangian was invariant under the transformation (21.1.1) with constant $\gamma(x)$, all dependence on $\gamma(x)$ dropped out except where $\gamma(x)$ is acted on by derivatives.

On the other hand, if the Lagrangian is invariant not only under constant G transformations but also under G transformations that depend on spacetime position, then the transformation (21.1.1) is a true symmetry of the theory, and all dependence on $\gamma(x)$ drops out of the Lagrangian, so that we can simply replace $\phi_n(x)$ everywhere with $\tilde{\phi}_n(x)$. This is a choice of *gauge*, fixed by imposing the condition (21.1.2) on $\tilde{\phi}(x)$ rather than by imposing conditions on the gauge fields themselves. (For instance, in the electrodynamics of a charged scalar field $\phi \equiv \phi_1 + i\phi_2$, we can choose a gauge like Lorentz or Coulomb gauge by imposing conditions $\partial_\mu A^\mu = 0$ or $\nabla \cdot \mathbf{A} = 0$, but we can also choose a gauge by imposing a condition on ϕ, as for instance that ϕ be real, or in other words by rotating the two-vector $\{\text{Re}\phi, \text{Im}\phi\}$ into the 1-direction.) The gauge defined by Eq. (21.1.2) is called *unitarity gauge*,[3a] because in this gauge it will be obvious that the theory does not have any degrees of freedom with negative probability, like timelike gauge bosons. More generally, the unitarity gauge makes manifest the menu of physical particles of the theory.

Eq. (21.1.2) shows that there are no Goldstone boson fields in unitarity gauge. Since the theory is gauge-invariant this means that there are no physical Goldstone bosons, whatever gauge we choose. What about the vector bosons? If the ϕ_n are elementary canonically normalized scalar fields, the Lagrangian will contain a term

$$\mathscr{L}_\phi = -\frac{1}{2} \sum_n \left(\partial_\mu \tilde{\phi}_n - i \sum_{m,\alpha} t^\alpha_{nm} A_{\alpha\,\mu} \tilde{\phi}_m \right)^2 , \tag{21.1.3}$$

where t^α runs over all the generators of the gauge group G. (From now on we shall drop tildes, it being understood henceforth in this section that we are already in unitarity gauge.) We are assuming that the symmetry G is broken by the vacuum expectation value v_n of ϕ_n, so in order to see the

* Here we are working with a real reducible or irreducible representation of the symmetry algebra, for which the matrices it_α are real. The transition to a complex representation is described below.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

nature of the particle spectrum, we define a shifted field ϕ':

$$\phi_n \equiv v_n + \phi'_n \,. \tag{21.1.4}$$

(It is sometimes convenient to take v_n in Eqs. (21.1.2) and (21.1.4) as the vacuum expectation value in the tree approximation. In order to generate a useful perturbation theory it is only necessary that v_n should agree with the true vacuum expectation value in lowest order.) Expanding Eq. (21.1.3) to second order in ϕ' and A, we have:

$$\mathscr{L}_{\phi,\mathrm{QUAD}} = -\frac{1}{2}\sum_n \left(\partial_\mu \phi'_n - i\sum_{m,\alpha} t^\alpha_{nm} A_{\alpha\mu} v_m\right)^2 . \tag{21.1.5}$$

Using Eq. (21.1.2), we see that the ϕ'–A cross term vanishes, yielding

$$\mathscr{L}_{\phi,\mathrm{QUAD}} = -\frac{1}{2}\sum_n \partial_\mu \phi'_n \partial^\mu \phi'_n - \frac{1}{2}\sum_{\alpha\beta} \mu^2_{\alpha\beta} A_{\alpha\mu} A^\mu_\beta \,, \tag{21.1.6}$$

where

$$\mu^2_{\alpha\beta} \equiv -\sum_{nm\ell} t^\alpha_{nm} t^\beta_{n\ell} v_m v_\ell \,. \tag{21.1.7}$$

Combining this with the quadratic terms in the Yang–Mills Lagrangian $-\frac{1}{4} F_{\alpha\mu\nu} F^{\mu\nu}_\alpha$, we see that the vector particles have a mass matrix $\mu^2_{\alpha\beta}$. In our notation the generators t^α_{nm} are proportional to gauge coupling constants, so Eq. (21.1.7) yields vector boson masses that are also proportional to these coupling constants.

Let's take a look at some of the algebraic properties of $\mu^2_{\alpha\beta}$. Since t^α_{nm} is imaginary and antisymmetric (and hence Hermitian) the matrix $\mu^2_{\alpha\beta}$ is real, symmetric, and positive. Also, if a certain real linear combination of generators $\sum_\alpha c_\alpha t_\alpha$ is unbroken, then

$$\sum_{\alpha,m} c_\alpha (t_\alpha)_{nm} v_m = 0 \,, \tag{21.1.8}$$

in which case Eq. (21.1.7) shows that, as noted by Kibble,[2]

$$\sum_\beta \mu^2_{\alpha\beta} c_\beta = 0 \,. \tag{21.1.9}$$

That is, we still have a massless gauge boson for every unbroken gauge symmetry. The converse is also true; from Eq. (21.1.7) we see that for arbitrary real constants c_α

$$\sum_{\alpha\beta} \mu^2_{\alpha\beta} c_\alpha c_\beta = \sum_n \left(\sum_{\alpha,m} c_\alpha \cdot i t^\alpha_{nm} v_m\right)^2 \geq 0 \,, \tag{21.1.10}$$

and this can vanish only if c_α satisfies Eq. (21.1.8).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

In particular, if there is just one unbroken symmetry $\sum_\alpha c_\alpha t_\alpha$, then the general gauge field can be written

$$A^\mu_\alpha = c_\alpha A^\mu + \cdots , \tag{21.1.11}$$

where '$\cdots$' denotes a linear combination of the gauge fields of definite non-zero mass, and c_α is the coefficient of t_α in the single unbroken generator q,

$$q = \sum_\alpha c_\alpha t_\alpha , \tag{21.1.12}$$

so that the field A^μ will appear with zero coefficient in the mass term $-\frac{1}{4}\sum_{\alpha\beta}\mu^2_{\alpha\beta}A^\mu_\alpha A_{\mu\beta}$. In order for the coefficient of $-\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2$ in the kinetic gauge Lagrangian $-\frac{1}{4}\sum_\alpha(\partial_\mu A_{\alpha\nu} - \partial_\nu A_{\alpha\mu})^2$ to have the canonical value 1, the c_α must also be normalized

$$\sum_\alpha c^2_\alpha = 1 . \tag{21.1.13}$$

Note that q *is* the charge to which A^μ is coupled, in the sense that

$$\sum_\alpha t_\alpha A^\mu_\alpha = q A^\mu + \cdots , \tag{21.1.14}$$

where '$\cdots$' again denotes terms involving massive gauge fields. We shall use these general results in studying the electroweak theory in Section 21.3.

These results have been derived here for scalar fields that form a real representation of the gauge group, but they can be straightforwardly converted to a form appropriate for complex representations. We saw in Section 19.6 that a complex scalar field $\chi(x)$ that transforms according to a representation of the gauge group with Hermitian generators T_α may be written as a set of real fields

$$\phi(x) = \begin{pmatrix} \mathrm{Re}\,\chi(x) \\ \mathrm{Im}\,\chi(x) \end{pmatrix} , \tag{21.1.15}$$

which furnish a real representation of the gauge group with generators

$$i t^\alpha = \begin{pmatrix} -\mathrm{Im}\,T^\alpha & -\mathrm{Re}\,T^\alpha \\ \mathrm{Re}\,T^\alpha & -\mathrm{Im}\,T^\alpha \end{pmatrix} . \tag{21.1.16}$$

Inserting Eqs. (21.1.15) and (21.1.16) in Eq. (21.1.7) gives the vector boson mass matrix (21.1.7) as

$$\mu^2_{\alpha\beta} = \mathrm{Re}\left(\langle\chi\rangle^\dagger_{\mathrm{VAC}}, T_\alpha T_\beta \langle\chi\rangle_{\mathrm{VAC}}\right) = \tfrac{1}{2}\left(\langle\chi\rangle^\dagger_{\mathrm{VAC}}, \{T_\alpha, T_\beta\}\langle\chi\rangle_{\mathrm{VAC}}\right) . \tag{21.1.17}$$

Now let's take a closer look at the vector field propagator. Including the quadratic term in the Yang–Mills Lagrangian, the part of the total

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Lagrangian that is quadratic in A is

$$\begin{aligned} &-\tfrac{1}{4}\sum_{\alpha}(\partial_\lambda A_{\alpha\nu}-\partial_\nu A_{\alpha\lambda})^2-\tfrac{1}{2}\sum_{\alpha\beta}\mu^2_{\alpha\beta}\,A_{\alpha\lambda}A_\beta^\lambda \\ &\quad=\tfrac{1}{2}\sum_{\alpha\beta}A_\alpha{}^\nu\mathscr{D}_{\alpha\nu,\beta\lambda}(\partial)A_\beta{}^\lambda+\text{ total derivatives}\,, \end{aligned} \tag{21.1.18}$$

where

$$\mathscr{D}_{\alpha\nu,\beta\lambda}(\partial)=\delta_{\alpha\beta}\left[\eta_{\nu\lambda}\Box-\partial_\nu\partial_\lambda\right]-\mu^2_{\alpha\beta}\,\eta_{\nu\lambda}\,. \tag{21.1.19}$$

Suppose for simplicity that all gauge symmetries are broken, so that $\mu^2_{\alpha\beta}$ has no zero eigenvalues. According to the general rules described in Section 9.4, the gauge-field propagator in momentum space is

$$\Delta_{\alpha\nu,\beta\lambda}(k)=-(\mathscr{D}^{-1})_{\alpha\nu,\beta\lambda}(ik)=\left[(k^2+\mu^2)^{-1}(\eta_{\nu\lambda}+\mu^{-2}k_\nu k_\lambda)\right]_{\alpha\beta}\,. \tag{21.1.20}$$

Because of the $k_\nu\,k_\lambda$ term, the propagator has an asymptotic behavior $\Delta(k)\sim O(k^0)$, which does not allow us to use the usual power-counting arguments to prove renormalizability. Fortunately, as we shall see in the next section, there is another gauge in which renormalizability is obvious, at the price of obscuring the particle content of the theory.

* * *

It is important to note that although Goldstone bosons suddenly reappear in the physical spectrum in the limit of zero gauge couplings, physical matrix elements are perfectly continuous in this limit. This is because in unitarity gauge the gauge bosons do not entirely decouple for zero gauge couplings. Consider the matrix element for a scattering process $A+B\to C+D$, where A and C belong to some representation of the gauge algebra, and B and D belong to some different representation of the gauge algebra. The S-matrix element for this process receives a contribution from vector boson exchange

$$S_{CD,AB}=i(2\pi)^4\delta^4(p_A+p_B-p_C-p_D)\langle C|J_{N\alpha}^\nu|A\rangle\,\Delta_{\alpha\nu,\beta\lambda}(k)\langle D|J_{N\beta}^\lambda|B\rangle\,, \tag{21.1.21}$$

where $k=p_A-p_C=p_D-p_B$, and $J_{N\alpha}^\nu$ is the current to which the gauge bosons are coupled, with the subscript N to remind us that in this gauge we omit the Goldstone boson pole term in this current. This current is proportional to gauge coupling constants, so the only terms in Eq. (21.1.21) that survive in the limit of zero gauge couplings are the ones involving the matrix μ^{-2}, which becomes singular in this limit. Hence for

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

zero gauge couplings, the S-matrix element is

$$S_{CD,AB} \to i(2\pi)^4\delta^4(p_A + p_B - p_C - p_D)k_\nu k_\lambda(\mu^{-2})_{\alpha\beta}\langle C|J_{N\alpha}^{\nu}|A\rangle \frac{1}{k^2}\langle D|J_{N\beta}^{\lambda}|B\rangle \,. \tag{21.1.22}$$

The gauge coupling constants in the currents are cancelled by the gauge coupling constants in the matrix μ^2.

Now let us compare this result with the contribution of Goldstone boson exchange. For vanishing gauge couplings there is a set of Goldstone bosons B_α associated with generator t_α (assuming for simplicity that all gauge symmetries are spontaneously broken) with the kinematic term in the Lagrangian given by $-\frac{1}{2}\partial_\mu\phi_\alpha\partial^\mu\phi_\beta\sum_n Z_{\alpha n}Z_{\beta n}$, where $Z_{\gamma n}$ is the component of the Goldstone boson B_α in the spinless field ϕ_n, defined by Eq. (19.2.39). In order that the Goldstone boson fields be canonically normalized, we must therefore have

$$\sum_n Z_{\alpha n}Z_{\beta n} = \delta_{\alpha\beta} \,. \tag{21.1.23}$$

According to Eq. (19.2.49), the Goldstone boson B_α couples to the currents $J_{N\beta}^{\mu}$ with a coupling constant $F_{\alpha\beta}^{-1}$, where $F_{\alpha\beta}$ is the coupling of the Goldstone boson associated with generator t_α to the current $J_{N\beta}^{\nu}$ defined by Eq. (19.2.38). Thus the exchange of Goldstone bosons would give a scattering matrix element

$$S_{CD,AB} = i(2\pi)^4\delta^4(p_A + p_B - p_C - p_D)k_\nu k_\lambda F_{\alpha\gamma}^{-1}F_{\beta\gamma}^{-1}\langle C|J_{N\alpha}^{\nu}|A\rangle \frac{1}{k^2}\langle D|J_{N\beta}^{\lambda}|B\rangle \,. \tag{21.1.24}$$

But Eqs. (19.2.40), (21.1.7), and (21.1.23) give the vector boson mass matrix as

$$\mu_{\alpha\beta}^2 = \sum_n F_{\alpha\gamma}Z_{\gamma n}F_{\beta\delta}Z_{\delta n} = F_{\alpha\gamma}F_{\beta\gamma} \,, \tag{21.1.25}$$

so in the limit of vanishing gauge boson coupling, the gauge boson exchange matrix element (21.1.22) is the same as the Goldstone boson exchange matrix element (21.1.24). This argument can be reversed; the requirement of continuity at zero gauge coupling can be used to derive a formula for the gauge boson masses even in the case where all other couplings are strong.

21.2 Renormalizable ξ-Gauges

In 1971 't Hooft[4] showed that path integrals in spontaneously broken gauge theories can be calculated in a gauge in which the vector boson

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

propagators vanish for momentum $k \to \infty$ as k^{-2}, so that these theories satisfy the power-counting test for renormalizability described in Section 12.2. Here we shall describe a larger class of renormalizable gauges, parameterized by an arbitrary constant ξ, that was introduced a little later by Fujikawa, Lee, and Sanda.[5]

In general gauges the kinematic term (21.1.3) in the Lagrangian for the scalar fields of a theory contains a cross term

$$i \sum_{nm\alpha} \partial_\mu \phi'_n \, t^\alpha_{nm} A_\alpha{}^\mu v_m \, ,$$

where v_m is the vacuum expectation value of ϕ_m, and ϕ'_n is the shifted field defined by Eq. (21.1.4). In unitarity gauge this term vanished as a consequence of the gauge condition (21.1.2). We will adopt a different approach here, similar to that of Sections 15.5 and 15.6. A functional $B[f]$ is introduced into the path integral, with

$$B[f] = \exp\left(\frac{-i}{2\xi} \int d^4x \sum_\alpha f_\alpha f_\alpha \right) . \tag{21.2.1}$$

This is equivalent to adding a gauge-fixing term in the Lagrangian

$$\mathscr{L}_{gf} = -\frac{1}{2\xi} \sum_\alpha f_\alpha f_\alpha \, . \tag{21.2.2}$$

Instead of taking $f_\alpha = \partial_\mu A^\mu_\alpha$ as in Section 15.5, we shall now take the gauge-fixing function as

$$f_\alpha = \partial_\mu A^\mu_\alpha - i\,\xi (t_\alpha)_{nm} \phi'_n v_m \, , \tag{21.2.3}$$

which is designed so that the above cross term in (21.1.3) is cancelled by the cross term in Eq. (21.2.2). Unitarity gauge is now a special case; for $\xi \to \infty$, the gauge-fixing functional (21.2.1) is infinitely sharply peaked at a ϕ' that satisfies the unitarity gauge condition (21.1.2). Another special case is provided by the limit $\xi \to 0$; here the gauge-fixing functional is peaked at a gauge field that satisfies the Landau gauge condition $\partial_\mu A^\mu_\alpha = 0$.

We also include in the Lagrangian a quartic polynomial $-P(\phi)$, subject to the gauge-invariance condition

$$\frac{\partial P(\phi)}{\partial \phi_n} (t_\alpha)_{nm} \phi_m = 0 \, . \tag{21.2.4}$$

We must of course also include in the Lagrangian a gauge-field term

$$\mathscr{L}_A = -\frac{1}{4} \sum_\alpha F^{\mu\nu}_\alpha F_{\alpha\mu\nu} \, . \tag{21.2.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

The total Lagrangian density of gauge fields and scalars is then

$$\begin{aligned}\mathscr{L}_{A,\phi} &\equiv \mathscr{L}_A + \mathscr{L}_\phi + \mathscr{L}_{gf} \\ &= -\frac{1}{4}\sum_\alpha F_\alpha^{\mu\nu}F_{\alpha\mu\nu} + \frac{1}{2}\sum_{n\alpha\beta}(t^\alpha\phi)_n(t^\beta\phi)_n A_\alpha^\mu A_{\beta\mu} - \frac{1}{2\xi}\sum_\alpha(\partial_\mu A_\alpha^\mu)\,(\partial_\nu A_\alpha^\nu) \\ &\quad -\frac{1}{2}\sum_n \partial_\mu\phi_n'\partial^\mu\phi_n' + \frac{\xi}{2}\sum_{\alpha nm}(t_\alpha v)_n(t_\alpha v)_m\phi_n'\phi_m' \\ &\quad -P(\phi) + i\sum_{\alpha nm}\partial_\mu\phi_n'(t_\alpha)_{nm}\phi_m' A_\alpha^\mu + \text{total derivatives}\,. \end{aligned} \tag{21.2.6}$$

As we saw in Section 15.6, the introduction of a gauge-fixing functional $B[f]$ requires also the introduction of a ghost field $\omega_\alpha(x)$, with Lagrangian depending on the gauge-transformation properties of f_α. Under a general gauge transformation (with an arbitrary function $\epsilon_\alpha(x)$)

$$\delta A_\alpha^\mu = -\sum_{\beta\gamma} C_{\alpha\beta\gamma}\epsilon_\beta A_\gamma{}^\mu + \partial^\mu\epsilon_\alpha\,, \tag{21.2.7}$$

$$\delta\phi_n = i\sum_{\alpha m}\epsilon_\alpha(t_\alpha)_{nm}\phi_m\,, \tag{21.2.8}$$

we have

$$\delta f_\alpha = \Box\,\epsilon_\alpha - \sum_{\beta\gamma} C_{\alpha\beta\gamma}\partial_\mu(\epsilon_\beta A_\gamma{}^\mu) + \xi\sum_{n\beta}(t_\alpha v)_n\,\epsilon_\beta(t_\beta\phi)_n\,. \tag{21.2.9}$$

According to the general results of Section 15.6, this yields a ghost Lagrangian

$$\mathscr{L}_\omega = \omega_\alpha^*\left[\Box\,\omega_\alpha - \sum_{\beta\gamma} C_{\alpha\beta\gamma}\partial_\mu(\omega_\beta A_\gamma^\mu) + \xi\sum_{n\beta}(t_\alpha v)_n\omega_\beta(t_\beta\phi)_n\right]. \tag{21.2.10}$$

Finally, if the theory involves spin $\frac{1}{2}$ fermions there will also be a general renormalizable term

$$\mathscr{L}_\psi = -\bar\psi(\partial\!\!\!/ - iA\!\!\!/_\alpha t_\alpha^{(\psi)} + m_0 + \Gamma_n\phi_n)\psi\,, \tag{21.2.11}$$

where $t_\alpha^{(\psi)}$ is the matrix representation of the generators of the gauge group for the fermions (including coupling-constant factors), and m_0 and Γ_n are constant matrices (in general, linear combinations of terms proportional to Dirac matrices 1 and γ_5) satisfying the gauge-invariance conditions

$$\left[t_\alpha^{(\psi)},\,\gamma_4 m_0\right] = 0\,, \tag{21.2.12}$$

$$\left[t_\alpha^{(\psi)},\,\gamma_4\Gamma_n\right] + \sum_m(t_\alpha)_{mn}\,\gamma_4\Gamma_m = 0\,. \tag{21.2.13}$$

(The factor $\gamma_4 \equiv i\gamma^0$ arises from the definition $\bar\psi \equiv \psi^\dagger\gamma_4$. It is relevant only if $t_\alpha^{(\psi)}$ involves terms proportional to γ_5.) The general theorem proved in Sections 15.5 and 15.6 guarantees that the S-matrix calculated from a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Lagrangian given by the sum of Eqs. (21.2.6), (21.2.10), and (21.2.11) is independent of the choice of the parameter ξ appearing in the gauge-fixing function (21.2.3), and so for any ξ will give the same results as the choice $\xi = \infty$ corresponding to unitarity gauge.

To derive the propagators for all these fields, we need the part of the Lagrangian that is quadratic in fields:

$$\begin{aligned}\mathscr{L}_{\text{QUAD}} = &-\frac{1}{4}\sum_{\alpha}(\partial^{\mu}A_{\alpha}{}^{\nu} - \partial^{\nu}A_{\alpha}{}^{\mu})(\partial_{\mu}A_{\alpha\nu} - \partial_{\nu}A_{\alpha\mu}) \\ &-\frac{1}{2}\sum_{\alpha\beta}\mu^2_{\alpha\beta}A^{\mu}_{\alpha}A_{\beta\mu} - \frac{1}{2\xi}\sum_{\alpha}(\partial_{\mu}A^{\mu}_{\alpha})(\partial_{\nu}A_{\alpha}{}^{\nu}) \\ &-\frac{1}{2}\sum_{n}(\partial_{\mu}\phi'_n)(\partial^{\mu}\phi'_n) - \frac{1}{2}\sum_{nm}M^2_{nm}\phi'_n\phi'_m \\ &-\bar{\psi}(\not{\partial} + m)\psi - \partial_{\mu}\omega^*_{\alpha}\,\partial^{\mu}\omega_{\alpha} - \xi\sum_{\alpha\beta}\mu^2_{\alpha\beta}\omega^*_{\alpha}\omega_{\beta} \\ &+ \text{ total derivatives}\,, \end{aligned} \tag{21.2.14}$$

where $\mu^2_{\alpha\beta}$ is the vector boson mass matrix (21.1.7):

$$\mu^2_{\alpha\beta} = -\sum_{n}(t^{\alpha}v)_n(t^{\beta}v)_n\,, \tag{21.2.15}$$

and M^2_{nm} and m are new scalar and fermion mass matrices:

$$\begin{aligned} M^2_{nm} &= \left.\frac{\partial^2 P(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=v} - \frac{\xi}{2}\sum_{\alpha}(t_{\alpha}v)_n(t_{\alpha}v)_m\,, \\ m &= m_0 + \sum_{n}\Gamma_n v_n\,. \end{aligned} \tag{21.2.16}$$

We see from Eq. (21.2.14) that the ghosts have gauge-dependent masses, equal to $\sqrt{\xi}$ times the corresponding vector boson masses.

These expressions give the particle masses in the zeroth order of perturbation theory. To this order, the vacuum expectation value v_n is just the location of the minimum of the polynomial 'potential' $P(\phi)$:

$$\left.\frac{\partial P(\phi)}{\partial\phi_n}\right|_{\phi=v} = 0\,. \tag{21.2.17}$$

Also, as we saw in Section 19.2, it follows from Eqs. (21.2.4) and (21.2.17) that

$$\sum_{m}\left.\frac{\partial^2 P(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=v}(t_{\alpha}v)_m = 0 \tag{21.2.18}$$

for all α. It follows then that in place of the Goldstone modes with mass zero, the scalar boson mass-squared matrix in Eq. (21.2.16) has eigenvalues

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

equal to $\sqrt{\xi}$ times the non-zero vector boson masses. That is, if $\mu^2_{\alpha\beta}$ has an eigenvector c_β with eigenvalue μ^2, then $\sum_\beta c_\beta t_\beta v$ is an eigenvector of M^2 with an eigenvalue $\xi\mu^2$:

$$\sum_m M^2_{nm}\left(\sum_\beta c_\beta t_\beta v\right)_m = \xi\sum_{\alpha\beta}\mu^2_{\alpha\beta}c_\beta(t_\alpha v)_n = \xi\mu^2\left(\sum_\alpha c_\alpha t_\alpha v\right)_n . \quad (21.2.19)$$

The other eigenvectors of M^2_{nm} are then orthogonal to all of these, and hence to all $t_\alpha v$, so these eigenvectors and the corresponding eigenvalues are just the same as for the matrix $(\partial^2 P(\phi)/\partial\phi_n\partial\phi_m)_{\phi=v}$. We see that for the unitarity gauge value $\xi\to\infty$, the Goldstone bosons are so heavy that they drop out of the theory, while the other boson masses are as usual.

The propagators are calculated by the usual rules: if the free-particle part of the Lagrangian takes the form (after integration by parts) $-\zeta^\dagger\mathscr{D}(\partial)\zeta$ for a complex field ζ or $-\frac{1}{2}\zeta^T\mathscr{D}(\partial)\zeta$ for a real field ζ, then the propagator of this field is $\mathscr{D}^{-1}(ik)$. This gives the propagators:

$$A: \quad \Delta_{\alpha\mu,\beta\nu}(k) = \left[\frac{1}{k^2+\mu^2}\left(\eta_{\mu\nu} - \frac{(1-\xi)k_\mu k_\nu}{k^2+\mu^2\xi}\right)\right]_{\alpha\beta}, \quad (21.2.20)$$

$$\phi: \quad \Delta_{nm}(k) = (k^2+M^2)^{-1}_{nm} + \xi\sum_{\alpha,\beta}(t_\alpha v)_n(t_\beta v)_m(k^2)^{-1}(k^2+\xi\mu^2)^{-1}_{\alpha\beta}, \quad (21.2.21)$$

$$\psi: \quad \Delta(k) = [-i\not{k}+m]\big/(k^2+m^2), \quad (21.2.22)$$

$$\omega: \quad \Delta_{\alpha\beta}(k) = (k^2+\xi\mu^2)^{-1}_{\alpha\beta}. \quad (21.2.23)$$

The poles in Eq. (21.2.20) at unphysical mass squares proportional to ξ are cancelled by the poles at the same masses in Eq. (21.2.21). Note that now for finite ξ all propagators have the same asymptotic behaviors as in the unbroken symmetry case, as required for renormalizability. In particular, the $k_\mu k_\nu$ term in the vector boson propagator no longer presents any problem for renormalizability, because it is accompanied with an extra factor $(k^2+\mu^2\xi)^{-1}$. We can even drop this term by choosing Feynman gauge, with $\xi=1$. It is only in the unitarity gauge case where $\xi\to\infty$ that this factor fails to give the propagator the asymptotic behavior needed for renormalizability.

Even with well-behaved propagators, it is still necessary to show that the ultraviolet divergences in these theories are constrained by the broken gauge symmetries in such a way that every infinity can be cancelled by the renormalization of a field or a parameter in the Lagrangian. This can be done by the same techniques as in Sections 17.2 and 17.3, but treating the vacuum expectation values of the spinless fields as external fields rather than fixed quantities that break the symmetries.[5a]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

21.3 The Electroweak Theory

The most important application of spontaneously broken gauge theories has been to the theory of weak and electromagnetic interactions.[3] Weak interactions at low energies are well described by an effective Lagrangian given by a sum of products of vector (including axial-vector) currents, as in Eq. (19.4.22). This suggests that these interactions may like electromagnetism be described by some sort of gauge theory. In order to insure the separate conservation of electronic-type and muonic-type leptons and baryons (or quarks) we may guess that the known electronic-type and muonic-type leptons and the quarks all form separate representations of the gauge group. With this assumption, there are only a few possibilities for the structure of the gauge group.

Let's first consider the electronic-type lepton fields. As far as we know, these consist only of the left- and right-handed parts of the electron field e:

$$e_L = \tfrac{1}{2}(1+\gamma_5)e\,, \qquad e_R = \tfrac{1}{2}(1-\gamma_5)e\,, \tag{21.3.1}$$

and a purely left-handed electron-neutrino field ν_{eL}:

$$\gamma_5 \nu_{eL} = \nu_{eL}\,. \tag{21.3.2}$$

The fields in any representation of the gauge group must all have the same Lorentz-transformation properties, so the representations of the gauge group here divide* into a left-handed doublet (ν_{eL}, e_L) and a right-handed singlet e_R. The largest possible gauge group is then

$$SU(2)_L \times U(1)_L \times U(1)_R\,,$$

under which the fields transform as

$$\delta \begin{pmatrix} \nu_e \\ e \end{pmatrix} = i\left[\vec{\epsilon}\cdot\vec{t} + \epsilon_L t_L + \epsilon_R t_R\right] \begin{pmatrix} \nu_e \\ e \end{pmatrix}\,, \tag{21.3.3}$$

where the generators are

$$\vec{t} = \frac{g}{4}(1+\gamma_5)\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\}\,, \tag{21.3.4}$$

$$t_L \propto (1+\gamma_5)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\,, \tag{21.3.5}$$

$$t_R \propto (1-\gamma_5)\,, \tag{21.3.6}$$

* If we allow gauge couplings that change electron-type lepton number, then it is possible to include the left-handed field $\overline{e_R}$ along with ν_{eL} and e_L in a representation of the gauge group. This was the basis for an early $SO(3)$ variant[6] of the electroweak theory, which has since been ruled out by experiment.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

with g a constant to be chosen later. It will be convenient instead of t_L and t_R to consider the generators

$$y \equiv g' \left[\left(\frac{1+\gamma_5}{4} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left(\frac{1-\gamma_5}{2} \right) \right] \tag{21.3.7}$$

and

$$n_e \equiv g'' \left[\left(\frac{1+\gamma_5}{2} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left(\frac{1-\gamma_5}{2} \right) \right] , \tag{21.3.8}$$

where g' and g'' are constants like g to be chosen later. The generator y appears along with t_3 in a linear combination that plays a special role in physics; it is the electric charge

$$q = e \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} = \frac{e}{g} t_3 - \frac{e}{g'} y . \tag{21.3.9}$$

Also, n_e is the electron-type lepton number. We want to include both charge-changing weak interactions (like beta decay) and electromagnetism in our theory, so we will assume that there are gauge fields $\vec{A}^\mu$ and B^μ coupled to $\vec{t}$ and y. In addition, we may or may not want to include a gauge field coupled to the one remaining independent linear combination of t_L and t_R, which can be taken as the electron-type lepton number (21.3.8). There are very stringent limits[7] on the long-range forces that would be produced by a massless gauge field coupled to n_e, so in order to include in our theory a gauge field coupled to n_e with a strength g'' comparable to the weak and electromagnetic interactions, we would have to assume that this gauge symmetry is spontaneously broken.** However, there is no experimental evidence for the weak interaction that would be produced by such a gauge coupling (and plenty of evidence by now against it) so we shall simply exclude n_e from the generators of the gauge group. The gauge group is then[8]

$$G = SU(2)_L \times U(1) \tag{21.3.10}$$

with generators $\vec{t}, y$ given by Eqs. (21.3.4) and (21.3.7) respectively. The coupling constants g and g' are to be adjusted so that the gauge fields $\vec{A}^\mu$ and B^μ coupled to these generators are canonically normalized. The most general gauge-invariant and renormalizable Lagrangian that involves just

** Note that this is possible without violating the global conservation law of electronic lepton conservation. We would have to assume that the Lagrangian is invariant under both a global phase transformation acting only on electron-type lepton fields, and also a local phase transformation that acts on electron-type lepton fields as well as on some scalar field that does not interact with leptons. The vacuum expectation value of this scalar would break the local symmetry, giving the gauge boson coupled to n_e a mass, without breaking the global symmetry.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

these gauge fields and electronic leptons is then

$$\mathscr{L}_{YM} + \mathscr{L}_e = -\tfrac{1}{4}\left(\partial_\mu \vec{A}_\nu - \partial_\nu \vec{A}_\mu + g\vec{A}_\mu \times \vec{A}_\nu\right)^2 - \tfrac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2$$
$$-\bar{\ell}(\not{\partial} - i\not{\vec{A}} \cdot \vec{t}_L - i\not{B}\, y)\ell \,. \tag{21.3.11}$$

(We here use the fact that the structure constants of $SU(2)_L$ and $U(1)$ are $C_{ijk} = -ig\,\epsilon_{ijk}$ and zero, respectively.)

Of course, of the four gauge fields coupled to $\vec{t}$ and y, only one linear combination, the electromagnetic field A^μ, is actually massless. We therefore must assume that $SU(2)_L \times U(1)$ is spontaneously broken to a subgroup $U(1)_{em}$, with generator given by the charge (21.3.9). The details of the symmetry-breaking mechanism will be considered a little later. However, whatever this mechanism may be, we know that the canonically normalized vector fields corresponding to particles of spin one and definite mass consist of one field of charge $+e$ with mass m_W:

$$W^\mu = \frac{1}{\sqrt{2}}(A_1^\mu + iA_2^\mu)\,, \tag{21.3.12}$$

another of charge $-e$ and the same mass:

$$W^{\mu *} = \frac{1}{\sqrt{2}}(A_1^\mu - iA_2^\mu)\,, \tag{21.3.13}$$

and two electrically neutral fields of mass m_Z and zero respectively, given by orthonormal linear combinations of A_3^μ and B^μ:

$$Z^\mu = \cos\theta\, A_3^\mu + \sin\theta\, B^\mu\,, \tag{21.3.14}$$

$$A^\mu = -\sin\theta\, A_3^\mu + \cos\theta\, B^\mu\,, \tag{21.3.15}$$

or equivalently

$$A_3^\mu = \cos\theta\, Z^\mu - \sin\theta\, A^\mu\,, \tag{21.3.16}$$

$$B^\mu = \sin\theta\, Z^\mu + \cos\theta\, A^\mu\,. \tag{21.3.17}$$

According to the general result (21.1.11)–(21.1.12), the generator of the unbroken symmetry, which is here electromagnetic gauge invariance, is given by a linear combination of generators in which the coefficients are the same as the coefficients of the corresponding massless field in the expansion of the canonically normalized gauge fields coupled to these generators. Inspecting Eqs. (21.3.16) and (21.3.17) shows that

$$q = -\sin\theta\, t_3 + \cos\theta\, y\,. \tag{21.3.18}$$

Comparing this with Eq. (21.3.9) gives then

$$g = -e/\sin\theta\,, \qquad g' = -e/\cos\theta\,. \tag{21.3.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

The complete lepton–gauge boson coupling can be expressed in terms of the couplings g and g':

$$
\begin{aligned}
i\mathscr{L}'_e &= -\overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}} \left[\sum_\alpha \not{A}_\alpha t_\alpha \right] \begin{pmatrix} \nu_e \\ e \end{pmatrix} \\
&= -\overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}} \left[\frac{1}{\sqrt{2}} \not{W}(t_{1L} - it_{2L}) + \frac{1}{\sqrt{2}} \not{W}^*(t_{1L} + it_{2L}) \right. \\
&\quad \left. + \not{Z}\,(t_{3L}\cos\theta + y\sin\theta) + \not{A}(-t_{3L}\sin\theta + y\,\cos\theta) \right] \begin{pmatrix} \nu_e \\ e \end{pmatrix} \\
&= \frac{g}{\sqrt{2}} \left(\bar{e}\, \not{W} \left(\frac{1+\gamma_5}{2} \right) \nu_e \right) + \frac{g}{\sqrt{2}} \left(\bar{\nu}_e\, \not{W}^* \left(\frac{1+\gamma_5}{2} \right) e \right) \\
&\quad -\frac{1}{2}\sqrt{g^2+g'^2}\,\bar{\nu}_e\, \not{Z} \left(\frac{1+\gamma_5}{2} \right) \nu_e + \frac{(g^2-g'^2)}{2\sqrt{g^2+g'^2}}\, \bar{e}\, \not{Z} \left(\frac{1+\gamma_5}{2} \right) e \\
&\quad + g'\bar{e}\, \not{Z} \left(\frac{1-\gamma_5}{2} \right) e - e\,(\bar{e}\not{A} e)\,.
\end{aligned}
\tag{21.3.20}
$$

To complete the theory, we must now make some assumption about the mechanism of symmetry breaking. We want this mechanism to give masses not only to the $W^\pm$ and Z^0, but to the electron as well. Now, the only way that this is possible in a renormalizable weakly-coupled theory is to have a scalar field coupled without derivatives to $\bar{\ell}_R$ and ℓ_L (and also $\bar{\ell}_L$ and ℓ_R). Then $SU(2)_L \times U(1)$ invariance requires that the scalar be an $SU(2)_L$ doublet like ℓ_L, but with a shifted value of y and hence of q. We thus assume a 'Yukawa' coupling

$$
\mathscr{L}_{\phi e} = -G_e \overline{\begin{pmatrix} \nu_e \\ e \end{pmatrix}}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R \;+\; \text{H.c.}\,, \tag{21.3.21}
$$

where (ϕ^+, ϕ^0) is a doublet, on which the $SU(2) \times U(1)$ generators are represented by the matrices:

$$
\vec{t}^{(\phi)} = \frac{g}{2} \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}, \tag{21.3.22}
$$

$$
y^{(\phi)} = -\,g'/2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{21.3.23}
$$

so that the charge matrix is

$$
q^{(\phi)} = \frac{e}{g} t_3^{(\phi)} - \frac{e}{g'} y^{(\phi)} = e \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{21.3.24}
$$

It is possible that there are other scalar multiplets in the theory, but for the moment let's suppose that this is the only one.

We must add a gauge-invariant term involving scalar and gauge fields to the Lagrangian. The most general form consistent with $SU(2) \times U(1)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

gauge invariance, Lorentz invariance, and renormalizability is:

$$\mathscr{L}_\phi = -\frac{1}{2}\left|(\partial_\mu - i\vec{A}_\mu \cdot \vec{t}^{(\phi)} - i\,B_\mu\, y^{(\phi)})\phi\right|^2 - \frac{\mu^2}{2}\,\phi^\dagger\phi - \frac{\lambda}{4}\,(\phi^\dagger\phi)^2\,, \quad (21.3.25)$$

where $\lambda > 0$, and

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}\,. \quad (21.3.26)$$

For $\mu^2 < 0$, there is a tree-approximation vacuum expectation value at the stationary point of the Lagrangian

$$\langle\phi\rangle^\dagger\langle\phi\rangle \equiv v^2 = |\mu^2|/\lambda\,. \quad (21.3.27)$$

We can always perform an $SU(2) \times U(1)$ gauge transformation to a unitarity gauge, in which $\phi^+ = 0$ and ϕ^0 is Hermitian, with a positive vacuum expectation value. (This is why we normalized the complex doublet ϕ so that an unconventional factor $\frac{1}{2}$ appears in the kinetic term in Eq. (21.3.25); $\mathrm{Re}\,\phi^0$ is the only physical scalar field, and Eq. (21.3.25) makes this a canonically normalized field.) In unitarity gauge the vacuum expectation values of the components of ϕ are

$$\langle\phi^+\rangle = 0\,, \qquad \langle\phi^0\rangle = v > 0\,. \quad (21.3.28)$$

The scalar Lagrangian (21.3.25) then yields a vector meson mass term

$$\begin{aligned} -\frac{1}{2}\left|\left(\vec{A}_\mu \cdot \vec{t}^{(\phi)} + B_\mu y^{(\phi)}\right)\langle\phi\rangle\right|^2 &= -\frac{1}{2}\left|\left(\frac{g}{2}\,\vec{A}_\mu \cdot \vec{\tau} - \frac{g'}{2}\,B_\mu\right)\begin{pmatrix} 0 \\ v \end{pmatrix}\right|^2 \\ &= -\frac{v^2 g^2}{4}\,W_\mu^\dagger W^\mu - \frac{v^2}{8}\,(g^2 + g'^2)\,Z_\mu Z^\mu\,. \end{aligned} \quad (21.3.29)$$

We see that as expected, the photon mass is zero, while the $W^\pm$ and Z^0 have the masses

$$m_W = \frac{v|g|}{2}\,, \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}\,. \quad (21.3.30)$$

Also, from Eqs. (21.3.21) and (21.3.28) we see that the electron is given a lowest-order mass

$$m_e = G_e v\,. \quad (21.3.31)$$

It is difficult to study reactions among electron-type leptons alone, though by now there are data on scattering processes like $\bar{\nu}_e + e^- \to \bar{\nu}_e + e^-$. For high precision data we have to consider reactions that also at least involve muonic-type leptons, such as the well-studied process of muon decay, $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$. It is trivial to extend the above model to include muon-type leptons — just add to the Lagrangian terms $\mathscr{L}_\mu$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

and $\mathscr{L}_{\phi\mu}$, like the last terms in Eqs. (21.3.11) and (21.3.21), with the fields e and ν_e replaced with the muon and muon-neutrino fields μ^- and ν_μ, and with G_e replaced with $G_\mu = G_e(m_\mu/m_e)$. Inspection of (21.3.20) and the corresponding term with e and ν_e replaced with μ and ν_μ shows that W exchange between low energy, e-type and μ-type leptons produces the effective interaction

$$\left(\frac{g}{\sqrt{2}}\right)^2 \frac{1}{m_W^2}\left(\bar{e}\gamma^\lambda\left(\frac{1+\gamma_5}{2}\right)\nu_e\right)\left(\bar{\nu}_\mu\gamma_\lambda\left(\frac{1+\gamma_5}{2}\right)\mu\right) + \text{H.c.} \quad (21.3.32)$$

This may be compared with the interaction of the effective '$V-A$' theory which is known to give a good description of muon decay

$$\frac{G_F}{\sqrt{2}}\left(\bar{e}\gamma^\lambda(1+\gamma_5)\nu_e\right)\left(\bar{\nu}_\mu\gamma_\lambda(1+\gamma_5)\mu\right) + \text{H.c.} \quad (21.3.33)$$

Here G_F is the conventional Fermi coupling constant, known from the muon decay rate to have the value $G_F = 1.16639(2) \times 10^{-5}\,\text{GeV}^{-2}$. Comparing these two expressions, we find

$$g^2/m_W^2 = 4\sqrt{2}\,G_F\,. \quad (21.3.34)$$

This allows an immediate determination of the vacuum expectation value v, given by Eq. (21.3.30) as

$$v = \frac{2m_W}{g} = \frac{1}{2^{1/4}G_F^{1/2}} = 247\,\text{GeV}\,. \quad (21.3.35)$$

Also, Eq. (21.3.31) shows that G_e has the very small value

$$G_e = \frac{0.511\,\text{MeV}}{247\,\text{GeV}} = 2.07 \times 10^{-6}\,. \quad (21.3.36)$$

From Eq. (21.3.30) we see that $m_Z > m_W$. We cannot use Eq. (21.3.30) to determine the actual values of m_Z and m_W without knowing something about g and g'. Using Eqs. (21.3.30) and (21.3.19), we can express m_Z and m_W in terms of the electroweak mixing angle θ:

$$m_W = \frac{ev}{2|\sin\theta|} = \frac{37.3\,\text{GeV}}{|\sin\theta|}\,,$$

$$m_Z = \frac{ev}{2|\sin\theta||\cos\theta|} = \frac{74.6\,\text{GeV}}{|\sin 2\theta|}\,.$$

These are the original results obtained in Ref. 3. Of course, there are radiative corrections of all sorts, most of which depend on details of the theory that have not yet been specified in this section. But there is one particularly large radiative correction that can be readily calculated without further information. The above values for m_W and m_Z were calculated using the conventionally defined electronic charge for e. However, as explained in Section 18.2, this is not precisely the appropriate value to use

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

in calculations of processes at energies $E \gg m_e$; we should instead use the electric charge e_μ defined at a sliding scale μ comparable to the energies of interest. For μ of the order of 90 GeV the effective fine structure constant $e_\mu^2/4\pi$ is about 1/129 (and quite insensitive to the precise value of μ), so the above values for m_W and m_Z should be multiplied with $\sqrt{137/129}$, giving

$$m_W = \frac{38.4 \text{ GeV}}{|\sin\theta|} , \tag{21.3.37}$$

$$m_Z = \frac{76.9 \text{ GeV}}{|\sin 2\theta|} . \tag{21.3.38}$$

Whatever the value of θ, these masses are too large for there to have been any hope of detecting the W or Z in the 1960s or early 1970s. Experimental evidence for the electroweak theory had to come instead from the discovery of the new class of weak interactions predicted by the theory, the neutral current processes produced by Z^0 exchange.[9] The first observation of a neutral current process was the 1973 bubble chamber detection of the purely leptonic process of ν_μ–e^- elastic scattering.[10] Although these processes are easy to deal with theoretically, the frequency of events is relatively low, because the cross section is proportional to the square of the center-of-mass energy.† It was years before the purely leptonic neutral current reactions could be used to give a reasonably precise value for the parameter $\sin^2\theta$. By 1994, the study of purely leptonic neutral current processes like $\nu_\mu + e^- \to \nu_\mu + e^-$ and $\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-$ had yielded the value 0.222 ± 0.011, which would give $m_W = 81.5$ GeV and $m_Z = 92.5$ GeV.

Even before the discovery of neutral currents, the electroweak theory had been extended to the weak and electromagnetic interactions of hadrons with each other and with leptons. By the mid-1960s, it had become understood that weak interaction processes in which charge is exchanged between leptons and hadrons are well described at low energy by the effective Lagrangian

$$\frac{G_F}{\sqrt{2}} \left[\bar{e}\gamma_\lambda(1+\gamma_5)\nu_e + \bar{\mu}\gamma_\lambda(1+\gamma_5)\nu_\mu\right] J^\lambda + \text{H.c.} , \tag{21.3.39}$$

where J^λ is an hadronic current. Within the quark model, the commutation and conservation properties of J^λ allowed it to be identified with the quark

† The cross section is proportional to G_F^2, so in order to have the dimensions of energy^{-2}, it must also be proportional to some energy squared. Where the center-of-mass energy is much larger than the electron mass, it is the only energy that can appear in this formula.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

current

$$J^\lambda = \bar{u}\gamma^\lambda(1+\gamma_5)d\cos\theta_c + \bar{u}\gamma^\lambda(1+\gamma_5)s\,\sin\theta_c\,. \tag{21.3.40}$$

Here u, d, s are the fields of the up, down, and strange quarks, and θ_c is another angle, known as the Cabibbo angle.[11] Experiments on processes like $O^{14} \to N^{14*} + e^+ + \nu_e$ and $K^+ \to \pi^0 + e^+ + \nu_e$ confirm that G_F has very nearly the same value as that measured in the purely leptonic process $\mu^+ \to \bar{\nu} + e^+ + \nu_e$, and give for θ_c the value[12] $\sin\theta_c = 0.220 \pm 0.003$. We naturally conclude that the quarks provide another $SU(2) \times U(1)$ doublet

$$\mathcal{2} = \left(\frac{1+\gamma_5}{2}\right)\begin{bmatrix} u \\ d\cos\theta_c + s\,\sin\theta_c \end{bmatrix}, \tag{21.3.41}$$

as well as right-handed singlets, with y values adjusted to give the quark charges $2e/3$ and $-e/3$. By itself this would lead to a serious difficulty. The Z^0 boson interacts with the quark neutral current

$$\sum_{\mathcal{2}} \bar{\mathcal{2}}\gamma^\mu(t_{3L}\cos\theta + y\sin\theta)\mathcal{2} = \sum_{\mathcal{2}} \bar{\mathcal{2}}\gamma^\mu(t_{3L}\sec\theta + q\tan\theta)\mathcal{2}\,, \tag{21.3.42}$$

with the sum running over all quark doublets $\mathcal{2}$ like (21.3.41). The charge matrix q is diagonal in quark flavors, but if (21.3.41) were the only quark doublet then the term involving the matrix t_{3L} would contain cross terms proportional to $\bar{s}\gamma^\mu(1+\gamma_5)d$ and $\bar{d}\gamma^\mu(1+\gamma_5)s$, leading to effective Z exchange interactions like $s + \bar{d} \leftrightarrow d + \bar{s}$ and $s + \bar{d} \leftrightarrow \mu^+ + \mu^-$ with the strength of ordinary first-order weak interactions. Such effects would lead to rates for processes like K^0–$\overline{K^0}$ oscillations and $K^0 \to \mu^+ + \mu^-$ many orders of magnitude greater than observed. Also, even without neutral current terms in the Lagrangian, the one-loop diagrams involving the interaction (21.3.39) with the charged current (21.3.40) would lead to an effective interaction $s + \bar{d} \to d + \bar{s}$ which is smaller than an ordinary first-order weak interaction only by a factor of order $\alpha/2\pi$, leading to a rate for K^0–$\overline{K^0}$ oscillations that is still much too large. In order to avoid this last difficulty, it was proposed[13] that there is another term in J^λ; in modern notation,

$$\bar{c}\gamma^\lambda(1+\gamma_5)\,[-d\,\sin\theta_c + s\cos\theta_c]\,, \tag{21.3.43}$$

where c is a fourth quark, like u with charge $2e/3$. Adding (21.3.43) to (21.3.40), the charged current may be written

$$J^\lambda = (\bar{u}\cos\theta_c - \bar{c}\sin\theta_c)\gamma^\lambda(1+\gamma_5)d + (\bar{u}\sin\theta_c + \bar{c}\cos\theta_c)\gamma^\lambda(1+\gamma_5)s\,.$$

The only reason that the interactions of the W with this current do not conserve strangeness is that the c and u have different masses, leading to transitions between $u\cos\theta_c - c\sin\theta_c$ and $u\sin\theta_c + c\cos\theta_c$. But this means that the loop diagrams for the effective interaction $s + \bar{d} \to d + \bar{s}$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

are suppressed by additional factors (since $m_u \ll m_c$) of m_c^2/m_W^2, bringing the rate for K^0–$\overline{K^0}$ oscillations into agreement with experiment.

It was subsequently noted[14] that this also solves the problem of the strangeness changing Z^0 interactions. In the context of the $SU(2) \times U(1)$ gauge theory the combination $-d_L \sin\theta_c + s_L \cos\theta_c$ cannot be a singlet, but must be part of another doublet

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} c \\ -d\,\sin\theta_c + s\,\cos\theta_c \end{bmatrix} . \tag{21.3.44}$$

Including this doublet in the weak neutral current (21.3.42), the strangeness non-conserving terms proportional to $\bar{s}\gamma^\mu(1+\gamma_5)d$ and $\bar{d}\gamma^\mu(1+\gamma_5)s$ cancel, removing the problem of excessive Z exchange contributions to processes like K^0–$\overline{K^0}$ oscillations and $K^0 \to \mu^+ + \mu^-$. Particles containing the c quark in a c–$\bar{c}$ bound state were discovered[15] in 1974, and indicated a mass $m_c \approx 1.5$ GeV.[††] This completed two generations of quarks and leptons: a (u, d) quark doublet mixed with a (c, s) quark doublet, together with two lepton doublets (ν_e, e) and (ν_μ, μ).

The first sign of a third generation was the discovery of a third charged lepton,[16] the τ. Later a fifth quark type, the b, was discovered,[17] with charge $-e/3$ and a mass of about 4.5 GeV. A sixth, the t with charge $2e/3$, then became theoretically necessary, and after a long interval it too was discovered,[18] with a mass quoted in 1995 as 181 ± 12 GeV.[19] Today the hadronic current J^λ in (21.3.39) is expressed as

$$J^\lambda = \overline{\begin{bmatrix} u \\ c \\ t \end{bmatrix}} \gamma^\lambda(1+\gamma_5)\, V \begin{bmatrix} d \\ s \\ b \end{bmatrix} , \tag{21.3.45}$$

where V is an incompletely known 3×3 unitary matrix, known as the Kobayashi–Maskawa matrix.[20] In the $SU(2) \times U(1)$ gauge theory, this means that there are three quark doublets:

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} u \\ V_{ud}d + V_{us}s + V_{ub}b \end{bmatrix} , \tag{21.3.46}$$

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} c \\ V_{cd}d + V_{cs}s + V_{cb}b \end{bmatrix} , \tag{21.3.47}$$

$$\left(\frac{1+\gamma_5}{2}\right) \begin{bmatrix} t \\ V_{td}d + V_{ts}s + V_{tb}b \end{bmatrix} . \tag{21.3.48}$$

[††] We do not observe quarks in isolation, so their masses are not precisely defined. The mass of the c quark quoted here is roughly half the mass of the J–ψ particle, interpreted as a c–$\bar{c}$ bound state. The b and t quarks are so heavy that their masses can be taken from the masses of the hadrons containing them with little ambiguity.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

It is important to recognize that this is just what we should naturally expect on general grounds for three quark doublets. The most general renormalizable $(SU(3)\times SU(2)\times U(1))$-invariant interactions of the scalar doublets ϕ_n with the quarks must in general take the form

$$\mathscr{L}_\phi = -\sum_{ijn} G^n_{ij} \overline{\begin{pmatrix} U_{iL} \\ D_{iL} \end{pmatrix}} \cdot \begin{pmatrix} \phi^0_n \\ \phi^-_n \end{pmatrix} U_{jR} - \sum_{ijn} H^n_{ij} \overline{\begin{pmatrix} U_{iL} \\ D_{iL} \end{pmatrix}} \cdot \begin{pmatrix} -\phi^{-\dagger}_n \\ \phi^{0\dagger}_n \end{pmatrix} D_{jR} + \text{H.c.}\,, \tag{21.3.49}$$

where U_i and D_i with $i = 1, 2, 3$ are three independent quark fields of charge $2e/3$ and $-e/3$, respectively, L and R denote the left- and right-handed parts of the quark fields, and G^n_{ij} and H^n_{ij} are unknown constants. The vacuum expectation values of the neutral scalars then produce a quark mass term

$$\mathscr{L}_m = -\sum_{ij} \overline{U_{iL}}\, m^U_{ij}\, U_{jR} - \sum_{ij} \overline{D_{iL}}\, m^D_{ij}\, D_{jR} + \text{H.c.}\,, \tag{21.3.50}$$

where

$$m^U_{ij} = \sum_n G^n_{ij} \langle \phi^0_n \rangle_{\text{VAC}}\,, \qquad m^D_{ij} = \sum_n H^n_{ij} \langle \phi^0_n \rangle^*_{\text{VAC}}\,. \tag{21.3.51}$$

The matrices m^U_{ij} and m^D_{ij} are not constrained in any way, and in particular may be complex and non-diagonal, in which case parity- and flavor-non-conserving terms appear in $\mathscr{L}_m$. But we can introduce new quark fields $U'_R = A^U_R U_R$, $U'_L = A^U_L U_L$, $D'_R = A^D_R D_R$, $D'_L = A^D_L D_L$, where the As are 3×3 matrices constrained only by the condition that they must be unitary in order to preserve the form of the kinematic term (19.4.1). Then the mass term (21.3.50) takes the same form when rewritten in terms of the primed quark fields, but with the matrices m^U and m^D replaced with

$$m^{U\prime} = A^U_L m^U A^{U\dagger}_R\,, \qquad m^{D\prime} = A^D_L m^D A^{D\dagger}_R\,. \tag{21.3.52}$$

Now it is a general theorem that for any matrix m, it is always possible to choose unitary matrices A and B such that AmB is real and diagonal. (Use the polar decomposition theorem to write $m = HU$, where H is Hermitian and U is unitary, and choose $A = S^\dagger$ and $B = U^\dagger S$, where S is the unitary matrix that diagonalizes H.) We can therefore choose the As so that $m^{U\prime}$ and $m^{D\prime}$ are real and diagonal, in which case the quark fields u, c, t, d, s, and b are to be identified with the components of $U'_L + U'_R$ and $D'_L + D'_R$. The weak doublets are now written as

$$Q_{iL} = \begin{pmatrix} (A^{U\,-1}_L U'_L)_i \\ (A^{D\,-1}_L D'_L)_i \end{pmatrix}\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

but we can just as well take the doublets as linear combinations $A_L^U Q_L$ that have charge $2e/3$ quarks of definite mass u, c, t as their top component, in which case these doublets take the form (21.3.46)–(21.3.48), with

$$V = A_L^U A_L^{D-1} . \tag{21.3.53}$$

Within 90% confidence limits, the best present (1995) values for the absolute values of the elements of the Kobayashi–Maskawa matrix are[20a]

$$\begin{pmatrix} 0.9745 \text{ to } 0.9757 & 0.219 \text{ to } 0.224 & 0.002 \text{ to } 0.005 \\ 0.218 \text{ to } 0.224 & 0.9736 \text{ to } 0.9750 & 0.036 \text{ to } 0.046 \\ 0.004 \text{ to } 0.014 & 0.034 \text{ to } 0.046 & 0.9989 \text{ to } 0.9993 \end{pmatrix} ,$$

with rows labelled u, c, and t, and columns labelled d, s, and b.

If there were only two quark doublets formed from u, d, c, and s quarks, it would be possible to choose the phases of the quark fields so that all V_{ij} are real,[‡] so that the V matrix is orthogonal, and the doublets (21.3.46) and (21.3.47) (with b omitted) take the form (21.3.41) and (21.3.44), respectively. In this case the gauge interactions would automatically conserve T and CP. The great importance of the third generation is that it is no longer always possible to choose quark phases so that the V matrix is real, and therefore the gauge interactions can violate T and CP conservation. But for unknown reasons the elements V_{ub}, V_{cb}, V_{td}, and V_{ts} that connect the third generation with the first two are all quite small, so the physics of the first two generations is hardly affected by the presence of the third, which explains in a more-or-less natural way why the Cabibbo assumption (21.3.40) works so well and why the violation of T and CP conservation is so weak. T and CP conservation can also be violated by scalar boson interactions if there are two or more scalar doublets;[20b] here the violation of T and CP conservation is expected to be weak because the scalar doublets couple weakly to light quarks. It is still unknown which of these mechanisms is responsible for the observed violation of T and CP conservation in K_2^0 decay, discussed in Section 3.3.

Neutral current processes involving hadrons, such as neutrino–nucleon deep-inelastic scattering, were discovered in 1973,[21] shortly after the detection of the purely leptonic process $\nu_\mu + e \to \nu_\mu + e$. Because of the much greater mass of the target particle here, it became possible before long to observe large numbers of events, and use them to confirm the electroweak theory and measure its parameters. Additional information on lepton–hadron neutral current interactions came from the observation of parity violation in atomic physics. By 1983 all direct measurements of

[‡] Adjust the phases of d and s so that V_{ud} and V_{us} are real. Unitarity then requires that V_{cd} and V_{cs} have the same phase, which can be eliminating by adjusting the phase of c.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

$\sin^2\theta$ had become consistent, and gave a combined value $\sin^2\theta = 0.23$, yielding the predictions $m_W = 80.1\,\text{GeV}$ and $m_Z = 91.4\,\text{GeV}$. Then in 1983 the W was discovered, with the Z following soon after.[22] Their measured masses are now (in 1995)

$$m_W = 80.410 \pm 0.180\ \text{GeV}^{23}\,, \qquad m_Z = 91.1887 \pm 0.0022\ \text{GeV}^{24}\,,$$

in satisfactory agreement with the predictions of the electroweak theory.

The very great accuracy of the measurement of the Z mass, which has been achieved by tuning the energy of e^+–e^- collisions to the Z resonance at LEP (CERN's Large Electron Positron collider) and the SLC (Stanford Linear Collider), has changed the way that electroweak data is analyzed. Instead of comparing predictions of W and Z masses with observed values, the Z mass is taken as an experimental input, along with the Fermi coupling constant $G_F = 1.16639(2) \times 10^{-5}\ \text{GeV}^{-2}$ taken from the rate of muon decay (including radiative corrections to order α), and the fine structure constant $\alpha(m_Z) = (128.87 \pm 0.12)^{-1}$, extrapolated from low energy measurements as described in Section 18.2. In this way $\sin^2\theta$ becomes a derived quantity; if defined by Eq. (21.3.38), it takes the value $\sin^2\theta = 0.2312 \pm 0.003$. With these inputs, the electroweak theory can be used to make predictions of other quantities like m_W with sufficient precision that it becomes necessary to take electroweak radiative corrections into account.[25] In one-loop order these radiative corrections involve the masses of the t quark and scalar ('Higgs') boson, and thus can be used to estimate these masses. For instance, before the top quark was discovered the agreement between theory and experiment set bounds on these radiative corrections which implied a top quark mass in the range 130–200 GeV,[26] in agreement with the value subsequently found experimentally. The W mass is predicted (in 1994) to be 80.29 GeV, with an uncertainty of ± 0.02 GeV from uncertainties in the inputs m_Z, G_F, and $\alpha(m_Z)$, and an uncertainty of ± 0.11 GeV from the range of possible values of m_t and m_{Higgs}. One 1995 study[27] concludes that $m_{\text{Higgs}} < 225$ GeV. The precise measurement of m_W expected at the LEP 2 electron–positron collider at CERN will allow a useful estimate of m_{Higgs}.

* * *

The most general renormalizable Lagrangian with the field content and $SU(3) \times SU(2) \times U(1)$ gauge symmetries of the electroweak theory automatically respects baryon and lepton conservation. This is obviously true for the gauge interactions and bare mass terms, because the quarks, antiquarks, leptons, and antileptons all belong to distinct representations of $SU(3) \times SU(2) \times U(1)$. With scalars all belonging to $SU(3)$ neutral $SU(2)$ doublets with $U(1)$ quantum number $\pm 1/2$, the only renormalizable interactions of scalars with fermions and/or antifermions are with quark–

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

antiquark and lepton–antilepton pairs, which of course conserve baryon and lepton number. (In much the same way, one can see that the charged hadronic currents with which leptons interact are necessarily linear combinations of the currents associated with the spontaneously broken $SU(3) \times SU(3)$ symmetry described in Section 19.7, as assumed without knowing the explanation in the original work on this broken symmetry.)

These results depend critically on the assumption that the standard model is renormalizable. But as we have repeatedly emphasized, the renormalizable Lagrangian of the standard model is expected to be accompanied with non-renormalizable terms of dimensionality $d > 4$, suppressed by $4 - d$ powers of some very large mass M. The leading corrections to the predictions of the renormalizable standard model come from terms with the smallest possible dimensionality greater than four.

The only Lorentz-invariant terms of dimensionality five that can be constructed out of the fermion and other fields of the standard model are at most bilinear in fermion fields and also contain either two scalars, or one scalar and one gauge-invariant derivative, or no scalars and two gauge-invariant derivatives (including their commutator, a field strength tensor). Color $SU(3)$ invariance requires that the fermion fields in such an interaction appear in either a quark–antiquark bilinear or a pair of lepton and/or antilepton fields, all of which operators conserve baryon number. There are a great number of such terms, but to violate lepton number conservation they must involve a product of two lepton fields or of their conjugates. The left-handed lepton doublets (ℓ^-_{Li}, ν_i) and right-handed charged lepton singlets ℓ^-_{Ri} (with $i = e$, μ, or τ) have $U(1)$ quantum numbers $1/2$ and $+1$, respectively, while the scalar doublet (or doublets) (ϕ^+, ϕ^0) have $U(1)$ quantum number $-1/2$, so we can construct $U(1)$-invariant interactions of dimensionality five out of two left-handed lepton doublets and two scalar doublets. With only a single type of scalar doublet, there is just one such term that satisfies $SU(2)$ and Lorentz invariance:[27a]

$$\sum_{ij} f_{ij}(\overline{\ell^c_{Li}}\phi^+ - \overline{\nu^c_i}\phi^0)(\ell_{Lj}\phi^+ - \nu_j\phi^0)\,, \tag{21.3.54}$$

where i and j are lepton flavor indices, and c denotes the charge conjugate field. At energies below the electroweak breaking scale, this yields an effective interaction

$$\sum_{ij} f_{ij}\,\overline{\nu^c_i}\nu_j\,\langle\phi^0\rangle^2\,. \tag{21.3.55}$$

We expect f_{ij} to be of order $1/M$, perhaps multiplied with small coupling constants, so this gives lepton number non-conserving neutrino masses at most of order[27b] $(300\text{ GeV})^2/M$. We shall see in Section 21.5 that M

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

is expected to be of order 10^{15}–10^{18} GeV, so we would expect neutrino masses in the range 10^{-4}–10^{-1} eV, or less if suppressed by small coupling constants. Such masses are too small for direct measurement, but there is no reason for the neutrino mass matrix to be diagonal, so neutrino masses might be detected in oscillations of one neutrino type into another over long flight paths.

A similar analysis shows that there are interactions of dimensionality six that violate both baryon and lepton number conservation, involving three quark fields and one lepton field.[27c] Such interactions would have coupling constants of order M^{-2}, and would lead to processes like proton decay, with rates proportional to M^{-4}.

21.4 Dynamically Broken Local Symmetries*

Our discussion of spontaneously broken local symmetries has so far been entirely within the context of perturbation theory. To some extent, this limitation is inevitable. Whereas for spontaneously broken global symmetries it is possible to prove exact theorems about the existence and interactions of massless Goldstone bosons, the spontaneous breakdown of a local symmetry does not lead to any such precise consequences. Even the existence of massive vector bosons is not really a general theorem; for sufficiently strong gauge coupling these particles decay so rapidly that they lose their identity as distinct resonances of definite spin $j = 1$.

On the other hand, if the gauge couplings like e or g or g' are sufficiently small then the theory with a spontaneously broken local symmetry must be very close to one with a spontaneously broken global symmetry, about which exact theorems can be proved. It is therefore possible to derive useful approximate results for such gauge theories, *even if the other non-gauge couplings are very strong*. One example is provided by the standard $SU(2) \times U(1)$ electroweak theory with a large scalar self-coupling λ (and hence a large scalar mass; see Eq. (21.3.27)). A more intriguing possibility is that the breakdown of electroweak symmetry is due to strong forces associated with some new gauge group acting on a set of new fermions. We will here consider the results that can be obtained for all such theories, without reference to the specific mechanism for spontaneous symmetry breaking.[28]

We assume that in the limit of zero gauge couplings, our theory is invariant under some group G of *global* symmetries, spontaneously broken

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

to a subgroup H. As discussed in Section 19.5, in this case the theory can be written in terms of a set of Goldstone boson fields ξ_a, plus other matter fields $\tilde{\psi}$, whose G-transformation properties are such that the Lagrangian is G-invariant if it is H-invariant and constructed solely from $\tilde{\psi}$ and covariant derivatives $D_{a\mu}$, $D_\mu\tilde{\psi}$ etc., given by Eqs. (19.6.14), (19.6.30), etc.

We now 'turn on' the gauge couplings. The gauge group $\mathscr{G}$ is of course required to be a subgroup $\mathscr{G} \subset G$ of the group G of all symmetries of the theory, and when G is spontaneously broken to H, $\mathscr{G}$ must be spontaneously broken to a subgroup $\mathscr{H}$, equal to the intersection of $\mathscr{G}$ with H. The generators $\mathscr{T}_\alpha$ of the gauge group $\mathscr{G}$ may be expressed as linear combinations of the generators T_A of the full group G:

$$\mathscr{T}_\alpha = \sum_A e_{\alpha A} T_A \,, \tag{21.4.1}$$

with coefficients $e_{\alpha A}$, the gauge couplings, that are taken very small. The index A runs over the labels i, a of the unbroken symmetry generators t_i and the broken symmetry generators x_a. (We are here taking the generators T_A to be conventionally normalized; that is, they are represented by matrices with elements of order unity. In particular, in contrast to the $\mathscr{T}_\alpha$, the structure constants of the x_a and t_i do not include factors of coupling constants.)

In the underlying theory in which G invariance is linearly realized, we introduce the coupling of gauge fields $\mathscr{A}_{\alpha\mu}$ to other fields ψ by replacing ordinary derivatives with gauge-covariant derivatives

$$\left(\partial_\mu - i\sum_\alpha \mathscr{T}_\alpha \mathscr{A}_{\alpha\mu}\right)\psi = \left(\partial_\mu - i\sum_A T_A A_{A\mu}\right)\psi \,, \tag{21.4.2}$$

where

$$A_{A\mu} \equiv \sum_\alpha e_{\alpha A}\mathscr{A}_{\alpha\mu} \,. \tag{21.4.3}$$

The resulting theory is then invariant under *formal* local transformations, under which the fields transform according to

$$\psi \to g\,\psi \,, \tag{21.4.4}$$

$$\sum_A T_A A_{A\mu} \to g\left(\sum_A T_A A_{A\mu}\right) g^{-1} - i(\partial_\mu g)g^{-1} \,, \tag{21.4.5}$$

where $g(x)$ is an arbitrary spacetime-dependent element of the group G. This is a purely formal invariance, because the gauge couplings in general actually break G, as shown by the fact that the transformation (21.4.5) does not in general preserve the form of the linear combination (21.4.3). Nevertheless, we can temporarily forget about Eq. (21.4.3), treating A_A^μ as an unconstrained classical external field, and analyze the structure

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

of the Lagrangian for the matter fields and their interaction with the gauge fields by requiring that it be invariant under local transformations (21.4.4), (21.4.5). In this way we will insure not only that the Lagrangian is invariant under the true local symmetry subgroup $\mathscr{G}$ (and for $e_{\alpha A} \to 0$ under the larger global symmetry group G), but also that the currents, the variational derivatives of the matter action with respect to $\mathscr{A}^\mu_\alpha$, will have the correct transformation under the broken global symmetry group G. Later we will restrict A^μ_A to the form (21.4.3), and treat the field $\mathscr{A}^\mu_\alpha$ as a quantum field, supplying a suitable kinematic term in the Lagrangian for this field.

In order to explore the implications of the spontaneous breakdown of the invariance group G to its subgroup H, we will proceed as we did in Section 19.6. First, replace ψ and A with new fields $\tilde{\psi}$, $\tilde{A}$:

$$\tilde{\psi} = \gamma^{-1}(\xi)\psi \,, \tag{21.4.6}$$

$$\tilde{A}^\mu_A = \sum_B D_{AB}\left(\gamma^{-1}(\xi)\right) A^\mu_B \,, \tag{21.4.7}$$

where $\gamma(\xi)$ is the standard G transformation that eliminates the Goldstone boson degrees of freedom in ψ, and $D(g)$ is the representation of G furnished by the gauge fields:

$$g\, T_A\, g^{-1} = \sum_B D_{BA}(g)\, T_B \,. \tag{21.4.8}$$

These Goldstone degrees of freedom reappear in the spacetime-dependent parameters ξ_a on which $\gamma(\xi)$ depends. By the same calculations as in Section 19.6, for local as well as global transformations the transformation rule (21.4.4) translates into the transformations

$$\xi^a \to \xi^{a'} = f^a(\xi, g) \,, \tag{21.4.9}$$

$$\tilde{\psi} \to \tilde{\psi}' = h(\xi, g)\tilde{\psi} \,, \tag{21.4.10}$$

where h and f are defined by

$$g\,\gamma(\xi) = \gamma(f(\xi, g))\, h(\xi, g) \,, \tag{21.4.11}$$

with h in the unbroken subgroup H. We also need to work out the transformation rule for $\tilde{A}^\mu_A$. Recall that, according to Eq. (21.4.5), under these local transformations the linear combinations of gauge fields in Eq. (21.4.2) transform as

$$\sum_A T_A A_{A\mu} \to \sum_A T_A A'_{A\mu} = g \left[\sum_A T_A A_{A\mu} - i\, g^{-1}\partial_\mu g\right] g^{-1} \,.$$

Multiplying on the left and right by $\gamma^{-1}(\xi')$ and $\gamma(\xi')$ respectively, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

using Eqs. (21.4.7), (21.4.8), and (21.4.11), we can write this as

$$i\sum_A T_A\tilde{A}'_{A\mu} = h(\xi,g)\left[i\sum_A T_A\tilde{A}_{A\mu} + \gamma^{-1}(\xi)[g^{-1}\partial_\mu g]\gamma(\xi)\right]h^{-1}(\xi,g)\,. \tag{21.4.12}$$

To see how to cancel the inhomogeneous $g^{-1}\partial_\mu g$ term, we note that by differentiating Eq. (21.4.11) and multiplying on the left with its inverse, we have

$$\begin{aligned}\gamma^{-1}(\xi')\partial_\mu\gamma(\xi') &= h(\xi,g)\gamma^{-1}(\xi)[g^{-1}\partial_\mu g]\gamma(\xi)h^{-1}(\xi,g)\\ &+ h(\xi,g)[\gamma^{-1}(\xi)\partial_\mu\gamma(\xi)]h^{-1}(\xi,g) + h(\xi,g)\partial_\mu h^{-1}(\xi,g)\,.\end{aligned} \tag{21.4.13}$$

So to cancel the inhomogeneous term, we must subtract Eq. (21.4.12) from Eq. (21.4.13):

$$\begin{aligned}&\gamma^{-1}(\xi')\partial_\mu\gamma(\xi') - i\sum_A T_A\tilde{A}'_{A\mu}\\ &\quad= h(\xi,g)\left[\gamma^{-1}(\xi)\partial_\mu\gamma(\xi) - i\sum_A T_A\tilde{A}_{A\mu}\right]h^{-1}(\xi,g)\\ &\qquad- \left[\partial_\mu h(\xi,g)\right]h^{-1}(\xi,g)\,.\end{aligned} \tag{21.4.14}$$

We therefore define new gauge-covariant quantities $\mathscr{D}$ and $\mathscr{E}$ by

$$\sum_a i\mathscr{D}_{a\mu}x_a + \sum_i i\mathscr{E}_{i\mu}t_i \equiv \gamma^{-1}(\xi)\partial_\mu\gamma(\xi) - i\sum_A T_A\tilde{A}_{A\mu} \tag{21.4.15}$$

with transformation properties

$$\mathscr{D}_{a\mu}\to\mathscr{D}'_{a\mu}\,,\qquad \mathscr{E}_{i\mu}\to\mathscr{E}'_{i\mu}\,, \tag{21.4.16}$$

where

$$\sum_a \mathscr{D}'_{a\mu}x_a = h(\xi,g)\left(\sum_a \mathscr{D}_{a\mu}x_a\right)h^{-1}(\xi,g)\,, \tag{21.4.17}$$

$$\sum_a \mathscr{E}'_{i\mu}t_i = h(\xi,g)\left(\sum_i \mathscr{E}_{i\mu}t_i\right)h^{-1}(\xi,g) + i\Big(\partial_\mu h(\xi,g)\Big)h^{-1}(\xi,g)\,, \tag{21.4.18}$$

just as in Eqs. (19.6.26) and (19.6.27). We can use $\mathscr{E}$ to construct fully covariant derivatives of matter fields

$$\mathscr{D}_\mu\tilde{\psi} = \partial_\mu\tilde{\psi} + i\sum_i t_i\mathscr{E}_{i\mu}\tilde{\psi} \tag{21.4.19}$$

as well as higher derivatives like $\mathscr{D}_\nu\mathscr{D}_\mu\tilde{\psi}$, etc. Because of the inhomogeneous term in Eq. (21.4.12), we cannot freely introduce $\tilde{A}_{A\mu}$ or covariant derivatives like (21.4.19) of $\tilde{A}_{\alpha\mu}$ into the Lagrangian. However, it is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

easy to construct a 'curl' that transforms covariantly under both local $\mathscr{G}$ transformations and global G transformations. It is

$$\tilde{F}_{A\mu\nu} \equiv \sum_B D_{AB}\Big(\gamma^{-1}(\xi)\Big)\left(\partial_\mu A_{B\nu} - \partial_\nu A_{B\mu} - \sum_{CD} C_{BCD} A_{C\mu} A_{D\nu}\right) . \tag{21.4.20}$$

This transforms under formal local G transformations as

$$\tilde{F}_{A\mu\nu} \to \tilde{F}'_{A\mu\nu} = \sum_B \mathscr{D}_{AB}\Big(h(\xi, g)\Big)\tilde{F}_{B\mu\nu} . \tag{21.4.21}$$

The Lagrangian is therefore invariant under formal local G transformations if it is constructed as an arbitrary function of $\tilde{\psi}$, $\mathscr{D}_{a\mu}$, $\mathscr{D}_\mu\tilde{\psi}$, $\tilde{F}_{A\mu\nu}$, *and higher covariant derivatives, that satisfies global H invariance.*

Now let us return to reality, and treat A^μ_A as a quantum field of the restricted form (21.4.3). Eqs. (21.4.15) and (21.4.20) now become

$$\begin{aligned} &\sum_a i\mathscr{D}_{a\mu}x_a + \sum_i i\mathscr{E}_{i\mu}t_i \\ &= \gamma^{-1}\partial_\mu\gamma(\xi) - i\sum_{AB\alpha} T_A D_{AB}(\gamma^{-1}(\xi))e_{\alpha B}\mathscr{A}_{\alpha\mu} \end{aligned} \tag{21.4.22}$$

and

$$\tilde{F}_{A\mu\nu} = \sum_{B\alpha} D_{AB}\Big(\gamma^{-1}(\xi)\Big)e_{\alpha B}\,\mathscr{F}_{\alpha\mu\nu} , \tag{21.4.23}$$

where

$$\mathscr{F}_{\beta\mu\nu} = \partial_\mu\mathscr{A}_{\beta\nu} - \partial_\nu\mathscr{A}_{\beta\mu} - \sum_{\gamma\delta}\mathscr{C}_{\beta\gamma\delta}\mathscr{A}_{\gamma\mu}\mathscr{A}_{\delta\nu} , \tag{21.4.24}$$

with $\mathscr{C}_{\beta\gamma\delta}$ the structure constant of the gauge group, related to the structure constant of G by

$$\sum_{CD} C_{BCD}\, e_{\gamma C}\, e_{\delta D} = \sum_\beta \mathscr{C}_{\beta\gamma\delta}\, e_{\beta B} . \tag{21.4.25}$$

We include in the Lagrangian as the kinematic term for this field the usual Yang–Mills term

$$\mathscr{L}_{\mathscr{A}} = -\frac{1}{4}\sum_\alpha \mathscr{F}_{\alpha\mu\nu}\mathscr{F}_\alpha{}^{\mu\nu} , \tag{21.4.26}$$

in which, by linear transformations of the $\mathscr{A}_{\alpha\mu}$ and correspondingly of the $e_{\alpha A}$, we have adjusted the coefficient of $\mathscr{F}_{\alpha\mu\nu}\mathscr{F}_\beta{}^{\mu\nu}$ to be just $\delta_{\alpha\beta}$. The linear term in $\mathscr{F}_{\alpha\mu\nu}$ is just $\partial_\mu\mathscr{A}_{\alpha\nu} - \partial_\nu\mathscr{A}_{\alpha\mu}$, so Eq. (21.4.26) has the effect of making $\mathscr{A}_{\alpha\mu}$ a canonically normalized vector field. The effective Lagrangian density is therefore to be taken as a function of $\tilde{\psi}$, $\mathscr{D}_\mu\tilde{\psi}$, and $\mathscr{D}_{a\mu}$ that is invariant under global H transformations, plus possible terms

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

that conserve $\mathscr{G}$ but not G, with coefficients proportional to two or more factors of the $e_{\alpha A}$.

Now let us see what sort of perturbation theory we can construct out of these ingredients. We know that the gauge bosons become massless in the limit $e_{\alpha A} \to 0$, where they decouple from the matter fields that experience the spontaneous symmetry breakdown. Let us therefore tentatively anticipate that their mass for small $e_{\alpha A}$ is of order eM, where e is a typical value of $e_{\alpha A}$ (the generators T_A being normalized to have structure constants of order unity) and M is an energy scale typical of the dynamics that leads to spontaneous symmetry breaking. We shall consider here a general Feynman diagram involving gauge and Goldstone bosons of energy or momentum $Q \lesssim eM$, and with all particles of higher energy or momentum and all heavier matter particles buried in corrections to the coupling constants in an effective field theory. Our perturbation theory will be an expansion in powers of e and Q/M. Following the same analysis as in Sections 19.4–6, the total number of powers of e and/or Q/M in any such diagram is

$$v = \sum_i V_i(d_i + e_i - 2) + 2L + 2 \,, \tag{21.4.27}$$

where V_i is the number of vertices of type i; d_i and e_i are the number of derivatives and factors of $e_{\alpha A}$, respectively, in an interaction of type i; and L is the number of loops. With the constraint (21.4.3), a field $A_{A\mu}$ or $\tilde{A}_{A\mu}$ contributes one factor of e. Inspection of Eq. (21.4.15) shows that each Goldstone boson covariant derivative $\mathscr{D}_{a\mu}$ contributes $+1$ to $d_i + e_i$, and inspection of Eqs. (21.4.19) and (21.4.15) shows that each additional covariant derivative of $\mathscr{D}_{a\mu}$ contributes another $+1$ to $d_i + e_i$. All allowed terms in the Lagrangian have $d_i + e_i \geq 2$, so the dominant contributions are those from tree graphs ($L = 0$) constructed entirely from interactions with $d_i + e_i = 2$. The only such interactions are the Goldstone boson kinematic term

$$\mathscr{L} = -\tfrac{1}{2} \sum_{ab} F_{ab}^2 \, \mathscr{D}_{a\mu} \mathscr{D}_b{}^{\mu} \,, \tag{21.4.28}$$

the Yang–Mills term (21.4.26), and possible symmetry-breaking non-derivative terms of second order in the $e_{\alpha A}$.

To see the physical significance of the field ξ_a, note that the linear term in $\mathscr{D}_{\alpha\mu}$ is

$$(\mathscr{D}_{a\mu})_{\text{LIN}} = \partial_\mu \xi_a - \sum_\alpha e_{\alpha a} \mathscr{A}_{\alpha\mu} \,. \tag{21.4.29}$$

As shown in the appendix to this chapter, we may always choose a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

'unitarity gauge' in which, for all α,

$$\sum_{ab} F_{ab}^2 \, \xi_a e_{\alpha b} = 0 \, , \tag{21.4.30}$$

which makes the cross term in Eq. (21.4.28) vanish. To clarify the significance of this condition, note that in the special case where all broken symmetries are gauge symmetries, any x_a can be written as a linear combination of gauge generators and unbroken generators

$$\begin{aligned} x_a &= \sum_\alpha c_{a\alpha} \mathscr{T}_\alpha + \sum_i c_{ai} t_i \, , \\ &= \sum_\alpha c_{a\alpha} \left(\sum_a e_{\alpha b} x_b + \sum_i e_{\alpha i} t_i \right) + \sum_i c_{ai} t_i \, , \end{aligned}$$

and therefore

$$\sum_\alpha c_{a\alpha} e_{\alpha b} = \delta_{ab} \, .$$

Contracting Eq. (21.4.30) with $c_{\alpha b}$, we see then that $\xi_a = 0$; there are no Goldstone bosons at all in this gauge. More generally, Eq. (21.4.30) leaves us with just those Goldstone bosons that do *not* correspond to gauge symmetries. Some of these are associated with elements of G that are broken by the gauge interactions, and therefore have masses of second order in gauge couplings; these are called pseudo-Goldstone bosons.

With ξ chosen to satisfy the unitarity gauge condition (21.4.30), the quadratic part of the Lagrangian (21.4.28) is simply

$$(\mathscr{L}_\xi)_{\text{QUAD}} = -\tfrac{1}{2} \sum_{ab} F_{ab}^2 \partial_\mu \xi_a \partial^\mu \xi_b - \tfrac{1}{2} \sum_{\alpha\beta} \mu_{\alpha\beta}^2 \mathscr{A}_{\alpha\mu} \mathscr{A}_\beta{}^\mu \, , \tag{21.4.31}$$

where

$$\mu_{\alpha\beta}^2 = \sum_{ab} F_{ab}^2 \, e_{\alpha a} e_{\beta b} \, . \tag{21.4.32}$$

This has two important implications. First, we note that ξ_a may be expressed in terms of a canonically orthonormalized field π_a, as

$$\xi_a = \sum_b F_{ab}^{-1} \, \pi_b \, , \tag{21.4.33}$$

where F_{ab} is the positive square root of the positive matrix F_{ab}^2. This shows that F_{ab}^{-1} are the factors analogous to F_π^{-1} that accompany the emission and absorption of low energy Goldstone bosons. Second, since $\mathscr{A}_{\alpha\mu}$ has been defined to be a canonically normalized vector field, Eq. (21.4.31) shows that $\mu_{\alpha\beta}^2$ is the square of the vector boson mass matrix. Eq. (21.4.32) is a universal formula for the vector boson mass matrix, valid to second order in the gauge couplings but to all orders in all other interactions. By using Eq. (21.4.1) in Eq. (21.1.7), it is easy to see that our

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

previous result Eq. (21.1.7) is a special case of Eq. (21.4.32), with

$$F^2_{ab} = -\sum_{nml}(x_a)_{nm}(x_b)_{nl}v_m v_l \ .$$

Eq. (21.4.32) may also be understood on the basis of the continuity arguments outlined (in a somewhat different notation) at the end of Section 21.1. It guarantees that the effects of gauge boson exchange that survive in the limit of zero gauge coupling are the same as the effects that would have been produced by Goldstone boson exchange if there were no gauge coupling.

In general, we cannot calculate the F^2_{ab} matrix, but we do know that it must be invariant under the unbroken subgroup H, in the sense that

$$\sum_d \left[C_{ibd}F^2_{dc} + C_{icd}F^2_{bd}\right] = 0 \ .$$

This condition allows us to put useful constraints on the gauge boson masses (21.4.32).

As an example, consider the case of the electroweak gauge group $SU(2) \times U(1)$, spontaneously broken to the $U(1)$ of electromagnetism. The three broken symmetry generators x_a can be taken as the three generators of $SU(2)$ (called t_1, t_2, t_3 in Section 21.3), without the coupling constant factor g, and the one unbroken symmetry generator t can be taken as the charge q, without the factor e. That is, the x_a and t are represented on the lepton doublets by the matrices

$$\vec{x} = \frac{1}{4}(1+\gamma_5)\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\} ,$$

$$t = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} .$$

The gauge generators are then given by

$$\vec{\mathscr{T}} = g\vec{x} \ , \qquad \mathscr{T}_y = g'(x_3 - t) \ . \tag{21.4.34}$$

That is, the non-zero coefficients $e_{\alpha a}$ of x_a in the gauge generator $\mathscr{T}_\alpha$ are

$$e_{11} = e_{22} = e_{33} = g \ , \qquad e_{y3} = g' \ .$$

Also, since t subjects the three-vector $\vec{x}$ to a rotation around the three-axis, this unbroken symmetry requires the matrix F^2_{ab} to have the non-zero components

$$F^2_{11} = F^2_{22} \equiv F^2_C \ , \qquad F^2_{33} \equiv F^2_N \ .$$

According to Eq. (21.4.32), the mass squared matrix of the gauge bosons then has the non-vanishing elements

$$\mu^2_{11} = \mu^2_{22} = g^2 F^2_C \ , \qquad \mu^2_{33} = g^2 F^2_N \ ,$$
$$\mu^2_{3y} = g\,g' F^2_N \ , \qquad \mu^2_{yy} = g'^2 F^2_N \ .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Its eigenvalues are

$$m_W^2 = g^2 F_C^2 , \qquad m_Z^2 = (g^2 + g'^2) F_N^2 , \qquad m_A^2 = 0 . \tag{21.4.35}$$

To go further, we need a relation between F_C and F_N. This is provided if the theory is invariant in the limit $g = g' = 0$ under a global symmetry group G larger than $SU(2) \times U(1)$, which breaks spontaneously to a subgroup H, that includes three-dimensional rotations under which $\vec{x}$ rotates as a three-vector. Such an unbroken symmetry would require F_{ab}^2 to be proportional to δ_{ab}, and hence

$$F_C = F_N .$$

Any such symmetry is known as a 'custodial' symmetry. It has the consequence that m_Z/m_W is given in terms of the gauge coupling constants by the successful formula discussed in Section 21.3

$$m_Z/m_W = \sqrt{1 + g'^2/g^2} = 1/\sin\theta . \tag{21.4.36}$$

For instance, in the absence of the gauge couplings, the Lagrangian (21.3.28) for the scalar doublet ϕ in the simplest version of the $SU(2)\times U(1)$ electroweak theory may be written

$$(\mathscr{L}_\phi)_{g=g'=0} = -\frac{1}{2}\,\partial_\mu \phi_n \partial^\mu \phi_n - \frac{\mu^2}{2}\,\phi_n \phi_n - \frac{\lambda}{4}(\phi_n \phi_n)^2 ,$$

where

$$\phi_1 \equiv \mathrm{Im}\phi^+ , \quad \phi_2 \equiv \mathrm{Re}\phi^+ , \quad \phi_3 \equiv \mathrm{Im}\phi^0 , \quad \phi_4 \equiv \mathrm{Re}\phi^0 .$$

This is automatically invariant under an 'accidental' $SO(4) \equiv SU(2) \times SU(2)$ global symmetry group, which is spontaneously broken by the vacuum expectation value of $\mathrm{Re}\phi^0$ down to an approximate unbroken $SO(3)$ custodial subgroup.[29] The result (21.4.36) applies even if there is more than one scalar doublet, because even though in this case the mass and interaction terms in the scalar Lagrangian will not in general respect the custodial symmetry, it is only the kinematic term that enters in the derivation, and this always has the full $SO(4)$ symmetry.

Custodial symmetries can be found in other theories. Consider for instance a theory with no scalar fields, but with new extra-strong vector gauge interactions,[30] called technicolor interactions, that act on a new $SU(2) \times U(1)$ doublet (U_r, D_r) of 'techniquarks' U_r and D_r, with r a technicolor index. As long as the left- and right-handed parts of both U_r and D_r all transform in the same way under the technicolor gauge group, the Lagrangian will be invariant in the limit of vanishing electroweak couplings under the group $SU(2) \times SU(2)$ of independent $SU(2)$ transformations on the left- and right-handed techniquark doublets. According to the arguments described in Section 19.9, the subgroup $SU(2)_V$ consisting

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

of simultaneous $SU(2)$ transformations on both the left- and right-handed techniquark doublets will not be spontaneously broken. It is reasonable to suppose that the technicolor interactions will produce a spontaneous breakdown of $SU(2) \times SU(2)$ to $SU(2)_V$, just as color interactions led to a spontaneous breakdown of the chiral $SU(2) \times SU(2)$ symmetry of quantum chromodynamics (for vanishing u and d quark masses) to its isospin subgroup. Under the unbroken $SU(2)_V$ symmetry the electroweak generator $\vec{x}$ or $\vec{\mathscr{T}}$ rotates as a three-vector, leading again to the relation $F_C = F_N$, and the consequent successful prediction of the relation between the W and Z masses.

The technicolor idea is attractive, because it provides a natural mechanism for breaking the electroweak symmetry at a characteristic scale which is very much smaller than what is often supposed to be the fundamental scale of physics, generally assumed (as in string theories) to be of the order of the Planck mass, or about 10^{18} GeV. It is only necessary to suppose that just below the fundamental scale there is an unbroken gauge group consisting of the $SU(3) \times SU(2) \times U(1)$ of the strong and electroweak interactions, plus a technicolor gauge group, all with comparable small coupling constants. If the technicolor gauge interactions are asymptotically free the technicolor coupling like the QCD color coupling will increase slowly with decreasing energy, becoming strong at an energy much less than the fundamental scale. This energy where the technicolor coupling becomes strong would set the scale for the parameters F_{ab} that appear in our formula, Eq. (21.4.32), for the gauge boson masses, and would therefore presumably be of the order of 300 GeV. Since the increase of coupling with decreasing energy is logarithmic, a moderate difference in the beta-functions for technicolor and color can easily give rise to the three orders of magnitude difference between the energy scales where color and technicolor forces become strong.

Unfortunately, although technicolor provides a very attractive picture of the spontaneous breaking of $SU(2) \times U(1)$, it does not by itself offer a mechanism for giving masses to the quarks and leptons. For this reason it has been suggested to add additional 'extended technicolor' gauge interactions with transformations that link quarks and techniquarks.[31] Such theories have potential problems with flavor-changing neutral current weak interactions, and though these problems may be surmounted, the added complications reduce their attractiveness. The question of elementary weakly coupled scalars versus dynamical symmetry breaking remains open.

21.5 Electroweak–Strong Unification

We saw in Section 15.2 that a gauge theory will have an independent coupling constant for each simple or $U(1)$ subgroup of the gauge group.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Thus the electroweak theory which is based on the gauge group $SU(2) \times U(1)$, has two independent couplings, g and g'. In order to reduce the number of free parameters it was suggested[32] that the $SU(2) \times U(1)$ gauge group might be embedded in a simple $SU(3)$ gauge group, which would give $g' = g/\sqrt{3}$, but this has been ruled out experimentally. After the advent of quantum chromodynamics, theorists confronted a gauge group $SU(3) \times SU(2) \times U(1)$, which has three independent gauge coupling constants: the coupling g_s of quantum chromodynamics, and the couplings g and g' of the electroweak interactions. To reduce these to just a single free parameter, it was proposed to embed $SU(3) \times SU(2) \times U(1)$ in various simple Lie groups:* $SU(4) \times SU(4)$,[33] $SU(5)$,[34] or $SO(10)$.[35] Such models are often known as grand-unified theories.

Fortunately, the consequences of these and a large class of other models for the ratios of the $SU(3) \times SU(2) \times U(1)$ coupling constants are independent of the details of the individual models.[36] This class of models is characterized by the fact that the observed generations of quarks and leptons are the only fermions in the models, or at least the only fermions that are not neutral under $SU(3) \times SU(2) \times U(1)$.

As shown in Section 15.2, for any simple compact Lie group there is a conventional choice of generators T_α with totally antisymmetric structure constants, which in each reducible or irreducible representation D satisfy the normalization condition:

$$\mathrm{Tr}\{T_\alpha T_\beta\} = N_D \delta_{\alpha\beta} \,. \tag{21.5.1}$$

We are assuming that all left-handed fermions form n_g generations:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \quad \cdots ,$$

$$\overline{e_R} \qquad \overline{\mu_R} \qquad \overline{\tau_R} \qquad \cdots ,$$

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \quad \cdots ,$$

$$\overline{u_R} \qquad \overline{c_R} \qquad \overline{t_R} \qquad \cdots ,$$

$$\overline{d_R} \qquad \overline{s_R} \qquad \overline{b_R} \qquad \cdots .$$

The $SU(3)$ generator $\frac{1}{2} g_s \lambda_3$ has eigenvalues: $+\frac{1}{2} g_s$ for the red quark doublets and the white antiquark singlets; $-\frac{1}{2} g_s$ for the white quark doublets and the red antiquark singlets; and zero for all other left-handed

* The group $SU(4) \times SU(4)$ is made simple by inclusion of a discrete symmetry operator that interchanges the two $SU(4)$s.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

fermions; so its square has the trace

$$\mathrm{Tr}\,(\tfrac{1}{2}g_s\lambda_3)^2 = 4n_g \times (\tfrac{1}{2}g_s)^2 + 4n_g \times (-\tfrac{1}{2}g_s)^2 = 2n_g g_s^2\,. \tag{21.5.2}$$

The $SU(2)$ generator t_3 has eigenvalues: $\frac{1}{2}g$ for the red, white, and blue quarks of charge $\frac{2}{3}$ and the neutrinos; $-\frac{1}{2}g$ for the red, white, and blue quarks of charge $-\frac{1}{3}$ and the charged leptons; and zero for all other fermions, so its square has the trace

$$\mathrm{Tr}\,(t_3)^2 = [3n_g + n_g] \times [(\tfrac{1}{2}g)^2 + (-\tfrac{1}{2}g)^2] = 2n_g g^2\,. \tag{21.5.3}$$

Finally, the $U(1)$ generator $y = t_3 - q$ has eigenvalues: $\frac{1}{2}g'$ for neutrinos and charged leptons; $-g'$ for charged antileptons; $-\frac{1}{6}g'$ for quarks; $\frac{2}{3}g'$ for antiquarks of charge $-\frac{2}{3}$; and $-\frac{1}{3}g'$ for antiquarks of charge $+\frac{1}{3}$, so its square has the trace

$$\begin{aligned}\mathrm{Tr}\,y^2 &= 2n_g(\tfrac{1}{2}g')^2 + n_g(-g')^2 + 6n_g(-\tfrac{1}{6}g')^2 + 3n_g(\tfrac{2}{3}g')^2 + 3n_g(-\tfrac{1}{3}g')^2 \\ &= \tfrac{10}{3}n_g g'^2\,.\end{aligned} \tag{21.5.4}$$

Eq. (21.5.1) requires the traces (21.5.2)–(21.5.4) to be all equal, so in this class of models the embedding of $SU(3) \times SU(2) \times U(1)$ in a simple Lie group imposes the coupling constant relations

$$g_s^2 = g^2 = \tfrac{5}{3}g'^2\,. \tag{21.5.5}$$

Now, Eq. (21.5.5) is in gross disagreement with the observed values of the coupling constants. The ratio $g'^2/g^2 = \frac{3}{5}$ implies an electroweak mixing angle with $\sin^2\theta \equiv g'^2/(g^2+g'^2) = \frac{3}{8}$, while the experimental value is $\sin^2\theta = 0.231$. Even worse, the strong coupling g_s^2 is of course much larger than g^2 or g'^2.

The solution[36] to this problem is that coupling-constant relations like Eq. (21.5.5) apply only to the couplings measured at an energy scale comparable to the typical mass[47] M of the gauge bosons that become massive in the spontaneous breakdown of the simple gauge group to $SU(3) \times SU(2) \times U(1)$. If the energy E at which the couplings is measured is very much less than M, then there will be large radiative corrections proportional to $\ln(M/E)$.

As emphasized in Chapter 19, there are no large logarithms in the relation between couplings measured at nearby energies μ and $\mu - d\mu$, so by integrating this relation from M down to E we can calculate the couplings at energies $E \ll M$ without encountering large logarithms. For this to be done it is only necessary that the couplings should stay small over this whole range. For the $SU(3) \times SU(2) \times U(1)$ couplings with n_g

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

fermion generations, Eq. (18.7.2) gives**

$$\mu\frac{d}{d\mu}g_s(\mu) = -\frac{g_s^3(\mu)}{4\pi^2}\left(\frac{11}{4} - \frac{n_g}{3}\right), \tag{21.5.6}$$

$$\mu\frac{d}{d\mu}g(\mu) = -\frac{g^3(\mu)}{4\pi^2}\left(\frac{11}{6} - \frac{n_g}{3}\right), \tag{21.5.7}$$

$$\mu\frac{d}{d\mu}g'(\mu) = -\frac{g'^3(\mu)}{4\pi^2}\left(-\frac{5n_g}{9}\right). \tag{21.5.8}$$

The solutions of these equations are

$$\frac{1}{g_s^2(\mu)} = \frac{1}{g_s^2(M)} - \frac{1}{8\pi^2}\left(11 - \frac{4n_g}{3}\right)\ln\left(\frac{M}{\mu}\right), \tag{21.5.9}$$

$$\frac{1}{g^2(\mu)} = \frac{1}{g^2(M)} - \frac{1}{8\pi^2}\left(\frac{22}{3} - \frac{4n_g}{3}\right)\ln\left(\frac{M}{\mu}\right), \tag{21.5.10}$$

$$\frac{1}{g'^2(\mu)} = \frac{1}{g'^2(M)} - \frac{1}{8\pi^2}\left(-\frac{20n_g}{9}\right)\ln\left(\frac{M}{\mu}\right). \tag{21.5.11}$$

Also, Eq. (21.5.5) should be interpreted to mean that

$$g_s^2(M) = g^2(M) = \tfrac{5}{3}g'^2(M). \tag{21.5.12}$$

We can therefore eliminate the couplings (21.5.12) *and* the number of generations by subtracting Eq. (21.5.10) from Eq. (21.5.9):

$$\frac{1}{g_s^2(\mu)} - \frac{1}{g^2(\mu)} = -\frac{11}{24\pi^2}\ln\left(\frac{M}{\mu}\right) \tag{21.5.13}$$

and by subtracting 3/5 of Eq. (21.5.11) from Eq. (21.5.10):

$$\frac{1}{g^2(\mu)} - \frac{3}{5g'^2(\mu)} = -\frac{11}{12\pi^2}\ln\left(\frac{M}{\mu}\right). \tag{21.5.14}$$

Taking the ratio of these two equations gives a formula for $\sin^2\theta \equiv g'^2/(g^2+g'^2)$:

$$\sin^2\theta = \frac{1}{6} + \frac{5e^2(m_Z)}{9g_s^2(m_Z)}. \tag{21.5.15}$$

** The second term in the brackets in Eqs. (21.5.6) and (21.5.7) is given by Eqs. (18.7.2) and (18.7.3) as $-n_f/6$, where n_f are the numbers of fermions in the defining representation of the respective $SU(N)$ gauge groups. But this was calculated under the assumption that left- and right-handed fermions are in the same representation of the gauge group. If we count only left-handed fermions (and antifermions), then the second term in the brackets in Eqs. (21.5.6) and (21.5.7) should be $-n_f/12$. For $SU(3)$ there are two left-handed quark triplets and two left-handed antiquark triplets per generation, so $n_f = 4n_g$, while for $SU(2)$ there are three left-handed quark doublets and one left-handed lepton doublet per generation, so again $n_f = 4n_g$. For $U(1)$, the beta-function is $g'/24\pi^2$ times the sum of the squares of the $U(1)$ charges for left-handed fermions and antifermions (compare Eq. (18.2.38)), which according to Eq. (21.5.4) is $(g'/24\pi^2)\times(10n_g g'^2/3)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

In this formula we have set μ equal to the typical energy of the processes used to measure $\sin^2\theta$, that is, $\mu \approx m_Z$. This has the advantage that we will be using the renormalization group equations (21.5.6)–(21.5.8) only above m_Z, where they are not strongly affected by the spontaneous breakdown of $SU(2) \times U(1)$. Eqs. (21.5.13) and (21.5.14) can also be combined to give a formula for the unification scale M:

$$\ln\left(\frac{M}{m_Z}\right) = \frac{4\pi^2}{11e^2}\left(1 - \frac{8e^2(m_Z)}{3g_s^2(m_Z)}\right), \tag{21.5.16}$$

where again, to avoid the effects of electroweak symmetry breaking on the renormalization group equations, we have taken μ as of order m_Z.

We saw in Section 18.2 that the value of $e(\mu)$ at $\mu \approx m_Z$ is given by $e(m_Z)^2/4\pi = (128.87 \pm 0.12)^{-1}$. This is the charge defined in the conventional (Gell-Mann–Low) manner, in terms of the vacuum polarization. For purposes of comparison with g_s, g', and g, it is better to use the coupling defined (as in Section 18.6) by modified minimum subtraction: $e(m_Z)^2/4\pi = (127.9 \pm 0.1)^{-1}$. The greatest uncertainty in Eqs. (21.5.13) and (21.5.14) is in the value of $g_s^2(m_Z)$. As discussed in Section 18.7, extrapolation of g_s from lower energy data gives $g_s^2(m_Z)/4\pi = 0.118 \pm 0.006$, while direct measurement from the rate of Z^0 decay into hadrons gives $g_s^2(m_Z)/4\pi = 0.120 \pm 0.0025$. For $g_s^2(m_Z)/4\pi = 0.118$ and $e(m_Z)^2/4\pi = 1/128$, Eqs. (21.5.15) and (21.5.16) give $\sin^2\theta = 0.203$ and $M \approx 1.1 \times 10^{15}$ GeV.

As mentioned in Section 21.3, there is no reason to expect that baryon and lepton number should be conserved by the suppressed non-renormalizable terms in the effective Lagrangian that describes physics at ordinary energies, so we may expect the presence of an $(SU(3) \times SU(2) \times U(1))$-conserving four fermion (three quarks and a lepton) interaction, with a coefficient that on dimensional grounds would be of order M^{-2}. The proton lifetime was first estimated on this basis, and found to be of order 10^{32} years.[36] Such baryon- and lepton-non-conserving four-fermion interactions are produced in models like those of Refs. 33–35 from the exchange of gauge bosons with masses of order M. More generally, once the standard model explained why baryon- and lepton-non-conserving processes are naturally suppressed, there ceased to be any reason to believe that either baryon or lepton number are exactly conserved.

We have seen that the prediction (21.5.15) comes quite close to the measured value 0.23 of $\sin^2\theta$, but the accuracy of the measurements and calculations has become good enough to make clear that they are not in precise agreement. The extra particles in supersymmetric theories appear to remove this discrepancy,[36a] and lead to a value of $M \simeq 2 \times 10^{16}$ GeV, about an order of magnitude larger.[36b] It is very interesting that this M is not so very different from the energy 10^{18} GeV at which the gravitational

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

interactions become strong. The larger value of M also has the effect of increasing the proton lifetime, which is proportional to M^4.

21.6 Superconductivity*

Superconductivity is quite different from the elementary particle phenomena that chiefly concern us in this book, but it is worth some consideration here, both as the earliest realistic example of a spontaneously broken gauge symmetry, and also as an exceptionally enlightening example of the power of effective field theories and of the use of topological arguments in field theory.

A superconductor is simply a material in which electromagnetic gauge invariance is spontaneously broken.** Detailed dynamical theories are needed to explain why and at what temperatures this symmetry breaking occurs, but they are not needed to derive the most striking aspects of superconductivity: exclusion of magnetic fields, flux quantization, zero resistivity, and alternating currents at a gap between superconductors held at different voltages. As we shall see here, these consequences of broken gauge invariance can be worked out, in a manner somewhat like our treatment of soft pions, solely on the basis of general properties of the Goldstone mode.[39]

The action for any system will be invariant under gauge transformations, which in cgs units take the form

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x)\,, \tag{21.6.1}$$

$$\psi_n(x) \to \exp\left(iq_n\Lambda(x)/\hbar\right)\psi_n(x)\,, \tag{21.6.2}$$

where $\Lambda(x)$ is arbitrary, and q_n is the electric charge destroyed by field ψ_n. All charges are assumed to be integral multiples of the electron charge $-e$, so this group is compact: the phases Λ and $\Lambda + 2\pi\hbar/e$ are regarded as identical. This symmetry group is assumed to be broken in a superconductor by non-vanishing expectation values of operators carrying charge $-2e$ (such as products of two electron fields), so there

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

** This is not the way that most experts have historically thought about superconductivity. Early phenomenological theories were known to violate electromagnetic gauge invariance, but this was regarded as more annoying than enlightening. Broken symmetry is never mentioned in the seminal paper by Bardeen, Cooper, and Schrieffer[37] that first gave us a microscopic theory of superconductivity. Anderson[38] subsequently stressed the important role of broken symmetry in superconductors, but even today most textbooks explain superconductivity in terms of detailed dynamical models, with broken symmetry rarely mentioned.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

is an unbroken Z_2 subgroup, consisting of gauge transformations with $\Lambda = 0$ and $\Lambda = \pi\hbar/e$.

We introduce a Goldstone boson field $\phi(x)$ by writing all charged fields as

$$\psi_n(x) = \exp\left(iq_n\phi(x)/\hbar\right)\tilde{\psi}_n(x)\,. \tag{21.6.3}$$

The field $\phi(x)$ parameterizes the coset space $U(1)/Z_2$, and so is given the gauge transformation property

$$\phi(x) \to \phi(x) + \Lambda(x)\,. \tag{21.6.4}$$

Because $\phi(x)$ parameterizes $U(1)/Z_2$ rather than $U(1)$, we must identify $\phi(x)$ and $\phi(x) + \pi\hbar/e$. All the $\tilde{\psi}_n(x)$ are gauge-invariant, and so when integrated out leave the Lagrangian as a gauge-invariant functional of ϕ and A^μ alone. It follows that the Lagrangian for the Goldstone and electromagnetic fields may be written as

$$L = -\tfrac{1}{4}\int d^3x\, F_{\mu\nu}F^{\mu\nu} + L_s\left[A_\mu - \partial_\mu\phi\right], \tag{21.6.5}$$

where L_s is an imperfectly known functional. The electric current and charge density here are

$$\mathbf{J}(x) = \frac{\delta L_s}{\delta \mathbf{A}(x)}\,, \tag{21.6.6}$$

$$J^0(x) = -\frac{\delta L_s}{\delta A^0(x)} = -\frac{\delta L_s}{\delta\dot{\phi}(x)}\,. \tag{21.6.7}$$

The equations of motion for the Goldstone boson field are then

$$\frac{\partial}{\partial t}\frac{\delta L_s}{\delta\dot{\phi}(x)} = \frac{\delta L_s}{\delta\phi(x)} = \nabla\cdot\frac{\delta L_s}{\delta\mathbf{A}(x)}\,, \tag{21.6.8}$$

which in light of Eqs. (21.6.6) and (21.6.7) is equivalent to the conservation law of electric charge

$$\nabla\cdot\mathbf{J} + \frac{\partial}{\partial t}J^0 = 0\,. \tag{21.6.9}$$

Now let us see how this formalism explains the remarkable properties of superconductors. About L_s we will only need to assume that the system is stable in the absence of Goldstone or external electromagnetic fields, so that the energy is at least at a local minimum at $A_\mu = \partial_\mu\phi$, with non-vanishing second derivatives with respect to $A_\mu - \partial_\mu\phi$.

One immediate consequence is that, deep in a large superconductor where boundary conditions are unimportant, the electromagnetic field is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

a pure gauge:

$$A_\mu = \partial_\mu \phi \,, \tag{21.6.10}$$

so that in particular the magnetic field must vanish. This is known as the *Meissner effect*. It is possible to be a little more quantitative about what we mean by 'deep in a large superconductor.' Since the energy is a minimum when Eq. (21.6.10) is satisfied, for small values of $|\mathbf{A} - \nabla\phi|$ it must be of order $|\mathbf{A} - \nabla\phi|^2 L^3/\lambda^2$, where λ is some length depending on the nature of the material, and L^3 is the superconductor volume. If a magnetic field of order B penetrated the superconductor, then we would have $|\mathbf{A} - \nabla\phi|$ of order BL, so the energy cost in allowing the magnetic field into the superconductor would be of order B^2L^5/λ^2. On the other hand, the energy cost in expelling a magnetic field B from a volume L^3 is of order B^2L^3. Hence a weak magnetic field will be expelled from a superconductor if $B^2L^5/\lambda^2 \gg B^2L^3$, or in other words if $L \gg \lambda$. For this reason λ is known as the *penetration depth* of the superconductor.

The same energetic considerations tell us that for any superconducting material, there is a critical magnetic field, above which superconductivity is extinguished. The existence of superconductivity at zero magnetic field means that the material in its normal state has an energy per unit volume which is higher than that in the superconducting state, say by an amount Δ. When a superconductor with linear dimensions much larger than λ is placed in a magnetic field B, the magnetic field is expelled from most of the material, at an energy cost per volume of $B^2/2$. Hence it is energetically favorable for the material to be in the superconducting state if and only if the magnetic field is below a critical value

$$B_c = \sqrt{2\Delta} \,. \tag{21.6.11}$$

(This is for uniform superconductors. As we shall see below, for certain kinds of superconductor it is possible to maintain superconductivity throughout most of the material for magnetic fields in a finite range above B_c by the formation of narrow vortex lines with normal metal at their cores.) A magnetic field $B < B_c$ will penetrate a superconductor to a depth λ, but this does not extinguish superconductivity in this layer; indeed, as shown by the field equation $\nabla \times \mathbf{B} = \mathbf{J}$, it is in this surface layer of the superconductor that electric currents can flow.

Now consider a thick superconducting wire, with thickness much larger than λ, bent into a closed ring. We can draw a closed contour $\mathscr{C}$ running deep inside the wire, along which $|\mathbf{A} - \nabla\phi|$ must vanish. This does not mean that either $\mathbf{A}$ or ϕ vanishes on this contour, but we do know that in going around the ring ϕ must return to an equivalent value, and can therefore only change by an amount $n\pi\hbar/e$, with n a positive or negative integer or zero. It follows then by Stokes's theorem that the magnetic flux

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

through the area $\mathscr{A}$ surrounded by $\mathscr{C}$ is subject to the *flux quantization* rule

$$\int_{\mathscr{A}} \mathbf{B} \cdot \mathbf{dS} = \oint_{\mathscr{C}} \mathbf{A} \cdot d\mathbf{x} = \oint_{\mathscr{C}} \nabla\phi \cdot d\mathbf{x} = \frac{n\pi\hbar}{e} . \tag{21.6.12}$$

The electric current that maintains the magnetic flux (21.6.12) flows in a layer of thickness λ just below the surface of the superconducting wire. The quantization of flux shows that this current cannot decay smoothly, but only in jumps at which the flux (21.6.12) drops by multiples of $\pi\hbar/e$, so there can be no ordinary electrical resistance in the superconductor.

The absence of resistance in a superconductor can be shown in a more general context than closed rings, by considering time-dependent effects in superconductors. Note that Eq. (21.6.7) can be interpreted as the statement that $-J^0$ is the canonical conjugate to ϕ. The Hamiltonian H_s is thus to be regarded as a functional of ϕ and J^0 rather than ϕ and $\dot{\phi}$, with the time dependence of ϕ given by the Hamiltonian equation

$$\dot{\phi}(x) = \frac{\delta H_s}{\delta(-J^0(x))} . \tag{21.6.13}$$

Now, the 'voltage' $V(x)$ at any point is just the change in energy density per change in the charge density at that point, so Eq. (21.6.13) gives the time dependence of the Goldstone boson field as

$$\dot{\phi}(x) = -V(x) . \tag{21.6.14}$$

It follows that a piece of superconducting wire that carries a steady current, with time-independent fields, must have zero voltage difference between its ends, because otherwise $\phi(x)$ would have a time-dependent gradient. A zero voltage difference at finite current is what we mean by zero resistance.

Now consider a gap between two pieces of superconducting material. In the absence of any gradients along the surface of the gap, or any vector potential, gauge invariance requires L_s to depend only on the difference $\Delta\phi$ between the Goldstone boson fields in the two superconductors:

$$L_{\text{junction}} = \mathscr{A} F(\Delta\phi) , \tag{21.6.15}$$

where $\mathscr{A}$ is the area of the junction. Furthermore, we can shift ϕ in either superconductor by any integer multiple of $\pi\hbar/e$ without physical effect, so the function F must be periodic[†]

$$F(\Delta\phi) = F(\Delta\phi + \pi\hbar n/e) . \tag{21.6.16}$$

[†] This function was calculated by Josephson,[40] who found it to be proportional to $\cos(2e\Delta\phi/\hbar)$, but this is an approximate result, while the periodicity is exact.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

A current flows through such a gap, which can be calculated by considering the junction in the presence of a vector potential $\mathbf{A}$. Gauge invariance then tells us that in place of $\Delta\phi$, the function F must depend on

$$\Delta_A\phi = \int \mathbf{dx}\cdot(\nabla\phi - \mathbf{A})\,,$$

the integral being taken over a line joining the two superconductors. Eq. (21.6.6) then shows that the current density is

$$\mathbf{J} = \frac{\delta L_{\text{junction}}}{\delta \mathbf{A}} = -\hat{\mathbf{n}}F'(\Delta_A\phi)\,,$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the gap. We can now take away the vector potential, and find the current

$$\mathbf{J} = -\hat{\mathbf{n}}F'(\Delta\phi)\,. \tag{21.6.17}$$

If we now suppose that the two superconductors are maintained at uniform voltages, with a voltage difference ΔV, then, according to Eq. (21.6.14), the difference in the Goldstone boson fields will have the time dependence

$$\Delta\phi = -t\Delta V + \text{constant}\,. \tag{21.6.18}$$

Using this in Eq. (21.6.17) and recalling Eq. (21.6.16), we see that the current oscillates at a frequency

$$\nu = e|\Delta V|/\pi\hbar\,. \tag{21.6.19}$$

This is the *ac Josephson effect*.[40] It is possible to measure frequencies and voltages with great accuracy, so this effect provides a very accurate method of measuring the constant $e/\hbar$.

As pointed out at the end of Section 19.6, the description of a system with broken symmetry in terms of Goldstone modes alone becomes inadequate when a system is brought close to the point where the broken symmetry becomes unbroken. Under these circumstances the Goldstone mode is accompanied with other modes that have nearly zero frequency in the long-wavelength limit, which together with the Goldstone mode forms a *linear* representation, usually irreducible, of the symmetry group, known as the order parameter. It is plausible to assume that a uniform superconductor in slowly varying external fields is described by a *local* order parameter, because any non-locality would be characterized by microscopic distance scales (such as the mean electron–electron spacing) that are much smaller than the scales of distances over which the electromagnetic and Goldstone boson fields are assumed to be varying. For superconductivity there is no doubt about the nature of this order parameter. The only non-trivial irreducible linear representation of the group $U(1)$ is a real two-vector ψ_n, or equivalently a Goldstone mode ϕ

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

and modulus field ρ, with

$$\psi_1 + i\psi_2 = \rho \exp\left(2ie\phi/\hbar\right) \equiv \psi \,. \tag{21.6.20}$$

The coefficient of $i\phi$ in the exponential must be $2e/\hbar$ in order that a gauge transformation with $\Lambda = \pi\hbar/e$ (and with no smaller Λ) should leave ψ_n invariant. For a nearly uniform time-independent system close to the symmetry-breaking transition, the order parameter is small and slowly varying in space, and so the Lagrangian in an external vector potential $\mathbf{A}$ may be approximated (from now on using natural units with $\hbar = 1$) by

$$L_s \simeq \int d^3x \left[-\tfrac{1}{2}\left(\nabla\psi_n - 2ie\, t_{nm}\mathbf{A}\psi_m\right)^2 + \tfrac{1}{2}m^2\psi_n\psi_n - \tfrac{1}{4}g(\psi_n\psi_n)^2 \right], \tag{21.6.21}$$

where t is the Hermitian $U(1)$ generator

$$t = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \,,$$

and g must be taken positive to give a bounded Hamiltonian. This is the *Ginzburg–Landau* theory of superconductivity.[41] It has been derived by Gor'kov[42] from the microscopic theory of superconductivity presented below, in the case of a short-range potential and a temperature close to the critical temperature at which the material loses its superconductivity.

In terms of ρ and ϕ, Eq. (21.6.21) becomes

$$L_s \simeq \int d^3x \left[-2e^2\rho^2\left(\nabla\phi - \mathbf{A}\right)^2 + \tfrac{1}{2}m^2\rho^2 - \tfrac{1}{4}g\rho^4 - \tfrac{1}{2}\left(\nabla\rho\right)^2 \right] . \tag{21.6.22}$$

The field equations are then

$$\nabla \times \mathbf{B} = 4e^2\rho^2(\nabla\phi - \mathbf{A}) \,, \tag{21.6.23}$$

$$\nabla^2\rho = -m^2\rho + g\rho^3 + 4e^2\rho(\nabla\phi - \mathbf{A})^2 \,. \tag{21.6.24}$$

The $U(1)$ symmetry is broken if these field equations are satisfied for $\rho \neq 0$, which in a field-free homogeneous material will be the case if $m^2 > 0$, in which case ρ takes the value $\langle\rho\rangle = m/\sqrt{g}$. The penetration depth λ was defined earlier as the inverse square root of the coefficient of $-\frac{1}{2}(\nabla\phi - \mathbf{A})^2$, so here

$$\lambda = \frac{1}{\sqrt{4e^2\langle\rho\rangle^2}} = \frac{\sqrt{g}}{2em} \,. \tag{21.6.25}$$

This is the distance that according to Eq. (21.6.23) characterizes variations in the magnetic field. On the other hand, variations in the modulus ρ are characterized by a distance scale known as the *correlation length*, given

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

according to Eq. (21.6.24) by[††]

$$\xi = 1/m\sqrt{2}\,. \tag{21.6.26}$$

Also, the superconducting state with $\rho = \langle\rho\rangle$ has an energy per unit volume that is less than that of the normal state with $\rho = 0$ by an amount

$$\Delta = \tfrac{1}{2}m^2\langle\rho\rangle^2 - \tfrac{1}{4}\langle\rho\rangle^4 = m^4/4g\,. \tag{21.6.27}$$

Eliminating the parameters m and g from Eqs. (21.6.25)–(21.6.27), we find one important approximate relation among the observable quantities λ, ξ, and Δ:

$$\Delta \simeq \frac{1}{8\,e^2\lambda^2\xi^2}\,. \tag{21.6.28}$$

The modular field becomes important in the dynamics of superconducting vortex lines. These arise when a superconductor of a certain type is placed in a magnetic field that is strong enough so that it is energetically favorable for tubes of magnetic flux known as vortex lines to penetrate the material.[43] (The conditions for the appearance of vortex lines are discussed below.) By drawing a closed curve $\mathscr{C}$ around the tube at a distance much larger than the penetration depth, where the magnetic field vanishes, and repeating the argument contained in Eq. (21.6.12), we see that the magnetic flux through the area $\mathscr{A}$ surrounded by $\mathscr{C}$ must be equal to the change of ϕ around the curve, and hence equal to an integer multiple of the flux quantum π/e, just as for the flux through a thick superconducting ring. Where this flux is not zero, there must be a line within each tube along which electromagnetic gauge invariance is *not* broken. To see this, note that if we shrink the curve $\mathscr{C}$ into the region of high magnetic field it becomes no longer true that $\nabla\phi = \mathbf{A}$, but the change of ϕ around the curve must remain an integer multiple of π/e, and so by continuity cannot change. Thus we must eventually encounter a line (conceivably of finite thickness) along which ρ vanishes, so that ϕ becomes ill-defined. (This is an elementary example of the sort of topological reasoning we shall use in Chapter 23.) Near this line we must take both ρ and ϕ into account as dynamical variables.

The quantization of magnetic flux shows that a superconducting vortex line of minimum flux π/e is stable. A vortex line of higher flux cannot simply disappear, but magnetic flux quantization alone does not prevent it from breaking up into vortex lines of smaller flux. Bogomol'nyi[43a] has shown that vortex lines of flux $n\pi/e$ with $n > 1$ are unstable against breakup into n vortex lines of flux π/e if and only if $\lambda > \xi$.

[††] The factor $\sqrt{2}$ is included along with m because at $\rho = \langle\rho\rangle$ the derivative of the function $-m^2\rho + g\rho^3$ in Eq. (21.6.24) is $2m^2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

For this and other reasons it is convenient to divide superconducting materials into two classes: *type I superconductors* (most pure metals, except niobium) have $\xi > \lambda$, while *type II superconductors* (niobium and most alloys) have $\lambda > \xi$. The corresponding distinction in the electroweak standard model is between theories where the scalar mass (analogous to $1/\xi$) is less than or greater than the W and Z masses (analogous to $1/\lambda$).

From the definitions of the correlation length ξ and penetration depth λ, it follows that the modular parameter will rise from zero at the central line of a vortex to its equilibrium value $\langle\rho\rangle$ in a distance of the order of the correlation length ξ, while the magnetic field will decay with distance from the central line in a distance of the order of the penetration depth λ. Hence a vortex solution in a type I superconductor with $\xi \gg \lambda$ would consist of a thin inner cylinder of nearly normal metal, within which the magnetic field drops to zero, surrounded by a much thicker outer cylinder within which the modular parameter rises to its asymptotic value $\langle\rho\rangle$. In contrast, a vortex in a type II superconductor with $\lambda \gg \xi$ consists of a thin inner cylinder of constant magnetic field, within which the modular parameter rises to its asymptotic value $\langle\rho\rangle$, surrounded by a much thicker outer cylinder of superconducting material within which the magnetic field falls to zero. Vortex solutions exist for both types of superconductor and for any magnetic field, but as we shall now see, it is only in type II superconductors and in a finite range of magnetic fields that vortex lines are energetically favored.

Because each vortex line has a cross-sectional area of order $\pi\xi^2$ within which the material is in its normal state or nearly so, the extra energy per volume required to create these vortex lines is of order $\mathcal{N}\pi\xi^2\Delta$, where $\mathcal{N}$ is the number of vortex lines per area. This vortex density is limited by the condition that $\mathcal{N} < 1/\pi\xi^2$, since otherwise the cylinders of normal metal would overlap, and the material would be considered to be in its normal state. The magnetic field must be expelled from a fraction $1 - \mathcal{N}\pi\lambda^2$ of the material if $\mathcal{N} < 1/\pi\lambda^2$, and from the whole material if $\mathcal{N} > 1/\pi\lambda^2$, so the energy per volume of the vortex state, relative to the superconducting state in the absence of magnetic fields, is

$$W_V \approx \mathcal{N}\pi\xi^2\Delta + \tfrac{1}{2}B^2 \times \begin{cases} 1 - \mathcal{N}\pi\lambda^2 & \mathcal{N} \leq 1/\pi\lambda^2 \\ 0 & \mathcal{N} \geq 1/\pi\lambda^2 \end{cases} . \quad (21.6.29)$$

(Numerical factors like $\frac{1}{2}$ and π are kept here to remind the reader of the origin of these expressions, but should not be taken literally.) For comparison, the energy per volume of the normal metal exceeds that in the superconducting state by $W_N = +\Delta$, and the energy per volume required to expel all magnetic fields from a superconductor is $W_S = B^2/2$. We can decide which state is present at a given magnetic field by checking which of W_S, W_N, or W_V is smallest.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

In type I superconductors we must distinguish between magnetic fields less or greater than the critical field $B_c \equiv \sqrt{2\Delta}$. Recalling that $\mathcal{N} < 1/\pi\xi^2$, and here $\xi > \lambda$, we also have $\mathcal{N} < 1/\pi\lambda^2$. Hence for $B < B_c$ Eq. (21.6.29) gives $W_V > \frac{1}{2}B^2 + \mathcal{N}\pi(\xi^2 - \lambda^2)\Delta > W_S$, so there can be no vortex lines. Also, for such fields $W_N > W_S$, so the material is superconducting. On the other hand, for $B > B_c$ Eq. (21.6.29) implies that $W_V > \Delta[1 + \mathcal{N}\pi(\xi^2 - \lambda^2)] > W_N$, so here again there are no vortex lines. Also for such fields $W_S > W_N$, so the material is in its normal state.

In type II superconductors we need to distinguish between magnetic fields in three ranges: $B < B_{c1}$, $B_{c1} < B < B_{c2}$, and $B > B_{c2}$, where B_{c1} and B_{c2} are a pair of critical fields, of order

$$B_{c1} \approx \sqrt{2\Delta}\,\xi/\lambda\,, \qquad B_{c2} \approx \sqrt{2\Delta}\,\lambda/\xi\,.$$

As we have seen, for type II superconductors the only stable vortex lines are those with the minimum flux[‡] π/e, so in a magnetic field B we should put the number of vortices per area $\mathcal{N}$ in Eq. (21.6.9) equal to eB/π. For $B < B_{c1}$ we can use Eq. (21.6.8) to show that $eB/\pi < 1/\pi\lambda^2$. The coefficient of the vortex density in Eq. (21.6.29) is therefore the positive quantity $\pi\xi^2\Delta - \frac{1}{2}B^2\lambda^2$, so here $W_V > W_S$, and there are no vortex lines for $B < B_{c1}$. Also, for such fields $W_N > W_S$, so the material is entirely superconducting. For $B > B_{c1}$, Eq. (21.6.28) shows that the vortex density $\mathcal{N} = eB/\pi$ is greater than $1/\pi\lambda^2$, so in the vortex state the magnetic field completely penetrates the superconductor, and the energy per volume is given by Eqs. (21.6.29) and (21.6.28) as

$$W_V \approx \mathcal{N}\pi\xi^2\Delta = eB\xi^2\Delta \approx (B/B_{c2})W_N \approx (B_{c1}/B)W_S\,.$$

Hence for $B_{c1} < B < B_{c2}$ we have $W_V < W_N$ and $W_V < W_S$, so the material is in its vortex state. For $B > B_{c2}$ we still have $W_V < W_S$ but now $W_N < W_V$, so the vortices disappear and all superconductivity is extinguished. The ability of type II superconductors with $\lambda \gg \xi$ to carry magnetic fields much higher than the critical value $B_c \approx \sqrt{\Delta}$ deduced

[‡] The derivation of flux quantization given above for an isolated vortex line is not fully applicable here. As we shall see, the separation of the vortex lines for $B > B_{c1}$ is less than the penetration depth λ, so it is not possible to find a contour $\mathscr{C}$ on which $\mathbf{A} - \nabla\phi = 0$ by simply drawing a circle around the vortex line at a distance from the vortex much greater than λ. Instead, it is necessary to appeal to considerations of continuity. (M. Tinkham, private communication.) Suppose we draw an arbitrary continuous curve between the centers of any two vortex lines. As shown below, on this curve the vector $\mathbf{A} - \nabla\phi$ will be very large close to either vortex, but pointing to opposite sides of the curve. Thus there is at least one point on each such curve where $\mathbf{A} - \nabla\phi = 0$. Because $\mathbf{A} - \nabla\phi$ is gauge-invariant, it must be continuous, so there is a closed contour $\mathscr{C}$ around each vortex line where $\mathbf{A} - \nabla\phi = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

earlier is important in technological applications of superconductivity, including magnets in high energy accelerators.

The Ginzburg–Landau theory holds only where the material is near the transition between its normal and superconducting states, so let us apply it near the center of a vortex line, where ρ drops to zero. Close to the center of a vortex line we can ignore its curvature, and assume cylindrical symmetry. The field $\mathbf{A}-\nabla\varphi$ is taken to have only an azimuthal component:

$$\left(\mathbf{A}-\nabla\phi\right)_\theta = A(r)\,, \tag{21.6.30}$$

so that the magnetic field has only an axial component:

$$B_z = \left(\nabla\times(\mathbf{A}-\nabla\phi)\right)_z = A'(r)+A(r)/r\,, \tag{21.6.31}$$

while ρ is a function only of r. The structure of the vortex line is thus governed by the pair of coupled differential equations:

$$A''(r)+r^{-1}A'(r)-r^{-2}A(r) = \frac{\rho^2(r)A(r)}{\lambda^2\langle\rho\rangle^2}\,, \tag{21.6.32}$$

$$\rho''(r)+r^{-1}\rho'(r)+\frac{1}{2\xi^2}\left(\rho(r)-\frac{\rho^3(r)}{\langle\rho\rangle^2}\right) = 4e^2\rho(r)A^2(r)\,. \tag{21.6.33}$$

When r is small compared with both the correlation length ξ and the penetration depth λ we can neglect the terms in Eqs. (21.6.32) and (21.6.33) proportional to $1/\xi^2$ and $1/\lambda^2$, and find the simplified equations

$$A''+A'/r-A/r^2 = 0\,, \tag{21.6.34}$$

$$\rho''+\rho'/r = 4e^2A^2\rho\,. \tag{21.6.35}$$

Eq. (21.6.34) has the general solution

$$A(r) = \frac{Br}{2}+\frac{C}{2er}\,, \tag{21.6.36}$$

where B is a constant that according to Eq. (21.6.31) is the magnetic field along the vortex line, and C is a real constant that is so far arbitrary. Using this in Eq. (21.6.35) shows that for $r\to 0$, the solution for $\rho(r)$ is a linear combination of $r^{|C|}$ and $r^{-|C|}$. The order parameter $\rho\exp(2ie\phi)$ must be a smooth function of position, so we can conclude that $|C|$ is a positive integer ℓ, with $\rho\propto r^\ell$ and $\phi = \pm\ell\varphi/2e+$ constant for $r\to 0$. Note that this solution for ϕ is consistent with Eq. (21.6.36) with $C=\pm\ell$; the azimuthal component of $\nabla\phi$ approaches $\pm\ell/2er$, while analyticity requires the azimuthal component of $\mathbf{A}$ to vanish as $r\to 0$. By a non-singular gauge transformation we can arrange that $\phi = \pm\ell\varphi/2e$ everywhere, so by the same reasoning as in Eq. (21.6.12) the magnetic flux carried by a single vortex line in a superconductor much larger than the penetration depth is $\pm\pi\ell/e$. We see not only that, as expected, the order parameter

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

vanishes at the center of the vortex line, but that it vanishes with a power of r equal to the magnitude of the magnetic flux in units of π/e.

This solution for ϕ obeys the 'quantization' condition that $\phi(2\pi) - \phi(0)$ is an integral multiple of π/e, which implies the quantization of magnetic flux. In the theory of Goldstone boson and electromagnetic fields based on Eq. (21.6.5) this condition on ϕ has to be imposed by hand on the solution of the field equations, while the Ginzburg–Landau equations 'know' about this condition, because these equations are based on an appropriate choice of order parameter.

* * *

Although the most dramatic properties of superconductors can be derived directly from the assumption that electromagnetic gauge invariance is spontaneously broken, a microscopic theory of superconductivity is needed to understand how and when this occurs. The derivation of the microscopic superconductivity theory of Bardeen, Cooper, and Schrieffer[37] has been recast[44,44a] in the language of effective field theory, using power-counting methods similar to those that we have used here in Sections 19.5 and 21.4. For this purpose, suppose we integrate out the degrees of freedom associated with the ions in a superconductor,[‡‡] leaving only an effective interaction among electrons. For simplicity, we shall work at zero temperature, and at first assume no external field, so that the Lagrangian is invariant under translations and the time reversal operation T. We shall also assume spin-independent forces, so that the Lagrangian is invariant under $SU(2)$ transformations that act on spin indices alone, but we shall not need to assume invariance under rotations acting on momenta. Electrons are then characterized by a momentum $\mathbf{p}$ and spin index $s = \pm \frac{1}{2}$, and are described by annihilation and creation operators $a(\mathbf{p}, s, t)$ and $a^\dagger(\mathbf{p}, s, t)$, with a Lagrangian of form

$$\begin{aligned} L = &- \sum_s \int d^3p\, a^\dagger(\mathbf{p}, s, t) \left[-i\frac{\partial}{\partial t} + E(\mathbf{p}) \right] a(\mathbf{p}, s, t) \\ &+ \sum_{s_1 s_2 s_3 s_4} \int d^3p_1\, d^3p_2\, d^3p_3\, d^3p_4\, V_{s_1 s_2 s_3 s_4}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) \\ &\quad \times a^\dagger(\mathbf{p}_1, s_1, t)\, a^\dagger(\mathbf{p}_2, s_2, t)\, a(\mathbf{p}_3, s_3, t)\, a(\mathbf{p}_4, s_4, t) \delta^3(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_3 - \mathbf{p}_4) \\ &+ \cdots , \end{aligned} \tag{21.6.37}$$

[‡‡] Strictly speaking, it is not possible to integrate out all degrees of freedom except electrons, because the phonon is a Goldstone boson whose frequency vanishes for very large wavelengths, like a massless particle in relativistic theories. However those effects of phonon interactions that cannot be represented as effective electron–electron interactions are suppressed by inverse factors of the ion masses.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

where '$\cdots$' represents terms with six or more creation and annihilation operators, and $E(\mathbf{p})$ is the electron energy minus the chemical potential. For free electrons, $E(\mathbf{p}) = \mathbf{p}^2/2m_e - E_F$ where E_F is the energy of electrons at the Fermi surface. Interactions will inevitably change this function, but it is natural to expect that since $E(\mathbf{p})$ vanishes for free electrons at momenta $\mathbf{p}$ on the sphere $|\mathbf{p}| = \sqrt{2m_e E_F}$, in the presence of interactions it will still vanish on some closed Fermi surface $\mathscr{S}$:

$$E(\mathbf{p}) = 0 \text{ for } \mathbf{p} \text{ on } \mathscr{S} . \tag{21.6.38}$$

For reasons that will become apparent, we shall integrate out all electrons except those within a thin shell of thickness κ around the Fermi surface.[‡] (Later we will be able to remove the cut-off κ by introducing a renormalized electron–electron potential.) The remaining degrees of freedom are then electrons with momenta of the form

$$\mathbf{p} = \mathbf{k} + \hat{\mathbf{n}}(\mathbf{k})\ell , \tag{21.6.39}$$

where $\mathbf{k}$ is on the Fermi surface $\mathscr{S}$, $\hat{\mathbf{n}}(\mathbf{k})$ is the unit vector normal to the surface at $\mathbf{k}$, and $0 \leq \ell \leq \kappa$. For such momenta,

$$E(\mathbf{p}) = v_F(\mathbf{k})\ell , \tag{21.6.40}$$

where

$$v_F(\mathbf{k}) \equiv \hat{\mathbf{n}}(\mathbf{k}) \cdot \Big(\nabla_{\mathbf{p}} E(\mathbf{p})\Big)_{\mathbf{p}=\mathbf{k}} . \tag{21.6.41}$$

The electron propagator in wave number–frequency space is then

$$\frac{1}{\omega - v_F(\mathbf{k})\ell + i\epsilon} . \tag{21.6.42}$$

Now let us consider how a general connected matrix element scales with κ as $\kappa \to 0$. We have an integral over frequencies for every loop, and a propagator for every internal line, so the integral of the product of propagators over all frequencies will have an ℓ dependence ℓ^{L-I}, where L is the number of loops and I is the number of internal lines. To count the number of ℓ integrals, it is important to note that for generic momenta the delta-function in the interaction term in Eq. (21.6.37) constrains the $\mathbf{k}$s, not the ℓs. (For instance, if momentum conservation constrains the momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ of two electron lines to have total momentum $\mathbf{P} \neq 0$, then the integral over $\mathbf{p}_1$ runs over the intersection of two closed shells of thickness κ, one centered on $\mathbf{P}$ and the other on zero. The intersection of these shells is a closed ring of thickness κ, so we have to integrate

[‡] For forces that are not spin-independent, there are two Fermi surfaces, one for each eigenvalue of the matrix $E_{s's}(\mathbf{p})$, and we integrate over all electrons in each spin eigenstate except those within a thin shell around the corresponding Fermi surface.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

over one **k**-component which gives position around the ring, and two ℓs which give position within the cross section of the ring.) Thus there are I integrals over ℓs, and since the integrand goes as ℓ^{L-I}, the matrix element will vary as

$$M \propto \kappa^L \,. \tag{21.6.43}$$

The number of loops is related to the number of internal lines and the numbers V_i of vertices of type i by the familiar relation

$$L = I - \sum_i V_i + 1 \,. \tag{21.6.44}$$

Also, the number of internal lines is related to the numbers of vertices and the number E of external lines by another familiar relation

$$2I + E = \sum_i n_i V_i \,, \tag{21.6.45}$$

where n_i is the number of electron operators in the interaction of type i. Eliminating I from Eqs. (21.6.44) and (21.6.45) gives

$$L = 1 - \tfrac{1}{2}E + \tfrac{1}{2}\sum_i V_i(n_i - 2) \,. \tag{21.6.46}$$

Terms in the action with $n_i = 2$ just serve to change the function $E(\mathbf{p})$, and hence to shift the Fermi surface. True interactions have $n_i > 2$, and so it appears from Eqs. (21.6.43) and (21.6.46) that they yield terms in the matrix element that make relatively negligible contributions for $\kappa \to 0$. In the language of Section 18.5, this would mean that all interactions are irrelevant operators. This is why electrons near the Fermi surface in normal metals behave pretty much like free particles.

There is, however, an exception to this conclusion. If a pair of electron lines disappears into the vacuum, then translation invariance requires the momenta of these two lines to be equal in magnitude and opposite in direction, so if one momentum is near the Fermi surface, then the other is also. (Time reversal invariance requires that $E(\mathbf{p})$ is even in $\mathbf{p}$, so even though the Fermi surface is generally not spherical, it is still true that if $E(\mathbf{p}) = 0$ then $E(-\mathbf{p}) = 0$.) The integral over the momentum $\mathbf{p}$ of one of these lines is then over a shell of thickness κ, or in other words over two ks and *one* ℓ. For each interaction involving two such lines, we have one rather than two integrals over ℓs, so that instead of reducing the order of magnitude of the matrix element by a factor κ as indicated in Eqs. (21.6.43) and (21.6.46), such an interaction has no effect on the order of magnitude of the matrix element. In other words, interactions involving four electron operators become marginal rather than irrelevant if they act between electron lines that eventually disappear into the vacuum.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

To see the consequences of this exception, it is convenient to use a trick, known as the *Hubbard–Stratonovich transformation*,[45] that was mentioned briefly in Section 10.7. We are now going to include slowly varying external electromagnetic fields, so it will be convenient to work in coordinate space. Gauge invariance tells us that the Lagrangian (21.6.37) then becomes

$$L = -\sum_s \int d^3x\, \psi_s^\dagger(\mathbf{x},t)\left[-i\frac{\partial}{\partial t} - A_0(\mathbf{x},t) + E\Big(-i\nabla + \mathbf{A}(\mathbf{x},t)\Big)\right]\psi_s(\mathbf{x},t)$$
$$+ \sum_{s_1\,s_2\,s_3\,s_4} \int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4\, V_{s_1\,s_2\,s_3\,s_4}(\mathbf{x}_1,\,\mathbf{x}_2,\,\mathbf{x}_3,\,\mathbf{x}_4)$$
$$\times\, \psi_{s_1}^\dagger(\mathbf{x}_1,t)\,\psi_{s_2}^\dagger(\mathbf{x}_2,t)\,\psi_{s_3}(\mathbf{x}_3,t)\,\psi_{s_4}(\mathbf{x}_4,t)\,, \tag{21.6.47}$$

where

$$\psi_s(\mathbf{x},t) \equiv (2\pi)^{-3/2}\int d^3p\ \exp(i\mathbf{p}\cdot\mathbf{x})\,a(\mathbf{p},s,t)\,. \tag{21.6.48}$$

We are now dropping terms with more than four electron operators, because they are all irrelevant. To this Lagrangian we add the term

$$\Delta L = \frac{1}{4}\int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4 \sum_{s_1 s_2 s_2 s_4} V_{s_1 s_2 s_2 s_4}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4)$$
$$\times\left[\Psi_{s_2 s_1}^\dagger(\mathbf{x}_2,\,\mathbf{x}_1,\,t) - \psi_{s_1}^\dagger(\mathbf{x}_1,\,t)\,\psi_{s_2}^\dagger(\mathbf{x}_2,\,t)\right]$$
$$\times\left[\Psi_{s_3 s_4}(\mathbf{x}_3,\,\mathbf{x}_4,\,t) - \psi_{s_3}(\mathbf{x}_3,\,t)\,\psi_{s_4}(\mathbf{x}_4,\,t)\right] \tag{21.6.49}$$

and do path integrals over the new 'pair' field $\Psi(\mathbf{x}, s, \mathbf{x}', s', t)$ as well as the electron field ψ. This is allowed because ΔL is quadratic in the pair field, with a field-independent coefficient of the second-order term, so, according to the appendix of Chapter 9, the integration over $\Psi_{ss'}(\mathbf{x}, \mathbf{x}', t)$ in path integrals just has the effect of setting $\Psi_{ss'}(\mathbf{x}, \mathbf{x}', t)$ equal to the stationary point $\Psi_{ss'}(\mathbf{x}, \mathbf{x}', t) = \psi_s(\mathbf{x}, t)\,\psi_{s'}(\mathbf{x}', t)$ of the Lagrangian, at which $\Delta L = 0$. This additional term is chosen so that the terms in $L + \Delta L$ that are quartic in electron fields cancel, leaving only terms quadratic in these fields:

$$L + \Delta L$$
$$= -\sum_s \int d^3x\, \psi_s^\dagger(\mathbf{x},t)\left\{-i\frac{\partial}{\partial t} - A_0(\mathbf{x},t) + E\Big(-i\nabla + \mathbf{A}(\mathbf{x},t)\Big)\right\}\psi_s(\mathbf{x},t)$$
$$-\frac{1}{4}\sum_{s_1\,s_2\,s_3\,s_4}\int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4\, V_{s_1\,s_2\,s_3\,s_4}(\mathbf{x}_1,\,\mathbf{x}_2,\,\mathbf{x}_3,\,\mathbf{x}_4)$$
$$\times\left[\psi_{s_1}^\dagger(\mathbf{x}_1,t)\,\psi_{s_2}^\dagger(\mathbf{x}_2,t)\,\Psi_{s_3 s_4}(\mathbf{x}_3,\,\mathbf{x}_4,\,t) + \Psi_{s_2 s_1}^\dagger(\mathbf{x}_2,\,\mathbf{x}_1,\,t)\psi_{s_3}(\mathbf{x}_3,t)\,\psi_{s_4}(\mathbf{x}_4,t)\right.$$
$$\left. - \Psi_{s_2 s_1}^\dagger(\mathbf{x}_2,\,\mathbf{x}_1,\,t)\,\Psi_{s_3 s_4}(\mathbf{x}_3,\,\mathbf{x}_4,\,t)\right]. \tag{21.6.50}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Now to the point. We have shown that the interaction V is irrelevant except where it acts on a pair of electron lines that disappear into the vacuum. When we calculate the quantum effective action $\Gamma[\Psi]$ in the presence of an external, slowly varying pair field $\Psi_{ss'}(\mathbf{x}, \mathbf{x}', t)$, this means that we drop all diagrams except those that become disconnected when we slice through any internal Ψ line. But the general definition of the quantum effective action requires that we drop all diagrams that *do* become disconnected when we slice through any internal line. *Therefore we must not include any internal* Ψ *lines at all.* Because the electron field enters quadratically in Eq. (21.6.50), the only graphs that survive when we integrate out the electron field are a one-vertex graph arising from the last term in the square brackets in Eq. (21.6.50), plus a one-loop graph given, by the same reasoning as in Section 16.2, by the logarithm of the determinant of the coefficient of the terms quadratic in electron fields:

$$\begin{aligned}\Gamma[\Psi] = \frac{1}{4}\sum_{s_1 s_2 s_3 s_4}\int dt \int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4\, V_{s_1 s_2 s_3 s_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \\ \times \Psi^\dagger_{s_2 s_1}(\mathbf{x}_2, \mathbf{x}_1, t)\, \Psi_{s_3 s_4}(\mathbf{x}_3, \mathbf{x}_4, t) \\ -\frac{i}{2}\ln \mathrm{Det}\begin{pmatrix} A & B \\ B^\dagger & -A^T \end{pmatrix} + \text{constant}\,. \end{aligned} \tag{21.6.51}$$

where A and B are the 'matrices':

$$A_{\mathbf{x}'s't',\mathbf{x}st} \equiv \delta_{s's}\left\{-i\frac{\partial}{\partial t} - A_0(\mathbf{x}, t) + E\left(-i\nabla + \mathbf{A}(\mathbf{x}, t)\right)\right\}\delta^3(\mathbf{x}' - \mathbf{x})\delta(t' - t)\,, \tag{21.6.52}$$

$$B_{\mathbf{x}'s't',\mathbf{x}st} \equiv -\Delta_{s's}(\mathbf{x}', \mathbf{x}, t)\delta(t' - t)\,, \tag{21.6.53}$$

and Δ is the *gap function*:

$$\Delta_{s's}(\mathbf{x}', \mathbf{x}, t) \equiv \frac{1}{2}\sum_{\sigma'\sigma}\int d^3y\, d^3y'\; V_{s's\sigma'\sigma}(\mathbf{x}', \mathbf{x}, \mathbf{y}', \mathbf{y})\, \Psi_{s's}(\mathbf{x}', \mathbf{x}, t)\,. \tag{21.6.54}$$

(Here and below, we ignore an additive constant in Γ arising from the modes that have been integrated out.) This is one problem in which the effective action can be calculated without any assumption about the smallness of the interaction in the Lagrangian.

In order to simplify our further discussion, let us now specialize to the case of a spin-singlet pair field

$$\Psi_{+-}(\mathbf{x}', \mathbf{x}, t) = -\Psi_{-+}(\mathbf{x}', \mathbf{x}, t) \equiv \Psi(\mathbf{x}', \mathbf{x}, t)\,, \tag{21.6.55}$$

where subscripts $+$ and $-$ stand for spin indices $+1/2$ and $-1/2$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

respectively.[¤¤] Using invariance under rotations of spin indices, the components of the potential needed here are then

$$V_{+-+-}(\mathbf{x}', \mathbf{x}, t) - V_{+--+}(\mathbf{x}', \mathbf{x}, t) = -V_{-++-}(\mathbf{x}', \mathbf{x}, t) + V_{-+-+}(\mathbf{x}', \mathbf{x}, t) \equiv 2V(\mathbf{x}', \mathbf{x}, t)\,. \tag{21.6.56}$$

Then Eq. (21.6.54) gives

$$\Delta_{+-}(\mathbf{x}', \mathbf{x}, t) = -\Delta_{-+}(\mathbf{x}', \mathbf{x}, t) \equiv \Delta(\mathbf{x}', \mathbf{x}, t)\,, \tag{21.6.57}$$

$$\Delta_{++}(\mathbf{x}', \mathbf{x}, t) = \Delta_{--}(\mathbf{x}', \mathbf{x}, t) = 0\,. \tag{21.6.58}$$

The quantum effective action (21.6.51) is now

$$\Gamma[\Psi] = \int dt \int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4\, V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \times \Psi^\dagger(\mathbf{x}_2, \mathbf{x}_1, t)\, \Psi(\mathbf{x}_3, \mathbf{x}_4, t) - i \ln \operatorname{Det} \begin{pmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{B}^\dagger & -\mathscr{A}^T \end{pmatrix}, \tag{21.6.59}$$

where

$$\mathscr{A}_{\mathbf{x}'t',\mathbf{x}t} \equiv \left\{ -i\frac{\partial}{\partial t} - A_0(\mathbf{x}, t) + E\Big(-i\nabla + \mathbf{A}(\mathbf{x}, t)\Big)\right\}\delta^3(\mathbf{x}' - \mathbf{x})\delta(t' - t)\,, \tag{21.6.60}$$

$$\mathscr{B}_{\mathbf{x}'t',\mathbf{x}t} \equiv -\Delta(\mathbf{x}', \mathbf{x}, t)\,\delta(t' - t)\,, \tag{21.6.61}$$

and

$$\Delta(\mathbf{x}', \mathbf{x}, t) \equiv \int d^3y\, d^3y'\; V(\mathbf{x}', \mathbf{x}, \mathbf{y}', \mathbf{y})\, \Psi(\mathbf{y}', \mathbf{y}, t)\,. \tag{21.6.62}$$

Let's first use these results to consider the translationally invariant case, with no external electromagnetic fields. Then the pair and gap fields can be expressed as Fourier transforms

$$\Psi(\mathbf{x}', \mathbf{x}) = \int d^3p\; e^{i\mathbf{p}\cdot(\mathbf{x}' - \mathbf{x})}\Psi(\mathbf{p})\,, \tag{21.6.63}$$

$$\Delta(\mathbf{x}', \mathbf{x}) = (2\pi)^{-3} \int d^3p\; e^{i\mathbf{p}\cdot(\mathbf{x}' - \mathbf{x})}\Delta(\mathbf{p})\,, \tag{21.6.64}$$

and the electron–electron potential appears here in the form

$$\int d^3x_1\, d^3x_2\, d^3x_3\, d^3x_4\; e^{i\mathbf{p}'\cdot(\mathbf{x}_1 - \mathbf{x}_2)}\, e^{i\mathbf{p}\cdot(\mathbf{x}_3 - \mathbf{x}_4)}\, V(x_1, x_2, x_3, x_4) \equiv \mathscr{V}_4\, V(\mathbf{p}', \mathbf{p})\,, \tag{21.6.65}$$

[¤¤] In liquid He^3 the non-vanishing components of the pair field form a spin triplet, with components $\Psi_1 = \Psi_{++}$, $\Psi_0 = \sqrt{2}\Psi_{+-} = \sqrt{2}\Psi_{-+}$, $\Psi_{-1} = \Psi_{--}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

where $\mathscr{V}_4$ is the spacetime volume

$$\mathscr{V}_4 \equiv \int d^3x \int dt\ 1\ . \tag{21.6.66}$$

By going over to wave number–frequency space, the 'matrices' (21.6.60) and (21.6.61) become diagonal, and the calculation of the determinant becomes trivial. The effective potential $V[\Psi]$ (not to be confused with the electron–electron potential) was defined in Section 16.1 as minus the effective action per spacetime volume:

$$\begin{aligned} V[\Psi] &\equiv -\Gamma[\Psi]/\mathscr{V}_4 \\ &= -\int d^3p\, d^3p'\ \Psi^*(\mathbf{p}')V(\mathbf{p}',\mathbf{p})\Psi(\mathbf{p}) \\ &\quad + \frac{i}{(2\pi)^4}\int d\omega\ d^3p\ \ln\left(1 - \frac{|\Delta(\mathbf{p})|^2}{\omega^2 - E^2(\mathbf{p}) + i\epsilon}\right) \end{aligned} \tag{21.6.67}$$

with

$$\Delta(\mathbf{p}) = -\int d^3p'\ V(\mathbf{p},\mathbf{p}')\,\Psi(\mathbf{p}')\ . \tag{21.6.68}$$

(The $i\epsilon$ term can be inferred from the Feynman rules for the electron loop graphs, and as shown in Section 9.2, it ultimately arises from the conditions on the electron field at $t \to \pm\infty$.) Wick rotating, integrating over ω, and expressing Ψ in terms of Δ, this becomes

$$\begin{aligned} V[\Delta] = &-\int d^3p\, d^3p'\ \Delta^*(\mathbf{p}')V^{-1}(\mathbf{p}',\mathbf{p})\Delta(\mathbf{p}) \\ &-\frac{1}{(2\pi)^3}\int d^3p\left[\sqrt{E^2(\mathbf{p}) + |\Delta(\mathbf{p})|^2} - E(\mathbf{p})\right]\ . \end{aligned} \tag{21.6.69}$$

As we saw in Section 16.1, the field equations are just the condition that the effective action is stationary, which in our case yields the famous *gap equation* for the equilibrium gap function $\Delta_0(\mathbf{p})$:

$$\begin{aligned} 0 &= \left.\frac{\delta V[\Delta]}{\delta\Delta^*(\mathbf{p})}\right|_{\Delta=\Delta_0} \\ &= -\int d^3p'\ V^{-1}(\mathbf{p},\mathbf{p}')\Delta_0(\mathbf{p}') - \frac{1}{2(2\pi)^3}\frac{\Delta_0(\mathbf{p})}{\sqrt{E^2(\mathbf{p}) + |\Delta_0(\mathbf{p})|^2}} \end{aligned}$$

or, in a more familiar form,

$$\Delta_0(\mathbf{p}) = -\frac{1}{2(2\pi)^3}\int d^3p'\frac{V(\mathbf{p},\mathbf{p}')\,\Delta_0(\mathbf{p}')}{\sqrt{E^2(\mathbf{p}') + |\Delta_0(\mathbf{p}')|^2}}\ . \tag{21.6.70}$$

All along, all integrals over momenta have been implicitly understood to be limited to momenta of the form (21.6.39) within a thin shell of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

thickness κ around the Fermi surface. The effective potential (21.6.69) is thus

$$V[\Delta] = -\kappa^2 \int_{\mathscr{S}} d^2k\, d^2k'\, \Delta^*(\mathbf{k}')V^{-1}(\mathbf{k}',\mathbf{k})\Delta(\mathbf{k}) - \frac{1}{(2\pi)^3}\int_0^\kappa \int_{\mathscr{S}} d^2k \left[\sqrt{\ell^2 v_F^2(\mathbf{k}) + |\Delta(\mathbf{k})|^2} - \ell\, v_F(\mathbf{k})\right]. \quad (21.6.71)$$

The effective potential is now to be understood as a functional only of the gap function Δ on the Fermi surface.

Since κ is arbitrary, the potential $V(\mathbf{k}',\mathbf{k})$ must be given a κ-dependence such that $V[\Delta]$ is κ-independent. Most applications[44] of renormalization group methods to superconductivity have followed the Wilson approach outlined in Section 12.4, deriving a differential equation for the dependence of $V(\mathbf{k}',\mathbf{k})$ on κ, and studying the behavior of its solutions for $\kappa \to 0$. For the sake of flexibility it is useful to note that one can just as well adopt the approach of Gell-Mann and Low, and introduce a renormalized electron–electron potential[44a]

$$V_\mu^{-1}(\mathbf{k}',\mathbf{k}) \equiv -\left.\frac{\delta^2 V[\Delta]}{\delta\Delta^*(\mathbf{k}')\delta\Delta(\mathbf{k})}\right|_{\Delta(\mathbf{k})=\Delta(\mathbf{k})^*=\mu}, \quad (21.6.72)$$

where μ is a sliding renormalization scale, like that introduced in Section 18.2. By expressing the original electron–electron potential in terms of V_μ, Eq. (21.6.71) becomes

$$V[\Delta] = -\int_{\mathscr{S}} d^2k\, d^2k'\, \Delta^*(\mathbf{k}')V_\mu^{-1}(\mathbf{k}',\mathbf{k})\Delta(\mathbf{k}) - \frac{1}{(2\pi)^3}\int_0^\kappa d\ell \int_{\mathscr{S}} d^2k \left[\sqrt{\ell^2 v_F^2(\mathbf{k}) + |\Delta(\mathbf{k})|^2} - \ell v_F(\mathbf{k}) - \frac{|\Delta(\mathbf{k})|^2}{2(\ell^2 v_F^2(\mathbf{k}) + \mu^2)^{1/2}} + \frac{\mu^2|\Delta(\mathbf{k})|^2}{4(\ell^2 v_F^2(\mathbf{k}) + \mu^2)^{3/2}}\right]. \quad (21.6.73)$$

The integral over ℓ now converges if we take the cutoff κ to infinity, and yields

$$V[\Delta] = -\int_{\mathscr{S}} d^2k\, d^2k'\, \Delta^*(\mathbf{k}')V_\mu^{-1}(\mathbf{k}',\mathbf{k})\Delta(\mathbf{k}) + \frac{1}{2(2\pi)^3}\int_{\mathscr{S}} d^2k\, \frac{|\Delta(\mathbf{k})|^2}{v_F(\mathbf{k})}\left[\ln\left(\frac{|\Delta(\mathbf{k})|}{\mu}\right) - 1\right]. \quad (21.6.74)$$

The condition that $V[\Delta]$ be stationary at $\Delta = \Delta_0$ yields the gap equation in the more useful form

$$\Delta_0(\mathbf{k}) = \frac{1}{2(2\pi)^3}\int_{\mathscr{S}} d^2k'\, V_\mu(\mathbf{k},\mathbf{k}')v_F^{-1}(\mathbf{k}')\Delta_0(\mathbf{k}')\ln\left(\frac{|\Delta_0(\mathbf{k}')|}{\mu}\right). \quad (21.6.75)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Using this in Eq. (21.6.74) shows that the energy density of the superconducting state is less than that of the normal state by an amount

$$\Delta \equiv V(0) - V(\Delta_0) = \int_{\mathscr{S}} d^2k \, \frac{|\Delta_0(\mathbf{k})|^2}{2(2\pi)^3 v_F(\mathbf{k})} \,. \tag{21.6.76}$$

Of course the μ dependence of the electron–electron potential V_μ must be chosen to satisfy a renormalization group equation that insures that the effective potential (21.6.74) is independent of the arbitrary renormalization scale μ:

$$\mu \frac{d}{d\mu} V_\mu^{-1}(\mathbf{k}', \mathbf{k}) = -\frac{\delta^2(\mathbf{k}' - \mathbf{k})}{2(2\pi)^3 v_F(\mathbf{k})}$$

or, equivalently,

$$\mu \frac{d}{d\mu} V_\mu(\mathbf{k}', \mathbf{k}) = \frac{1}{2(2\pi)^3} \int d^2k'' \; V_\mu(\mathbf{k}', \mathbf{k}'') \, v_F^{-1}(\mathbf{k}'') \, V_\mu(\mathbf{k}'', \mathbf{k}) \,. \tag{21.6.77}$$

This can be usefully rewritten in terms of the Hermitian kernel

$$K_\mu(\mathbf{k}', \mathbf{k}) \equiv \frac{1}{2(2\pi)^3} \, v_F^{-1/2}(\mathbf{k}') \, v_F^{-1/2}(\mathbf{k}) \, V_\mu(\mathbf{k}', \mathbf{k}) \,, \tag{21.6.78}$$

as

$$\mu \frac{d}{d\mu} K_\mu = K_\mu^2 \,. \tag{21.6.79}$$

The eigenvectors $u_n(\mathbf{k})$ of $K_\mu(\mathbf{k}, \mathbf{k}')$ are then independent of μ, while the eigenvalues take the form $1/\ln(\Lambda_n/\mu)$, where the Λ_n are integration constants, like the Λ of quantum chromodynamics. The potential therefore takes the form

$$V_\mu(\mathbf{k}', \mathbf{k}) = 2(2\pi)^3 \, v_F^{1/2}(\mathbf{k}) \, v_F^{1/2}(\mathbf{k}') \sum_n \frac{u_n(\mathbf{k}) \, u_n^*(\mathbf{k}')}{\ln(\Lambda_n/\mu)} \,, \tag{21.6.80}$$

where we have chosen the eigenvectors to be orthonormal, so that the completeness relation takes the form

$$\sum_n u_n(\mathbf{k}) u_n^*(\mathbf{k}') = \delta^2(\mathbf{k} - \mathbf{k}') \,. \tag{21.6.81}$$

The effective potential (21.6.74) may then be written as

$$V[\Delta] = \frac{1}{2(2\pi)^3} \int d^2k \, d^2k' \sum_n \frac{\Delta^*(\mathbf{k}') \, u_n^*(\mathbf{k}') \, \Delta(\mathbf{k}) \, u_n(\mathbf{k})}{\sqrt{v_F(\mathbf{k}') \, v_F(\mathbf{k})}} \left[\ln\left(\frac{|\Delta(\mathbf{k})|}{\Lambda_n} \right) - 1 \right] \,. \tag{21.6.82}$$

For a material with local rotational invariance the Fermi surface is a sphere and the eigenvectors $u_n(\mathbf{k})$ are spherical harmonic functions of the coordinates on this sphere, but Eq. (21.6.82) holds without any assumption of rotational invariance.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

We can now use this formalism to get an idea of when superconductivity occurs. The logarithm in Eq. (21.6.82) is large and negative for very small Δ, and large and positive for very large Δ, so as the overall scale of Δ increases from zero to infinity, $V[\Delta]$ drops from zero to negative values and then rises to infinity. It therefore always has a minimum at a non-vanishing value of Δ for any electron–electron potential. But this result is subject to an important qualification. When we take the cut-off κ to infinity, the integral in Eq. (21.6.73) is effectively cut off at ℓ of order $|\Delta|/v_F$, while the Fermi surface has a radius of order k_F, where $8\pi k_F^3/3(2\pi)^3$ equals the electron number density. Since we have been assuming that the electrons taken into account here are in a *thin* shell around the Fermi surface, this derivation is only valid if $|\Delta| \ll \omega_D$, where ω_D is the Debye frequency $k_F v_F$. In particular, in the case of rotational invariance and a direction-independent gap function, the effective potential reaches a local minimum for a gap function of the order of $\Lambda_{s\text{ wave}}$, so the symmetry is spontaneously broken as long as $\Lambda_{s\text{ wave}} \ll \omega_D$. Another way of stating this result is that for s-wave superconductivity, the s-wave projection of the electron–electron potential (21.6.80) must be attractive if renormalized at a scale $\mu \approx \omega_D$. *But it does not matter how strong this renormalized attractive potential is.*

This feature of superconductivity, that a Goldstone boson forms for an attractive potential however weak the potential may be, is a consequence of the existence of a Fermi surface, which enhances long-range effects. In quantum field theories without elementary spinless fields like those of Section 21.5, we would not normally expect a spontaneous symmetry breakdown in empty space unless the interactions are sufficiently strong.

Now let us return to the case of an external electromagnetic field. As usual, we can introduce the Goldstone boson field $\phi(\mathbf{x}, t)$ by writing each charged field of the theory, which here is just the gap field $\Delta(\mathbf{x}, \mathbf{x}', t)$ or equivalently the pair field $\Psi(\mathbf{x}, \mathbf{x}', t)$, as a gauge transformation with gauge parameter $\phi(\mathbf{x}, t)$ acting on the corresponding gauge-invariant field, distinguished with a tilde:

$$\Psi(\mathbf{x}, \mathbf{x}', t) = \exp\Big(-i\phi(\mathbf{x}, t)\Big)\tilde{\Psi}(\mathbf{x}, \mathbf{x}', t)\exp\Big(-i\phi(\mathbf{x}', t)\Big)\,. \qquad (21.6.83)$$

The effective action is then given by using Eq. (21.6.83) in Eq. (21.6.59). By a gauge transformation, we can then remove the ϕ dependence in Eq. (21.6.83), provided we replace $A_\mu(x)$ in Eq. (21.6.60) with $A_\mu(x) - \partial_\mu\phi(x)$. When the material is not only superconducting but far from the transition between normal and superconducting states, we may also integrate out the gauge-invariant degrees of freedom associated with $\tilde{\Psi}(\mathbf{x}, \mathbf{x}', t)$, which simply means that we replace it with its equilibrium value $\Psi_0(\mathbf{x}, \mathbf{x}', t)$. The part of the effective action that depends on the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

Goldstone and external electromagnetic fields is then

$$\Gamma[\phi, A] = \Gamma_{\Delta=0}[A] - i \ln \operatorname{Det} \begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{B}^\dagger & -\mathscr{A}^T \end{bmatrix} + i \ln \operatorname{Det} \begin{bmatrix} \mathscr{A} & 0 \\ 0 & -\mathscr{A}^T \end{bmatrix}, \tag{21.6.84}$$

where now

$$\mathscr{A}_{\mathbf{x}'t',\mathbf{x}t} = \left[-i\frac{\partial}{\partial t} + eA_0(\mathbf{x},t) - e\dot{\phi}(\mathbf{x},t) + E\Big(-i\nabla + e\mathbf{A}(\mathbf{x},t) - e\nabla\phi(\mathbf{x},t) \Big) \right] \times \delta^3(\mathbf{x}' - \mathbf{x})\delta(t' - t)\,, \tag{21.6.85}$$

$$\mathscr{B}_{\mathbf{x}'t',\mathbf{x}t} = \Delta_0(\mathbf{x}' - \mathbf{x})\delta(t' - t)\,. \tag{21.6.86}$$

Quantitative properties of the superconductor such as the penetration depth can be read off[46] from the expansion of Eq. (21.6.84) in powers of $A_0(\mathbf{x},t) - \dot{\phi}(\mathbf{x},t)$ and $\mathbf{A}(\mathbf{x},t) - \nabla\phi(\mathbf{x},t)$.

Appendix General Unitarity Gauge

In this appendix we shall show that in general spontaneously broken gauge theories it is always possible to adopt a 'unitarity' gauge in which the Goldstone boson fields satisfy Eq. (21.4.30):

$$\sum_{ab} F_{ab}^2\, \xi_a e_{\alpha b} = 0\,. \tag{21.A.1}$$

Using the exponential parameterization for all groups, we note first that any element of G, in at least a finite neighborhood of the identity, may be put in the form

$$g = \exp\left(-i \sum_\alpha \theta_\alpha \mathscr{T}_\alpha \right) \exp\left(i \sum_a \phi_a x_a \right) \exp\left(i \sum_i \mu_i t_i \right)\,, \tag{21.A.2}$$

with ϕ_a subject to the linear constraint that for all α

$$\sum_{ab} F_{ab}^2\, \phi_a\, e_{\alpha b} = 0\,. \tag{21.A.3}$$

This is easy to see when g is infinitesimally close to the identity. Any such g may be written

$$g = 1 + i \sum_a \phi_a^0\, x_a + i \sum_i \mu_i^0\, t_i\,, \tag{21.A.4}$$

with ϕ_a^0 and μ_i^0 infinitesimal. Equivalently,

$$g = 1 + i \sum_i \phi_a x_a + i \sum_i \mu_i t_i - i \sum_\alpha \theta_\alpha \mathscr{T}_\alpha\,, \tag{21.A.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

where θ_α is an arbitrary infinitesimal, and

$$\phi_a(\theta) \equiv \phi_a^0 + \sum_\alpha \theta_\alpha e_{\alpha a} \,, \tag{21.A.6}$$

$$\mu_i(\theta) \equiv \mu_i^0 + \sum_\alpha \theta_\alpha e_{\alpha i} \,. \tag{21.A.7}$$

For any given ϕ_a^0, we can choose θ_α to minimize the positive quantity

$$\sum_{ab} F_{ab}^2 \phi_a(\theta)\phi_b(\theta) \,. \tag{21.A.8}$$

At this minimum, the quantity (21.A.8) is stationary with respect to variations of θ_α, so $\phi_a(\theta)$ satisfies Eq. (21.A.3). For ϕ, μ, and θ infinitesimal, Eq. (21.A.5) is the same as Eq. (21.A.2), so we see that the set of all gs of the form (21.A.2) (with ϕ_a satisfying Eq. (21.A.3)) includes all gs infinitesimally close to the identity. It follows then from continuity that the same is true of all gs in at least some finite neighborhood of the identity.

Next, consider the particular group element

$$g = \gamma(\xi) = \exp\left(i \sum_a \xi_a x_a\right) \tag{21.A.9}$$

and write it in the form (21.A.2):

$$\gamma(\xi) = \exp\left(-i \sum_\alpha \theta_\alpha(\xi) \mathscr{T}_\alpha\right) \gamma\left(\phi(\xi)\right) \exp\left(i \sum_i \mu_i(\xi) t_i\right) \,, \tag{21.A.10}$$

with $\phi_a(\xi)$ subject to Eq. (21.A.3). This just says that the gauge transformation $\exp\left(i \sum_\alpha \theta_\alpha(\xi) \mathscr{T}_\alpha\right)$ transforms ξ_a into

$$\xi_a' = \phi_a(\xi) \,. \tag{21.A.11}$$

Now dropping the prime, we have thus succeeded in constructing a gauge in which ξ_a satisfies Eq. (21.A.1), as was to be shown.

Problems

1. Calculate the effective ghost Lagrangian in a 'generalized unitarity gauge,' with $B[f]$ given as usual by Eq. (15.5.22), but now where $f_\alpha(x) = i\phi_n(x)(t_\alpha)_{nm}\langle\phi_m(0)\rangle_{VAC}$, with ϕ_n real scalar fields, and t_α imaginary antisymmetric matrices representing the Lie algebra of the gauge group. What is the ghost propagator? Is this part of the Lagrangian renormalizable?

2. What would be the effect in the $SU(2) \times U(1)$ electroweak theory if the gauge symmetry were broken by the vacuum expectation value

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

of a field ϕ_3 belonging to a real triplet $\vec{\phi} = (\phi^+, \phi^0, \phi^-)$ instead of the usual complex doublet (ϕ^0, ϕ^-)?

3. Consider the standard electroweak theory, with a single scalar doublet. In one-loop order, calculate the effect of Z^0 and neutral scalar exchange on the anomalous magnetic moment of the muon.

4. What is the lowest-order magnetic moment of the W^+ and Z^0 particles in the standard electroweak theory?

5. What would be the effect of the discovery of a fourth generation of quarks and leptons on the predictions of the unified theories of strong and electroweak interactions discussed in Section 21.5?

6. Suppose that several fields with incommensurate values of the electric charge had nonvanishing vacuum expectation values in a superconductor. What effect would this have on the properties of superconductors discussed in Section 21.6?

References

1. J. Goldstone. *Nuovo Cimento* **9**, 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127**, 965 (1962).

2. P. W. Higgs, *Phys. Lett.* **12**, 132 (1964); *Phys. Rev. Lett.* **13**, 508 (1964); *Phys. Rev.* **145**, 1156 (1966); F. Englert and R. Brout, *Phys. Rev. Lett.* **13**, 321 (1964); G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Phys. Rev. Lett.* **13**, 585 (1964); T. W. B. Kibble, *Phys. Rev.* **155**, 1554 (1967); S. Weinberg, *Phys. Rev. Lett.* **18**, 507 (1967): footnote 7.

3. S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967). A. Salam, in *Elementary Particle Physics*, ed. N. Svartholm (Almqvist and Wiksells, Stockholm,1968): p. 367.

3*a*. The Feynman rules for general gauge theories in unitarity gauge were given by S. Weinberg, *Phys. Rev.* **D7**, 1068 (1973).

4. G. 't Hooft, *Nucl. Phys.* **B35**, 167 (1971); B. W. Lee, *Phys. Rev.* **D5**, 823 (1972).

5. K. Fujikawa , B. W. Lee, and A. Sanda, *Phys. Rev.* **D6**, 2923 (1972).

5*a*. B. W. Lee and J. Zinn-Justin, *Phys. Rev.* **D5**, 3121, 3137 (1972); *Phys. Rev.* **D7**, 1049 (1972); G. 't Hooft and M. Veltman, *Nucl. Phys.* **B50**, 318 (1972); B. W. Lee, *Phys. Rev.* **D9**, 933 (1974).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

6. H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **28**, 1494 (1972).

7. T. D. Lee and C. N. Yang, *Phys. Rev.* **98**, 101 (1955).

8. Models with an incomplete $SU(2) \times U(1)$ symmetry were proposed by S. L. Glashow, *Nucl. Phys.* **22**, 579 (1961); A. Salam and J. Ward, Phys. Lett. **13**, 168 (1964). Also see J. Schwinger, *Ann. Phys. (N.Y.)* **2**, 407 (1957).

9. Speculations about possible neutral currents go back to work of G. Gamow and E. Teller, *Phys. Rev.* **51**, 288 (1937); N. Kemmer, *Phys. Rev.* **52**, 906 (1937); G. Wentzel, *Helv. Phys. Acta* **10**, 108 (1937); S. Bludman, *Nuovo Cimento* **9**, 433 (1958); J. Leites-Lopes, *Nucl. Phys.* **8**, 234 (1958).

10. Neutral currents were first observed in a single $\bar{\nu}_\mu$–e^- scattering event at CERN, by F. J. Hasert *et al.* (Aachen–Brussels–CERN–Ecole Polytechnique–Milan–Orsay–London collaboration), *Phys. Lett.* **46**, 121 (1973).

11. N. Cabibbo, *Phys. Rev. Lett.* **10**, 531 (1963).

12. I. S. Towner, E. Hagberg, J. C. Hardy, V. T. Koslowsky, and G. Savard, Chalk River preprint nucl-th/9507005 (1995).

13. S. L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev.* **D2**, 1285 (1970).

14. S. Weinberg, *Phys. Rev. Lett.* **27**, 1688 (1971); *Phys. Rev.* **5**, 1413 (1972).

15. J. J. Aubert *et al.*, *Phys. Rev. Lett.* **33**, 1404 (1974); J. E. Augustin *et al.*, *Phys. Rev. Lett.* **33**, 1406 (1974).

16. M. L. Perl *et al.*, *Phys. Rev. Lett.* **35**, 1489 (1975).

17. S. W. Herb *et al.*, *Phys. Rev. Lett.* **39**, 252 (1977).

18. F. Abe *et al.*, *Phys. Rev. Lett.* **74**, 2626 (1995); S. Abachi *et al.*, *Phys. Rev. Lett.* **74**, 2632 (1995).

19. The values given by the CDF and D0 collaborations at Fermilab are respectively $176 \pm 8 \pm 10$ GeV (F. Abe *et al.*, Ref. 18.) and $199^{+19}_{-21} \pm 22$ GeV (S. Abachi *et al.*, Ref. 18), with the first and second errors statistical and systematic, respectively. These results are interpreted to give a combined value of 181 ± 12 GeV by J. Ellis, G. L. Fogli, and E. Lisi, CERN-BARI preprint hep-ph/9507424.

20. M. Kobayashi and K. Maskawa, *Prog. Theor. Phys.* **49**, 282 (1972).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

20*a*. F. J. Gilman, K. Kleinknecht, and B. Renk, Carnegie Mellon–Mainz preprint CMU-HUP95-19–DOE-ER/40682- 107 (1995), to be published in the 1995 *Review of Particle Properties*.

20*b*. S. Weinberg, *Phys. Rev. Lett.* **37**, 657 (1976). For earlier models in which scalar fields are responsible for CP and T non-conservation see T. D. Lee, *Phys. Rev.* **D8**, 1226 (1973); *Phys. Rep.* **9C**, 143 (1974).

21. F. J. Hasert *et al.*, *Phys. Lett.* **46B**, 138 (1973); P. Musset, *Jour. de Physique* **11/12**, T34 (1973). Neutral current events were seen at about the same time by the Harvard–Pennsylvania–Wisconsin–Fermilab group at Fermilab, but publication of their paper was delayed, so they took the opportunity to rebuild their detector, and at first did not find the same signal. Evidence for neutral currents was published by this group in A. Benvenuti *et al.*, *Phys. Rev. Lett.* **32**, 800 (1974).

22. G. Arnison *et al. Phys. Lett.* **122B**, 103 (1983); **126B**, 398 (1983); **129B**, 273 (1983); **134B**, 469 (1984); **147B**, 241 (1984).

23. F. Abe et al. (CDF collaboration), *Phys. Rev. Lett.* **75**, 11 (1995). The W mass is measured here in observations of the decays $W \to \mu + \nu$ and $W \to e + \nu$.

24. CERN report LEPEWWG/95-01, unpublished (1995).

25. M. Green and M. Veltman, *Nucl. Phys.* **B169**, 137 1980); V. A. Novikov, L. B. Okun', and M. I. Vysotsky, *Nucl. Phys.* **B397**, 35 (1992). For later studies, see P. Langacker, in *Precision Tests of the Standard Model* (World Scientific, Singapore, 1994); P. Bamert, C. P. Burgess, and I. Maksymyk, McGill–Neuchatal–Texas preprint hep-ph/9505339 (1995); W. Hollik, Karlsruhe preprint hep-ph/9507406 (1995).

26. S. Fanchotti, B. Kniehl, and A. Sirlin, *Phys. Rev.* **D48**, 307 (1973), and references cited therein.

27. J. Ellis, G. L. Fogli, and E. Lisi, Ref. 19.

27*a*. S. Weinberg, *Phys. Rev. Lett.* **43**, 1566 (1979).

27*b*. For instance, in the so-called see-saw mechanism, a neutrino mass of this order would be produced by exchange of a heavy neutral lepton of mass M; see M. Gell-Mann, P. Ramond, and R. Slansky, in *Supergravity*, ed. P. van Nieuwenhuizen and D. Freedman (North Holland, Amsterdam, 1979): p. 315; T. Yanagida, *Prog. Theor. Phys.* **B135**, 66 (1978).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

27*c*. S. Weinberg, Ref. 27*a*; F. Wilczek and A. Zee, *Phys. Rev. Lett.* **43**, 1571 (1979). Also see S. Weinberg, *Phys. Rev.* **D22**, 1694 (1980).

28. This general formalism was described by S. Weinberg, *Phys. Rev.* **D13**, 974 (1976). For earlier work on dynamical symmetry breaking in gauge theories, see R. Jackiw and K. Johnson, *Phys. Rev.* **D8**, 2386 (1973); J. M. Cornwall and R. E. Norton, *Phys. Rev.* **D8**, 3338 (1973).

29. This interpretation of the results of the standard model is due to L. Susskind, Ref. 30.

30. S. Weinberg, ref. 28 and *Phys. Rev.* **D19**, 1277 (1979); L. Susskind, *Phys. Rev.* **D19**, 2619 (1979). The term 'technicolor' is due to Susskind.

31. S. Dimopoulos and L. Susskind, *Nucl. Phys.* **B15**, 237 (1979); E. Eichten and K. Lane, *Phys. Lett.* **90B**, 125 (1980); S. Weinberg, unpublished work quoted by Eichten and Lane. For a review, see E. Farhi and L Susskind, *Phys. Rep.* **74**, 277 (1981).

32. S. Weinberg, *Phys. Rev.* **D5**, 1962 (1972).

33. J. Pati and A. Salam, *Phys. Rev. Lett.* **31**, 275 (1973).

34. H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **32**, 438 (1974).

35. H. Georgi, in *Particles and Fields — 1974*, ed. C. Carlson (Amer. Inst. of Physics, New York, 1975).

36. H. Georgi, H. R. Quinn, and S. Weinberg, *Phys. Rev. Lett.* **33**, 451 (1974).

36*a*. S. Dimopoulos and H. Georgi, *Nucl. Phys.* **B193**, 150 (1981); J. Ellis, S. Kelley, and D. V. Nanopoulos, *Phys. Lett.* **B260**, 131 (1991); U. Amaldi, W. de Boer, and H. Furstmann, *Phys. Lett.* **B260**, 447 (1991). For more recent analyses of the data, see P. Langacker and N. Polonsky, *Phys. Rev.* **D47**, 4028 (1993); **D49**, 1454 (1994); L. J. Hall and U. Sarid, *Phys. Rev. Lett.* **70**, 2673 (1993).

36*b*. S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Rev.* **D24**, 1681 (1981).

37. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).

38. P. M. Anderson, *Phys. Rev.* **130**, 439 (1963).

39. S. Weinberg, *Prog. Theor. Phys. Suppl.* No. 86, 43 (1986).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

40. B. D. Josephson, *Phys. Lett.* **1**, 25 (1962).

41. V. L. Ginzburg and L. D. Landau, *JETP (USSR)* **20**, 1064 (1950).

42. L. P. Gor'kov, *Soviet Phys. JETP* **9**, 1364 (1959).

43. A. A. Abrikosov, *Soviet Phys. JETP* **5**, 1174 (1957).

43*a*. E. B. Bogomol'nyi, *Sov. J. Nucl. Phys.* **24**, 449 (1976). For numerical calculations, see E. B. Bogomol'nyi and A. I. Vainshtein, *Sov. J. Nucl. Phys.* **23**, 588 (1976).

44. G. Benfatto and G. Gallavotti, *J. Stat. Phys.* **59**, 541 (1990); *Phys. Rev.* **42**, 9967 (1990); J. Feldman and E. Trubowitz, *Helv. Phys. Acta* **63**, 157 (1990); **64**, 213 (1991); **65**, 679 (1992); R. Shankar, *Physica* **A177**, 530 (1991); *Rev. Mod. Phys.* **66**, 129 (1993); J. Polchinski, in *Recent Developments in Particle Theory, Proceedings of the 1992 TASI*, eds. J. Harvey and J. Polchinski (World Scientific, Singapore, 1993).

44*a*. S. Weinberg, *Nucl. Phys.* **B413**, 567 (1994).

45. R. L. Stratonovich, *Sov. Phys. Dokl.* **2**, 416 (1957); J. Hubbard, *Phys. Rev. Lett.* **3**, 77 (1959).

46. A. A. Abrikosov, L. P. Gor'kov, I. Ye. Dzyaloshinskii, *Quantum Field Theoretical Methods in Statistical Physics*, 2nd edn., transl. by D. E. Brown (Pergamon Press, Oxford, 1965).

47. In order to give a precise meaning to the mass M it is necessary to take into account both two-loop corrections to the renormalization group equations (21.5.6)–(21.5.8), and one-loop 'threshold corrections' to condition (21.5.5). See S. Weinberg, *Phys. Lett.* **82B**, 387 (1979).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.009

22
Anomalies

There are subtleties in the implications of symmetries in quantum field theory that have no counterpart in classical theories. Even in renormalizable theories, the infinities in quantum field theory require that some sort of regulator or cut-off be used in actual calculations. The regulator may violate symmetries of the theory, and even when this regulator is removed at the end of the calculation it may leave traces of this symmetry violation. This problem first emerged in trying to understand the decay rate of the neutral pion, in the form of an anomaly that violates a *global* symmetry of the strong interactions. Anomalies can also violate *gauge* symmetries, but in this case the theory becomes inconsistent, so that the condition of anomaly cancellation may be used as a constraint on physical gauge theories. The importance of anomalies will become even more apparent in the next chapter, where we shall study the non-perturbative effects of anomalies in the presence of topologically non-trivial field configurations.

22.1 The π^0 Decay Problem

By the mid-1960s the picture of the pion as a Goldstone boson associated with a spontaneously broken $SU(2) \otimes SU(2)$ symmetry of the strong interactions had scored a number of successes, outlined here in Chapter 19. However, this picture also had a few outstanding failures. The most disturbing had to do with the rate of the dominant decay mode of the neutral pion, $\pi^0 \to 2\gamma$. It was the solution of this problem that led to the discovery of symmetry-breaking anomalies.

After integrating out all heavy and trapped particles, we would expect the effective Lagrangian for $\pi^0 \to 2\gamma$ to be given by the unique gauge- and Lorentz-invariant term with no more than two derivatives:

$$\mathscr{L}_{\pi\gamma\gamma} = g\pi^0 \epsilon^{\mu\nu\rho\lambda} F_{\mu\nu} F_{\rho\lambda} \,, \tag{22.1.1}$$

where g is an unknown constant with the dimensions $[mass]^{-1}$. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

methods of Section 3.4 can then be used to calculate the rate for $\pi^0 \to 2\gamma$

$$\Gamma(\pi^0 \to 2\gamma) = \frac{m_\pi^3 g^2}{\pi} . \tag{22.1.2}$$

One might naively expect g to be of order

$$g \approx \frac{e^2}{8\pi^2 F_\pi} , \tag{22.1.3}$$

where $F_\pi \simeq 190$ MeV is used as a typical strong interaction mass scale, and a factor $1/8\pi^2$ is inserted because the graphs responsible for $\pi^0 \to 2\gamma$ would have to include at least one loop. For instance, in 1949, using the pre-QCD theory of pions and nucleons with interaction Lagrangian $iG_{\pi N}\vec{\pi} \cdot \bar{N}2\vec{t}\gamma_5 N$, Steinberger[1] calculated that the contribution to g from triangle graphs with a single proton loop is

$$g_\triangle = \frac{e^2 G_{\pi N}}{32\pi^2 m_N} . \tag{22.1.4}$$

This is numerically not very different from Eq. (22.1.3), because the Goldberger–Treiman relation (see Section 19.4) gives $G_{\pi N} = 2m_N g_A/F_\pi$.

This estimate of the amplitude for $\pi^0 \to 2\gamma$ leaves out the special constraints imposed by the $SU(2) \otimes SU(2)$ symmetry. The electromagnetic interaction violates most of this symmetry, but at least formally it has no effect on the $U(1) \times U(1)$ subgroup generated by the electrically neutral generators of $SU(2) \otimes SU(2)$. Acting on quarks, the infinitesimal pseudoscalar element of this subgroup has the effect:

$$\delta u = i\epsilon\gamma_5 u , \qquad \delta d = -i\epsilon\gamma_5 d , \tag{22.1.5}$$

so that the electric current is invariant:

$$\delta\left[\frac{2}{3}\bar{u}\gamma^\mu u - \frac{1}{3}\bar{d}\gamma^\mu d\right] = 0 . \tag{22.1.6}$$

(This argument was made by Sutherland[2] and Veltman[3] before the advent of quantum chromodynamics, with the proton and neutron in place of the u and d quarks.) Because the π^0 is the Goldstone boson associated with this symmetry, a non-derivative interaction like (22.1.1) can arise only from the breaking of this symmetry by quark masses, and so must be proportional to $m_\pi^2 \propto m_u + m_d$. With this in mind, we would expect the constant g to be suppressed[2,3] by an extra factor m_π^2/m_N^2:

$$g \approx \frac{e^2}{8\pi^2 F_\pi}\left(\frac{m_\pi^2}{m_N^2}\right) . \tag{22.1.7}$$

(In place of m_N we might have used the chiral-symmetry-breaking scale $2\pi F_\pi = 1200$ MeV discussed in Section 19.3, with little change in our results). There are also chiral-invariant effective $\pi^0\gamma\gamma$ interactions involving

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

the derivative of the π^0 field. Lorentz invariance requires that these involve at least two additional derivatives. Using the homogeneous Maxwell equation $\partial_\mu F_{\nu\lambda} - \partial_\nu F_{\mu\lambda} = -\partial_\lambda F_{\mu\nu}$ and integrating by parts, we see that there is only one independent chiral-invariant coupling with just two extra derivatives, given by inserting a d'Alembertian acting on the π^0 field in Eq. (22.1.1). On the pion mass shell this is just the same as an interaction (22.1.1) with an extra factor m_π^2, leading to the same estimate (22.1.7). It was sometimes said that chiral symmetry would forbid the decay $\pi^0 \to 2\gamma$, but it is more accurate to say that chiral symmetry would make the rate for this process go as m_π^7 instead of m_π^3 for $m_\pi \to 0$.

The difficulty is that the observed rate for $\pi^0 \to 2\gamma$ is much larger than would be expected from Eq. (22.1.7), and is in fact much closer to what would be inferred from the naive result Eq. (22.1.3). To be specific, from Eqs. (22.1.7) and (22.1.2) we would expect a decay rate

$$\Gamma(\pi^0 \to 2\gamma) \approx \frac{m_\pi^7 \alpha^2}{4\pi^3 F_\pi^2 m_N^4} = 1.9 \times 10^{13}\ \text{s}^{-1}\,, \tag{22.1.8}$$

while using Eq. (22.1.3) in place of Eq. (22.1.7) would give

$$\Gamma(\pi^0 \to 2\gamma) \approx \frac{m_\pi^3 \alpha^2}{4\pi^3 F_\pi^2} = 4.4 \times 10^{16}\,\text{s}^{-1}\,. \tag{22.1.9}$$

The observed rate is $\Gamma(\pi^0 \to 2\gamma) = (1.19 \pm 0.08) \times 10^{16}\ \text{s}^{-1}$, which is in satisfactory agreement with the naive rough estimate (22.1.9), and almost three orders of magnitude larger than the 'improved' result (22.1.8)! One is forced to the conclusion that something anomalous here is invalidating the chiral symmetry that led us to introduce the additional factor m_π^2/m_N^2 in g. Similar problems arise in trying to understand the rates of some other processes, such as $\eta^0 \to \gamma + \gamma$.

In 1969 the source of this anomaly was traced by Bell and Jackiw[4] to the violation of chiral symmetry by the regulator that is needed in order to derive consequences of the conservation of the neutral axial vector current for one-loop Feynman diagrams. Their result was confirmed, generalized, and extended to higher orders by Adler,[5] who independently discovered chiral anomalies in a study of axial vector Ward identities in quantum electrodynamics. It was subsequently realized in 1979 by Fujikawa[6] that in the path-integral formulation of field theory the chiral-symmetry-breaking anomaly enters only in the measure used to define the path integral over fermion fields. As we shall see in the next section, this approach makes it simple to evaluate the amplitude for $\pi^0 \to 2\gamma$ produced by this anomaly to all orders of perturbation theory. After that, we shall return to the direct calculation of anomalies in more general theories, and discuss various applications.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

22.2 Transformation of the Measure: The Abelian Anomaly

We now turn to a calculation of anomalies of the sort relevant to π^0 decay. For this purpose we adopt Fujikawa's interpretation[6] of the anomaly as a symptom of the impossibility of defining a suitably invariant measure for integrations over fermionic field variables. Fujikawa's analysis of this problem was based on the use of path integrals in Euclidean space, and on an expansion of the fermionic variables of integration in eigenfunctions of the gauge-invariant Dirac operator, which is Hermitian in four-dimensional Euclidean space. Here we shall first present a less rigorous derivation, based on the familiar path integrals in Minkowski space, which will allow us to obtain the correct answer with a minimum of work. The Euclidean approach will be taken up briefly at the end of this section, and used to derive a famous index theorem.

We will start by evaluating the anomaly in the transformation of the measure under an arbitrary local matrix transformation $\psi(x) \to U(x)\psi(x)$ of a column $\psi_n(x)$ of massless complex spin 1/2 fermion fields that interact non-chirally with a set of gauge fields $A^\mu_\alpha(x)$ (such as the u and d quark fields interacting with the electromagnetic vector potential $A^\mu(x)$ in the problem of calculating the rate for $\pi^0 \to 2\gamma$). These are fermionic variables, so the measure is transformed not with the determinant of the transformation matrix, but with its inverse:

$$[d\psi][d\bar{\psi}] \to (\mathrm{Det}\,\mathscr{U}\ \mathrm{Det}\,\bar{\mathscr{U}})^{-1}[d\psi][d\bar{\psi}]\,, \tag{22.2.1}$$

where

$$\mathscr{U}_{xn,ym} \equiv U(x)_{nm}\delta^4(x-y)\,, \tag{22.2.2}$$

$$\bar{\mathscr{U}}_{xn,ym} \equiv [\gamma_4 U(x)^\dagger \gamma_4]_{nm}\delta^4(x-y) \tag{22.2.3}$$

and $\gamma_4 \equiv i\gamma^0$ is the matrix used to define $\bar{\psi} = \psi^\dagger\gamma_4$. The indices n, m run over flavor labels and Dirac spin indices.

The reader may wonder at this point why we bother to include factors of γ_4 in Eq. (22.2.3), since they just amount to a unitary transformation that should not affect the determinants. The answer is that in order for calculations to be meaningful we will find it necessary to regulate the sum over fermion modes in calculating propagators and determinants, and we shall find that the γ_4 factors do affect the regulated determinants. Whether or not we include the γ_4 factors thus depends on the method of regularization that is used to make these hand-waving manipulations meaningful. We include the factors of γ_4 because we wish to regulate in such a way as to preserve Lorentz invariance, and in the cases of interest here it is $\gamma_4\, U(x)^\dagger\, \gamma_4$ rather than $U(x)^\dagger$ that transforms as a scalar.

First, let us consider the case where $U(x)$ is a unitary non-chiral trans-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

formation

$$U(x) = \exp[i\alpha(x)t] \,, \tag{22.2.4}$$

with t an ordinary Hermitian matrix (not involving γ_5, but not necessarily traceless) and $\alpha(x)$ an arbitrary real function of x. In this case $\mathscr{U}$ is pseudounitary:

$$\bar{\mathscr{U}}\mathscr{U} = 1 \,, \tag{22.2.5}$$

so the measure is invariant under this sort of transformation. In particular, the symmetry under the gauge group itself, where t is one of the non-chiral generators t_α, is not spoiled by any anomalies.

Next, consider a local chiral transformation, with

$$U(x) = \exp[i\gamma_5\alpha(x)t] \,, \tag{22.2.6}$$

with t again an ordinary Hermitian matrix and $\alpha(x)$ again an arbitrary real function of x. In this case $\mathscr{U}$ is pseudo-Hermitian:

$$\bar{\mathscr{U}} = \mathscr{U} \,. \tag{22.2.7}$$

The measure is not invariant under the chiral transformation; rather, we have

$$[d\psi][d\bar{\psi}] \to (\mathrm{Det}\,\mathscr{U})^{-2}[d\psi][d\bar{\psi}] \,. \tag{22.2.8}$$

Now let us specialize to the case of an *infinitesimal* local chiral transformation. Taking $\alpha(x)$ to be infinitesimal in Eq. (22.2.6), we have here

$$[\mathscr{U} - 1]_{nx,my} = i\alpha(x)[\gamma_5\, t]_{nm}\, \delta^4(x - y) \,. \tag{22.2.9}$$

Using the identity $\mathrm{Det}M = \exp \mathrm{Tr} \ln M$ and the limiting formula $\ln(1 + x) \to x$ for $x \to 0$, the measure now has the transformation property

$$[d\psi][d\bar{\psi}] \to \exp\left\{ i \int d^4x\, \alpha(x)\mathscr{A}(x) \right\} [d\psi][d\bar{\psi}] \,, \tag{22.2.10}$$

where $\mathscr{A}$ is the anomaly function:

$$\mathscr{A}(x) = -2\,\mathrm{Tr}\,\{\gamma_5 t\}\, \delta^4(x - x) \,, \tag{22.2.11}$$

with 'Tr' here denoting a trace over both Dirac and species indices. The measure $[d\psi][d\bar{\psi}]$ appears in the path integral weighted with a factor $\exp\{i \int d^4x\, \mathscr{L}(x)\}$, so the factor $\exp\left\{ i \int d^4x\, \alpha(x)\mathscr{A}(x) \right\}$ in the transformation rule (22.2.10) for the measure has the same effect as if the Lagrangian density $\mathscr{L}(x)$ were not invariant under these transformations, but instead $\mathscr{L}(x) \to \mathscr{L}(x) + \alpha(x)\mathscr{A}(x)$. Hence when we use an effective Lagrangian with the fermions integrated out, to take account of the anomaly we must

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

be sure to include a non-invariant term so that

$$\mathscr{L}_{\text{eff}}(x) \to \mathscr{L}_{\text{eff}}(x) + \alpha(x)\mathscr{A}(x) . \tag{22.2.12}$$

It now remains to calculate the anomaly function $\mathscr{A}(x)$.

At first sight it does not look like we can get any definite result for the anomaly. The delta-function is infinite, but the trace vanishes. To do better, we must introduce a regulator to make $\delta^4(x-x)$ meaningful. We can do this in a gauge-invariant manner by inserting a differential operator $f(-\not{D}_x^2/M^2)$ acting on the delta-function before taking its argument to zero:

$$\mathscr{A}(x) = -2\left[\text{Tr}\left\{\gamma_5 t f(-\not{D}_x^2/M^2)\right\}\delta^4(x-y)\right]_{y\to x} . \tag{22.2.13}$$

Here D_x is the Dirac differential operator in the presence of the gauge field $A_\alpha^\mu(x)$:

$$(D_x)_\mu \equiv \frac{\partial}{\partial x^\mu} - it_\alpha A_{\alpha\mu}(x) . \tag{22.2.14}$$

Also, M is some large mass, eventually to be taken to infinity, and $f(s)$ is a smooth function, subject only to the condition* that as s goes from 0 to ∞, $f(s)$ must drop smoothly from 1 to 0:

$$f(0) = 1 , \qquad f(\infty) = 0 , \tag{22.2.15}$$

$$sf'(s) = 0 \quad \text{at} \quad s = 0 \ \text{and} \ s = \infty . \tag{22.2.16}$$

Note that we do not take the regulator function to be a function of $\not{\partial}$, because we want to preserve gauge invariance, and we do not take it to be a function of $D^\mu D_\mu$, because we need it to regulate not only the determinant but also the fermion propagator $\not{D}^{-1}$.

To evaluate the expression (22.2.13), we use the Fourier representation of the delta-function, and write the anomaly function as

$$\begin{aligned}\mathscr{A}(x) &= -2\int\frac{d^4k}{(2\pi)^4}\left[\text{Tr}\left\{\gamma_5 t f(-\not{D}_x^2/M^2)\right\}e^{ik\cdot(x-y)}\right]_{y=x} \\ &= -2\int\frac{d^4k}{(2\pi)^4}\,\text{Tr}\left\{\gamma_5 t f\Big(-[i\not{k}+\not{D}_x]^2/M^2\Big)\right\} .\end{aligned}$$

(A derivative with respect to x gives zero when acting on the extreme right in the second expression, but not when acting on $A_{\alpha\mu}(x)$.) Rescaling the momentum k^μ by a factor M, this is

$$\mathscr{A}(x) = -2M^4\int\frac{d^4k}{(2\pi)^4}\,\text{Tr}\left\{\gamma_5 t f\Big(-[i\not{k}+\not{D}_x/M]^2\Big)\right\} . \tag{22.2.17}$$

* For instance, we might take $f(s) = e^{-s^2}$, as done originally by Fujikawa, or $f(s) = 1/(1+s)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

The argument of the cut-off function may be written

$$-\left[i\,\not{k}+\frac{\not{D}_x}{M}\right]^2=k^2-\frac{ik\cdot D_x}{M}-\left(\frac{\not{D}_x}{M}\right)^2\,. \tag{22.2.18}$$

Eq. (22.2.17) receives contributions in the limit $M\to\infty$ only from terms in the expansion of $f(-[i\not{k}+\not{D}_x/M]^2)$ that have no more than four factors of $1/M$, and also only from terms that contain at least four Dirac gamma matrices, because otherwise the trace over Dirac indices vanishes. This leaves us with only the terms of second order in $\not{D}_x^2$:

$$\mathscr{A}(x)=-\int\frac{d^4k}{(2\pi)^4}\,f''(k^2)\,\mathrm{Tr}\left\{\gamma_5\,t\,\not{D}_x^4\right\}\,, \tag{22.2.19}$$

which are now independent of the regulator mass M.

To evaluate the integral over k, we perform the same sort of rotation of the k^0 contour of integration as in evaluating Feynman diagrams, so that in effect k^0 is replaced with ik^4, with k^4 running from $-\infty$ to $+\infty$. (This is a step that can really only be justified by working with Euclidean path integrals from the beginning.) The integral is then in effect

$$\int d^4k\,f''(k^2)=i\int_0^\infty 2\pi^2\kappa^3\,d\kappa\,f''(\kappa^2)\,. \tag{22.2.20}$$

By repeated integration by parts using Eq. (22.2.16) and then Eq. (22.2.15), this is

$$\int d^4k\,f''(k^2)=i\pi^2\int_0^\infty ds\,s\,f''(s)=-i\pi^2\int_0^\infty ds\,f'(s)=i\pi^2\,. \tag{22.2.21}$$

To calculate the trace, we write

$$\begin{aligned}\not{D}_x^2&=\tfrac{1}{4}\left\{(D_x)^\mu,(D_x)^\nu\right\}\left\{\gamma_\mu,\gamma_\nu\right\}+\tfrac{1}{4}\left[(D_x)^\mu,(D_x)^\nu\right]\left[\gamma_\mu,\gamma_\nu\right]\\&=D_x^2-\tfrac{1}{4}it_\alpha F_\alpha^{\mu\nu}\left[\gamma_\mu,\gamma_\nu\right]\,.\end{aligned} \tag{22.2.22}$$

The only term in $\not{D}_x^4$ that contributes to the trace over Dirac indices is the one involving a product of four Dirac matrices, which gives

$$\mathrm{tr_D}\left\{\gamma_5\left[\gamma_\mu,\gamma_\nu\right]\left[\gamma_\rho,\gamma_\sigma\right]\right\}=16\,i\,\epsilon_{\mu\nu\rho\sigma}\,, \tag{22.2.23}$$

where '$\mathrm{tr_D}$' denotes a trace over Dirac indices only, and as usual $\epsilon_{\mu\nu\rho\sigma}$ is the totally antisymmetric tensor with $\epsilon^{0123}=+1$. Using Eqs. (22.2.21)–(22.2.23) in Eq. (22.2.19) then gives the anomaly function as

$$\mathscr{A}(x)=-\frac{1}{16\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\,F_\alpha^{\mu\nu}(x)\,F_\beta^{\rho\sigma}(x)\,\mathrm{tr}\left\{t_\alpha\,t_\beta\,t\right\}\,, \tag{22.2.24}$$

with 'tr' here denoting a trace only over indices labelling the various fermion species. In the special case in which t is the unit matrix, the quantity (22.2.24) is known as the *Chern–Pontryagin density*.

This result can be expressed in terms of the current associated with the anomalous symmetry. If we assume for simplicity that the action itself

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

is invariant under the symmetry transformation $\psi(x) \to \psi(x) + it\gamma_5\alpha\psi(x)$ with constant infinitesimal parameter α, then as discussed in Section 7.3, the change in the action when we make such a transformation with a spacetime-*dependent* parameter $\alpha(x)$ may be written as $\delta I = \int d^4x\, J_5^\mu(x)\partial_\mu\alpha(x)$, where $J_5^\mu(x)$ is the current that becomes conserved when the field operators satisfy the dynamical equations that make the action stationary with respect to arbitrary variations in the fields. When we make the change of variables $\delta\psi(x) = it\gamma_5\alpha(x)\psi(x)$, the change in the path integral over the fermionic fields is

$$\delta \int [d\psi][d\bar{\psi}]\, e^{iI} = i \int d^4x \int [d\psi][d\bar{\psi}][\mathscr{A}(x)\alpha(x) + J_5^\mu(x)\partial_\mu\alpha(x)]e^{iI} \,. \tag{22.2.25}$$

But this is a mere change of variables, and so for arbitrary $\alpha(x)$ it cannot affect the path integral. Therefore for arbitrary gauge fields

$$\langle \partial_\mu J_5^\mu \rangle_A = \mathscr{A} = -\frac{1}{16\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\, F_\alpha^{\mu\nu}\, F_\beta^{\rho\sigma}\, \mathrm{tr}\,\{t_\alpha\, t_\beta\, t\}\,, \tag{22.2.26}$$

where for any operator $\mathcal{O}$, $\langle \mathcal{O} \rangle_A$ is the quantum average of $\mathcal{O}$ in a fixed background field $A^\mu(x)$:

$$\langle \mathcal{O} \rangle_A \equiv \frac{\int [d\psi][d\bar{\psi}]\, e^{iI}\, \mathcal{O}}{\int [d\psi][d\bar{\psi}]\, e^{iI}}\,. \tag{22.2.27}$$

Incidentally, it is possible to rewrite Eq. (22.2.26) as a conservation condition. Consider the special case where $\mathrm{tr}\,\{t_\alpha\, t_\beta\, t\}$ is proportional to $\delta_{\alpha\beta}$:

$$\mathrm{tr}\,\{t_\alpha\, t_\beta\, t\} = N\delta_{\alpha\beta}\,. \tag{22.2.28}$$

Define a current known as the *Chern–Simons class*:

$$\begin{aligned} G^\mu &\equiv 2\epsilon^{\mu\nu\lambda\rho}\left[A_{\gamma\nu}\partial_\lambda A_{\gamma\rho} + \frac{1}{3}C_{\alpha\beta\gamma}A_{\alpha\nu}A_{\beta\lambda}A_{\gamma\rho}\right] \\ &= \epsilon^{\mu\nu\lambda\rho}\left[A_{\gamma\nu}\, F_{\gamma\lambda\rho} - \frac{1}{3}C_{\alpha\beta\gamma}A_{\alpha\nu}A_{\beta\lambda}A_{\gamma\rho}\right]\,, \end{aligned} \tag{22.2.29}$$

which satisfies the identity

$$\partial_\mu G^\mu = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}\, F_{\gamma\mu\nu}\, F_{\gamma\lambda\rho}\,. \tag{22.2.30}$$

Eq. (22.2.30) allows us to rewrite Eq. (22.2.26) as the condition

$$\partial_\mu K^\mu = 0\,, \tag{22.2.31}$$

where

$$K^\mu \equiv \left\langle J_5^\mu \right\rangle_A + \frac{N}{8\pi^2}\, G^\mu\,. \tag{22.2.32}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

However, we cannot use the conservation of the current K^μ to argue, as we did in the previous section, that the process $\pi^0 \to 2\gamma$ is suppressed, because in making that argument we assumed not only the chiral symmetry associated with axial-vector current conservation but also electromagnetic gauge invariance, and as we see from Eq. (22.2.29), the current K^μ though conserved is not gauge-invariant.

Our derivation of the formula (22.2.24) for the anomaly shows that if we had evaluated the anomaly function using a differential operator $f(-\not{\partial}_x^2/M^2)$ in place of $f(-\not{D}_x^2/M^2)$ in Eq. (22.2.13), we would have obtained a vanishing anomaly function. In fact, with this regulator procedure the axial-vector current is K^μ, not J_5^μ. The trouble with this procedure, as mentioned above, is that the regulator operator is no longer gauge-invariant, resulting in the presence of a non-gauge-invariant term in K^μ. There is no regulator procedure for fermion propagators as well as determinants that is both gauge- and chiral-invariant.

We may now return to the problem that gave anomalies their start, and use our results to calculate the actual rate of the process $\pi^0 \to 2\gamma$. The symmetry of interest here is generated by charge-neutral chiral transformations of the light quark fields:

$$\delta u = i\alpha\gamma_5 u\,, \qquad \delta d = -i\alpha\gamma_5 d\,. \tag{22.2.33}$$

This symmetry is anomaly-free in pure quantum chromodynamics, because the u and d belong to the same representation of the color gauge group, so that their contributions to the gluon–gluon terms in the anomaly of the symmetry (22.2.33) cancel. On the other hand, in the presence of an electromagnetic field $A^\mu(x)$, this symmetry has an anomaly

$$\mathscr{A}(x) = -\frac{1}{16\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\,F^{\mu\nu}(x)\,F^{\rho\sigma}(x)\,\mathrm{tr}\left\{q^2\tau_3\right\}\,,$$

where q is the quark charge matrix and τ_3 is the diagonal 2×2 matrix with elements $+1$ for u and -1 for d. If we assume as usual that there are N_c u quarks of charge $2e/3$ and an equal number of d quarks of charge $-e/3$, the trace is

$$\mathrm{tr}\left\{q^2\tau_3\right\} = N_c\left(\frac{2e}{3}\right)^2(+1) \;+\; N_c\left(\frac{-e}{3}\right)^2(-1) = \frac{N_c e^2}{3}\,,$$

so the anomaly function here is

$$\mathscr{A}(x) = -\frac{N_c e^2}{48\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\,F^{\mu\nu}(x)\,F^{\rho\sigma}(x)\,. \tag{22.2.34}$$

We must then include in the effective Lagrangian terms that under the chiral transformation (22.2.33) give it the transformation (22.2.12), that is,

$$\delta\mathscr{L}_{\mathrm{eff}}(x) = \alpha\mathscr{A}(x) = -\frac{N_c e^2}{48\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\,F^{\mu\nu}(x)\,F^{\rho\sigma}(x)\alpha\,. \tag{22.2.35}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Under the transformation (22.2.33) the pion field transforms as

$$\delta\pi^0 = \alpha F_\pi \,, \tag{22.2.36}$$

where $F_\pi = 184$ MeV is the pion decay amplitude introduced in Chapter 19. (The normalization convention for this constant is fixed by our definition of the symmetry generator as $\gamma_5\tau_3 = 2\gamma_5 t_3$.) It follows that we must include in the effective Lagrangian a term

$$\frac{\pi^0(x)\mathscr{A}(x)}{F_\pi} = -\frac{N_c e^2}{48\pi^2 F_\pi}\,\epsilon_{\mu\nu\rho\sigma}\,F^{\mu\nu}(x)\,F^{\rho\sigma}(x)\,\pi^0(x)\,. \tag{22.2.37}$$

Comparing this with the general formula (22.1.1) for the effective Lagrangian for the decay $\pi^0 \to 2\gamma$, we see that the constant g in Eq. (22.1.1) must have the value[5]

$$g = \frac{N_c e^2}{48\pi^2 F_\pi}\,. \tag{22.2.38}$$

(This shows that our previous crude order-of-magnitude estimate (22.1.3) was too large by just a factor $6/N_c$.) The rate (22.1.2) for pion decay is thus predicted to be

$$\Gamma(\pi^0 \to 2\gamma) = \frac{N_c^2\alpha^2 m_\pi^3}{144\pi^3 F_\pi^2} = \left(\frac{N_c}{3}\right)^2 \times 1.11 \times 10^{16}\ \mathrm{s}^{-1}\,. \tag{22.2.39}$$

The observed rate is $\Gamma(\pi^0 \to 2\gamma) = (1.19 \pm 0.08) \times 10^{16}\ \mathrm{s}^{-1}$, in good agreement with the theoretical result (22.2.39) if and only if $N_c = 3$. The success of this calculation was one of the first pieces of firm evidence that there are three colors of quarks.

Remarkably, as we have seen in the previous section, shortly after the discovery of the π^0 Steinberger calculated g from a single proton-loop diagram, and obtained the result $g = e^2 G/32\pi^2 m_N$, where G is the pseudoscalar pion–nucleon coupling constant. This result would precisely agree with Eq. (22.2.38) for $N_c = 3$ if we used the Goldberger–Treiman relation with $g_A = 1$ to set $G = 2m_N/F_\pi$. The correct result is larger than the Steinberger result by a factor $g_A^2 = 1.56$. The reason that Steinberger got nearly the right result is because the answer is determined by the triangle anomaly, which is proportional to $\mathrm{tr}\{q^2\tau_3\}$. For one proton, this trace has the value e^2, which is the same as the quark value found above for three colors of quarks.

* * *

As mentioned earlier, a more rigorous derivation of the anomaly can be given by using path integrals in Euclidean spacetime. (The use of Euclidean path integrals is briefly discussed in Appendix A of Chapter 23.) We introduce a Euclidean fourth coordinate $x_4 = ix^0 = -ix_0$, and correspondingly $\partial_4 = -i\partial_0$, $\gamma_4 \equiv i\gamma^0$, and $A_{4\alpha} = iA_\alpha^0$. The spacetime

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

volume is then expressed as $d^4x = -i(d^4x)_E$, where $(d^4x)_E$ is the Euclidean volume element $(d^4x)_E = dx_1 dx_2 dx_3 dx_4$. In Euclidean spacetime the fields $\psi(x)$ and $\bar{\psi}(x)$ must be regarded as entirely independent, with local chiral transformations defined by $\delta\psi(x) = i\alpha(x)t\gamma_5\psi(x)$, $\delta\bar{\psi}(x) = -i\alpha(x)\bar{\psi}(x)t\gamma_5$. The transformation of the measure is then again given by Eq. (22.2.10), with the anomaly function $\mathscr{A}(x)$ given by Eq. (22.2.11). Introducing a regulator function as before, this again gives the formula (22.12.13) for $\mathscr{A}(x)$. One great advantage of the Euclidean approach is that with x_4 and $A_{4\alpha}$ real, the Dirac operator $i\not{D}$ in Eq. (22.2.13) is Hermitian:

$$i\not{D} = \left[i\partial_i + t_\alpha A_{i\alpha}\right]\gamma_i \,, \tag{22.2.40}$$

with i, j, etc. here summed over the values 1, 2, 3, 4. It therefore has orthonormal spinor eigenfunctions $\varphi_\kappa(x)$:

$$i\not{D}\varphi_\kappa = \lambda_\kappa \varphi_\kappa \,, \tag{22.2.41}$$

$$\int (d^4x)_E \; \varphi_\kappa(x)^\dagger \varphi_{\kappa'}(x) = \delta_{\kappa\kappa'} \,, \tag{22.2.42}$$

with real eigenvalues λ_κ. We also are assuming, as throughout this section, that t commutes with $i\not{D}$, so that we can choose the φ_κ so that $t\varphi_\kappa = t_\kappa \varphi_\kappa$. These eigenfunctions satisfy a completeness relation

$$\sum_\kappa \varphi_\kappa(x)\varphi_\kappa^\dagger(y) = \delta^4(x-y)1 \,, \tag{22.2.43}$$

where '1' is the 4×4 unit matrix. Therefore the anomaly function may now be written as the limit of a manifestly convergent sum:

$$\begin{aligned}\mathscr{A}(x) &= -2\lim_{M\to\infty} \mathrm{Tr}\left\{\gamma_5 \, t\, f(-\not{D}^2/M^2) \sum_\kappa \varphi_\kappa(x)\varphi_\kappa^\dagger(x)\right\} \\ &= -2\lim_{M\to\infty} \sum_\kappa t_\kappa \, f(\lambda_\kappa^2/M^2)\left(\varphi_\kappa^\dagger(x)\gamma_5\varphi_\kappa(x)\right) \,. \end{aligned} \tag{22.2.44}$$

In just the same way that we derived formula (22.2.24) for the anomaly function, we can show here that

$$\mathscr{A}(x) = \frac{1}{16\pi^2}\, \epsilon^E_{ijkl}\, F_{ij\alpha}\, F_{kl\beta}\, \mathrm{tr}\{t_\alpha t_\beta t\} \,, \tag{22.2.45}$$

where ϵ^E_{ijkl} is the totally antisymmetric tensor with $\epsilon^E_{1234} = +1$. (The difference of sign in Eqs. (22.2.24) and (22.2.45) arises because Eq. (22.2.45) is missing two factors of i as compared with Eq. (22.2.24): one factor i from Eq. (22.2.23), because

$$\mathrm{tr}_D\{\gamma_5[\gamma_i,\gamma_j][\gamma_k,\gamma_l]\} = 16\,\epsilon^E_{ijkl} \,,$$

together with another factor i from the replacement of d^4k with $(d^4k)_E$ in Eq. (22.2.20).)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Now, given any eigenfunction $\varphi_\kappa(x)$ of $i\not{D}$ and t with eigenvalue $\lambda_\kappa \neq 0$, there is another normalized eigenfunction $\varphi_{\kappa'}(x)$ with eigenvalues $\lambda_{\kappa'} = -\lambda_\kappa$ and $t_{\kappa'}$, given by $\varphi_{\kappa'}(x) = \gamma_5\varphi_\kappa(x)$. (Recall that in the notation used throughout this book, γ_5 is the Hermitian matrix $\gamma_5 = -i\gamma_1\gamma_2\gamma_3\gamma_0 = \gamma_1\gamma_2\gamma_3\gamma_4$.) Because $\phi_\kappa(x)$ and $\phi_{\kappa'}(x)$ are eigenvectors of a Hermitian operator with different eigenvalues they are orthogonal, so $\int d^4x \left(\varphi_\kappa^\dagger(x)\gamma_5\varphi_\kappa(x)\right) = 0$. This leaves just a sum over the eigenfunctions with $\lambda_\kappa = 0$. These are not generally paired; rather, since γ_5 anticommutes with $i\not{D}$, they can be chosen as simultaneous orthornormal eigenfunctions φ_u, φ_v of $i\not{D}$ with eigenvalue zero *and* of γ_5 with eigenvalues $+1$ and -1, respectively.

$$\begin{aligned} i\not{D}\varphi_u &= 0\,, & \gamma_5\varphi_u &= \varphi_u\,, \\ i\not{D}\varphi_v &= 0\,, & \gamma_5\varphi_v &= -\varphi_v\,. \end{aligned} \tag{22.2.46}$$

Using the fact that $f(0) = 1$, Eq. (2.2.44) then becomes

$$\mathscr{A}(x) = -2\left[\sum_u t_u\left(\varphi_u^\dagger(x)\varphi_u(x)\right) - \sum_v t_v\left(\varphi_v^\dagger(x)\varphi_v(x)\right)\right]. \tag{22.2.47}$$

Now, since the φ_u and φ_v are normalized as in Eq. (22.2.42), the integral of Eq. (22.2.47) gives

$$\int (d^4x)_E \mathscr{A}(x) = -2\left[\sum_u t_u - \sum_v t_v\right], \tag{22.2.48}$$

with the sums over u and v running over left- and right-handed zero modes of the operator $i\not{D}$, respectively. In particular, for the case where t is the unit matrix, by using Eq. (22.2.45) this may be expressed as a relation between a functional of the gauge field and the numbers of the zero modes of the Dirac operator with definite chiralities:

$$-\frac{1}{32\pi^2}\int (d^4x)_E\, \epsilon^E_{ijkl} F_{\alpha ij} F_{\beta kl}\, \mathrm{tr}\left[t_\alpha t_\beta\right] = n_+ - n_-\,, \tag{22.2.49}$$

where here $n_\pm$ are the numbers of zero modes of $\not{D}$ that have eigenvalues ± 1 for γ_5. This is the celebrated *Atiyah–Singer index theorem*.[6a] Among other things, it shows that under variations in the gauge field the integral on the left-hand side of Eq. (22.2.49) cannot change smoothly, but only by integers, so this integral can only depend on the topology of the gauge field. Its dependence on this topology will be described in Section 23.5.

22.3 Direct Calculation of Anomalies: The General Case

We saw in Section 22.2 how to use the elegant Fujikawa approach to calculate anomalies of chiral symmetries in gauge theories like quantum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

chromodynamics, where the gauge interactions are non-chiral and fermion number is conserved. This method can also be used to deal with more general problems, but there it becomes less straightforward.[7] In this section we will treat the anomaly by direct calculation, as it was originally done. This will provide a useful additional perspective on the anomalies, and will incidentally allow us to discuss anomalies in general theories with very little extra trouble.

To deal with the general case, we will unite *all* left-handed fermion fields (including antifermions, where that distinction is meaningful) in a single column χ. For instance, if ψ is a column containing all quark and lepton (as opposed to antiquark and antilepton) fields, then

$$\chi \equiv \begin{bmatrix} \frac{1}{2}(1+\gamma_5)\psi \\ \frac{1}{2}[\beta\mathscr{C}(1-\gamma_5)\psi]^* \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(1+\gamma_5)\psi \\ \frac{1}{2}(1+\gamma_5)\beta\mathscr{C}\psi^* \end{bmatrix}, \tag{22.3.1}$$

where $\mathscr{C}$ is the matrix defined in Section 5.4 by

$$\mathscr{C}\gamma_\mu^{\mathrm{T}}\mathscr{C}^{-1} = -\gamma_\mu ,$$

which is needed in order that all components of χ should belong to the same $(1/2,0)$ representation of the Lorentz group. Under an infinitesimal fermion number (e. g., baryon number or baryon number minus lepton number) conserving gauge transformation:

$$\delta\psi = i\theta_\alpha \left[\tfrac{1}{2}(1+\gamma_5)t_\alpha^L + \tfrac{1}{2}(1-\gamma_5)t_\alpha^R \right] \psi \tag{22.3.2}$$

this column undergoes the transformation

$$\delta\chi = i\epsilon_\alpha T_\alpha \chi , \tag{22.3.3}$$

where

$$T_\alpha = \begin{bmatrix} t_\alpha^L & 0 \\ 0 & -t_\alpha^{R*} \end{bmatrix} = \begin{bmatrix} t_\alpha^L & 0 \\ 0 & -(t_\alpha^R)^{\mathrm{T}} \end{bmatrix} . \tag{22.3.4}$$

We shall not limit ourselves here to theories that conserve fermion number, so the T_α will now be any Hermitian representations of the gauge algebra, not necessarily of the block-diagonal form (22.3.4). For the moment we will consider only massless fermions, returning to the effect of fermion masses a little later.

We shall be concerned here with the one-loop three-point function

$$\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z) \equiv \langle T\{j^\mu_\alpha(x),\, j^\nu_\beta(y),\, j^\rho_\gamma(z)\}\rangle_{\mathrm{VAC}} , \tag{22.3.5}$$

where j^μ_α is the fermionic current, calculated in terms of free fields:

$$j^\mu_\alpha = -i\bar{\chi}T_\alpha\gamma^\mu\chi . \tag{22.3.6}$$

The two Feynman diagrams of Figure 22.1 give

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

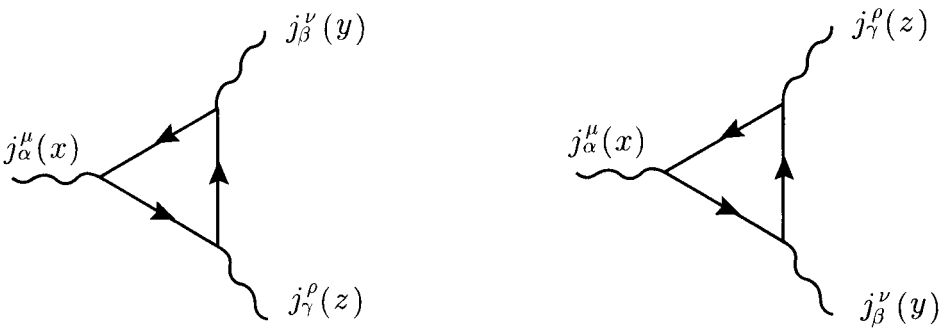

Figure 22.1. Two triangle diagrams for the anomaly in the current $j^{\mu}_{\alpha}(x)$. Solid lines are fermions; wavy lines indicate fictitious gauge fields, coupled to the currents.

$$\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z) = -i\mathrm{Tr}\,[S(x-y)T_\beta\gamma^\nu P_L S(y-z)T_\gamma\gamma^\rho P_L S(z-x)T_\alpha\gamma^\mu P_L] \\ -i\mathrm{Tr}\,[S(x-z)T_\gamma\gamma^\rho P_L S(z-y)T_\beta\gamma^\nu P_L S(y-x)T_\alpha\gamma^\mu P_L]\ , \tag{22.3.7}$$

where P_L is the projection operator on the left-handed fermion fields

$$P_L = \left(\frac{1+\gamma_5}{2}\right) \tag{22.3.8}$$

and $S(x)$ is the propagator of a massless fermion field:

$$S(x) = \frac{-i}{(2\pi)^4}\int d^4p\,\left(\frac{-i\not{p}}{p^2-i\epsilon}\right)e^{ip\cdot x}\ . \tag{22.3.9}$$

(For further comments on this use of the Feynman rules, see the end of this section.) Collecting factors, Eq. (22.3.7) now reads

$$\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z) = \frac{i}{(2\pi)^{12}}\int d^4k_1\,d^4k_2\,e^{-i(k_1+k_2)\cdot x}\,e^{ik_1\cdot y}\,e^{ik_2\cdot z}\int d^4p \\ \times\left\{\mathrm{tr}\left[\frac{\not{p}-\not{k}_1+\not{a}}{(p-k_1+a)^2-i\epsilon}\gamma^\nu\frac{\not{p}+\not{a}}{(p+a)^2-i\epsilon}\gamma^\rho\frac{\not{p}+\not{k}_2+\not{a}}{(p+k_2+a)^2-i\epsilon}\gamma^\mu\frac{1+\gamma_5}{2}\right]\right. \\ \times\mathrm{tr}\,[T_\beta T_\gamma T_\alpha] \\ +\mathrm{tr}\left[\frac{\not{p}-\not{k}_2+\not{b}}{(p-k_2+b)^2-i\epsilon}\gamma^\rho\frac{\not{p}+\not{b}}{(p+b)^2-i\epsilon}\gamma^\nu\frac{\not{p}+\not{k}_1+\not{b}}{(p+k_1+b)^2-i\epsilon}\gamma^\mu\frac{1+\gamma_5}{2}\right] \\ \left.\times\mathrm{tr}\,[T_\gamma T_\beta T_\alpha]\right\}, \tag{22.3.10}$$

where 'tr' here denotes a trace over either Dirac or group indices, depending on its argument. We have introduced the arbitrary constant four-vectors a and b because, although the expression (22.3.10) is convergent and hence independent of the labelling of the momenta carried

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

by internal lines, the evaluation of $\partial_\mu \Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}$ involves the manipulation of divergent integrals that do depend on this labelling. We will see that the freedom to choose a^μ and b^μ corresponds to a freedom to move the anomaly in these integrals from one current to another, but does not allow us to eliminate the anomaly altogether.

In taking the divergence of Eq. (22.3.10), we use the identities

$$\not{k}_1 + \not{k}_2 = (\not{p} + \not{k}_2 + \not{a}) - (\not{p} - \not{k}_1 + \not{a}) = (\not{p} + \not{k}_1 + \not{b}) - (\not{p} - \not{k}_2 + \not{b})$$

and find

$$\begin{aligned}
\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z) = \frac{1}{(2\pi)^{12}} \int d^4k_1\, d^4k_2\, e^{-i(k_1+k_2)\cdot x}\, e^{ik_1\cdot y}\, e^{ik_2\cdot z} \int d^4p \\
\times \Bigg\{ \mathrm{tr}\left[T_\beta\, T_\gamma\, T_\alpha\right]\ \mathrm{tr}\left[\frac{\not{p} - \not{k}_1 + \not{a}}{(p-k_1+a)^2 - i\epsilon}\gamma^\nu \frac{\not{p} + \not{a}}{(p+a)^2 - i\epsilon}\gamma^\rho \frac{1+\gamma_5}{2}\right] \\
-\mathrm{tr}\left[T_\beta\, T_\gamma\, T_\alpha\right]\ \mathrm{tr}\left[\frac{\not{p} + \not{a}}{(p+a)^2 - i\epsilon}\gamma^\rho \frac{\not{p} + \not{k}_2 + \not{a}}{(p+k_2+a)^2 - i\epsilon}\gamma^\nu \frac{1+\gamma_5}{2}\right] \\
+\mathrm{tr}\left[T_\gamma\, T_\beta\, T_\alpha\right]\ \mathrm{tr}\left[\frac{\not{p} - \not{k}_2 + \not{b}}{(p-k_2+b)^2 - i\epsilon}\gamma^\rho \frac{\not{p} + \not{b}}{(p+b)^2 - i\epsilon}\gamma^\nu \frac{1+\gamma_5}{2}\right] \\
-\mathrm{tr}\left[T_\gamma\, T_\beta\, T_\alpha\right]\ \mathrm{tr}\left[\frac{\not{p} + \not{b}}{(p+b)^2 - i\epsilon}\gamma^\nu \frac{\not{p} + \not{k}_1 + \not{b}}{(p+k_1+b)^2 - i\epsilon}\gamma^\rho \frac{1+\gamma_5}{2}\right] \Bigg\} .
\end{aligned} \tag{22.3.11}$$

At this point it is convenient to separate the three-point function into terms symmetric and antisymmetric in the group indices, by writing

$$\mathrm{tr}\left[T_\beta\, T_\gamma\, T_\alpha\right] = D_{\alpha\beta\gamma} + \tfrac{1}{2} i N\, C_{\alpha\beta\gamma}$$

and

$$\mathrm{tr}\left[T_\gamma\, T_\beta\, T_\alpha\right] = D_{\alpha\beta\gamma} - \tfrac{1}{2} i N\, C_{\alpha\beta\gamma} ,$$

where $D_{\alpha\beta\gamma}$ is the totally symmetric quantity

$$D_{\alpha\beta\gamma} = \tfrac{1}{2}\mathrm{tr}\left[\{T_\alpha,\, T_\beta\}\, T_\gamma\right] \tag{22.3.12}$$

and the coefficient of the structure constant $C_{\alpha\beta\gamma}$ is defined by

$$\mathrm{tr}[T_\alpha T_\beta] = N\, \delta_{\alpha\beta} .$$

The terms that are antisymmetric in group indices are in general non-zero, but they do not represent any breakdown of the symmetry. Just as in the derivation of the Ward identity in Section 10.4, in a formal calculation of the divergence of the matrix element (22.3.5) we encounter contributions from the time-derivatives of the theta-functions in the time-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

ordered product, equal to

$$\left[\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\right]_{\text{formal}} = -iC_{\alpha\beta\delta}\delta^4(x-y)\langle j^\nu_\delta(y)\, j^\rho_\gamma(z)\rangle_{\text{VAC}} - iC_{\alpha\gamma\delta}\delta^4(x-z)\langle j^\nu_\beta(y) j^\rho_\delta(z)\rangle_{\text{VAC}} . \quad (22.3.13)$$

It is not hard to see that the antisymmetric terms in Eq. (22.3.11) just reproduce Eq. (22.3.13). The anomaly is contained in the symmetric part of Eq. (22.3.11):

$$\begin{aligned}\left[\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\right]_{\text{anom}} &= \frac{1}{(2\pi)^{12}} D_{\alpha\beta\gamma}\int d^4k_1\, d^4k_2\, e^{-i(k_1+k_2)\cdot x}\, e^{ik_1\cdot y}\, e^{ik_2\cdot z} \\ \times \int d^4p \Bigg\{ &\operatorname{tr}\left[\frac{\not p - \not k_1 + \not a}{(p-k_1+a)^2 - i\epsilon}\gamma^\nu \frac{\not p + \not a}{(p+a)^2 - i\epsilon}\gamma^\rho \frac{1+\gamma_5}{2}\right] \\ &-\operatorname{tr}\left[\frac{\not p + \not a}{(p+a)^2 - i\epsilon}\gamma^\rho \frac{\not p + \not k_2 + \not a}{(p+k_2+a)^2 - i\epsilon}\gamma^\nu \frac{1+\gamma_5}{2}\right] \\ &+\operatorname{tr}\left[\frac{\not p - \not k_2 + \not b}{(p-k_2+b)^2 - i\epsilon}\gamma^\rho \frac{\not p + \not b}{(p+b)^2 - i\epsilon}\gamma^\nu \frac{1+\gamma_5}{2}\right] \\ &-\operatorname{tr}\left[\frac{\not p + \not b}{(p+b)^2 - i\epsilon}\gamma^\nu \frac{\not p + \not k_1 + \not b}{(p+k_1+b)^2 - i\epsilon}\gamma^\rho \frac{1+\gamma_5}{2}\right]\Bigg\} . \end{aligned} \quad (22.3.14)$$

Grouping together the first and fourth traces, and the third and second traces, this may be put in the form

$$\begin{aligned}\left[\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\right]_{\text{anom}} &= \frac{1}{(2\pi)^{12}} D_{\alpha\beta\gamma}\int d^4k_1\, d^4k_2\, e^{-i(k_1+k_2)\cdot x}\, e^{ik_1\cdot y}\, e^{ik_2\cdot z} \\ &\times \Bigg\{ \operatorname{tr}\left[\gamma^\kappa\gamma^\nu\gamma^\lambda\gamma^\rho \frac{1+\gamma_5}{2}\right] I_{\kappa\lambda}(a-b-k_1, b, b+k_1) \\ &+ \operatorname{tr}\left[\gamma^\kappa\gamma^\rho\gamma^\lambda\gamma^\nu \frac{1+\gamma_5}{2}\right] I_{\kappa\lambda}(b-a-k_2, a, a+k_2)\Bigg\} , \end{aligned} \quad (22.3.15)$$

where

$$I_{\kappa,\lambda}(k,c,d) \equiv \int d^4p \left[f_{\kappa\lambda}(p+k,c,d) - f_{\kappa\lambda}(p,c,d)\right] , \quad (22.3.16)$$

$$f_{\kappa\lambda}(p,c,d) \equiv \frac{(p+c)_\kappa (p+d)_\lambda}{[(p+c)^2 - i\epsilon][(p+d)^2 - i\epsilon]} . \quad (22.3.17)$$

To evaluate these integrals consider the expansion of the function $f_{\kappa\lambda}(p+k,c,d)$ in powers of k:

$$f_{\kappa\lambda}(p+k,c,d) = \sum_{n=0}^{\infty} \frac{1}{n!} k^{\mu_1}\cdots k^{\mu_n} \frac{\partial^n f_{\kappa\lambda}(p,c,d)}{\partial p^{\mu_1}\cdots\partial p^{\mu_n}} .$$

The zeroth-order term $f_{\kappa\lambda}(p,c,d)$ clearly cancels in Eq. (22.3.16). All other terms in Eq. (22.3.16) are integrals of derivatives with respect to p, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

hence after Wick rotation may be written as surface integrals over a large three-sphere, say of radius P. The nth derivative of f then yields the surface integral of a function that behaves as $P^{-2-(n-1)}$, while the area of the three-sphere of radius P goes as P^3, so the only terms that can contribute for $P \to \infty$ are those with $n = 1$ and $n = 2$:

$$I_{\kappa\lambda}(k,c,d) = k^\mu \int d^4p \frac{\partial f_{\kappa\lambda}(p,c,d)}{\partial p^\mu} + \tfrac{1}{2}k^\mu k^\nu \int d^4p \, \frac{\partial^2 f_{\kappa\lambda}(p,c,d)}{\partial p^\mu \partial p^\nu} \,. \quad (22.3.18)$$

A straightforward calculation then gives

$$I_{\kappa\lambda}(k,c,d) = \tfrac{1}{6}i\pi^2\Big[2k_\lambda c_\kappa + 2k_\kappa d_\lambda - k_\lambda d_\kappa - k_\kappa c_\lambda - \eta_{\kappa\lambda}\, k\cdot(k+c+d)\Big] \,. \quad (22.3.19)$$

We must now separately consider the terms in the traces in Eq. (22.3.15) arising from the 1 and the γ_5 in the projection matrix $\frac{1}{2}(1+\gamma_5)$. The terms arising from the 1 involve $\mathrm{tr}[\gamma^\kappa\gamma^\nu\gamma^\lambda\gamma^\rho]$, which is symmetric in κ and λ, and also in ν and ρ, so the integrals here appear in the combination

$$\begin{aligned} &I_{\kappa\lambda}(a-b-k_1,b,b+k_1) + I_{\lambda\kappa}(a-b-k_1,b,b+k_1) \\ &+I_{\kappa\lambda}(b-a-k_2,a,a+k_2) + I_{\lambda\kappa}(b-a-k_2,a,a+k_2) \,. \end{aligned}$$

Using Eq. (22.3.19), it is not difficult to see that this vanishes if and only if we choose the arbitrary constant vectors so that

$$a = -b \,. \quad (22.3.20)$$

Furthermore, this is a choice that avoids the non-chiral anomaly for all three currents, because in $(\partial/\partial y^\nu)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$, a and b would be replaced with $a' = k_2 + a$ and $b' = -k_2 + b$, while in $(\partial/\partial z^\rho)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$, a and b would be replaced with $a'' = k_1 + a$ and $b'' = -k_1 + b$, so that taking $a = -b$ also insures that $a' = -b'$ and $a'' = -b''$.

We are left with the term in the trace involving γ_5. This is totally antisymmetric

$$\mathrm{tr}\left[\gamma^\kappa\gamma^\nu\gamma^\lambda\gamma^\rho\gamma_5\right] = -4i\epsilon^{\kappa\nu\lambda\rho} \,, \quad (22.3.21)$$

where $\epsilon^{\kappa\nu\lambda\rho}$ is the totally antisymmetric tensor with $\epsilon^{0123} = 1$. Using this in Eq. (22.3.15) with $b = -a$ gives

$$\begin{aligned} \left[\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\right]_{\text{anom}} &= \frac{2}{(2\pi)^{12}} D_{\alpha\beta\gamma} \int d^4k_1 \, d^4k_2 \, e^{-i(k_1+k_2)\cdot x} \\ &\quad \times e^{ik_1\cdot y}\, e^{ik_2\cdot z}\, \pi^2\epsilon^{\kappa\nu\lambda\rho} a_\kappa \, (k_1+k_2)_\lambda \,. \quad (22.3.22) \end{aligned}$$

We could eliminate the anomaly (22.3.22) in the current $J^\mu_\alpha(x)$ by taking $a \propto k_1 + k_2$. Although this is possible, it would not eliminate the anomaly altogether; it would appear in $(\partial/\partial y^\nu)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$ or $(\partial/\partial z^\rho)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$. The symmetry of the problem indicates that the anomaly will be absent in $(\partial/\partial y^\nu)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$ if and only if $(k_2 + a) - (-k_2 + b) \propto k_1$, or in other

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

words $a+k_2 \propto k_1$, and will be absent in $(\partial/\partial z^\rho)\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$ if and only if $(-k_1+a)-(k_1+b)\propto k_2$, or in other words $a-k_1\propto k_2$. It is possible to choose a^μ to satisfy any two of these conditions, so that the anomaly can be removed from any two of the currents, but for non-parallel k_1 and k_2 it is evidently impossible to simultaneously satisfy all three conditions — $a\propto k_1+k_2$ and $a+k_2\propto k_1$ and $a-k_1\propto k_2$. (For instance, the first two conditions would imply that $a=-k_1-k_2$, in contradiction with the third condition.) We therefore conclude that although we have some freedom in deciding which current exhibits the anomaly, *the non-vanishing of $D_{\alpha\beta\gamma}$ shows definitely that there is an anomaly in at least one of the currents $J^\mu_\alpha(x)$, $J^\nu_\beta(y)$, or $J^\rho_\gamma(z)$*. This is one of the chief results of these calculations, and will be exploited in the next section as a source of constraints on the matter content of gauge theories.

We have seen that the evaluation of the anomalies depends on the choice we make of the shift vector a^μ. Unfortunately, there is no one choice that is uniformly satisfactory, so we have to choose a^μ in accordance with the special features of the problem at hand.

In one class of problems of great importance, $J^\mu_\alpha(x)$ is the current of a global symmetry, while $J^\nu_\beta(y)$ and $J^\rho_\gamma(z)$ are currents of gauge symmetries, that is, currents to which gauge fields are coupled. (We dealt with such a problem in the previous section.) In such cases, we *must* choose a^μ so that the anomaly is solely in $J^\mu_\alpha(x)$, not in $J^\nu_\beta(y)$ or $J^\rho_\gamma(z)$. As we have seen, this requires that $a+k_2\propto k_1$ and $a-k_1\propto k_2$, which leads to the unique result that

$$a=k_1-k_2\,. \tag{22.3.23}$$

Adopting this value of a^μ, the anomaly (22.3.22) is

$$\begin{aligned}\left[\frac{\partial}{\partial x^\mu}\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\right]_{\text{anom}} &= \frac{1}{(2\pi)^{12}}D_{\alpha\beta\gamma}\int d^4k_1\,d^4k_2\,e^{-i(k_1+k_2)\cdot x}\,e^{ik_1\cdot y}\,e^{ik_2\cdot z}\\ &\quad\times 4\pi^2\epsilon^{\kappa\nu\lambda\rho}k_{1\kappa}k_{2\lambda}\\ &= -\frac{1}{4\pi^2}D_{\alpha\beta\gamma}\epsilon^{\kappa\nu\lambda\rho}\frac{\partial\delta^4(y-x)}{\partial y^\kappa}\frac{\partial\delta^4(z-x)}{\partial z^\lambda}\,.\end{aligned} \tag{22.3.24}$$

We note in passing that a result like this can only arise from a theory that involves massless particles.[7a] Otherwise, we would expect the Fourier transform of $\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)$ to have a power series expansion around zero momentum. The only terms in such an expansion that could possibly lead to a current divergence of the form (22.3.24) are pseudotensors of first order in momenta

$$\begin{aligned}\left[\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(k_1,k_2)\right]_{\text{anom}} &\equiv \int d^4y\,d^4z\,e^{-ik_1\cdot y}e^{-ik_2\cdot z}\,\left[\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(0,y,z)\right]_{\text{anom}}\\ &= \epsilon^{\mu\nu\rho\sigma}\left[A_{\alpha\beta\gamma}k_{1\sigma}+B_{\alpha\beta\gamma}k_{2\sigma}\right],\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

where $A_{\alpha\beta\gamma}$ and $B_{\alpha\beta\gamma}$ are constants. According to Eq. (22.3.24),

$$(k_1+k_2)_\mu \left[\Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(k_1,k_2)\right]_{\text{anom}} = \frac{i}{4\pi^2} D_{\alpha\beta\gamma}\epsilon^{\kappa\nu\lambda\rho}k_{1\kappa}k_{2\lambda}$$

so

$$A_{\alpha\beta\gamma} - B_{\alpha\beta\gamma} = \frac{i}{4\pi^2} D_{\alpha\beta\gamma} \ .$$

But the symmetry of the amplitude among the three currents requires that

$$A_{\alpha\beta\gamma}k_{1\sigma} + B_{\alpha\beta\gamma}k_{2\sigma} = -A_{\alpha\gamma\beta}k_{2\sigma} - B_{\alpha\gamma\beta}k_{1\sigma} = -A_{\gamma\alpha\beta}(-k_{1\sigma}-k_{2\sigma}) - B_{\gamma\alpha\beta}k_{2\sigma}$$

so, equating the coefficients of $k_{1\sigma}$ and $k_{2\sigma}$,

$$A_{\alpha\beta\gamma} = -B_{\alpha\gamma\beta} = A_{\gamma\alpha\beta} \ , \qquad B_{\alpha\beta\gamma} = -A_{\alpha\gamma\beta} = A_{\gamma\alpha\beta} - B_{\gamma\alpha\beta} \ .$$

Taking the difference of these equations gives

$$A_{\alpha\beta\gamma} - B_{\alpha\beta\gamma} = A_{\alpha\gamma\beta} - B_{\alpha\gamma\beta} = B_{\gamma\alpha\beta} \ .$$

Since this is proportional to $D_{\alpha\beta\gamma}$ it must be totally symmetric, so $B_{\alpha\beta\gamma}$ and hence also $A_{\alpha\beta\gamma}$ are totally symmetric. But then the conditions for the symmetry of the three-point function read $A = -B = B - A$, which implies that $A_{\alpha\beta\gamma} = B_{\alpha\beta\gamma} = 0$, in contradiction to Eq. (22.3.24). By the same reasoning, it is not possible to cancel the anomaly by adding a local term to the interaction whose contribution to the divergence of the current J^μ_α cancels that given by Eq. (22.3.24).

Returning now to Eq. (22.3.24), we can express this result in terms of the vacuum expectation value of the current J^μ_α in the presence of gauge fields coupled to the currents J^ν_β and J^ρ_γ. The triangle graphs make a contribution to the current in the presence of the gauge fields

$$\langle J^\mu_\alpha(x)\rangle_\triangle = -\tfrac{1}{2}\int d^4y\, d^4z\ \Gamma^{\mu\nu\rho}_{\alpha\beta\gamma}(x,y,z)\, A^\beta_\nu(y)\, A^\gamma_\rho(z) \ . \tag{22.3.25}$$

Using Eq. (22.3.24), this has the anomalous divergence

$$[\langle \partial_\mu J^\mu_\alpha(x)\rangle_\triangle]_{\text{anom}} = -\frac{1}{8\pi^2} D_{\alpha\beta\gamma}\, \epsilon^{\kappa\nu\lambda\rho}\, \partial_\kappa A^\beta_\nu(x)\, \partial_\lambda A^\gamma_\rho(x) \ . \tag{22.3.26}$$

There are additional diagrams, shown in Figure 22.2, that also exhibit anomalies. Gauge invariance requires that the diagrams of Figure 22.2 should add up to give the gauge-invariant result

$$[\langle \partial_\mu J^\mu_\alpha(x)\rangle]_{\text{anom}} = -\frac{1}{32\pi^2} D_{\alpha\beta\gamma}\, \epsilon^{\kappa\nu\lambda\rho}\, F^\beta_{\kappa\nu}(x)\, F^\gamma_{\lambda\rho}(x) \ . \tag{22.3.27}$$

As a check, consider a fermion-conserving theory, with generators T_α of the form (22.3.4). The constant $D_{\alpha\beta\gamma}$ in the anomaly (22.3.26) is here given by

$$D_{\alpha\beta\gamma} = \tfrac{1}{2}\text{tr}[\{t^L_\alpha, t^L_\beta\}t^L_\gamma] - \tfrac{1}{2}\text{tr}[\{t^R_\alpha, t^R_\beta\}t^R_\gamma] \ . \tag{22.3.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

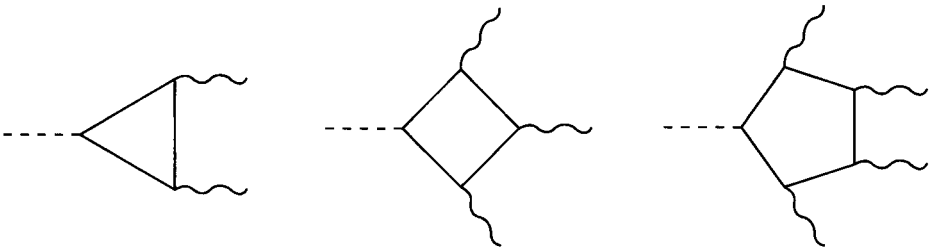

Figure 22.2. One-loop diagrams for the anomaly in a current indicated by the dashed line. Solid lines are fermions; wavy lines are gauge fields with which they interact.

More specifically, in Section 22.2 we calculated the divergence of an axial-vector current J_5^μ, with $t^L = -t^R \equiv t$, due to gauge-field interactions with vector currents J_β^ν and J_γ^ρ (called J_α^ν and J_β^ρ in Section 22.2) with $t_\beta^L = t_\beta^R \equiv t_\beta$, and likewise for t_γ. Hence in this case $D_{\alpha\beta\gamma}$ is replaced with $\mathrm{tr}[\{t, t_\beta\}t_\gamma] = \mathrm{tr}[\{t_\beta, t_\gamma\}t]$, and Eq. (22.3.27) becomes

$$[\langle \partial_\mu J_5^\mu(x)\rangle]_{\text{anom}} = -\frac{1}{32\pi^2}\mathrm{tr}[\{t_\beta, t_\gamma\}t]\,\epsilon^{\kappa\nu\lambda\rho}\,F_{\kappa\nu}^\beta(x)\,F_{\lambda\rho}^\gamma(x)\,, \qquad (22.3.29)$$

in agreement with Eq. (22.2.26).

Where no gauge fields are coupled to any of the currents $J_\alpha^\mu(x)$, $J_\beta^\nu(y)$, or $J_\gamma^\rho(z)$, the choice of the shift vector a^μ is a matter of convenience. When some but not all currents are associated with symmetries that are spontaneously broken, we keep the unbroken symmetries manifest if we choose a^μ so that there are no anomalies in the currents corresponding to the unbroken symmetries. (This will be important in Section 22.7.) For instance, in quantum chromodynamics and similar theories, where the generators T_α of global symmetries like chiral $SU(3) \times SU(3)$ take the form (22.3.4), all currents are either vector, with $t_\alpha^R = t_\alpha^L$, corresponding to unbroken symmetries, or axial-vector, with $t_\alpha^R = -t_\alpha^L$, corresponding to broken symmetries. We see from Eq. (22.3.28) that in this case the only triangle graphs with anomalies are those with one axial-vector and two vector currents, or three axial-vector currents. In the case of one axial-vector and two vector currents, we choose a^μ so that there is no anomaly that interferes with the conservation of the vector currents. Thus, if $J_\alpha^\mu(x)$ is the axial-vector current and $J_\beta^\nu(y)$ and $J_\gamma^\rho(z)$ the two vector currents, then as in Eq. (22.3.23) we must take $a^\mu = k_1^\mu - k_2^\mu$, so that the anomaly is given by Eq. (22.3.24). On the other hand, for three axial-vector currents, there is no reason to require that any one of them should be free of anomalies. Instead, it is natural to give a^μ a value that respects the symmetry among the three currents. Guided by Lorentz invariance, suppose we try $a = \alpha k_1 + \beta k_2$, where α and β are constants.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Then symmetry would require that the momentum of *each* internal line should be p plus α times the momentum flowing out at the end of the line plus β times the momentum flowing out at the beginning of the line. That is, if we require that $a = \alpha k_1 + \beta k_2$, then symmetry requires also that $a - k_1 = -\alpha(k_1 + k_2) + \beta k_1$ and $a + k_2 = \alpha k_2 - \beta(k_1 + k_2)$. These three relations are satisfied for non-parallel k_1 and k_2 if and only if $\alpha = -\beta = 1/3$, so that

$$a = \tfrac{1}{3}(k_1 - k_2) \ . \tag{22.3.30}$$

Using this in Eq. (22.3.22) and comparing with Eq. (22.3.23), we see that the anomaly in the axial-vector current in a Feynman amplitude for three axial-vector currents is one-third what it would be for one axial-vector and two vector currents.

The divergence of the current contains additional anomalies from the graphs of Figure 22.2. Gauge invariance is no guide here, because even the triangle graph does not yield conserved currents. The total anomaly has been calculated by Bardeen[8] for the chiral $SU(3) \times SU(3)$ symmetry of the strong interactions, with momentum shift vectors chosen so that vector currents are conserved and graphs with all axial currents are symmetric among these currents. Although this $SU(3) \times SU(3)$ symmetry (which acts on quark flavors rather than colors) is not gauged in quantum chromodynamics, it is convenient to express the anomaly as a failure of gauge invariance in a functional $\Gamma[V, A]$ of fictitious weakly coupled gauge fields: an octet of vector fields $V_a^\mu(x)$ and an octet of axial-vector fields $A_a^\mu(x)$. We also introduce infinitesimal gauge-transformation operators*

$$i\mathscr{Y}_a(x) = -\frac{\partial}{\partial x^\mu}\frac{\delta}{\delta V_{a\mu}(x)} - f_{abc}V_{b\mu}(x)\frac{\delta}{\delta V_{c\mu}(x)} - f_{abc}A_{b\mu}(x)\frac{\delta}{\delta A_{c\mu}(x)} \tag{22.3.31}$$

and

$$i\mathscr{X}_a(x) = -\frac{\partial}{\partial x^\mu}\frac{\delta}{\delta A_{a\mu}(x)} - f_{abc}V_{b\mu}(x)\frac{\delta}{\delta A_{c\mu}(x)} - f_{abc}A_{b\mu}(x)\frac{\delta}{\delta V_{c\mu}(x)} \ , \tag{22.3.32}$$

where f_{abc} are the $SU(3)$ structure constants. As mentioned above, the labelling of internal line momenta is chosen so that the vector current is not anomalous:

$$\mathscr{Y}_a \Gamma[V, A] = 0 \ , \tag{22.3.33}$$

* The operators $\mathscr{Y}_a$ and $\mathscr{X}_a$ here are what in Ref. 8 were called X_a and Y_a. This switch in notation is made to maintain consistency with the notation of Chapter 19, where broken symmetry generators are consistently labelled X_a.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

but then an anomaly appears in the axial-vector currents

$$\mathscr{X}_a \Gamma[V,A] = \frac{i}{32\pi^2}\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\left\{\lambda_a\left[V_{\mu\nu}V_{\rho\sigma} + \tfrac{1}{3}A_{\mu\nu}A_{\rho\sigma} - \tfrac{32}{3}A_\mu A_\nu A_\rho A_\sigma + \tfrac{8}{3}i(A_\mu A_\nu V_{\rho\sigma} + A_\mu V_{\rho\sigma}A_\nu + V_{\rho\sigma}A_\mu A_\nu)\right]\right\}, \tag{22.3.34}$$

where λ_a are the $SU(3)$ matrices given by Eq. (19.7.2), and

$$V_\mu \equiv \tfrac{1}{2}\lambda_a V_{\mu a}\,, \qquad A_\mu \equiv \tfrac{1}{2}\lambda_a A_{\mu a}\,, \tag{22.3.35}$$

$$V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu - i[V_\mu, V_\nu] - i[A_\mu, A_\nu]\,, \tag{22.3.36}$$

$$A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[V_\mu, A_\nu] - i[A_\mu, V_\nu]\,. \tag{22.3.37}$$

The factor 1/3 accompanying the second term on the right-hand side in Eq. (22.3.34) has already been explained as a consequence of the different choice of a^μ in the AVV and AAA graphs. In Section 22.6 we will describe consistency conditions that allow the cubic and quartic terms in Eq. (22.3.34) to be calculated from the quadratic terms.

For anomalies involving symmetries that are *all* spontaneously broken, there is no reason to choose a^μ in a way that distinguishes among the different currents, and instead it is natural to label internal fermion lines in a way that is symmetric among the attached gauge boson lines. As we have seen, this means that the triangle graph is to be calculated with the choice $a = \frac{1}{3}(k_1 - k_2)$. This gives a triangle anomaly one-third of the value given by (22.3.26). When square and pentagon graphs are included, this result becomes[8a]

$$[\langle D_\mu J_\alpha^\mu\rangle]_{\text{anom}} = -\frac{1}{24\pi^2}\epsilon^{\kappa\nu\lambda\rho}\mathrm{Tr}\left\{T_\alpha\left[\partial_\kappa A_\nu\,\partial_\lambda A_\rho - \tfrac{1}{2}i\partial_\kappa A_\nu\,A_\lambda A_\rho + \tfrac{1}{2}iA_\kappa\,\partial_\nu A_\lambda\,A_\rho - \tfrac{1}{2}iA_\kappa A_\nu \partial_\lambda A_\rho\right]\right\}, \tag{22.3.38}$$

where $A^\mu \equiv A_\alpha^\mu T_\alpha$. We will not derive this result here, because we have already seen that in such cases the terms quadratic in A_μ are given by one-third the corresponding terms in Eq. (22.3.26), and in Section 22.6 we shall be able to use consistency conditions to derive the remaining terms in Eq. (22.3.38) from the quadratic terms.

Now we must consider possible corrections to these results. A careful derivation of the anomaly to all orders of perturbation theory was given by Adler and Bardeen;[9] what follows is intended to give only the gist of their analysis.

The argument that led to Eq. (22.3.22) may be repeated in any order of perturbation theory, and shows that the anomaly in general arises from

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

momentum-space integrals that can be reexpressed as surface terms. In consequence, as we saw here in Eq. (22.3.18), the only graphs for the current divergence that contribute to the anomaly are those for which the integral over the momentum circulating in the fermion loop has dimensionality (in powers of momentum) zero or greater. Interactions of the fermions in the loop with virtual gauge bosons would reduce the dimensionality of the integral over the fermion loop momentum enough to eliminate the anomaly, so the anomaly receives no contribution from such radiative corrections. (It is true that the integral over the momenta of the virtual gauge bosons as well as the fermion loop would have non-negative dimensionality, but the gauge boson propagators can be regulated, as the fermion propagator cannot, without interfering with the chiral symmetry in question.) The anomaly is affected by interactions of the gauge bosons attached to the fermion loop with other gauge bosons and fermion loops, but these just serve to renormalize operators like $\epsilon^{\kappa\nu\lambda\rho} F^{\beta}_{\kappa\nu}(x) F^{\gamma}_{\lambda\rho}(x)$. By the same argument, any fermion mass that respects the symmetries in question (if this were possible) would not change the anomaly, because extracting factors of this mass would lower the dimensionality of the momentum-space integral.

The last remark of the previous paragraph raises the question of whether we can calculate the anomaly without knowing all the fermions in a theory, heavy fermions as well as light or massless ones. Yes, we can; we shall now show that *no fermion that is allowed by a given symmetry to have a mass can contribute to the anomaly for that symmetry*. In the general class of theories considered here, a mass term in the Lagrangian density would take the form

$$\mathscr{L}_{\text{mass}} = -\sum_{nn'\sigma\sigma'} \chi_{\sigma n} \epsilon_{\sigma\sigma'} M_{nn'} \chi_{\sigma' n'} + \text{H.c.}\,, \tag{22.3.39}$$

where σ is the two-component spinor index of the $(\frac{1}{2},0)$ representation of the Lorentz group, $\epsilon_{\sigma\sigma'}$ is the antisymmetric matrix with $\epsilon_{\frac{1}{2},-\frac{1}{2}} = +1$, needed for Lorentz invariance, and M is a symmetric mass matrix.** Now, in order for $\mathscr{L}_{\text{mass}}$ to respect gauge invariance, the mass matrix must

** In fermion-number-conserving theories where χ has the form (22.3.1), M is related to the usual mass matrix m by

$$M = \frac{1}{2}\begin{pmatrix} 0 & m \\ m^{\mathrm{T}} & 0 \end{pmatrix}.$$

With all inversion phases equal to unity, parity, charge conjugation, and time-reversal invariance respectively have the further consequences that m is Hermitian, symmetric, or real.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

satisfy

$$-T_\alpha^{\mathrm{T}} M = M T_\alpha \,. \tag{22.3.40}$$

The index n may be replaced with an index r labelling different irreducible representations of the gauge group, and an index s labelling components within each irreducible representation, so that

$$(T_\alpha)_{rs,r's'} = \delta_{rr'}(T_\alpha^{(r)})_{s,s'} \,, \tag{22.3.41}$$

and we write

$$M_{rs,r's'} = (M^{(rr')})_{ss'} \,. \tag{22.3.42}$$

Eq. (22.3.40) then becomes

$$-T_\alpha^{(r)\mathrm{T}} M^{(r,r')} = M^{(r,r')} T_\alpha^{(r')} \,. \tag{22.3.43}$$

(This is not summed over r or r'.) Schur's lemma[10] tells us that whenever the matrices of a pair of irreducible representations are related by such an equation, the matrix that relates them is either zero or non-singular. (See Section 5.5.) Thus, either (i) $M^{(r,r')} = 0$, or else (ii) $-T_\alpha^{(r)\mathrm{T}}$ and $T_\alpha^{(r')}$ are related by a similarity transformation, and likewise for T_β and T_γ. In the latter case the contributions to the anomaly constant (22.3.12) from fermions belonging to the individual irreducible representations r and r' are related by

$$D_{\alpha\beta\gamma}^{(r)} = -D_{\alpha\beta\gamma}^{(r')} \,, \tag{22.3.44}$$

so the anomaly either vanishes (for $r = r'$) or cancels between the two representations (for $r \neq r'$).The anomalies in a given set of symmetries therefore are unaffected by the possible presence of fermions with a mass allowed by these symmetries.

* * *

There is a fine point in the use of the Feynman rules to calculate the three-point function (22.3.7), to which we now return. In using the conventional fermion propagator (22.3.9), we have in effect doubled the number of fermion fields; in addition to the purely left-handed fields $\chi(x)$ given by Eq. (22.3.1), the propagator (22.3.9) includes right-handed modes (unrelated to the fields $(1-\gamma_5)\psi$ of fermion-number-conserving theories) that do not interact with the gauge fields. Combining these non-interacting right-handed fields with the interacting left-handed fields in a single spinor Ψ, the fermion Lagrangian density is $-\bar{\Psi}\not{D}\Psi$, where $\not{D}$ is now

$$\not{D} = \not{\partial} - i\not{A}_\alpha T_\alpha \left(\frac{1+\gamma_5}{2}\right) \,. \tag{22.3.45}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

The one-loop vacuum functional for spin $\frac{1}{2}$ fields interacting with real or fictitious gauge fields is just Det $\not{D}$. The failure of gauge invariance in this determinant can be blamed on the fact that the operator (22.3.45) is the sum of two terms

$$\not{D} = \not{\phi}\left(\frac{1-\gamma_5}{2}\right) + \not{D}\left(\frac{1+\gamma_5}{2}\right) \tag{22.3.46}$$

of which only the second is gauge-invariant.

There is another way of looking at anomalies,[10a] which provides a framework for calculations of a large class of anomalies, of coordinate as well as gauge symmetries in spaces of any dimensionality. Instead of working with the operator (22.3.45), which has a well-defined determinant, one deals instead with just the second term in Eq. (22.3.46)

$$\not{D}_L \equiv \not{D}\left(\frac{1+\gamma_5}{2}\right) . \tag{22.3.47}$$

This is perfectly gauge-invariant, but does not have a well-defined determinant, because this operator does not map the space of fermion fields of one handedness into itself, but rather into the space of fermion fields of the other handedness. One can try to define a gauge-invariant vacuum functional Det $\not{D}_L$ by writing differential equations for Det $\not{D}_L$ in the space of gauge fields, modulo gauge transformations, but then one may encounter obstructions. There are local obstructions due to violations of the necessary integrability conditions, which correspond to the anomalies we have been discussing. Even where these local obstructions are absent, in cases where the infinite-dimensional space of gauge-field configurations, modulo gauge transformations, is not simply connected, topological obstructions can prevent the definition of single-valued functionals on this space. Such a global anomaly was found by Witten[10b] for the gauge group $SU(2)$ (which is free of local anomalies) with an odd number of massless left-handed fermions in $SU(2)$ doublets.

22.4 Anomaly-Free Gauge Theories

We have calculated the effect of anomalies on the conservation of a general current J^μ_α. Where this current is itself coupled to a gauge field, gauge invariance requires that the anomaly be absent. We saw in the previous section that the anomaly is proportional to the completely symmetric constant factor $D_{\alpha\beta\gamma}$ defined by Eq. (22.3.12), so for gauge currents we must have[11]

$$D_{\alpha\beta\gamma} \equiv \tfrac{1}{2}\mathrm{tr}\left[\{T_\alpha, T_\beta\}\, T_\gamma\right] = 0 , \tag{22.4.1}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

where T_α is the representation of the gauge algebra on the set of all left-handed fermion and antifermion fields, and 'tr' again denotes a sum over these fermion and antifermion species. This condition may be satisfied for any gauge group if the fermion fields furnish a suitable reducible or irreducible representation of the group. In addition, there are some gauge groups for which Eq. (22.4.1) is satisfied for fermions in *any* representation of the group.[12] (The Batalin–Vilkovisky formalism is used in Section 22.6 to give a purely algebraic proof that for such gauge groups the anomaly is absent to all orders of perturbation theory.)

The condition (22.4.1) is obviously satisfied if the left-handed fermion (and antifermion) fields furnish a representation of the gauge algebra that is equivalent to its complex conjugate, in the sense that

$$(iT_\alpha)^* = S\,(iT_\alpha)\,S^{-1}$$

or equivalently (since we always take T_α Hermitian)

$$T_\alpha^T = -S\,T_\alpha S^{-1}\,. \tag{22.4.2}$$

(Inserting Eq. (22.4.2) in Eq. (22.3.12) gives $D_{\alpha\beta\gamma} = -D_{\alpha\beta\gamma}$.) Such a representation may be either *real*, in which case it is possible by a similarity transformation $T'_\alpha = R\,T_\alpha\,R^{-1}$ to convert the representation to a form in which T'_α is imaginary and antisymmetric, or *pseudoreal*, in which case this is impossible. (For instance, the three-dimensional irreducible representation of $SU(2)$ is real, while the two-dimensional representation is pseudoreal.) There is no anomaly for gauge algebras that have only real or pseudoreal representations, namely[13] $SO(2n+1)$ (including $SU(2) \equiv SO(3)$), $SO(4n)$ for $n \geq 2$, $USp(2n)$ for $n \geq 3$, G_2, F_4, E_7, and E_8, and all of their direct sums. A few other algebras also have only representations for which $D_{\alpha\beta\gamma}$ vanishes, even though some representations are neither real nor pseudoreal.[12] These are $SO(4n+2)$ (except for $SO(2) \equiv U(1)$ and $SO(6) \equiv SU(4)$) and E_6, and their direct sums with each other and the above algebras. Anomalies are thus only possible for gauge algebras that include $SU(n)$ (for $n \geq 3$) or $U(1)$ factors. As it happens, these are among the most important gauge algebras in today's physics. The standard model is based on the gauge group $SU(3) \times SU(2) \times U(1)$, so we must rely on cancellations among the quarks and leptons to make the theory free of anomalies.

Table 22.1 gives a classification of the left-handed spinor fields of the first generation of the standard model according to the representations they furnish of the color $SU(3)$ group and the electroweak $SU(2)$ group, and the value of the $U(1)$ quantum number $y/g' = t_3/g - q/e$.

We can now check whether $D_{\alpha\beta\gamma}$ vanishes when T_α, T_β, and T_γ run over all of the generators of $SU(3) \times SU(2) \times U(1)$. We need only consider those combinations of generators for which the product of T_α, T_β, and T_γ

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Table 22.1. First-generation left-handed fermion and antifermion fields of the standard model.

Fermions	$SU(3)$	$SU(2)$	$U(1)$ $[y/g']$
$\begin{pmatrix} u \\ d \end{pmatrix}_L$	3	2	$-1/6$
u_R^*	$\bar{3}$	1	$+2/3$
d_R^*	$\bar{3}$	1	$-1/3$
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	1	2	$1/2$
e_R^*	1	1	-1

is neutral under $SU(3) \times SU(2) \times U(1)$, since $D_{\alpha\beta\gamma}$ obviously vanishes for all the others. We can make invariants out of either zero, two, or three $SU(3)$ generators (because 8×8 and $8 \times 8 \times 8$ both contain singlets), zero, two, or three $SU(2)$ generators, and any number of $U(1)$ generators, so we only need to check the following cases:

***SU*(3)–*SU*(3)–*SU*(3)**: Here $D_{\alpha\beta\gamma}$ vanishes because the left-handed fermions furnish a representation $3 + 3 + \bar{3} + \bar{3} + 1 + 1 + 1$ of $SU(3)$, which is real.

***SU*(3)–*SU*(3)–*U*(1)**: Here the anomaly is proportional to

$$\sum_{3,\bar{3}} y/g' = -\frac{1}{6} - \frac{1}{6} + \frac{2}{3} - \frac{1}{3} = 0 .$$

***SU*(2)–*SU*(2)–*SU*(2)**: There is no anomaly here because $SU(2)$ only has real or pseudoreal representations.

***SU*(2)–*SU*(2)–*U*(1)**: Here the anomaly is proportional to

$$\sum_{\text{doublets}} y/g' = 3\left(-\frac{1}{6}\right) + \frac{1}{2} = 0 .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

$U(1)$–$U(1)$–$U(1)$: Here the anomaly is proportional to

$$\sum(y/g')^3 = 6\left(-\frac{1}{6}\right)^3 + 3\left(\frac{2}{3}\right)^3 + 3\left(-\frac{1}{3}\right)^3 + 2\left(\frac{1}{2}\right)^3 + (-1)^3 = 0\,.$$

We see that all anomalies cancel for the gauge symmetries of the standard model.[13a] This result can be neatly understood by noting that $SU(3) \times SU(2) \times U(1)$ may be embedded in $SO(10)$.[14] All of the representations of $SO(10)$ are anomaly-free, so the same property is inherited by any reducible representation of $SU(3) \times SU(2) \times U(1)$ that furnishes a complete representation of $SO(10)$. As it happens, the left-handed fields of a single generation of quarks, leptons, antiquarks, and antileptons plus one additional $(SU(3) \times SU(2) \times U(1))$-singlet forms a complete 16-dimensional representation of $SO(10)$ (the fundamental spinor representation), so for this set of left-handed fermions there are no $SU(3) \times SU(2) \times U(1)$ anomalies. The singlet would not contribute to such anomalies anyway, so there are no anomalies in the gauge symmetries of the standard model.

There is one more anomaly that needs to be evaluated. All species of fermions interact with gravitation in the same way. Calculation of the fermion loop graph for the expectation value of the current $\bar{\chi}T\gamma^\mu\chi$ (where χ is a column of all left-handed fermion and antifermion fields) in the presence of an external gravitational field yields an anomaly[15] $\partial_\mu(\bar{\chi}T\gamma^\mu\chi)$ proportional to

$$\mathrm{tr}\{T\}\epsilon^{\mu\nu\rho\sigma}R_{\mu\nu\kappa\lambda}R_{\rho\sigma}{}^{\kappa\lambda}\,.$$

In particular, to avoid a gravitational violation of a gauge symmetry like (22.3.3), the generators must satisfy

$$\mathrm{tr}\{T_\alpha\} = 0\,. \tag{22.4.3}$$

Like the pure gauge anomaly, this vanishes for gauge generators satisfying Eq. (22.4.2), so this anomaly receives no contribution from fermions that form real or pseudoreal representations of the gauge group, and it can therefore be calculated taking into account only those fermions whose masses arise from a breaking of the gauge symmetry. Also, this condition is automatically satisfied by the generators of any *simple* subalgebras like $SU(2)$ or $SU(3)$ of the gauge algebra. ($\mathrm{tr}\{T_\alpha\}$ is just a number which commutes with all the T_β, so if it is non-zero then the algebra is not simple.) Hence we only need to check that Eq. (22.4.3) is satisfied by the $U(1)$ generators of the gauge algebra. In the standard model, the sum of the values of the weak hypercharge y for all left-handed fermions is

$$\sum y/g' = 6\left(-\frac{1}{6}\right) + 3\left(\frac{2}{3}\right) + 3\left(-\frac{1}{3}\right) + 2\left(\frac{1}{2}\right) + (-1) = 0\,,$$

so there are no gravitational anomalies in the standard model currents.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

The requirement of vanishing of anomalies may be used as a guide in formulating realistic theories. For instance, the values of the weak hypercharges y for various $SU(3) \times SU(2)$ multiplets were originally taken from experiment, but one may wonder why these weak hypercharges (and the corresponding average electric charges in each multiplet) take the observed values. To answer this question, suppose we assign arbitrary weak hypercharges a, b, c, d, and e to the multiplets (u_L, d_L), u_R^*, d_R^*, (ν_L, e_L), and e_R^*, respectively. The conditions for anomaly cancellation tell us that:

$\boldsymbol{SU(3)}$–$\boldsymbol{SU(3)}$–$\boldsymbol{U(1)}$:

$$\sum_{3,\bar{3}} y = 2a + b + c = 0 \ ;$$

$\boldsymbol{SU(2)}$–$\boldsymbol{SU(2)}$–$\boldsymbol{U(1)}$:

$$\sum_{\text{doublets}} y = 3a + d = 0 \ ;$$

$\boldsymbol{U(1)}$–$\boldsymbol{U(1)}$–$\boldsymbol{U(1)}$:

$$\sum y^3 = 6a^3 + 3b^3 + 3c^3 + 2d^3 + e^3 = 0 \ ;$$

graviton–graviton–$\boldsymbol{U(1)}$:

$$\sum y = 6a + 3b + 3c + 2d + e = 0 \ .$$

Aside from the possibility of interchanging u_R and d_R, these equations have only two solutions, which we may call $U(1)$ and $U(1)'$:

$$\begin{aligned} U(1): &\quad b/a = -4 \ , \quad c/a = 2 \ , \quad d/a = -3 \ , \quad e/a = 6 \ , \\ U(1)': &\quad b = -c \ , \quad a = c = d = e = 0 \ . \end{aligned}$$

Furthermore, these solutions are exclusive; we cannot suppose that *both* $U(1)$ and $U(1)'$ are local symmetries, because then we would encounter a $U(1)'$–$U(1)'$–$U(1)$ anomaly proportional to $(-4) + (+2) \neq 0$ and a $U(1)'$–$U(1)$–$U(1)$ anomaly proportional to $(-4)^2 - (+2)^2 \neq 0$. The $U(1)$ generator is just the weak hypercharge of the standard electroweak theory (with the overall constant factor a absorbed into the definition of g'), while the $U(1)'$ symmetry resembles nothing observed in nature. This little calculation provides a rational explanation of the assignment of y values, or equivalently of electric charges, in the standard model, and it shows that if all gauge anomalies must cancel within a single generation of quarks and leptons, then it is not possible to couple a gauge boson to any other $U(1)$ quantum number, in addition to weak hypercharge.

On the other hand, although it is reasonable to guess that the $SU(3) \times SU(2) \times U(1)$ gauge bosons of the standard model may couple only to the known quarks and leptons (both to explain why no other fermions have

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

been discovered and to preserve the beautiful cancellation of anomalies in the standard model), there may be other $U(1)'$ gauge bosons that couple to other undetected $(SU(3) \times SU(2) \times U(1))$-neutral fermions as well as to the known quarks and leptons. Suppose we denote the $U(1)'$ quantum numbers y' of the multiplets (u_L, d_L), u_R^*, d_R^*, (ν_L, e_L), and e_R^* as a', b', c', d', and e', respectively. Since we don't know anything about possible $(SU(3) \times SU(2) \times U(1))$-neutral fermions, the requirement of cancellation of the $U(1)'$–$U(1)'$–$U(1)'$ and graviton–graviton–$U(1)'$ anomalies does not help us to constrain a', b', c', d', or e'. The remaining conditions of anomaly cancellation tell us that:

$\boldsymbol{SU(3)}$–$\boldsymbol{SU(3)}$–$\boldsymbol{U(1)'}$:

$$\sum_{3,\bar{3}} y' = 2a' + b' + c' = 0 \ ;$$

$\boldsymbol{SU(2)}$–$\boldsymbol{SU(2)}$–$\boldsymbol{U(1)'}$:

$$\sum_{\text{doublets}} y' = 3a' + d' = 0 \ ;$$

$\boldsymbol{U(1)}$–$\boldsymbol{U(1)}$–$\boldsymbol{U(1)'}$:

$$\sum y^2 y' = 6a' + 3(-4)^2 b' + 3(2)^2 c' + 2(-3)^2 d' + (6)^2 e' = 0 \ ;$$

$\boldsymbol{U(1)}$–$\boldsymbol{U(1)'}$–$\boldsymbol{U(1)'}$:

$$\sum y y'^2 = 6a'^2 + 3(-4)b'^2 + 3(2)c'^2 + 2(-3)d'^2 + (6)e'^2 = 0 \ .$$

The general solution has y' a linear combination of y and a quantum number $B - L$ (where B and L are the conventional baryon and lepton numbers) that takes values $1/3$, $-1/3$, $-1/3$, -1, and $+1$ for the multiplets (u_L, d_L), u_R^*, d_R^*, (ν_L, e_L), and e_R^*, respectively. If $B - L$ *is* a local symmetry, with a coupling that is not many orders of magnitude smaller than e, then it must be spontaneously broken, since ordinary bodies have macroscopic values of $B-L$. To avoid conflict with observations of neutral currents, the characteristic scale F of the symmetry breaking would have to be larger than that of the electroweak interactions, but not necessarily many orders of magnitude larger. Thus a neutral vector boson somewhat heavier than the Z^0 and coupled to $B - L$ seems like the most plausible addition to the standard model.

This has all been for a single generation of the standard model. With three generations there are many more anomaly-free symmetries. One class of symmetries that are not broken by anomalies or (as far as we know) by anything else consists of the differences in the numbers of leptons of the various flavors. Together with $B - L$, this will be important in Section 23.5 in classifying the baryon- and lepton-non-conserving processes produced by anomalies.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

22.5 Massless Bound States*

It is sometimes speculated that quarks and leptons may be bound states of more fundamental particles. If these hypothetical fundamental particles had asymptotically free gauge interactions, like those in quantum chromodynamics, then we would expect them to be trapped, which would explain why they are not observed. However, there is a difficulty with this picture. No internal structure of the quarks and leptons has ever been observed, so the characteristic energy scale Λ' (analogous to the $\Lambda \approx 200$ MeV of quantum chromodynamics) of these gauge interactions must be very high. For instance, as remarked in Section 12.3, the agreement between theory and experiment for the magnetic moment of the muon indicates that $\Lambda' > 3$ TeV. But, aside from Goldstone bosons, we would normally expect the bound states in such a theory to have masses of order Λ', or perhaps $2\pi\Lambda'$, just as in quantum chromodynamics the proton mass is of order $2\pi\Lambda_{\rm QCD}$. This expectation is of course in sharp contradiction with the fact the observed masses of the quarks and leptons are much less than Λ'. To put this problem another way, if the leptons and quarks are bound states, then why is their size (as measured by anomalous magnetic moments, etc.) so much smaller than their Compton wavelengths?

One way to answer this question is to suppose that, unlike quantum chromodynamics, this theory has unbroken chiral symmetries that keep the quarks and leptons massless, aside from small corrections from other interactions. In general, a chiral symmetry is any symmetry for which the massless elementary fields of some given helicity (including the complex conjugates of the fields of opposite helicity) furnish a complex representation. By operating on the vacuum with products of the elementary fields we can construct other states of definite helicity that also furnish complex representations of these symmetries. If any of these states are actual composite particles then they must be massless, because all helicity components of the state of a massive particle must furnish the same representation of any symmetry that commutes with rotations, so a given helicity component of a particle together with the antiparticle of the opposite helicity component would together furnish a real representation. Of course, it may not be so easy to tell which of the states constructed in this way correspond to actual composite massless particles, but if they do, their masslessness is natural, because in order for some massless particles of a given helicity that belong to a complex representation of the symmetry

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

group to become massive as we change some parameter of the theory, their symmetry properties would have to change discontinuously from a complex to a real representation.

Although this reasoning shows that there are theories in which it is natural to have massless or very light composite particles, it does not give any indication of when this actually happens. The question is an interesting one, quite apart from the problem of understanding quarks and leptons as possible composite particles. 't Hooft[16] has suggested a powerful way of answering this question, based on considerations of anomalies. Briefly, if the underlying theory has global chiral symmetries (not broken by gauge anomalies, and not spontaneously broken) consisting of transformations on the elementary left-handed spin $\frac{1}{2}$ fermions (and antifermions) χ with symmetry generators T_α, T_β, etc., and if the anomaly constant $\mathrm{tr}[\{T_\alpha, T_\beta\}T_\gamma]$ of these global symmetries is non-zero, then the spectrum of bound states must include spin $\frac{1}{2}$ massless particles on whose left-handed states the same symmetries induce transformations with generators $\mathscr{T}_\alpha$, $\mathscr{T}_\beta$, etc., with the same anomaly constant

$$\mathrm{tr}[\{\mathscr{T}_\alpha, \mathscr{T}_\beta\}\mathscr{T}_\gamma] = \mathrm{tr}[\{T_\alpha, T_\beta\}T_\gamma] \,. \tag{22.5.1}$$

't Hooft's argument was as follows. Imagine that some weakly-interacting gauge bosons are coupled to generators T_α, T_β, etc. of the global symmetries of the underlying theory. Suppose also that although some of the coefficients $D_{\alpha\beta\gamma} \equiv \mathrm{tr}[\{T_\alpha, T_\beta\}T_\gamma]$ are non-zero, this anomaly is cancelled by anomalies due to other 'spectator' massless fermions, which do not feel the strong forces that trap the constituents of the composite particles. Physical processes at energies much less than the characteristic energy scale Λ' of the trapping interactions will be described by an effective Lagrangian in which the trapped fermions do not appear. If the symmetries with generators T_α, T_β, etc. are not spontaneously broken, then there are no Goldstone bosons, so the only particles in this effective Lagrangian will be the weakly coupled gauge bosons and spectator fermions, plus any massless bound states of the trapped fermions and strongly interacting gauge bosons. The consistency of the effective field theory requires that it must be anomaly-free, but the spectator fermions by assumption have an anomaly constant equal to $-D_{\alpha\beta\gamma}$, so there must be massless bound states, to provide an anomaly constant equal and opposite to this, and hence equal to that of the original trapped fermions.** Note that this argument

** It remains to show that these particles must have spin $\frac{1}{2}$. We are assuming that there are no Goldstone bosons here, and that other spinless particles would not naturally be massless. The elementary gauge bosons of the theory are assumed neutral under the anomalous symmetry transformations, so they could not contribute to the anomaly. Composite particles of spin $j \geq 1$ are ruled out by a different argument.[16a] This is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

works no matter how weak are the gauge interactions with generators T_α, T_β, etc., so these gauge bosons and the untrapped fermions do not have to be real particles to reach the conclusion that *the massless spin $\frac{1}{2}$ bound states reproduce the anomalies of the trapped elementary spin $\frac{1}{2}$ fermions of which they are composed.*

As a simple example, suppose that the underlying theory contains n 'flavors' of massless fermions, each of which has both left- and right-handed parts in the defining representation N of an asymptotically free $SU(N)$ gauge group. We require N to be odd so that there can be untrapped $SU(N)$-neutral fermionic bound states. Just as in quantum chromodynamics, this theory has an automatic global $SU_L(n)\times SU_R(n)\times U_V(1)$ symmetry, with the left- and right-handed massless fermions respectively in its $(n,1)$ and $(1,n)$ representations, both having equal values for the $U_V(1)$ quantum number, which can be taken as unity. There are non-vanishing anomaly constants in the underlying theory for the $SU(n)_L$–$SU(n)_L$–$U(1)_Y$ and $SU(n)_R$–$SU(n)_R$–$U(1)_Y$ current triplets, which have the values

$$D_{aL,bL,0} = D_{aR,bR,0} = N\delta_{ab} ,$$

where a, b, etc. label the $SU(n)$ generators λ_a, normalized so that in the defining n-component representation $\mathrm{tr}\{\lambda_a\lambda_b\} = \frac{1}{2}\delta_{ab}$. For $n > 2$ there are also non-vanishing anomaly constants for the $SU(n)_L$–$SU(n)_L$–$SU(n)_L$ and $SU(n)_R$–$SU(n)_R$–$SU(n)_R$ currents, equal to

$$D_{aL,bL,cL} = D_{aR,bR,cR} = N\mathrm{tr}[\{\lambda_a,\lambda_b\},\lambda_c] .$$

We suppose here that the $SU_L(n) \times SU_R(n) \times U(1)$ symmetry is not spontaneously broken. Because of trapping, the only fermionic bound states in the physical spectrum will be contained in representations of this symmetry that can be formed from m_L and m_R elementary fermions of helicity $+\frac{1}{2}$ and $-\frac{1}{2}$, respectively, and $\bar{m}_L$ and $\bar{m}_R$ of their antiparticles, with

$$m_L + m_R - \bar{m}_L - \bar{m}_R = kN , \tag{22.5.2}$$

where k is any positive or negative odd integer. In consequence, the only irreducible representations (r,s) of $SU(n)_L \times SU_R(n)$ encountered are those for which r is in the direct product of m_L of the defining representations of

a theory in which the anomalous currents can be constructed as Lorentz four-vector functions of the elementary spin $\frac{1}{2}$ fields, and in order to contribute to the anomaly these currents would have to have non-vanishing matrix elements between any massless composite particles of spin $j = 1$, which would violate Lorentz invariance. Massless composite particles of spin $j \geq 3/2$ are ruled out because in this theory it is also possible to construct a conserved energy-momentum tensor, which would have to have non-vanishing matrix elements between these massless composite particles, which for $j \geq 3/2$ would also violate Lorentz invariance.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

$SU(n)$ and $\bar{m}_L$ of their complex conjugates, while s is in the direct product of m_R of the defining representations of $SU(n)$ and $\bar{m}_R$ of their complex conjugates, and the $U(1)_V$ quantum number is kN, with k, m_L, m_R, $\bar{m}_L$, and $\bar{m}_R$ subject to Eq. (22.5.2). Let $p(r,s,k)$ be the number of times an irreducible representation (r,s) of $SU_L(n) \times SU_R(n)$ with $U(1)_V$ quantum number kN appears among the helicity $+\frac{1}{2}$ bound states. Eq. (22.5.1) then reads

$$\sum_{r,s,k} p(r,s,k) d_s \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}\mathscr{T}_c] = N \operatorname{tr}[\{\lambda_a, \lambda_b\}\lambda_c] , \tag{22.5.3}$$

$$\sum_{r,s,k} p(r,s,k) d_s k \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}] = \operatorname{tr}[\{\lambda_a, \lambda_b\}] , \tag{22.5.4}$$

where $\operatorname{tr}^{(r)}$ denotes the trace in the irreducible representation r of $SU(n)$, and d_s is the dimensionality of representation s of $SU(N)$. The only other constraint on the $p(r,s,k)$ is that they must be positive integers.

The complex conjugate $(\bar{r},\bar{s},-k)$ of any representation (r,s,k) of $SU(2)_L \times SU(2)_R \times U(1)_V$ has values of the traces $\operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}\mathscr{T}_c]$ and $k \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}]$ opposite to those of the representation (r,s,k), so Eqs. (22.5.3) and (22.5.4) only restrict the values of

$$\ell(r,s,k) \equiv p(r,s,k) - p(\bar{r},\bar{s},-k) .$$

Recall that in the notation we have been using here, these traces are over all massless bound states of helicity $+\frac{1}{2}$, including the antiparticles of the massless bound states of helicity $-\frac{1}{2}$, which transform according to the complex conjugate representations. The complex conjugate of any representation of $SU(2)_L \times SU(2)_R \times U(1)_V$ has values of $\operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}\mathscr{T}_c]$ and $k \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}]$ opposite to those of that representation. Thus we can sum in (22.5.3) and (22.5.4) only over the representations with $U(1)_V$ quantum number $kN > 0$

$$\sum_{r,s,k>0} \ell(r,s,k)\, d_s \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}\mathscr{T}_c] = N \operatorname{tr}[\{\lambda_a, \lambda_b\}\lambda_c] , \tag{22.5.5}$$

$$\sum_{r,s,k>0} \ell(r,s,k)\, d_s k \operatorname{tr}^{(r)}[\{\mathscr{T}_a, \mathscr{T}_b\}] = \operatorname{tr}[\{\lambda_a, \lambda_b\}] , \tag{22.5.6}$$

with $\ell(r,s,k)$ equal to the number of times an irreducible representation (r,s) of $SU_L(n) \times SU_R(n)$ with $U(1)_V$ quantum number $kN > 0$ appears among the helicity $+\frac{1}{2}$ bound states, *minus* the number of times the same representation appears among the helicity $-\frac{1}{2}$ bound states. (If parity is unbroken then a representation r,s,k must occur among the helicity $-\frac{1}{2}$ bound states as often as the representation s,r,k occurs among the helicity $+\frac{1}{2}$ bound states, so in this case $\ell(r,s,k) = -\ell(s,r,k)$.)

First, consider the case of $n = 2$ flavors. There is no way of coupling three three-vectors together symmetrically to make an $SU(2)$-invariant,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

so both sides of (22.5.5) vanish automatically, leaving only the condition (22.5.6). The defining two-component representation of $SU(2)$ is contained in the product of any odd number of these two-component representations, so we can always find a solution of (22.5.6) with $\ell(r,s,k) = 0$ for all $SU(2)_L \times SU(2)_R \times U(1)_V$ representations with r non-trivial except the one in which r is the defining representation, s is the trivial representation, and $k = 1$, for which we take $\ell = 1$. Unfortunately, this solution is far from unique — there is an infinite number of ways of reproducing the anomalies of the underlying theory.

For general n and N it is usual to find many solutions of Eqs. (22.5.5) and (22.5.6), but there are some cases where no solutions can be found. In such theories we can reach the conclusion that the $SU(n)_L \times SU(n)_R \times U(1)_V$ symmetry *must* be partly or completely spontaneously broken. This has particularly interesting implications for quantum chromodynamics, so let us now specialize to the case of an $SU(3)$ gauge group.

To be concrete, we will focus on the representations, with $k = 1$ and $\bar{m}_L = \bar{m}_R = 0$, that can be formed from just three elementary fermions and no antifermions. These are:[†]

(a) r is the symmetric third-rank $SU(n)$ tensor; s is the trivial representation.

(b) r is the antisymmetric third-rank $SU(n)$ tensor; s is the trivial representation.

(c) r is the third-rank $SU(n)$ tensor of mixed symmetry; s is the trivial representation.

(d) r is the symmetric second-rank $SU(n)$ tensor; s is the $SU(n)$ vector.

(e) r is the antisymmetric second-rank $SU(n)$ tensor; s is the $SU(n)$ vector.

(f) r is the $SU(n)$ vector; s is the symmetric second-rank $SU(n)$ tensor.

(g) r is the $SU(n)$ vector; s is the antisymmetric second-rank $SU(n)$ tensor.

(h) r is the trivial representation; s is the symmetric third-rank $SU(n)$ tensor.

(i) r is the trivial representation; s is the antisymmetric third-rank $SU(n)$ tensor.

(j) r is the trivial representation; s is the third-rank $SU(n)$ tensor of mixed symmetry.

[†] All $SU(N)$ vectors and tensors here are understood to be contravariant.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

For $n > 2$, Eqs. (22.5.5) and (22.5.6) here read [††]

$$\tfrac{1}{2}(n+3)(n+6)\ell_a + \tfrac{1}{2}(n-3)(n-6)\ell_b + (n^2-9)\ell_c + n(n+4)\ell_d$$
$$+n(n-4)\ell_e + \tfrac{1}{2}n^2(n+1)\ell_f + \tfrac{1}{2}n^2(n-1)\ell_g = 3 \qquad (22.5.7)$$

and

$$\tfrac{1}{2}(n+2)(n+3)\ell_a + \tfrac{1}{2}(n-2)(n-3)\ell_b + (n^2-3)\ell_c + n(n+2)\ell_d$$
$$+n(n-2)\ell_e + \tfrac{1}{2}n(n+1)\ell_f + \tfrac{1}{2}n(n-1)\ell_g = 1\,. \qquad (22.5.8)$$

There is no problem in satisfying Eq. (22.5.7), but note that if n is a multiple of three then for all values of the ℓs, each term on the left-hand side of Eq. (22.5.8) is also a multiple of three, which makes it impossible to satisfy this condition. We conclude in particular that *the $SU(3)_L \times SU(3)_R \times U(1)_V$ symmetry of quantum chromodynamics with three flavors of massless quarks must be spontaneously broken.* This result is not limited to the representations (**a**)–(**j**), but applies to any representations of $SU(3)_L \times SU(3)_R \times U(1)_V$ that can be formed from color-neutral combinations of quarks and antiquarks.

* * *

Aside from special cases such as an $SU(3)$ gauge group with $n = 3$ elementary fermion $SU(3)$-triplets, the 't Hooft anomaly matching condition is not very restrictive; in general it allows a wide variety of massless bound states when chiral symmetry is unbroken. 't Hooft also proposed a *decoupling condition*, which requires that when one or more of the elementary fermion flavors become very heavy there should be no unbroken chiral symmetries that prevent composite particles that contain heavy elementary fermions from acquiring masses. For instance, in the case of an $SU(3)$ color gauge group, if we give one of the n quark types a large mass then those three-fermion bound states that contain a single massive quark will furnish representations (r', s') of the group $SU(n-1) \times SU(n-1)$ that are either

(v) r' is the symmetric second rank $SU(n-1)$ tensor; s' is the trivial representation,

(w) r' is the antisymmetric second rank $SU(n-1)$ tensor; s' is the trivial representation,

(x) r' and s' are $SU(n)$ vectors,

[††] In Ref. 16, 't Hooft assumed that parity is not spontaneously broken, so he gave these formulas for the case where $\ell_a = -\ell_h$, $\ell_b = -\ell_i$, $\ell_c = -\ell_j$, $\ell_d = -\ell_f$, and $\ell_e = -\ell_g$. As we see here, the main conclusion does not depend on parity conservation.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

(y) r' is the trivial representation, s' is the symmetric second rank $SU(n-1)$ tensor,

(z) r' is the trivial representation, s' is the antisymmetric second rank $SU(n-1)$ tensor.

In order for the three-fermion mass state to acquire a large mass, it is necessary that $\ell'(r',s') = 0$, where $\ell'(r',s')$ is the number of times that an irreducible representation (r',s') occurs among the helicity $+\frac{1}{2}$ bound states minus the number of times the same representation appears among the helicity $-\frac{1}{2}$ bound states. By inspecting the list of three-fermion representations **(a)**–**(j)** of $SU(n) \times SU(n)$ to see what representations of $SU(n-1) \times SU(n-1)$ they yield, we see then that the 't Hooft decoupling condition requires that[†]

$$\begin{aligned} 0 &= \ell'_v = \ell_a + \ell_c + \ell_d \,, \\ 0 &= \ell'_w = \ell_b + \ell_c + \ell_e \,, \\ 0 &= \ell'_x = \ell_f + \ell_g + \ell_d + \ell_e \,, \\ 0 &= \ell'_y = \ell_f + \ell_h + \ell_j \,, \\ 0 &= \ell'_z = \ell_g + \ell_i + \ell_j \,. \end{aligned} \tag{22.5.9}$$

Unfortunately, in most cases there are still an infinite number of solutions, though there are no solutions in which the ℓs are n-independent integers.

The decoupling condition seems quite plausible, but its use by 't Hooft was questioned[17] on the ground that, as one or more fermion masses increase, one usually encounters phase transitions that make the mass spectrum different from what it would be for small fermion masses. There is a stronger condition, known as the *persistent mass condition*, which requires that when one or more of the elementary fermion flavors acquire *any* mass, there should be no unbroken chiral symmetries that prevent composite particles that contain these massive elementary fermions from acquiring some masses.[17] If valid, the persistent mass condition would lead to the same consequences, such as Eqs. (22.5.9), that were described by 't Hooft, with no chance of being invalidated by phase transitions.

It is easy to construct non-realistic models in which the persistent mass condition is violated, such as theories with spontaneously broken *non-chiral* symmetries, which lead to massless Goldstone bosons formed as composites of massive fermions.[17] (These models also exhibit phase transitions with increasing fermion masses, which invalidate 't Hooft's conclusions from the decoupling condition.) But the work of Vafa and

[†] In the parity conserving case studied by 't Hooft, the fourth and fifth equations are identical to the first and second, while the third is empty.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Witten discussed in Section 19.9 shows that, in a variety of more realistic QCD-like theories, non-chiral symmetries cannot be spontaneously broken, so this should not be taken as a serious objection to the persistent mass condition.

22.6 Consistency Conditions

The numerical coefficient appearing in the anomaly for any symmetry depends on the matter content of the theory. On the other hand, the *form* of the anomaly is largely independent of the details of the theory, because it is governed by consistency conditions, first presented in 1971 by Wess and Zumino.[18]

Even when we are interested in anomalies in global symmetry currents, in order to derive the consistency conditions it is convenient to imagine that all symmetry currents are coupled to gauge fields, which for non-Abelian symmetries also couple to each other, in such a way that these symmetries become *local* symmetries of the Lagrangian density. We can always return to the case of global symmetry by letting the corresponding gauge coupling constants become infinitesimal. Apart from anomalies, the effective action $\Gamma[A]$ in a background gauge field $A_{\alpha\mu}(x)$ will in this formalism be invariant under infinitesimal transformations $A_{\beta\mu}(y) \to A_{\beta\mu}(y) + i\int d^4x\, \epsilon_\alpha(x)\, \mathscr{T}_\alpha(x)\, A_{\beta\mu}(y)$ on the gauge field, where in order to reproduce the transformation (15.1.9) we must take

$$-i\mathscr{T}_\alpha(x) = -\frac{\partial}{\partial x^\mu}\frac{\delta}{\delta A_{\alpha\mu}(x)} - C_{\alpha\beta\gamma}A_{\beta\mu}(x)\frac{\delta}{\delta A_{\gamma\mu}(x)}\,. \tag{22.6.1}$$

Taking anomalies into account, $\mathscr{T}_\alpha(x)$ no longer annihilates $\Gamma[A]$, but rather

$$\mathscr{T}_\alpha(x)\Gamma[A] = G_\alpha[x;A]\,, \tag{22.6.2}$$

where $G_\alpha[x;A]$ represents the effect of the anomaly. Eq. (22.6.2) may also be written as a formula for the covariant divergence of the expectation value of the current:

$$D_\mu\langle J^\mu_\alpha(x)\rangle = -iG_\alpha[x;A]\,, \tag{22.6.3}$$

where

$$\langle J^\mu_\alpha(x)\rangle \equiv \frac{\delta}{\delta A_{\alpha\mu}(x)}\Gamma[A] \tag{22.6.4}$$

and D_μ is the gauge-covariant derivative (15.1.10), taken here in the adjoint representation with $(t_\beta)_{\gamma\alpha} = -iC_{\alpha\beta\gamma}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

The Wess–Zumino consistency conditions follow from the commutation relations*

$$[\mathscr{T}_\alpha(x), \mathscr{T}_\beta(y)] = iC_{\alpha\beta\gamma}\delta^4(x-y)\mathscr{T}_\gamma(x)\,. \tag{22.6.5}$$

From Eqs. (22.6.2) and (22.6.5), we derive the general consistency condition

$$\mathscr{T}_\alpha(x)G_\beta[y;A] - \mathscr{T}_\beta(y)G_\alpha[x;A] = iC_{\alpha\beta\gamma}\delta^4(x-y)G_\gamma[y;A]\,. \tag{22.6.6}$$

These consistency conditions were originally derived by Wess and Zumino for the chiral $SU(3) \times SU(3)$ symmetry of the strong interactions, a special case of physical as well as historical importance. Here the generators $\mathscr{T}_\alpha(x)$ acting on the gauge fields consist of the even parity generators $\mathscr{Y}_a(x)$ defined by Eq. (22.3.31), and the odd parity generators $\mathscr{X}_a(x)$ defined by (22.3.32).** They satisfy the commutation relations

$$[\mathscr{Y}_a(x), \mathscr{Y}_b(y)] = i\delta^4(x-y)f_{abc}\mathscr{Y}_c(x)\,,$$

$$[\mathscr{Y}_a(x), \mathscr{X}_b(y)] = i\delta^4(x-y)f_{abc}\mathscr{X}_c(x)\,,$$

$$[\mathscr{X}_a(x), \mathscr{X}_b(y)] = i\delta^4(x-y)f_{abc}\mathscr{Y}_c(x)\,,$$

where f_{abc} are the $SU(3)$ structure constants. Since the $SU(3)$ subgroup generated by the $\mathscr{Y}_a$ is not spontaneously broken, it is convenient to treat the integration over fermion momenta in such a way as to preserve invariance under gauge transformations generated by $\mathscr{Y}_a$, so that

$$\mathscr{Y}_a(x)\,\Gamma = 0\,,$$

leaving us with the non-zero anomaly

$$\mathscr{X}_a(x)\,\Gamma = G_a(x)\,.$$

The non-trivial consistency conditions are then

$$\mathscr{Y}_a(x)G_b(y) = i\delta^4(x-y)f_{abc}G_c(x)$$

and

$$\mathscr{X}_a(x)G_b(y) - \mathscr{X}_b(x)G_a(y) = 0\,.$$

The first of these simply says that $G_a(x)$ transforms like an octet under ordinary $SU(3)$ transformations. The second condition imposes other strong constraints on $G_a(x)$. The reader may check that this condition is

* The factor $-i$ was inserted in Eq. (22.6.1) in order to provide the conventional factor $+i$ accompanying the structure constant $C_{\alpha\beta\gamma}$ in this commutation relation. A reminder: we are using a basis for the Lie algebra in which the structure constants are totally antisymmetric, and so we do not distinguish between upper and lower gauge indices.

** Another reminder: as mentioned in Section 22.3, the $\mathscr{X}_a$ and $\mathscr{Y}_a$ used here are what were called Y_a and X_a in Ref. 8.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

satisfied by the Bardeen formula (22.3.34) for $G_a(x)$. We will not go into this here, but will instead use as an illustration a general gauge theory with all currents treated symmetrically. In Section 22.3 we quoted the formula (22.3.38) for the anomaly in this case, but did not derive the terms in this formula of higher than second order in the gauge fields. Here we shall show that these terms are dictated by the consistency conditions (22.6.6).

For this purpose, and also to allow for further generalizations, it is very convenient to reformulate the set of Wess–Zumino consistency conditions as a condition of invariance under the BRST transformations described in Section 15.7. Let us introduce a ghost field ω_α, and define the nilpotent BRST operator s for a general gauge theory by

$$sA_{\alpha\mu} = \partial_\mu \omega_\alpha + C_{\alpha\beta\gamma} A_{\beta\mu} \omega_\gamma \,, \tag{22.6.7}$$

$$s\omega_\alpha = -\tfrac{1}{2} C_{\alpha\beta\gamma} \omega_\beta \omega_\gamma \,, \tag{22.6.8}$$

it being understood that s satisfies the distributive rule $s(AB) = (sA)B \pm A(sB)$, the sign being negative where A is a fermionic quantity like ω_α, and otherwise positive. In place of the anomaly function $G_\alpha[x;A]$, we shall work with a functional

$$G[\omega, A] = \int \omega_\alpha(x)\, G_\alpha[x;A]\, d^4x \,. \tag{22.6.9}$$

Then (keeping in mind that ω_α is fermionic)

$$\begin{aligned} sG[\omega,A] &= -\tfrac{1}{2} C_{\alpha\beta\gamma} \int d^4x\, \omega_\beta(x)\, \omega_\gamma(x)\, G_\alpha[x;A] \\ &\quad - \int d^4x\, \omega_\alpha(x) \int d^4y \left[\frac{\partial \omega_\beta(y)}{\partial y^\mu} + C_{\beta\gamma\delta} A_{\gamma\mu}(y) \omega_\delta(y) \right] \frac{\delta G_\alpha[x;A]}{\delta A_{\beta\mu}(y)} \\ &= \int d^4x \int d^4y\, \omega_\alpha(x)\, \omega_\beta(y) \Big[-\tfrac{1}{2} C_{\alpha\beta\gamma} \delta^4(x-y) G_\gamma[x;A] \\ &\quad + i\mathscr{T}_\beta(y) G_\alpha[x;A] \Big] . \end{aligned}$$

Because the ghost fields anticommute with each other, this can be written

$$\begin{aligned} sG[\omega,A] &= -\tfrac{1}{2} i \int d^4x \int d^4y\, \omega_\alpha(x)\, \omega_\beta(y) \\ &\quad \times \Big[iC_{\alpha\beta\gamma} \delta^4(x-y) G_\gamma[x;A] + \mathscr{T}_\beta(y) G_\alpha[x;A] - \mathscr{T}_\alpha(x) G_\beta[y;A] \Big] . \end{aligned}$$

We see that the consistency condition (22.6.6) will hold if and only if $G[\omega, A]$ is BRST-invariant

$$sG[\omega, A] = 0 \tag{22.6.10}$$

for all ghost fields $\omega_\alpha(x)$.

Now consider the possibility that the anomaly $G[\omega; A]$ could be written

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

as the BRST operator s acting on a local functional $F[A]$:

$$G[\omega;A] = sF[A] \,. \tag{22.6.11}$$

(Note that the functional F is necessarily independent of the ghost field, because the operator s adds one ghost field factor, and the anomaly functional G is already linear in the ghost field.) The BRST operator satisfies $s^2 = 0$, so such an anomaly would satisfy the consistency condition $sG = 0$. If $F[A]$ is a *local* functional† of the gauge field, then it could be subtracted from the action, thus cancelling the anomaly. The same is true of any term in the anomaly that can be written as the BRST operator s acting on a local functional; such a term satisfies the consistency condition by itself, and can be cancelled by adding a local term to the action. The possible anomalies that interest us are thus the local functionals $G[\omega;A]$ of ghost number unity that satisfy the consistency condition (22.6.10), modulo terms that can be expressed as s acting on some local functional of ghost number zero. In accordance with the usual terminology for nilpotent operators, the equivalence classes of such functionals form what is called the *cohomology* of the s operator at ghost number unity.

We can also express this in terms of the local densities themselves. We can write the anomaly (or any term in the anomaly) as $G = \int d^4x\, \mathscr{G}(x)$, where $\mathscr{G}(x)$ is a power series in the gauge and ghost fields and their derivatives at the spacetime point x. The condition $sG = 0$ then is equivalent to the statement that

$$s\mathscr{G}(x) = \partial_\mu \mathscr{J}^\mu(x) \tag{22.6.12}$$

for some function $\mathscr{J}^\mu(x)$ of fields and field derivatives. Likewise, the terms in $\mathscr{G}$ that can be cancelled by adding local terms to the action are those that are of the form $s\mathscr{F}$ up to possible derivatives. Thus the anomalies that interest us are the local functions of ghost number unity that satisfy the consistency condition (22.6.12), modulo terms that can be expressed as s acting on some local functional of ghost number zero, modulo possible derivatives. This is known as the cohomology of s at ghost number unity in the space of local functions, modulo derivatives, and denoted $H^1(s|d)$.

Algebraic methods have been used to work out the cohomology of the BRST operator s, and thereby deduce the form of the anomaly in general gauge theories.[19] This approach leaves unknown only constant coefficients that depend on the matter content of the theory and need to be calculated by the methods of Sections 22.2 or 22.3. Since we have already calculated the terms in the anomaly of second order in the gauge fields for general gauge theories, including their constant coefficients, here we will use the

† By a local functional is meant the integral of a local function, that is, a function of fields and field derivatives at a given point.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

consistency condition (22.6.12) to calculate the terms of higher order in the fields.

We saw in Section 22.3 that when all currents are treated symmetrically, the terms of second order in the gauge fields are one-third the expression (22.3.26). This is an operator of dimensionality four (in units of mass), while the Wess–Zumino consistency condition (22.6.6) relates only operators of the same dimensionality, so to satisfy this condition we should add only terms of higher order in the gauge field that have the same dimensionality. We therefore seek a solution of the consistency conditions in the (not necessarily unique) form

$$G_\alpha = i[\langle \partial_\mu J^\mu_\alpha \rangle]_{\text{anom}} = -\frac{i}{24\pi^2}\epsilon^{\kappa\nu\lambda\rho}\,\text{Tr}\left\{T_\alpha\left[\partial_\kappa A_\nu\,\partial_\lambda A_\rho + ic_1\partial_\kappa A_\nu\,A_\lambda A_\rho\right.\right.$$
$$\left.\left. + ic_2 A_\kappa\,\partial_\nu A_\lambda\,A_\rho + ic_3 A_\kappa A_\nu \partial_\lambda A_\rho - c_4 A_\kappa A_\nu A_\lambda A_\rho\right]\right\}, \qquad (22.6.13)$$

where $A_\mu \equiv A_{\alpha\mu}T_\alpha$, and the c_i are constants to be determined.

In order to save a great deal of effort, it will be convenient to rewrite this in the language of differential forms. (See Section 8.8.) We introduce a set of c-number parameters dx^μ that are taken to anticommute with themselves and all fermionic fields, such as the ghost field ω_α. The dx^μ then also anticommute with the BRST operator s. Because $dx^\kappa dx^\nu dx^\lambda dx^\rho$ is totally antisymmetric, it may be written as

$$dx^\kappa dx^\nu dx^\lambda dx^\rho = \epsilon^{\kappa\nu\lambda\rho}d^4x\,, \qquad d^4x = dx^0dx^1dx^2dx^3\,. \qquad (22.6.14)$$

We also introduce the exterior derivative

$$d \equiv dx^\mu \frac{\partial}{\partial x^\mu}\,,$$

which since derivatives commute is nilpotent as well as anticommuting with s:

$$d^2 = 0\,, \qquad ds + sd = 0\,. \qquad (22.6.15)$$

Finally, we introduce the anticommuting quantities

$$A \equiv iA_\mu dx^\mu = iA_{\alpha\mu}T_\alpha dx^\mu\,, \qquad \omega \equiv i\omega_\alpha T_\alpha\,. \qquad (22.6.16)$$

In this notation, Eq. (22.6.13) reads

$$G[\omega, A] = \frac{1}{24\pi^2}\int \text{Tr}\left\{\omega\left[(dA)^2 + c_1(dA)A^2\right.\right.$$
$$\left.\left. + c_2 A(dA)A + c_3 A^2(dA) + c_4 A^4\right]\right\}. \qquad (22.6.17)$$

In order to implement the consistency condition (22.6.10), we note

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

that the BRST transformation rules (22.6.7) and (22.6.8) may be written as

$$sA = -d\omega + \{A, \omega\}\,, \tag{22.6.18}$$

$$s\omega = \omega^2\,. \tag{22.6.19}$$

Now, the BRST transformation of the last term in Eq. (22.6.17) is given by

$$\begin{aligned} s\mathrm{Tr}\left[\omega A^4\right] &= \mathrm{Tr}\left[\omega^2 A^4 - \omega\{A,\omega\}A^3 + \omega A\{A,\omega\}A^2\right. \\ &\quad \left. -\omega A^2\{A,\omega\}A + \omega A^3\{A,\omega\}\right] + \omega d\omega A^3 \text{ terms} \\ &= \mathrm{Tr}\left[\omega^2 A^4\right] + \omega d\omega A^3 \text{ terms}\,. \end{aligned}$$

There is no other contribution to sG proportional to $\mathrm{Tr}\left[\omega^2 A^4\right]$, so the consistency condition (22.6.10) can only be satisfied if $c_4 = 0$. With $c_4 = 0$, a straightforward calculation gives

$$\begin{aligned} sG = \frac{1}{24\pi^2}\int \mathrm{Tr}\Big\{ & -(dA)^2\omega^2 + \omega\, d\omega\, A\, dA - d\omega\, \omega\, dA\, A \\ & -A\omega\, dA\, d\omega - \omega A\, d\omega\, dA \\ & +c_1\left[\omega\, dA\, d\omega\, A - \omega\, dA\, A\, d\omega\right] \\ & +c_2\left[\omega\, d\omega\, dA\, A - \omega A\, dA\, d\omega\right] \\ & +c_3\left[\omega\, d\omega\, A\, dA - \omega A\, d\omega\, dA\right] \\ & -c_1\left[-\omega A\, d\omega\, A^2 + \omega\, d\omega\, A^3 + \omega\, dA\, A^2\omega\right] \\ & -c_2\left[\omega A^2\, d\omega\, A - \omega A\, d\omega\, A^2 + \omega A\, dA\, d\omega\right] \\ & -c_3\left[-\omega A^3\, d\omega + \omega A^2\, d\omega\, A + \omega A^2\, dA\, \omega\right]\Big\}\,. \end{aligned}$$

We do not need to assume that the integrand vanishes, but only that it is the derivative of some local function, so that its integral vanishes. This condition must be satisfied separately for the terms involving two derivatives and those involving one derivative, since no cancellation can occur between terms with different numbers of derivatives. It is not hard to see that the terms in the integrand involving just one derivative are of the form $d\mathscr{F}$ if we take $c_1 = -c_2 = +c_3 \equiv c$. The remaining terms are a total derivative if $c = -1/2$, thus justifying the previously quoted result

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

(22.3.38). This result is often expressed more compactly as

$$G[\omega, A] = \frac{1}{24\pi^2}\int \mathrm{Tr}\left\{\omega\, d\left[A\, dA - \frac{1}{2}A^3\right]\right\}$$
$$= \frac{1}{24\pi^2}\int \mathrm{Tr}\left\{\omega\, d\left[AF + \frac{1}{2}A^3\right]\right\}, \qquad (22.6.20)$$

where F is the matrix-valued field-strength two-form

$$F \equiv \tfrac{1}{2}it_\alpha F_{\alpha\mu\nu}\, dx^\mu\, dx^\nu = dA - A^2\,. \qquad (22.6.21)$$

The anomaly does not have to be put in the form (22.6.13), so Eq. (22.6.20) is not the unique result for $G[\omega, A]$. Results quoted in the next section show that, for any subgroup H of G for which $\mathrm{Tr}\left\{t_i\{t_j, t_k\}\right\} = 0$ for all generators of H, it is possible to add local terms to the action in such a way that the anomaly G_i vanishes when t_i is any generator of H.

There is an elegant algebraic tool, known as the *Stora–Zumino descent equations*,[18a] for constructing a solution of the consistency conditions. It is just as easy to describe this method in a spacetime of any even dimensionality as in four spacetime dimensions, so we will take the spacetime dimensionality to be $2n$. To start, one must imagine that at least two additional variables have been added to the $2n$ coordinates of space and time, in order to give meaning to the $(2n+2)$-form $\mathrm{Tr}\, F^{n+1}$. Note that

$$dF = -d(A^2) = -(dA)A + A(dA) = [A, F]\,, \qquad (22.6.22)$$

so that $\mathrm{Tr}\, F^{n+1}$ is closed:

$$d\mathrm{Tr}\, F^{n+1} = (n+1)\mathrm{Tr}\left\{(dF)F^n\right\} = \mathrm{Tr}\left\{[A, F^{n+1}]\right\} = 0\,. \qquad (22.6.23)$$

As long as the extended spacetime is simply connected, Poincaré's theorem then tells us that $\mathrm{Tr}\left\{F^{n+1}\right\}$ is exact, in the sense that there is a $(2n+1)$-form Ω_{2n+1} (known as the *Chern–Simons form*), for which

$$\mathrm{Tr}\left\{F^{n+1}\right\} = d\,\Omega_{2n+1}\,. \qquad (22.6.24)$$

Further, $\mathrm{Tr}\left\{F^{n+1}\right\}$ is manifestly gauge-invariant and depends only on the gauge field, so it is BRST-invariant:

$$s\mathrm{Tr}\left\{F^{n+1}\right\} = 0\,. \qquad (22.6.25)$$

The 'differentials' dx^μ are understood to anticommute with fermionic fields like the ghost field ω_α, so the operator d anticommutes with the operator s defined by Eqs. (22.6.7) and (22.6.8):

$$sd + ds = 0\,.$$

Because s is nilpotent, it follows then that $s\,\Omega_{2n+1}$ is also closed:

$$d(s\,\Omega_{2n+1}) = -s\mathrm{Tr}\left\{F^{n+1}\right\} = 0\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

Again using Poincaré's theorem, this means that there must be a $2n$-form Ω^1_{2n}, of first order in ω_α, for which

$$s\,\Omega_{2n+1} = d\Omega^1_{2n}\,. \tag{22.6.26}$$

Furthermore, $d(s\,\Omega^1_{2n}) = -s^2\Omega_{2n+1} = 0$, so there is also a $(2n-1)$-form Ω^2_{2n-1}, of second order in the ghost field, for which

$$s\,\Omega^1_{2n} = d\Omega^2_{2n-1}\,. \tag{22.6.27}$$

It follows then that the integral of Ω^1_{2n} over the $2n$ dimensions of space and time is BRST-invariant:

$$s\int_{\text{spacetime}} \Omega^1_{2n} = 0\,, \tag{22.6.28}$$

even though Ω^1_{2n} is not itself BRST-invariant. We can thus find a candidate $\int \Omega^1_{2n}$ for the anomaly functional $G[\omega, A]$ by integrating the two first-order differential equations $d\Omega_{2n+1} = \mathrm{Tr}\{F^{n+1}\}$ and $d\Omega^1_{2n} = s\,\Omega_{2n+1}$. General (not unique) solutions of these equations are

$$\Omega_{2n+1} = (n+1)\int_0^1 dt\ \mathrm{Tr}\left\{A\,F_t^n\right\}, \tag{22.6.29}$$

$$\Omega^1_{2n} = -(n+1)\sum_{r=0}^{n-1}\int_0^1 dt\,(1-t)\,\mathrm{Tr}\left\{\omega\,d(F_t^r\,A\,F_t^{n-1-r})\right\}, \tag{22.6.30}$$

where $F_t \equiv tF + (t-t^2)A^2$. Evaluation of the integral (22.6.30) shows that Eq. (22.6.20) gives a result for $G[\omega, A]$ proportional to $\int \Omega^1_4$ in the case of four spacetime dimensions.

We can continue this descent, and derive other useful results. In particular, it follows from Eq. (22.6.27) and the nilpotence of s that $d(s\Omega^2_{2n-1}) = 0$, so Poincaré's theorem tells us that $s\Omega^2_{2n-1}$ is of the form $d\Omega^3_{2n-2}$, so the integral of Ω^2_{2n-1} over the $2n-1$ coordinates of *space* is BRST-invariant:

$$s\int_{\text{space}} \Omega^2_{2n-1} = 0\,. \tag{22.6.31}$$

Such BRST-invariant functionals of second order in the ghost fields are candidates[18b] for so-called *Schwinger terms*.[18c] Schwinger terms of the sort that concern us here arise as anomalous terms $S_{\alpha,\beta}(\mathbf{x},\mathbf{y})$ in the equal-time commutation relations of the time components of two symmetry currents:

$$\left[J^0_\alpha(\mathbf{x},t),\,J^0_\beta(\mathbf{y},t)\right] = iC_{\alpha\beta\gamma}J^0_\gamma(\mathbf{x},t)\delta^{2n-1}(\mathbf{x}-\mathbf{y}) + S_{\alpha\beta}(\mathbf{x},\mathbf{y},t)\,. \tag{22.6.32}$$

(All operators in this paragraph are taken at the same time t, which will henceforth not be shown explicitly.) From the antisymmetry of the commutator we have $S_{\alpha\beta}(\mathbf{x},\mathbf{y}) = -S_{\beta\alpha}(\mathbf{y},\mathbf{x})$, so all information about

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

$S_{\alpha\beta}(\mathbf{x},\mathbf{y})$ is contained in the functional

$$S[\omega] \equiv \int d^{2n-1}x\, d^{2n-1}y\; \omega_\alpha(\mathbf{x})\,\omega_\beta(\mathbf{y})\,S_{\alpha\beta}(\mathbf{x},\mathbf{y})\,. \tag{22.6.33}$$

Note that $S_{\alpha\beta}(\mathbf{x},\mathbf{y})$ depends in general on various matter and gauge fields, so $S[\omega]$ generally depends on these fields as well as on the ghost field $\omega_\alpha(\mathbf{x})$. Taking the commutator of Eq. (22.6.32) with a third current $J^0_\gamma(\mathbf{z})$, contracting with $\omega_\alpha(\mathbf{x})\,\omega_\beta(\mathbf{y})\,\omega_\alpha(\mathbf{z})$, integrating over $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, and using the Jacobi identities, we find

$$0 = \int d^{2n-1}x\, d^{2n-1}y\, d^{2n-1}z\; \omega_\alpha(\mathbf{x})\,\omega_\beta(\mathbf{y})\,\omega_\gamma(\mathbf{z}) \times \left[iC_{\alpha\beta\delta}\,S_{\gamma\delta}(\mathbf{z},\mathbf{x})\delta^{2n-1}(\mathbf{x}-\mathbf{y}) + [J^0_\gamma(\mathbf{z}),\, S_{\alpha\beta}(\mathbf{x},\mathbf{y})]\right].$$

On functionals of gauge and matter fields like $S_{\alpha\beta}(\mathbf{x},\mathbf{y})$, the action of the BRST operator s is the same as that of a gauge transformation with transformation parameter ω_α, so

$$s\,S_{\alpha\beta}(\mathbf{x},\mathbf{y}) = i\Big[\int d^{2n-1}z\;\omega_\gamma(\mathbf{z})\,J^0_\gamma(\mathbf{z}),\; S_{\alpha\beta}(\mathbf{x},\mathbf{y})\Big]\,.$$

Recalling Eq. (22.6.8), we find that the functional (22.6.33) is BRST-invariant

$$s\,S[\omega] = 0\,. \tag{22.6.34}$$

Also, by adding terms to the currents we can change $S[\omega]$ by terms of the form $sT[\omega]$, so the set of possible Schwinger terms that could *not* be removed by adding terms to the currents is given by the cohomology of the BRST operator s at ghost number two — that is, by BRST-invariant functionals S of second order in the ghost field that are not themselves of the form sT. Eq. (22.6.31) shows that $\int \Omega^2_{2n-1}$ is a candidate for such a functional.

* * *

The analysis of anomalies given so far in this section is strictly applicable only to anomalies in one-loop order. It is true that a theory with one-loop anomalies in currents to which quantum gauge fields are coupled is inconsistent, and therefore does not need to be studied in higher orders. But the converse does not hold; if a theory with quantum gauge fields is anomaly-free in one-loop order, we still need to show that anomalies are absent in higher orders. Also, there is nothing inconsistent in theories with anomalies in global symmetries, such as the chiral symmetries of quantum chromodynamics, and for these we need to know whether higher-order corrections affect the anomalies.

Since BRST transformations act non-linearly on the fields, even in the absence of anomalies we would not necessarily expect the quantum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

effective action $\Gamma[\omega, A]$ to be BRST-invariant beyond the one-loop approximation. As we saw in Section 17.1, beyond this approximation we need to consider functionals not only of gauge and ghost fields but also of their antifields. (The introduction of antifields is sometimes important even in one-loop order for another reason: a local functional of fields alone that satisfies the Wess–Zumino consistency conditions, and that is not expressible as the BRST operator acting on a local functional of fields alone, will not be a candidate anomaly if it *can* be expressed as the antibracket of the action with a local functional of fields and antifields, because in this case the anomaly can be cancelled by subtracting this term from the action. This corresponds to a change in the action of the fields combined with a change in the gauge symmetry obeyed by this action.)

The study of anomalies with antifields included in the action turns out also to have a cohomological formulation.[20,21] The analysis of this problem is based on the Zinn-Justin version (17.1.10) of what Batalin and Vilkovisky called the master equation. In the absence of anomalies $(\Gamma, \Gamma) = 0$, so that in one-loop order $(S, \Gamma_1) = 0$, where S is the zeroth-order action and Γ_1 is the one-loop contribution to the quantum effective action. In the presence of anomalies we have instead

$$(S, \Gamma_1) = G_1 \,, \tag{22.6.35}$$

where G_1 is some functional of fields and antifields which, since S and Γ_1 have ghost number zero, must have ghost number unity. The action is assumed to satisfy the classical master equation $(S, S) = 0$, so the antibracket operation in Eq. (22.6.35) is nilpotent, and therefore $(S, G_1) = 0$. But if $G_1 = (S, F_1)$ for some local functional F_1 of ghost number zero then we can cancel the anomaly to one-loop order by subtracting the term F_1 (which is treated as a quantum correction of order $\hbar$) from the action and hence from Γ_1. (Of course, $G_1 = (S, \Gamma_1)$, but in the presence of massless particles Γ_1 is not a local functional.) Thus the candidate anomalies are those local functionals G_1 of ghost number unity that are closed, in the sense that they satisfy $(S, G_1) = 0$, but that are not exact, that is, that cannot be expressed as $G_1 = (S, F_1)$ for local functionals F_1 of ghost number unity. In other words, the candidate anomalies correspond to the cohomology of the antibracket operation $X \mapsto (S, X)$ at ghost number unity on the space of local functionals of the fields and antifields.

This is just like the result found earlier in this section, but with the antibracket $(S, \cdots)$ replacing the BRST operator s. If we set the antifields equal to zero in Eq. (22.6.35) and recall that $(\delta\Gamma_1/\delta\chi^n)_{\chi^\ddagger=0} = s\chi^n$, we find that the condition (22.6.35) in fact yields the condition $s\Gamma_1 = 0$, which as we have seen is equivalent to the requirement that Γ_1 satisfies the Wess–Zumino consistency conditions. But there is a sense in which the analysis based on Eq. (22.6.35) can be extended to higher orders.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

To see this, suppose that it is found that the antibracket operation $X \mapsto (S, X)$ has an empty cohomology at ghost number unity in the space of local functionals, and that we redefine the action as described above to make $G_1 = (S, \Gamma_1) = 0$. An anomaly that violates the master equation $(\Gamma, \Gamma) = 0$ in two-loop order is represented by a function G_2 for which

$$\left(\Gamma_1, \Gamma_1\right) + 2\left(S, \Gamma_2\right) = G_2 \, .$$

But since $(S, \Gamma_1) = 0$ and $(S, S) = 0$, any such G_2 would satisfy $(S, G_2) = 0$. On the assumption of an empty cohomology, this means that it can be expressed as $G_2 = (S, F_2)$ with F_2 a local functional of ghost number zero, so in this order the anomaly can be cancelled by subtracting F_2 from the action.

This argument can be extended to all orders. Suppose we have cancelled the anomalies in the master equation up to order $N - 1$, so that

$$0 = G_M = \sum_{L=0}^{M} \left(\Gamma_L, \Gamma_{M-L}\right) = 2\left(S, \Gamma_M\right) + \sum_{L=1}^{M-1} \left(\Gamma_L, \Gamma_{M-L}\right) ,$$

for all $M < N$. The Nth-order term in the antibracket (Γ, Γ) is likewise

$$G_N = 2\left(S, \Gamma_N\right) + \sum_{M=1}^{N-1} \left(\Gamma_M, \Gamma_{N-M}\right) ,$$

so, using the Jacobi identity (15.9.21) (in which for three bosonic operators all signs are $-$) and the above formula for (S, Γ_M), we find:

$$\left(S, G_N\right) = -2 \sum_{M=1}^{N-1} \left(\left(S, \Gamma_M\right), \Gamma_{N-M}\right) = \sum_{M=1}^{N-1} \sum_{L=1}^{M-1} \left(\left(\Gamma_L, \Gamma_{M-L}\right), \Gamma_{N-M}\right) .$$

This can be written in a more symmetric way

$$\left(S, G_N\right) = - \sum_{M_1=1}^{N-2} \sum_{M_2=1}^{N-2} \sum_{M_3=1}^{N-2} \delta_{N, M_1+M_2+M_3} \left(\Gamma_{M_1}, \left(\Gamma_{M_2}, \Gamma_{M_3}\right)\right) .$$

Since the ranges of M_1, M_2, and M_3 are the same, we can write the double antibracket in this sum as a sum over the 3! permutations over these indices, which vanishes according to the Jacobi identity (15.9.21), leading to the conclusion that $(S, G_N) = 0$. If as assumed the cohomology is empty then this implies that there is a local functional F_N for which $G_N = (S, F_N)$, so by subtracting F_N from the action the anomaly can be cancelled in order N, as was to be shown.

Using purely algebraic methods Barnich, Brandt, and Henneaux[22] have succeeded in showing that for Yang–Mills theories (in four spacetime dimensions) based on semisimple gauge groups, the cohomology of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

antibracket operation $X \mapsto (S, X)$ (at ghost number unity on the space of local functionals) consists entirely of a linear combination of terms of the form (22.6.20), one for each simple subgroup of the gauge group, with unknown coefficients.[††] This shows without any reference to the matter content of the theory that in one-loop order, where the anomaly G_1 automatically satisfies $(S, G_1) = 0$, the anomaly for a semi-simple gauge group must be a linear combination of terms of the form (22.6.20), with only the constant coefficient for each simple subgroup left to be determined by detailed calculations that take into account the matter content of the theory. Further, we saw in Section 22.4 that there are gauge groups for which the trace in Eq. (22.6.20) vanishes automatically for any menu of fermion fields. (These are the semisimple gauge groups with no $SU(n)$ factors with $n \geq 3$.) In such cases the theorem of Ref. 22 shows that the cohomology of the antibracket operation $X \mapsto (S, X)$ at ghost number unity is zero. As we have seen, this means that in such theories there is no anomaly in any order of perturbation theory.

There is another sense in which the anomaly is related to the cohomology of the antibracket operation.[23] In deriving the Slavnov–Taylor identity (16.4.6), we assumed that the measure $\prod_{n,x} d\chi^n(x)$ is invariant under the symmetry transformation in question. In Section 15.9 the Zinn-Justin equation was derived from the Slavnov–Taylor identity for the symmetry transformation $\chi^n \to \chi^n + \theta \delta S / \delta \chi^{\ddagger}_n$, so the derivation of the Zinn-Justin equation given in Section 15.9 breaks down unless $\prod_{n,x} d\chi^n(x)$ is invariant under this transformation, or in other words unless $\Delta S = 0$, where Δ is the operator (15.9.34). Where $\Delta S \neq 0$, it still may be possible to save the Zinn-Justin equation by adding local functionals to S that violate the classical master equation $(S, S) = 0$ in such a way as to cancel the effect of the non-invariance of the measure. It turns out that the condition for this cancellation is nothing but the quantum master equation (15.9.35). To construct an action S that satisfies this equation, we start with a zeroth-order action S_0 that satisfies the classical master equation, $(S_0, S_0) = 0$, and add quantum corrections. In the case in which the cohomology of the operation $X \to (S_0, X)$ (at ghost number unity in the space of local functionals) is empty, the same proof that was used above to show the absence in this case of anomalies to all orders can be used to show that a local functional can be added to S_0 so that the quantum master equation is satisfied to all orders.

[††] It is not necessary to specify the representation of the gauge algebra in which the trace in Eq. (22.6.20) is to be calculated, because this trace is the same up to a constant coefficient for all representations of a simple Lie algebra,[19a] and the constant coefficient is not determined anyway by this cohomology theorem.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

22.7 Anomalies and Goldstone Bosons

In the same 1971 paper in which they introduced the consistency conditions, Wess and Zumino[18] also noted that the possibility of anomalies has important consequences for the interactions of Goldstone bosons. To understand their point, it is helpful to apply the 'anomaly matching' argument developed by 't Hooft[16] in 1979, which has already been used in Section 22.5.

Consider a broken global symmetry group G that is realized linearly in some underlying theory of trapped massless fermions, as for instance the global chiral $SU(3) \times SU(3)$ symmetry in quantum chromodynamics with three massless quarks. In this underlying theory introduce fictitious gauge fields, so that aside from possible anomalies the global symmetry G becomes local. In general this local symmetry *will* be broken by anomalies, since in the real world the symmetry is purely global, and there is no reason why a global symmetry should admit an extension to an anomaly-free local symmetry. However these anomalies can be cancelled by adding suitable massless spectator fermions. As long as the gauge couplings introduced in this way are sufficiently small, and the spectator fermions have only these very weak gauge interactions, the dynamics of the theory will not be substantially changed by these modifications.

Next consider the effective field theory that describes physics at low energy, where the trapped fermions are unobservable. The only degrees of freedom in this theory will be the massless particles: the fictitious gauge bosons and spectator fermions, and a set of Goldstone bosons with fields ξ_a, one for each independent broken symmetry. Since the underlying theory had been made gauge-invariant and anomaly-free, the same must be true of the effective field theory. But the spectator fermions produce an anomaly which had previously cancelled the anomaly due to the trapped massless fermions in the underlying theory, so in order for them to cancel the anomaly due to the Goldstone bosons, *the gauged effective field theory of the Goldstone bosons must have an anomaly for the fictitious local symmetries which is equal to that produced by the trapped fermions in the underlying theory.* That is, in place of Eq. (22.6.2), the effective action $\Gamma[\xi, A]$ of the fictitious gauge fields and Goldstone bosons is subject to the condition

$$\mathscr{T}_\beta(x)\Gamma[\xi, A] = G_\beta[x; A] \,, \tag{22.7.1}$$

where $G_\beta[x; A]$ is the anomaly function of the underlying theory, in which there are no Goldstone bosons, and $\mathscr{T}_\beta$ is a generator of the gauge group G, now acting on both gauge and Goldstone boson fields. (The index β on $\mathscr{T}_\beta$ runs over values i labelling a complete set of independent generators $\mathscr{Y}_i$ of the unbroken symmetry subgroup H, together with values

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

a labelling a set of independent broken symmetry generators $\mathscr{X}_a$; there is one Goldstone boson field ξ_a for each $\mathscr{X}_a$.)

Of course, Eq. (22.7.1) may also be used to study the interactions of Goldstone bosons with real weakly coupled gauge fields. For instance, where the underlying theory includes electroweak gauge bosons coupled to the quarks, we would identify some of what we have called 'fictitious' gauge fields with the electroweak gauge fields. In such cases some of the 'spectator' fermions must also be real, in order to cancel the anomalies produced by loops of trapped fermions in the gauge symmetries of the real weakly coupled gauge fields, as for instance the leptons cancel the electroweak anomalies produced by the quarks.

Now let us turn to the implications of Eq. (22.7.1). To calculate the gauge symmetry generator $\mathscr{T}_\beta(x)$ in this context, note that under a general group transformation $g = \exp(-i\int \zeta_\beta(x)\mathscr{T}_\beta(x)\,d^4x)$, the Goldstone boson fields $\xi_a(x)$ transform into fields $\xi'_a(x)$ given by Eq. (19.6.18), and the gauge fields $A^\mu_\beta(x)$ transform into gauge-transformed fields $A'^\mu_\beta(x)$, so that

$$\mathscr{T}_\beta(x) = \mathscr{T}^A_\beta(x) + \mathscr{T}^\xi_\beta(x)\,. \tag{22.7.2}$$

Here $\mathscr{T}^A_\beta(x)$ acts on the gauge fields and is given by Eq. (22.6.1):

$$-i\mathscr{T}^A_\beta(x) = -\frac{\partial}{\partial x^\mu}\frac{\delta}{\delta A_{\beta\mu}(x)} - C_{\beta\gamma\alpha}A_{\gamma\mu}(x)\frac{\delta}{\delta A_{\alpha\mu}(x)}\,, \tag{22.7.3}$$

where $C_{\alpha\beta\gamma}$ are the totally antisymmetric structure constants of the gauge group G, while $\mathscr{T}^\xi_\beta(x)$ acts on the Goldstone fields and is given by the infinitesimal limit of Eq. (19.6.17), which (for the exponential parameterization $\gamma(\xi) = \exp(i\xi_a X_a)$) reads:*

$$T_\beta \exp(i\xi_a(x)X_a) = -\mathscr{T}^\xi_\beta(x)\exp(i\xi_a(x)X_a) + \exp(i\xi_a(x)X_a)\,\theta_{\beta i}(x)Y_i\,. \tag{22.7.4}$$

Here T_β are matrices representing the generators of G in any representation; they are divided into sets X_a and Y_i, which represent the broken and unbroken symmetry generators, respectively. Also the $\theta_{\beta i}(x)$ are ξ-dependent functions whose form will not concern us here.

About the anomaly functions $G_\beta[x;A]$, we will assume only the consistency conditions (22.6.6)

$$\mathscr{T}_\alpha(x)G_\beta[y;A] - \mathscr{T}_\beta(y)G_\alpha[x;A] = iC_{\alpha\beta\gamma}\delta^4(x-y)G_\gamma[y;A] \tag{22.7.5}$$

* The minus sign in the first term on the right-hand side of Eq. (22.7.4) and the factor $-i$ on the left-hand side of Eq. (22.7.3) appear because it is $\exp[-i\int \Lambda_\beta(x)\mathscr{T}_\beta(x)]$ that induces a gauge transformation (15.1.17) with gauge parameter Λ_β, as can be seen by requiring the $\mathscr{T}_\beta(x)$ to satisfy the commutation relations (22.6.5).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

and the absence of anomalies in the unbroken symmetries

$$G_i[x;A] = 0\,. \tag{22.7.6}$$

As mentioned in the previous section, as long as the trace $\mathrm{Tr}\,[T_i\{T_j, T_k\}]$ vanishes for the generators of unbroken symmetry subgroup (as it does for the non-chiral generators of $SU(3) \times SU(3)$), it is always possible to add a local functional to the action so that Eq. (22.7.6) is satisfied.

Under the assumptions (22.7.5) and (22.7.6), it is always possible to find a solution of the anomalous Slavnov–Taylor identities (22.7.1):

$$\Gamma[\xi, A] = -i\int_0^1 dt \int \xi_b(y)\, G_b[y; A_{-t\xi}]\, d^4y\,, \tag{22.7.7}$$

where $[A_{-t\xi}(x)]_\mu$ is the result of acting on $A_\mu \equiv T_\beta A_{\mu\beta}$ with a gauge transformation (15.1.17) having $\Lambda_a = -t\xi_a$ and $\Lambda_i = 0$:

$$\begin{aligned}[A_{-t\xi}(x)]_\mu = {} & \exp\left(-itX_a\xi_a(x)\right) A_\mu(x) \exp\left(itX_a\xi_a(x)\right) \\ & -i\Big[\partial_\mu \exp\left(-itX_a\xi_a(x)\right)\Big] \exp\left(itX_a\xi_a(x)\right)\,. \end{aligned} \tag{22.7.8}$$

In contrast with the case in which unbroken symmetries have anomalies, Eqs. (22.7.7) and (22.7.8) define a *local* (though complicated) functional of gauge and Goldstone boson fields. Any other solution of Eq. (22.7.1) will differ from this one by an anomaly-free functional.

Here is an outline of the proof[24] that the action (22.7.7) satisfies Eq. (22.7.1). Instead of working with the local generator $\mathscr{T}_\beta(x)$, it is convenient to introduce an arbitrary function $\eta_\beta(x)$, and define

$$\mathscr{T}[\eta] = \int d^4x\, \eta_\beta(x)\mathscr{T}_\beta(x)\,. \tag{22.7.9}$$

To evaluate $\mathscr{T}[\eta]\xi_b(x)$, we introduce the matrix

$$\eta_{-t\xi}(x) \equiv \exp\Big(-iX_a\xi_a(x)t\Big)\Big[\eta(x)+\mathscr{T}[\eta]\Big] \exp\Big(iX_a\xi_a(x)t\Big) \equiv [\eta_{-t\xi}(x)]_\beta T_\beta\,, \tag{22.7.10}$$

where $\eta(x) \equiv \eta_\beta(x)T_\beta$. Then

$$\frac{\partial}{\partial t}\eta_{-t\xi}(x) = -i\Big[X_a\xi_a(x), \eta_{-t\xi}(x)\Big] + i\Big(\mathscr{T}[\eta]\xi_a(x)\Big)X_a\,,$$

so that

$$\mathscr{T}[\eta]\xi_b(x) = -i\frac{\partial}{\partial t}[\eta_{-t\xi}(x)]_b + iC_{a\gamma b}\,\xi_a(x)\,[\eta_{-t\xi}(x)]_\gamma\,. \tag{22.7.11}$$

To evaluate $\mathscr{T}[\eta]G_b[y, A]$, we apply $\mathscr{T}[\eta]$ to the gauge field, and after a straightforward calculation find that

$$\mathscr{T}[\eta][A_{-t\xi}(x)]_\mu = \Big(\mathscr{T}^A[\eta_{-t\xi}]A_\mu(x)\Big)_{A\to A_{-t\xi}}\,, \tag{22.7.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

so, using the consistency condition (22.7.5),

$$\begin{aligned}\mathscr{T}[\eta]G_b[y;A_{-t\xi}] &= \int d^4x\,[\eta_{-t\xi}(x)]_\gamma \left(\mathscr{T}^A_\gamma(x)\,G_b[y;A]\right)_{A\to A_{-t\xi}} \\ &= \int d^4x\,[\eta_{-t\xi}(x)]_a \left(\mathscr{T}^A_b(y)\,G_a[x;A]\right)_{A\to A_{-t\xi}} \\ &\quad + iC_{\gamma ba}\,[\eta_{-t\xi}(y)]_\gamma G_a[y;A_{-t\xi}]\,. \end{aligned} \tag{22.7.13}$$

The structure constant terms in Eqs. (22.7.11) and (22.7.13) cancel, yielding

$$\begin{aligned}\mathscr{T}[\eta]\Gamma[\xi,A] = \int_0^1 dt \int d^4y \Bigg\{ &-\left[\frac{\partial}{\partial t}[\eta_{-t\xi}(y)]_b\right] G_b[y;A_{-t\xi}] \\ &-i[\eta_{-t\xi}(y)]_a \left(\mathscr{T}[\xi]\,G_a[y;A]\right)_{A\to A_{-t\xi}}\Bigg\}\,. \end{aligned} \tag{22.7.14}$$

Another straightforward calculation shows that

$$\frac{\partial}{\partial t}[A_{-t\xi}(x)]_\mu = i\left(\mathscr{T}^A(\xi)A_\mu(x)\right)_{A\to A_{-t\xi}}\,, \tag{22.7.15}$$

so the terms in the integrand in Eq. (22.7.14) add up to a t-derivative

$$\begin{aligned}\mathscr{T}[\eta]\Gamma[\xi,A] &= -\int_0^1 dt \int d^4y\,\frac{\partial}{\partial t}\left\{[\eta_{-t\xi}(y)]_b\,G_b[y;A_{-t\xi}]\right\} \\ &= -\int d^4y \left[[\eta_{-t\xi}(y)]_b\,G_b[y;A_{-t\xi}]\right]_{t=0}^{t=1}\,. \end{aligned} \tag{22.7.16}$$

At $t=1$ we have

$$\eta_{-\xi}(x) = \exp\left(-iX_a\xi_a(x)\right)\left[\eta(x)+\mathscr{T}[\eta]\right]\exp\left(iX_a\xi_a(x)\right)\,.$$

From Eq. (22.7.4) we see that this is a linear combination of the generators of the unbroken symmetry subgroup H, so the coefficient $[\eta_{-\xi}(y)]_b$ of any broken symmetry generator X_b vanishes. Also, Eqs. (22.7.10) and (22.7.8) show immediately that at $t=0$, $\eta_{-t\xi}(y)=\eta(y)$ and $[A_{-t\xi}(y)]_\mu = A_\mu(y)$, so Eq. (22.7.16) yields

$$\mathscr{T}[\eta]\Gamma[\xi,A] = \int d^4y\,[\eta_0(y)]_b\,G_b[y;A_0] = \int d^4y\,\eta_b(y)\,G_b[y;A]\,, \tag{22.7.17}$$

which with Eq. (22.7.6) is equivalent to the desired result (22.7.1), as was to be proved.

The solution (22.7.7) of Eq. (22.7.1) is not unique, but it *is* the unique

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

solution that vanishes for $\xi = 0$. To see this, note that

$$\exp\left[-i\int \eta_\beta(x)\mathscr{T}_\beta(x)\,d^4x\right]\Gamma[\xi,A] = \Gamma[\xi',A']\,, \tag{22.7.18}$$

where the primes indicate a gauge transformation with transformation parameter η_β. It is convenient to represent the exponential here as

$$\exp(z) = 1 + \int_0^1 dt\,\exp(zt)\,z\,,$$

so Eqs. (22.7.1) and (22.7.18) yield

$$\Gamma[\xi,A] - i\int_0^1 dt\,\exp\left[-it\int \eta_\beta(x)\mathscr{T}^A_\beta(x)\,d^4x\right]\int \eta_b(y)G_b[y;A]d^4y = \Gamma[\xi',A']\,. \tag{22.7.19}$$

In particular, if we take $\eta_a = -\xi_a$ and $\eta_i = 0$, then $\xi'_a = 0$, in which case by assumption Eq. (22.7.19) has a vanishing right-hand side, and therefore yields the formula:

$$\Gamma[\xi,A] = -i\int_0^1 dt\,\exp\left[it\int \xi_a(x)\mathscr{T}^A_a(x)\,d^4x\right]\int \xi_b(y)G_b[y;A]\,d^4y\,. \tag{22.7.20}$$

The functional operator $\exp[it\int \xi_a(x)\mathscr{T}^A_a(x)\,d^4x]$ in Eq. (22.7.20) simply produces a gauge transformation (15.1.17) with gauge parameter $\Lambda_\beta(x) = -t\xi_a(x)$, so that Eq. (22.7.20) may be written as in Eq. (22.7.7).

Eq. (22.7.7) may be applied to study the electroweak interactions of the octet of pseudoscalar Goldstone bosons, but it has important implications for the interactions of the Goldstone bosons themselves, in the absence of real gauge fields. In the case where $A = 0$, Eq. (22.7.8) becomes a 'pure gauge' field

$$\begin{aligned}[A_{-t\xi}(x)]_\mu &= -i\left[\partial_\mu \exp\left(-itX_a\xi_a(x)\right)\right]\exp\left(itX_a\xi_a(x)\right)\\ &= -i\left[\partial_\mu V\left(t\xi(x)\right)\right]\,V^{-1}\left(t\xi(x)\right)\,,\end{aligned} \tag{22.7.21}$$

where

$$V\left(t\xi(x)\right) \equiv \exp\left(-itX_a\xi_a(x)\right)\,. \tag{22.7.22}$$

With $A^\mu_\alpha(x) = 0$, Eq. (22.7.1) gives

$$\mathscr{T}_\beta(x)\Gamma[\xi,0] = 0\,, \tag{22.7.23}$$

so the result of using Eq. (22.7.21) in Eq. (22.7.7) is a G-invariant local functional of the Goldstone boson field $\xi_a(x)$, though, as we shall see, it is not in general the integral over spacetime of a G-invariant function of $\xi_a(x)$ and its derivatives.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

The simplest example is the case of a completely broken symmetry group. Here the condition (22.7.6) is empty, and we can use the symmetric form (22.3.38) for the anomaly:

$$G_a[x;A] = \frac{i}{24\pi^2}\epsilon^{\kappa\nu\lambda\rho}\,\mathrm{Tr}\left\{T_a\left[\partial_\kappa A_\nu(x)\,\partial_\lambda A_\rho(x) - \tfrac{1}{2}i\partial_\kappa A_\nu(x)\,A_\lambda(x)A_\rho(x)\right.\right.$$
$$\left.\left. + \tfrac{1}{2}iA_\kappa(x)\,\partial_\nu A_\lambda(x)\,A_\rho(x) - \tfrac{1}{2}iA_\kappa(x)A_\nu(x)\partial_\lambda A_\rho(x)\right]\right\}, \quad (22.7.24)$$

where now T_a is the specific representation of the group generator furnished by the left-handed fermions (including antifermions, where the distinction is relevant) of the theory. In this case when we use Eq. (22.7.21) in Eq. (22.7.24), we find that the only terms in the trace in (22.7.24) that survive when contracted with $\epsilon^{\kappa\nu\lambda\rho}$ are then all proportional to

$$\mathrm{Tr}\left\{T_a(\partial_\kappa V)V^{-1}(\partial_\nu V)V^{-1}(\partial_\lambda V)V^{-1}(\partial_\rho V)V^{-1}\right\},$$

with coefficients -1, $+\frac{1}{2}$, $-\frac{1}{2}$ and $+\frac{1}{2}$, respectively, so Eq. (22.7.7) becomes here

$$\Gamma[\xi,0] = -\frac{1}{48\pi^2}\epsilon^{\kappa\nu\lambda\rho}\int d^4y\;\xi_a(y)\int_0^1 dt\;\mathrm{Tr}\left\{T_a\left[\partial_\kappa V\left(t\xi(y)\right)\right]\right.$$
$$\times V^{-1}\left(t\xi(y)\right)\left[\partial_\nu V\left(t\xi(y)\right)\right]V^{-1}\left(t\xi(y)\right)\left[\partial_\lambda V\left(t\xi(y)\right)\right]$$
$$\left.\times V^{-1}\left(t\xi(y)\right)\left[\partial_\rho V\left(t\xi(y)\right)\right]V^{-1}\left(t\xi(y)\right)\right\}. \quad (22.7.25)$$

As promised, this is not the integral of an invariant function of fields and field derivatives. For instance, for small Goldstone boson fields Eq. (22.7.25) becomes

$$\Gamma[\xi,0] = -\frac{1}{240\pi^2}\epsilon^{\kappa\nu\lambda\rho}\,\mathrm{Tr}\,\{T_aT_bT_cT_dT_e\}\int d^4y\;\xi_a\,\partial_\kappa\xi_b\,\partial_\nu\xi_c\,\partial_\lambda\xi_d\,\partial_\rho\xi_e$$
$$+O(\xi^6)\,. \quad (22.7.26)$$

Any function of the covariant derivatives of Goldstone boson fields would have a term of lowest order in the fields which would be simply a product of partial derivatives of the Goldstone fields, and therefore could not have a lowest-order term of the form (22.7.26).

For an example of greater practical importance, consider the $SU(3) \times SU(3)$ chiral symmetry of quantum chromodynamics with massless u, d, and s quarks, spontaneously broken to the diagonal $SU(3)$ subgroup of Gell-Mann and Ne'eman. In order to use the results of this section, we *must* label internal momenta in the fermion loop integral so that the vector currents of the diagonal $SU(3)$ subgroup are anomaly-free, in which case

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

the anomaly takes the Bardeen form (22.3.34):

$$G_a[V,A] = \frac{in}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\left\{t_a\left[V_{\mu\nu}V_{\rho\sigma} + \tfrac{1}{3}A_{\mu\nu}A_{\rho\sigma} - \tfrac{32}{3}A_\mu A_\nu A_\rho A_\sigma \right.\right.$$
$$\left.\left. + \tfrac{8}{3}i(A_\mu A_\nu V_{\rho\sigma} + A_\mu V_{\rho\sigma}A_\nu + V_{\rho\sigma}A_\mu A_\nu)\right]\right\}, \qquad (22.7.27)$$

where V_μ, A_μ, $V_{\mu\nu}$, and $A_{\mu\nu}$ are defined by Eqs. (22.3.35)–(22.3.37); t_a is here half the Gell-Mann matrix λ_a, given by Eq. (19.7.2); and n is the number of quark types ('colors') of each flavor. (We are now using a lower case t for the matrices representing the group generators, because in this trace we sum only over the left-handed quarks, not the left-handed antiquarks.) For a pure gauge field like (22.7.21) the field strengths $V_{\mu\nu}$ and $A_{\mu\nu}$ vanish, so that using Eq. (22.7.27) in Eq. (22.7.7) yields the anomalous effective Goldstone boson action

$$\Gamma[\xi,0] = -\frac{2n}{3\pi^2}\epsilon^{\mu\nu\rho\sigma}\int_0^1 dt \int d^4x\ \mathrm{Tr}\left\{\xi_a t_a [A_{-t\xi}]_\mu [A_{-t\xi}]_\nu [A_{-t\xi}]_\rho [A_{-t\xi}]_\sigma\right\}, \qquad (22.7.28)$$

where 'A' now denotes the axial-vector as opposed to the vector gauge field. To find $[A_{-t\xi}]_\mu$ we use Eq. (22.7.21), which here reads

$$[V_{-t\xi}(x)]_\mu + \gamma_5[A_{-t\xi}(x)]_\mu = -i\left[\partial_\mu \exp\left(-it\gamma_5 t_a \xi_a(x)\right)\right]\exp\left(it\gamma_5 t_a \xi_a(x)\right). \qquad (22.7.29)$$

Multiplying with $(1+\gamma_5)/2$ and $(1-\gamma_5)/2$ and taking the difference yields the axial-vector term

$$\begin{aligned}[A_{-t\xi}(x)]_\mu &= -\tfrac{1}{2}i\left[\partial_\mu \exp\left(-it\xi_a t_a\right)\right]\exp\left(it\xi_a t_a\right) \\ &\quad + \tfrac{1}{2}i\left[\partial_\mu \exp\left(it\xi_a t_a\right)\right]\exp\left(-it\xi_a t_a\right) \\ &= \tfrac{1}{2}i\exp\left(it\xi_a t_a\right)U^{-1}\left(t\xi(x)\right)\left[\partial_\mu U\left(t\xi(x)\right)\right]\exp\left(-it\xi_a t_a\right),\end{aligned} \qquad (22.7.30)$$

where

$$U\left(t\xi\right) \equiv \exp\left(2it_a\xi_a t\right). \qquad (22.7.31)$$

Using this in Eq. (22.7.28) yields the effective anomalous action

$$\Gamma[\xi,0] = -\frac{n}{24\pi^2}\epsilon^{\mu\nu\rho\sigma}\int_0^1 dt\int d^4x\ \mathrm{Tr}\left\{\xi_a t_a U^{-1}\left(t\xi(x)\right)\left[\partial_\mu U\left(t\xi(x)\right)\right]\right.$$
$$\times U^{-1}\left(t\xi(x)\right)\left[\partial_\nu U\left(t\xi(x)\right)\right]U^{-1}\left(t\xi(x)\right)\left[\partial_\rho U\left(t\xi(x)\right)\right]U^{-1}\left(t\xi(x)\right)$$
$$\left.\times\left[\partial_\sigma U\left(t\xi(x)\right)\right]\right\}. \qquad (22.7.32)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

As noted by Witten,[25] this may be expressed in a convenient five-dimensional form. We take t as our fifth coordinate, and define $\xi_a(x,t) \equiv t\xi_a(x)$. Then Eq. (22.7.32) becomes

$$\Gamma[\xi,0] = -\frac{in}{240\pi^2}\epsilon^{ijklm}\int d^5z\ \mathrm{Tr}\left\{U^{-1}\Big(\xi(z)\Big)\Big[\partial_i U\Big(\xi(z)\Big)\Big]\right.$$
$$\times U^{-1}\Big(\xi(z)\Big)\Big[\partial_j U\Big(\xi(z)\Big)\Big]U^{-1}\Big(\xi(z)\Big)\Big[\partial_k U\Big(\xi(z)\Big)\Big]U^{-1}\Big(\xi(z)\Big)$$
$$\left.\times\Big[\partial_l U\Big(\xi(z)\Big)\Big]U^{-1}\Big(\xi(z)\Big)\Big[\partial_m U\Big(\xi(z)\Big)\Big]\right\}, \tag{22.7.33}$$

where i, j, etc. run over the values 1, 2, 3, 0, 5, with $z^i = x^i$ for $i = 1, 2, 3, 0$ and $z^5 = t$, and the integral is over the region $0 \leq z^5 \leq 1$. (An extra factor $1/5$ appears in Eq. (22.7.33) to take account of the fact that any one of the five indices i, j, k, l, or m may take the value 5.) Since $\xi_a(z)$ takes a fixed value zero for $z^5 = 0$ and all values of the other components z^μ of z^i, we can identify these values of z^i as a single point, and consider the region of integration in Eq. (22.7.33) as a five-dimensional ball, with the four-dimensional boundary $z^5 = 1$ taken as ordinary spacetime. Eq. (22.7.33) is thus a special case of the 'Wess–Zumino–Witten' action given by Eqs. (19.8.1) and (19.8.3), which was also proportional to an integer n; the only difference is that n is now identified as the number of colors. We saw in Section 19.8 that the integral (19.8.3) depends only on the values of $\xi_a(z)$ on the spacetime boundary of the five-ball, so in deriving Eq. (22.7.33) we have shown that Eq. (19.8.2) applies for any continuation of $\xi_a(x)$ into the interior of the five-ball, not just for $\xi_a(x,t) = t\xi(x)$.

* * *

More generally, consider an arbitrary gauge group G that is spontaneously broken to a subgroup H that by itself is anomaly-free — that is, for which the D-symbol (22.3.12) vanishes for any three generators of H. Chu, Ho, and Zumino[26] have shown that we can add a local functional $B[A]$ to the action, in such a way that the currents of the subgroup H are anomaly free, even when the gauge fields of the broken symmetries are taken into account. This functional is

$$B[A] = \frac{1}{48\pi^2}\epsilon_{\mu\nu\rho\sigma}\int d^4x\ \mathrm{Tr}\left\{[A_h^\mu, A^\nu](F^{\rho\sigma} + F_h^{\rho\sigma}) + A^\mu A_h^\nu A_h^\rho A_h^\sigma\right.$$
$$\left.-A_h^\mu A^\nu A^\rho A^\sigma + \frac{1}{2}A_h^\mu A^\nu A_h^\rho A^\sigma\right\}, \tag{22.7.34}$$

where again $A^\mu \equiv T_\alpha A_\alpha^\mu$ and $F^{\mu\nu} \equiv T_\alpha F_\alpha^{\mu\nu}$, while $A_h^\mu \equiv T_i A_i^\mu$ and $F_h^{\mu\nu} \equiv T_i F_i^{\mu\nu}$ are the terms in A^μ and $F^{\mu\nu}$ in the algebra of the unbroken symmetry

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

subgroup H. The total anomaly is then

$$G'_\beta[x;A] = G_\beta[x;A] + \mathscr{T}_\beta(x)B[A] \tag{22.7.35}$$

(where $G_\beta[x;A]$ is the symmetrized anomaly (22.3.38)) and satisfies the desired condition

$$G'_i[x;A] = 0\,. \tag{22.7.36}$$

The anomalous part of the effective action for Goldstone bosons and gauge fields is given by using Eq. (22.7.35) in place of G_β in Eq. (22.7.7). In this way, Chu *et al.*[26] found an anomalous effective action

$$\Gamma'[\xi,A] = \Gamma[\xi,A] - B[A_{-\xi}] + B[A]\,, \tag{22.7.37}$$

where $\Gamma[\xi,A]$ is the previously derived effective action (22.7.7), and $A_{-\xi}(x)$ is obtained by setting $t=1$ in Eq. (22.7.8). In particular, where the gauge field vanishes $A_{-\xi}$ is the pure gauge field

$$[A_{-\xi}(x)]_\mu = -i\Big[\partial_\mu V\Big(\xi(x)\Big)\Big]V^{-1}\Big(\xi(x)\Big)\,,$$
$$V\Big(\xi(x)\Big) = \exp\Big(-iX_a\xi_a(x)\Big)\,,$$

so here the Goldstone boson action is

$$\Gamma'[\xi,0] = \Gamma[\xi,0] - \frac{1}{48\pi^2}\,\epsilon_{\mu\nu\rho\sigma}\int d^4x\ \mathrm{Tr}\Big\{A^\mu_{-\xi}\,A^\nu_{-\xi\,h}\,A^\rho_{-\xi\,h}\,A^\sigma_{-\xi\,h}$$
$$-A^\mu_{-\xi\,h}\,A^\nu_{-\xi}\,A^\rho_{-\xi}\,A^\sigma_{-\xi} + \frac{1}{2}A^\mu_{-\xi\,h}\,A^\nu_{-\xi}\,A^\rho_{-\xi\,h}\,A^\sigma_{-\xi}\Big\}\,. \tag{22.7.38}$$

This result is not unique; in particular, in parity-conserving theories like quantum chromodynamics we can add additional local terms to the effective action to cancel any parity-non-conserving terms in Eq. (22.7.38).

Problems

1. Calculate the rate of the decay process $\eta \to \gamma + \gamma$, to leading order in m_s, with $m_u = m_d = 0$.

2. Consider a chiral $SU(3)$ symmetry under which the left-handed parts of the spin $\frac{1}{2}$ fields of a fermion-number-conserving theory form N defining representations **3** of $SU(3)$, while the right-handed parts are all singlets. Evaluate the anomaly in the $SU(3)$ symmetry. What is the anomaly if we add M fermion fields whose left-handed parts are singlets, and whose right-handed parts transform as symmetric traceless second-rank $SU(3)$ tensors?

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

3. Find a solution of the 't Hooft anomaly matching conditions (22.5.5) and (22.5.6) for the case of $n = 4$ flavors. Find a solution for $n = 2$ flavors other than the one given in the text.

4. Derive the Zinn-Justin equation from the quantum master equation, without assuming that $\Delta S = 0$.

References

1. J. Steinberger, *Phys. Rev.* **76**, 1180 (1949); Also see R. J. Finkelstein, *Phys. Rev.* **72**, 415 (1949); H. Fukuda and Y. Miyamoto, *Prog. Theor. Phys.* **4**, 347 (1949); J. Schwinger, *Phys. Rev.* **82**, 664 (1951); L. Rosenberg, *Phys. Rev.* **129**, 2786 (1963). A rough estimate of this decay rate was made by S. Sakata and Y. Tanikawa, *Phys. Rev.* **57**, 548 (1940), before the experimental discovery of the π^0 meson.

2. D. G. Sutherland, *Nucl. Phys.* **B2, 433** (1967)

3. M. Veltman, *Proc. Roy. Soc.* **A301**, 107 (1967).

4. J. S. Bell and R. Jackiw, *Nuovo Cimento* **60A**, 47 (1969). Also see R. Jackiw, in *Lectures on Current Algebra and its Applications*, (Princeton Press, Princeton, 1972).

5. S. Adler, *Phys. Rev.* **177**, 2426 (1969). Also see S. Adler, in *Lectures on Elementary Particles and Quantum Field Theory – 1970 Brandeis University Summer Institute in Theoretical Physics*, eds. S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, MA, 1970): Volume I.

6. K. Fujikawa, *Phys. Rev. Lett.* **42**, 1195 (1979).

6*a*. M. F. Atiyah and I. M. Singer, *Proc. Nat. Acad. Sci.* **81**, 2597 (1984).

7. L. Alvarez-Gaumè and P. Ginsparg, *Ann. Phys.* **161**, 423 (1985).

7*a*. For a more detailed argument, see Y. Frishman, A. Schwimmer, T. Banks, and S. Yankielowicz, *Nucl. Phys.* **B177**, 157 (1981).

8. W. A. Bardeen, *Phys. Rev.* **184**, 1848 (1969).

8*a*. See, e.g., B. Zumino, Y–S. Wu, and A. Zee, *Nucl. Phys.* **B239**, 477 (1984).

9. S. L. Adler and W. A. Bardeen, *Phys. Rev.* **182**, 1517 (1969).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

10. See, e.g., H. Georgi, *Lie Algebras in Particle Physics* (Benjamin–Cummings, Reading, MA, 1982): pp. 15, 198. The original reference is I. Schur, *Sitz. Preuss. Akad.*, p. 406 (1905).

10*a*. M. F. Atiyah and I. M. Singer, Ref. 6a; B. Zumino, in *Relativity, Groups, and Topology II*, eds. B. S. De Witt and R. Stora (North-Holland, Amsterdam, 1984); L. Alvarez-Gaumé and P. Ginsparg, Ref. 7; W. Bardeen and B. Zumino, *Nucl. Phys.* **B244**, 421 (1984); R. Stora, in *Progress in Gauge Field Theory*, eds. G. 't Hooft *et al.* (Plenum, New York, 1984): 543; O. Alvarez, I. Singer, and B. Zumino, *Comm. Math. Phys* **96**, 409 (1984); L. Alvarez-Gaumé and P. Ginsparg, *Nucl. Phys.* **B262**, 439 (1985); J. Mañes, R. Stora, and B. Zumino, *Commun. Math. Phys.* **102**, 157 (1985); B. Zumino, *Nucl. Phys.* **B253**, 477 (1985); J. M. Bismut and D. S. Freed, *Commun. Math. Phys.* **106**, 159 (1986); **107**, 103 (1986); D. S. Freed, *Commun. Math. Phys.* **107**, 483 (1986).

10*b*. E. Witten, *Phys. Lett.* **117 B**, 324 (1982).

11. D. J. Gross and R. Jackiw, *Phys. Rev.* **96**, 477 (1969).

12. H. Georgi and S. L. Glashow, *Phys. Rev.* **D6**, 429 (1972).

13. M. L. Mehta, *J. Math. Phys.* **7**, 1824 (1966); M. L. Mehta and P. K. Srivastava, *J. Math. Phys.* **7**, 1833 (1966).

13*a*. The cancellation of anomalies in the four-quark version of the standard electroweak theory was shown by C. Bouchiat, J. Iliopoulos, and Ph. Meyer, *Phys. Lett* **38B**, 519 (1972); S. Weinberg, in *Fundamental Interactions in Physics and Astrophysics*, eds. G. Iverson *et al.* (Plenum Press, New York, 1973): p. 157.

14. H. Georgi, in *Particles and Fields — 1974*, ed. C. Carlson (Amer. Inst. of Physics, New York, 1975).

15. R. Delbourgo and A. Salam, *Phys. Lett.* **40B**, 381 (1972); T. Eguchi and P. Freund, *Phys. Rev. Lett.* **37**, 1251 (1976). Also see N. K. Nielsen, M. T. Grisaru, R. Romer, and P. van Nieuwenhuizen, *Nucl. Phys.* **B140**, 477 (1978); M. J. Perry, *Nucl. Phys.* **B143**, 114 (1978); S. W. Hawking and C. Pope, *Nucl. Phys.* **B146**, 381 (1978); S. M. Christensen and M. J. Duff, *Phys. Lett.* **76B**, 571 (1978); R. Critchley, *Phys. Lett.* **78B**, 410 (1978); A. J. Hanson and H. Romer, *Phys. Lett.* **80B**, 58 (1978). For a general review, see T. Eguchi, P. B. Gilkey, and A. J. Hanson, *Phys. Rep.* **66**, 213 (1980). In $4n + 2$ dimensions there is also an anomaly in the divergence of the energy momentum tensor; see L. Alvarez-Gaumè and E. Witten, *Nuc. Phys.* **B234**, 269 (1984).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

16. G. 't Hooft, lecture given at the Cargèse Summer Institute, 1979, in *Recent Developments in Gauge Theories*, eds. G 't Hooft *et al.* (Plenum, New York, 1980), reprinted in *Dynamical Gauge Symmetry Breaking*, eds. E. Farhi and R. Jackiw (World Scientific, Singapore, 1982), and in G. 't Hooft, *Under the Spell of the Gauge Principle* (World Scientific, Singapore, 1994). Also see S. Dimopoulos, S. Raby, and L. Susskind, *Nucl. Phys.* **B173**, 208 (1980); S. Coleman and E. Witten, *Phys. Rev. Lett.* **45**, 1000 (1980); Y. Frishman, A. Schwimmer, T. Banks, and S. Yankielowicz, *Nucl. Phys.* **B177**, 157 (1981); A. Zee, Univ. of Pennsylvania report, 1980 (unpublished); R. Barbieri, L. Maiani, and R. Petronzio, *Phys. Lett.* **96B**, 63 (1980); G. Farrar, *Phys. Lett.* **96B**, 273 (1980); R. Chanda and P. Roy *Phys. Lett.* **99B**, 453 (1981).

16*a*. S. Weinberg and E. Witten, *Phys. Lett.* **96B**, 59 (1980).

17. J. Preskill and S. Weinberg, *Phys. Rev.* **D24**, 1059 (1981).

18. J. Wess and B. Zumino, *Phys. Lett.* **37B**, 95 (1971).

18*a*. R. Stora, in *Progress in Gauge Field Theory*, eds. G 't Hooft *et al.* (Plenum, New York, 1984):543 ; B. Zumino, in *Relativity, Groups and Topology II*, eds. B. S. De Witt and R. Stora (Elsevier, Amsterdam, 1984):1293; J. Mañes, R. Stora, and B. Zumino, *Commun. Math. Phys.* **102**, 157 (1985).

18*b*. L.D. Faddeev, *Phys. Lett.* **145B**, 81 (1984); B. Zumino, *Nucl. Phys.* **B253**, 477 (1985).

18*c*. J. Schwinger, *Phys. Rev. Lett.* **3**, 296 (1959).

19. J. A. Dixon, unpublished preprints (1976–9); *Commun. Math. Phys.* **139**. 495 (1991); F. Brandt, N. Dragon, and M. Kreuzer, *Nucl. Phys.* **B332**, 224 (1990); M. Dubois-Violette, M. Henneaux, M. Talon, and C. M. Viallet, *Phys. Lett.* **B289**, 361 (1992); M. Dubois-Violette, M. Talon, and C. M. Viallet, *Phys. Lett.* **B158**, 231 (1985); *Commun. Math. Phys.* **102**, 105 (1985); J. Mañes and B. Zumino, in *Supersymmetry and its Applications: Superstrings, Anomalies, and Supergravity*, eds. G. W. Gibbons, S. W. Hawking, and P. K. Townsend (Cambridge University Press, Cambridge, 1986).

19*a*. See, e.g., H. Georgi, Ref. 10: Eq. (XXV11.9).

20. R. Stora, in *New Directions in Quantum Field Theory and Statistical Mechanics – Lectures at the 1976 Cargese Summer School*, eds. M. Lévy and P. Mitter (Plenum, New York, 1977).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

21. J. A. Dixon, Ref. 19.

22. G. Barnich and M. Henneaux, *Phys. Rev. Lett.* **72**, 1588 (1994); G. Barnich, F. Brandt, and M. Henneaux, *Commun. Math. Phys.* **174**, 57, 93 (1995).

23. W. Troost, P. van Nieuwenhuizen, and A. Van Proeyen, *Nucl. Phys.* **B333**, 727 (1990).

24. I learned about this proof from B. Zumino, private communication.

25. E. Witten, *Nucl. Phys.* **B223**, 422 (1983).

26. C.-S. Chu, P.-M. Ho, and B. Zumino, *Nucl. Phys.* **B475**, 484 (1996).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.010

23

Extended Field Configurations

Most of this book has been devoted to applications of quantum field theory that can at least be described in perturbation theory, whether or not the perturbation series actually works well numerically. In using perturbation theory, we expand the action around the usual spacetime-independent vacuum values of the fields, keeping the leading quadratic term in the exponential $\exp(iI)$, and treating all terms of higher order in the fields as small corrections. Starting in the mid-1970s, there has been a growing interest in effects that arise because there are extended spacetime-dependent field configurations, such as those known as instantons,[1] that are also stationary 'points' of the action. In principle, we must include these configurations in path integrals and sum over fluctuations around them. (In Section 20.7 we have already seen an example of an instanton configuration, applied in a different context.) Although such non-perturbative contributions are often highly suppressed, they are large in quantum chromodynamics, and produce interesting exotic effects in the standard electroweak theory.

There are also extended field configurations that occur, not only as correction terms in path integrals for processes involving ordinary particles, but also as possible components of actual physical states. These configurations include some that are particle-like, such as magnetic monopoles[2] and skyrmions,[3] which are concentrated around a point in space or, equivalently, around a world line in spacetime. There are also string-like configurations,[4] similar to the vortex lines in superconductors discussed in Section 21.6, which are concentrated around a line in space or, equivalently, around a world sheet in spacetime. Then there are configurations that are sheet-like, like the domain walls[5] between spatial regions in which discrete symmetries are broken in different ways. In contrast, the instantons mentioned above are *event-like*, concentrated about a point in spacetime, and therefore never appear as components of actual physical states.

Some extended field configurations are stabilized because of boundary conditions that are imposed by the nature of the problem in which they appear. An example is the 'bounce' solution, which appears in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

analysis of vacuum decay,[6] and will be discussed in Section 23.8. Other configurations are stable because they carry a quantum number whose conservation forbids any possible decay mode.[7]

In this chapter we shall mostly be concerned with extended field configurations that are stabilized by their topology. In analyzing all such configurations we use the same topological tools, chiefly homotopy theory, so we shall begin by considering all topologically stabilized configurations together, in a space or spacetime of arbitrary dimensionality d.

23.1 The Uses of Topology

It often happens that the space of all possible field configurations may be given a non-trivial topology by the condition that some functional S of the various fields is finite. In classical field theory S is the potential energy (or in some cases the potential energy per unit area or per unit length); no finite perturbation can produce a configuration where this is infinite. In classical statistical mechanics S is the Hamiltonian, and in quantum field theory formulated in Euclidean spacetime S is the Euclidean action or is proportional to it. (Euclidean path integrals and some of their applications are discussed in Appendix A of this chapter.) We construct a perturbation theory by starting with some equilibrium field configuration for which the Euclidean action or Hamiltonian is finite, and then integrating over fluctuations which leave it finite.

Two field configurations are said to be topologically equivalent if it is possible to deform one of them continuously into the other without passing through forbidden configurations with S infinite. This is evidently an equivalence relation (in the sense of being reflexive, symmetric, and transitive), and therefore divides the set of all field configurations into equivalence classes, each consisting of configurations of the same topology. For example, if S is the potential energy (the Hamiltonian for time-independent fields) in d space dimensions, then topologically different field configurations are forbidden by an infinite energy barrier from being transformed into one another. In particular, extended configurations with a different topology from the usual spatially uniform vacuum fields cannot spread out to become spatially uniform.

The topological classification is also useful when we are looking for a local minimum of S. If we can find a configuration that minimizes S for all configurations of a given topological type, then that field configuration must be at least a local minimum of S for *all* configurations of any type, since no small variation of the fields can change their topological type. Such a configuration is therefore a solution of the field equations, which are equivalent to the condition that S is stationary. This sort of problem

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

comes up not only in stability problems, where S is the Hamiltonian, but also in finding field configurations around which we may expand the field variables in path integrals in Euclideanized d-dimensional spacetime. Here S is the negative of the Euclidean action I, and we must look for a local minimum so that the leading term in the expansion around this configuration will be a quadratic free-field action with second-order terms of the right sign.

Here are some examples:[8]

(a) Skyrmions, etc. Consider the real Goldstone boson fields π_a associated with the spontaneous breakdown of a continuous global symmetry group G to a subgroup H. As we saw in Chapter 19, the potential energy for these Goldstone bosons in a Euclidean space of dimensionality $d > 2$ will take the form

$$S[\pi] = \int d^d x \left[\tfrac{1}{2} \sum_{ab} g_{ab}(\pi) \partial_i \pi_a \partial_i \pi_b + \ldots \right] , \tag{23.1.1}$$

where g_{ab} is a positive-definite matrix, and '$+\ldots$' denotes possible terms of higher order in the derivatives of π. Alternatively, Eq. (23.1.1) can be regarded as minus the action for a Goldstone boson field in a d-dimensional Euclidean *spacetime*.

Field configurations of finite S must have $\partial_i \pi_a(\mathbf{x})$ vanishing at infinity faster than $|\mathbf{x}|^{-d/2}$ (where $|\mathbf{x}| \equiv \sqrt{x_i x_i}$), so that $\pi_a(\mathbf{x})$ must approach a constant $\pi_{a\infty}$ as $\mathbf{x} \to \infty$ with a remainder vanishing faster than $|\mathbf{x}|^{(2-d)/2}$. The Goldstone boson fields π_a at any point form a homogeneous space, the coset space G/H, for which it is possible to transform any one field value to any other by a transformation of G, so by a global G transformation it is always possible to arrange that the asymptotic limit $\pi_{a\infty}$ takes any specific value, say $\pi_{a\infty} = 0$. The field $\pi_a(\mathbf{x})$ thus represents a mapping of the whole d-dimensional space, with the sphere $r = \infty$ taken as a single point, into the manifold G/H of all field values.

Now, a d-dimensional Euclidean space with the $(d-1)$-dimensional spherical surface at infinity identified as a single point is topologically the same as S_d, the d-dimensional sphere (that is, the surface of a $(d+1)$-dimensional ball), in the sense that either can be continuously mapped into the other. The fields $\pi(\mathbf{x})$ that approach zero as $\mathbf{x} \to \infty$ may therefore be classified according to the topologically distinct mappings of S_d into the manifold G/H of the field variables, for which the point at infinity is mapped into zero. The set of classes of such topologically distinct mappings $S_d \mapsto \mathcal{M}$ with one point of S_d mapped into a fixed point of $\mathcal{M}$ is known as $\pi_d(\mathcal{M})$, the dth *homotopy group* of the manifold $\mathcal{M}$. These homotopy groups will be discussed (and their group structure explained) in the next section, and a list of homotopy groups for various manifolds

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

is given in Appendix B of this chapter. For the present it will be enough to mention that, although when the manifold $\mathcal{M}$ is a linear space the homotopy group $\pi_d(\mathcal{M})$ is trivial (in the sense that any field configuration $\pi(\mathbf{x})$ that approaches a constant value as $\mathbf{x} \to \infty$ can be continuously deformed into one in which the field takes that value everywhere), the manifold $\mathcal{M} = G/H$ of the Goldstone boson fields often has a non-trivial homotopy group. In the cases relevant to quantum chromodynamics, of $SU(2) \otimes SU(2)$ broken to $SU(2)$ or of $SU(3) \otimes SU(3)$ broken to $SU(3)$, the manifold G/H is the same as $SU(2)$ or $SU(3)$, respectively, for which according to Appendix B the homotopy groups $\pi_3(H)$ are non-trivial. The topologically non-trivial fields at local minima of the potential energy with $d = 3$ are known as *skyrmions*.[3] Baryons like the proton may in some respects be regarded as skyrmion solutions in a pure meson theory.

The functional (23.1.1) does not have skyrmion stationary points, unless terms involving higher powers of $\partial_i \pi_a$ are included in the integrand. In the absence of such terms, any topologically non-trivial field configurations will have a continuum of values for S, extending down to a lower bound $S = 0$ at which π becomes singular, so topology cannot stabilize such configurations. This is generally known as *Derrick's theorem*.[9] To prove this theorem, note that for any field configuration $\pi_a(\mathbf{x})$ we may introduce another configuration with the same topology,

$$\pi_a^R(\mathbf{x}) \equiv \pi_a(\mathbf{x}/R)$$

with R an arbitrary real positive scale factor. Then for the terms shown explicitly in Eq. (23.1.1)

$$S[\pi^R] = R^{d-2} S[\pi] \,.$$

For $d > 2$ this is a decreasing function of R as $R \to 0$, so there is a continuum of values of $S[\pi^R]$ extending down to a value $S = 0$. Furthermore $S[\pi] > 0$, because $S[\pi]$ can only vanish if $\pi(\mathbf{x})$ is constant, which is not possible because we assumed that π is topologically non-trivial. Hence this lower bound is attained only at $R = 0$, at which $\pi_a^R(\mathbf{x})$ becomes singular.

Goldstone boson field configurations can be stabilized by adding higher derivative terms to S. For instance, if we take $S[\pi] = T[\pi] + D[\pi]$ with

$$T[\pi] \equiv \int d^d x \; \tfrac{1}{2} \sum_{ab} g_{ab}(\pi) \partial_i \pi_a \partial_i \pi_b \geq 0 \,,$$

$$D[\pi] \equiv \int d^d x \, f_{abcd}(\pi) \, \nabla \pi_a \cdot \nabla \pi_b \, \nabla \pi_c \cdot \nabla \pi_d \geq 0 \,,$$

then $D[\pi^R] = R^{d-4} D[\pi]$, while as before $T[\pi^R] = R^{d-2} T[\pi]$, so $S[\pi^R]$ reaches a minimum at a finite R if $2 < d < 4$, and in particular for the physically interesting case $d = 3$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

The problem with the theory of skyrmions is not that we have to include higher-derivative terms like $D[\pi]$ in the action. As discussed in Section 19.5, we expect such terms in the action of any effective field theory of Goldstone bosons. The problem is that there is no rationale for excluding an infinite number of other higher-derivative terms, all of which are generically of the same order of magnitude for configurations that are stabilized by a balance between terms with different numbers of derivatives, which makes realistic calculations impossible.

(b) Domain boundaries. When a *discrete* symmetry is broken, we have the possibility that the symmetry is broken in different ways in different domains, separated by domain boundaries in which the vacuum fields make a transition from one minimum of the potential to another. For instance, consider a flat domain boundary in the y–z plane, and suppose that the energy per unit area is given by

$$S[\phi] = \int_{-\infty}^{\infty} dx \left[\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 + V(\phi)\right], \tag{23.1.2}$$

where $\phi(x)$ is a real scalar field that is assumed to depend only on the distance x along the direction normal to the boundary, and $V(\phi)$ is a potential satisfying the reflection symmetry $\phi \to -\phi$, with minima only at field values $\pm\bar{\phi}$. For convenience, we will adjust an additive constant in $V(\phi)$ so that the minimum value of $V(\phi)$ is zero, in which case $V(\phi) \geq 0$, and $V(\phi) = 0$ only at $\phi = \pm\bar{\phi}$. To keep S finite, it is necessary for ϕ to approach either $+\bar{\phi}$ or $-\bar{\phi}$ as $x \to \infty$, and also to approach either $+\bar{\phi}$ or $-\bar{\phi}$ as $x \to -\infty$. We then have four topologically distinct configurations, in two of which ϕ approaches the same limits as $x \to \pm\infty$, so that the configuration can be smoothly deformed into the vacuum configurations with $\phi(x)$ constant everywhere, and in two of which ϕ approaches opposite limits as $x \to \pm\infty$, which are topologically stable. Here we are classifying field configurations according to $\pi_0(G)$, where $\pi_0(\mathcal{M})$ for any manifold $\mathcal{M}$ is conventionally defined as the set of connected components of $\mathcal{M}$, and G is the symmetry group, which in our case is the Z_2 group of reflections $\phi \to -\phi$.

This is a good place to introduce a trick due to Belavin *et al.*[1] that will prove useful in Sections 23.3 and 23.5 in dealing with the more complicated cases of monopoles and instantons. Rewrite Eq. (23.1.2) in the form

$$S[\phi] = \frac{1}{2}\int_{-\infty}^{\infty} dx \left(\frac{d\phi}{dx} \mp \sqrt{2V(\phi)}\right)^2 \pm \int_{\phi(-\infty)}^{\phi(\infty)} \sqrt{2V(f)}\, df\,. \tag{23.1.3}$$

The integral in the second term on the right-hand side of Eq. (23.1.3) can be regarded as a 'topological charge,' which depends only on the values taken by the field at $x \to \pm\infty$. For configurations that approach the same limit as $x \to +\infty$ and $x \to -\infty$, this integral vanishes, and the minimum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

value of S is zero, reached for constant fields. For a field $\phi(x)$ that takes different values at $x = \pm\infty$, we can choose the $\pm$ sign in Eq. (23.1.3) to yield a lower bound

$$S[\phi] \geq \int_{-\bar{\phi}}^{\bar{\phi}} \sqrt{2V(f)}\, df \,. \tag{23.1.4}$$

This bound is reached when the first term in Eq. (23.1.3) vanishes, or in other words, when

$$x = \pm \int_0^{\phi(x)} \frac{df}{\sqrt{2V(f)}} + x_0 \,, \tag{23.1.5}$$

where x_0 is an integration constant, which evidently gives the position of the center of the domain boundary. Note that Derrick's theorem is no obstacle to a solution here, because for domain boundaries $d = 1$, so for a rescaled field $\phi(x/R)$ the integrals of the two terms in the integrand in Eq. (23.1.2) go as R^{-1} and R^{+1}, respectively.

Eq. (23.1.5) could have been obtained more directly, by deriving a second-order differential equation for $\phi(x)$ from the condition that Eq. (23.1.2) must be stationary under small variations in $\phi(x)$, and then using this differential equation to show that the quantity $\frac{1}{2}(d\phi/dx)^2 - V(\phi)$ is constant in x. The advantage of the derivation based on the formula (23.1.3) is that it shows immediately that the solution (23.1.5) is stable against small perturbations that maintain the flatness of the boundary, aside from the 'zero-mode' associated with changes in the boundary location x_0. By adding a term $\frac{1}{2}(d\phi/dy)^2 + \frac{1}{2}(d\phi/dz)^2$ in the integrand of Eq. (23.1.2), we can see that this solution is also stable against *any* perturbation $\delta\phi(x, y, z)$, provided $\delta\phi(x, y, z) \to 0$ for $x \to \pm\infty$ with fixed y and z.

If there are discrete spontaneously broken symmetries then domain boundaries would have formed when these symmetries became broken in the early universe. If the domain boundaries did not disappear they would produce gross distortions of the observed isotropy and homogeneity of the present universe.[5] We do not now know of any exact discrete symmetries except CPT, or of any spontaneously broken approximate or exact discrete symmetries, so for the present this is not a problem.

(c) Instantons, etc. Now consider a gauge theory, with

$$S[A] = \tfrac{1}{4} \int d^d x\, F_{\alpha ij} F_{\alpha ij} \,, \tag{23.1.6}$$

where $F_{\alpha ij}$ is the usual field strength tensor, and we take $d \geq 4$. This can either be regarded as the action for quantum gauge fields in a Euclidean

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

d-dimensional spacetime, or the potential energy for classical gauge fields in temporal gauge, with $A_\alpha^0 = 0$, in $(d+1)$-dimensional spacetime.

In order for $S[A]$ to be finite, $F_{\alpha ij}(\mathbf{x})$ must vanish as $\mathbf{x} \to \infty$. This can be achieved by having $A_{\alpha i}(\mathbf{x})$ vanish sufficiently rapidly as $\mathbf{x} \to \infty$, but even for $d \geq 4$ it is also possible for $S[A]$ to be finite for a field $A_{\alpha i}(\mathbf{x})$ that vanishes as slowly as $1/|\mathbf{x}|$, as long as the field approaches a pure gauge as $|\mathbf{x}| \to \infty$

$$it_\alpha A_{\alpha i}(\mathbf{x}) \to g^{-1}(\hat{\mathbf{x}})\, \partial_i g(\hat{\mathbf{x}})\,, \tag{23.1.7}$$

where $g(\hat{\mathbf{x}})$ is a direction-dependent element of the gauge group G. Furthermore $A_{\alpha i}(\mathbf{x})$ is unaffected if we replace $g(\hat{\mathbf{x}})$ with $g_0 g(\hat{\mathbf{x}})$ for any fixed group element $g_0 \in G$, so by choosing $g_0 = g^{-1}(\hat{\mathbf{x}}_1)$ we can arrange that $g(\hat{\mathbf{x}}_1) = 1$ for any one direction $\hat{\mathbf{x}}_1$. Each gauge field of finite $S[A]$ therefore defines a mapping from the unit sphere $|\hat{\mathbf{x}}| = 1$ to the group manifold, with the point $\hat{\mathbf{x}}_1$ mapped into the unit element of G. (In the case in which the gauge field vanishes faster than $1/|\mathbf{x}|$ as $|\mathbf{x}| \to \infty$, this mapping takes all points on the unit sphere into the identity element of the gauge group.) The set of classes of such topologically distinct mappings $S_{d-1} \mapsto G$, with one point of S_{d-1} mapped into a fixed element of G, is known as $\pi_{d-1}(G)$, the $(d-1)$th homotopy group of the group manifold. As indicated in Appendix B of this chapter, $\pi_3(G)$ is non-trivial for any semisimple Lie group G. The topologically non-trivial stationary points of $S[A]$ for $d = 4$ are known as *instantons*.[1] Their importance in quantum chromodynamics is discussed in Sections 23.5 and 23.6.

In order for $S[A]$ to be stationary at a field $A(\mathbf{x})$, it is necessary that $A(\mathbf{x})$ should satisfy the field equations

$$\partial_i F_{\alpha ij} = 0\,. \tag{23.1.8}$$

A simple scaling argument again limits the values of the dimensionality d where we can hope to find a topologically non-trivial local minimum of $S[A]$. Define $A^R(\mathbf{x}) \equiv A(\mathbf{x}/R)/R$. Then

$$S[A^R] = R^{d-4} S[A]\,,$$

so for $d \neq 4$ there can be no topologically non-trivial stationary point of $S[A]$ unless $S[A] = 0$. But if $S[A] = 0$ then $F_{\alpha ij} = 0$ everywhere, so by a gauge transformation we can make $A_{\alpha i}$ also vanish everywhere.

As we shall see in Section 23.5, for $d = 4$ it *is* possible to find instanton solutions where $S[A]$ (here identified as $-I[A]$) is stationary, with $F_{\alpha ij}$ not equal to zero except at infinity. The scaling argument above shows that if $A(\mathbf{x})$ is such an instanton solution then so is $A(\mathbf{x}/R)/R$, but this degeneracy is removed by quantum corrections.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

(d) Monopoles, vortex lines, etc. Now consider a theory of gauge fields together with scalars that furnish a linear representation of the gauge group, with

$$S[\phi, A] = \int d^d x \left[\tfrac{1}{2} \sum_{ab} g_{ab}(\phi) D_i \phi_a D_i \phi_b + \tfrac{1}{4} F_{\alpha ij} F_{\alpha ij} + U(\phi) \right], \quad (23.1.9)$$

where $g_{ab}(\phi)$ is a positive-definite matrix (usually ϕ-independent), $U(\phi)$ is bounded below, and shifted by a constant term so that its minimum value is zero, and $F_{\alpha ij}$ and D_i are the usual field strength and gauge-covariant derivative. We require that $U(\phi)$ is a scalar and $g_{ab}(\phi)$ is a tensor under transformations in the gauge group G. Again, Eq. (23.1.9) gives either the action for a quantum field theory in Euclidean d-dimensional spacetime, or the potential energy for a classical field theory in temporal gauge in $(d+1)$-dimensional spacetime.

For $S[\phi, A]$ to be finite it is necessary for $U(\phi(\mathbf{x}))$ to vanish as $\mathbf{x} \to \infty$. The set of ϕs at which $U(\phi)$ vanishes is invariant under G and may be discrete or continuous. In case (b) above we have dealt with an example where this set is discrete. Let us now consider the broken symmetry case where the zeros of $U(\phi)$ form a continuous manifold $\mathscr{M}_0$ consisting of fields related by transformations $g \in G$. In this case each $\phi(\hat{\mathbf{x}})$ may be obtained by a transformation $\gamma(\hat{\mathbf{x}}) \in G$ acting on the value $\phi(\hat{\mathbf{x}}_1)$ of the field in any one direction $\hat{\mathbf{x}}_1$. We may therefore consider the field $\phi(\mathbf{x})$ to define a mapping $S_{d-1} \mapsto G/H$ into a coset space G/H; in other words, into the group G with elements g_1 and g_2 identified if they differ only by right multiplication with some element h of the subgroup $H \subset G$ that leaves $\phi(\hat{\mathbf{x}}_1)$ invariant, that is, if $g_1 = g_2 h$. In particular, the point $\hat{\mathbf{x}}_1$ is mapped into the subgroup H, in order that $\gamma(\hat{\mathbf{x}}_1)$ acting on $\phi(\hat{\mathbf{x}}_1)$ should yield $\phi(\hat{\mathbf{x}}_1)$ itself. The fields that approach values on the manifold $\mathscr{M}_0$ as $\mathbf{x} \to \infty$ may therefore be classified according to the topologically distinct mappings of S_{d-1} into G/H that map the point $\hat{\mathbf{x}}_1$ into the fixed 'unit' element H of G/H. The set of classes of such topologically distinct mappings $S_{d-1} \mapsto G/H$ with one point of S_{d-1} mapped into a fixed element of G/H is known as $\pi_{d-1}(G/H)$, the $(d-1)$th homotopy group of the manifold G/H.

In this case $\partial_i \phi(\mathbf{x})$ goes as $1/|\mathbf{x}|$ for $\mathbf{x} \to \infty$. In order for $S[\phi]$ to be finite $D_i \phi$ must vanish faster than $|\mathbf{x}|^{-d/2}$ for $\mathbf{x} \to \infty$, so it is necessary for $it_\alpha A_{\alpha i}(\mathbf{x})$ to approach $\gamma^{-1}(\hat{\mathbf{x}}) \partial_i \gamma(\hat{\mathbf{x}})$ faster than $|\mathbf{x}|^{-d/2}$ for $\mathbf{x} \to \infty$. This is a pure gauge field, so the field strength tensor $F_{\alpha ij}(\mathbf{x})$ vanishes faster than $|\mathbf{x}|^{-d/2-1}$, which is fast enough to make $\int d^d x F_{\alpha ij} F_{\alpha ij}$ converge.

Derrick's theorem does not apply for the gauge theory defined by Eq. (23.1.9), but it is interesting to see where the same reasoning takes us. For any given fields $\phi(\mathbf{x})$ and $A(\mathbf{x})$, again define $\phi^R(\mathbf{x}) \equiv \phi(\mathbf{x}/R)$, and now also $A^R(\mathbf{x}) \equiv A(\mathbf{x}/R)/R$. The three terms in the Hamiltonian (23.1.9)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

now have the scaling properties

$$T[\phi^R, A^R] = R^{d-2} T[\phi, A] , \quad K[A^R] = R^{d-4} K[A] , \quad V[\phi^R] = R^d V[\phi] ,$$

where $T[\phi, A] \equiv \frac{1}{2} \int d^d x \sum_{ab} g_{ab}(\phi) D_i \phi_a D_i \phi_b$, $K[A] \equiv \frac{1}{4} \int d^d x F_{\alpha ij} F_{\alpha ij}$, and $V[\phi] \equiv \int d^d x \, U(\phi)$. Now, for $d > 4$, $S[\phi^R, A^R]$ has no minimum at any finite value of R, so there is no stable configuration with non-trivial topology. For $0 < d < 4$ there is no difficulty in finding a finite value of R at which $S[\phi^R, A^R]$ is a minimum.

In the physically interesting case $d = 3$, topologically non-trivial field configurations are classified according to the homotopy group $\pi_2(G/H)$, which is non-trivial for a simply connected group G (such as $SU(2)$) broken to an H containing the $U(1)$ of electromagnetism. The topologically non-trivial classical field configurations with $d = 3$ are known as *magnetic monopoles*.[2] As we shall see in Section 23.3, their magnetic pole strength is quantized, the different values corresponding to different elements of $\pi_2(G/H)$.

For $d = 2$, topologically non-trivial configurations correspond to elements of $\pi_1(G/H)$, which is non-trivial when G is a non-simply connected group like $U(1)$ or $SO(3)$, broken either completely or to a discrete subgroup. The topologically non-trivial classical field configurations with $d = 2$ are the cross sections of *vortex lines*. One example is provided by superconductivity, where $G = U(1)$ is spontaneously broken to $H = Z_2$. We have seen in Section 21.6 that vortex lines occur in type II superconductors for magnetic field strengths in a certain range, and that the magnetic flux carried by a vortex line is quantized, the different values of the flux corresponding to different elements of $\pi_1(U(1)/Z_2)$. Vortex lines can also occur in relativistic quantum field theories,[4] and may be produced in symmetry-breaking transitions in the early universe, in which case they are known as *cosmic strings*.[11]

Monopoles and vortex lines share a remarkable feature that can be deduced on purely topological grounds. In both cases the forms of the Goldstone boson fields $\pi_a(\mathbf{x})$ on large spheres (S_1 for vortex lines, S_2 for monopoles) surrounding the configurations are twisted, in such a way that they cannot smoothly be deformed into constants. In particular, it is impossible smoothly to reduce the radii of these spheres to zero without encountering some sort of singularity, because a non-singular field $\pi_a(\mathbf{x})$ on a sphere would have to become a constant as the radius of the sphere shrinks to zero. The singularity in both cases occurs in a core (a line or perhaps a tube for vortex lines, a point or perhaps a ball for monopoles) within which the group G is no longer broken, so that the system is no longer described by Goldstone boson fields, but by an order parameter that transforms linearly under G.

For $d = 4$, the function $S[\phi^R, A^R]$ of R can only have a minimum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

at some finite R if $T[\phi, A] = V[\phi] = 0$, which would require that $\phi(\mathbf{x})$ everywhere takes a value at which $U(\phi) = 0$. Assuming that these values form a continuum related by transformations in the gauge group G, by a gauge transformation they may be made constants, $\phi(\mathbf{x}) = \phi_0$. Then in this gauge the condition $T[\phi, A] = 0$ implies that $A_a(\mathbf{x}) = 0$ for all broken symmetries, for which $t_a\phi_0 \neq 0$. Both $T[\phi, A]$ and $V[\phi]$ are stationary at such a field configuration, so in order for $S[\phi, A]$ to be stationary $K[\phi, A]$ must also be stationary, which means that the non-zero gauge fields $A_{i\mu}$ (which belong to the subgroup $H \subset G$ that is unbroken by ϕ_0) must satisfy the Euclidean Yang–Mills field equations

$$\partial_\mu F_{i\mu\nu} = 0 \,. \tag{23.1.10}$$

This case therefore reduces to case (c), but with the gauge group G replaced with its unbroken subgroup H.

23.2 Homotopy Groups

We learned in the previous section to classify field configurations, at which the Hamiltonian or other functionals are finite, in correspondence with the elements of appropriate homotopy groups. But we have not yet explained in what sense the homotopy groups are *groups*, nor have we given any physical significance to the group structure. As we shall see, there is a natural definition of the multiplication rule for elements of homotopy groups, according to which two extended configurations of fields forming a manifold $\mathcal{M}$ in d dimensions, that belong to different elements c_1 and c_2 of $\pi_d(\mathcal{M})$, can only fuse continuously to form a configuration belonging to the element $c_1 \times c_2$ of $\pi_d(\mathcal{M})$.

We will begin by defining the first homotopy group $\pi_1(\mathcal{M})$ of an arbitrary manifold $\mathcal{M}$, also known as the *fundamental group of the manifold*. As we have seen, the existence of a non-trivial $\pi_1(G/H)$ for some coset space G/H is the condition for the topological stability of a vortex line in three dimensions (or a monopole in two dimensions). After considering $\pi_1(\mathcal{M})$, we will then move on to more general homotopy groups.

A connected manifold $\mathcal{M}$ is said to be multiply connected if there is some closed curve of points $p(z)$ on the manifold, parameterized by a single variable z with $0 \leq z \leq 1$ and $p(0) = p(1)$, which cannot be contracted to a point by a continuous deformation. Since on a connected manifold we can always continuously deform any such closed curve so that any one point on the curve is anywhere we like on the manifold, we may restrict our attention only to curves for which $p(0) = p(1) = p_0$, where p_0 is any fixed point of the manifold, known as the *base point*. Two such closed curves $p_1(z)$ and $p_2(z)$ are said to be homotopically equivalent

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

if they *can* be deformed into each other, that is, if there is a *continuous* function $p(z,t)$ for $0 \leq t \leq 1$, with

$$p(z,0) = p_1(z)\,, \qquad p(z,1) = p_2(z)\,,$$

$$p(0,t) = p(1,t) = p_0\,.$$

The relation of homotopic equivalence is an equivalence relation in the sense of being symmetric, reflexive, and transitive, so it divides the space of closed curves on the manifold into equivalence classes: two closed curves are in the same class if and only if they are homotopically equivalent. The set of these equivalence classes is known as the first homotopy group of the manifold, $\pi_1(\mathscr{M})$.

To define the multiplication rule for $\pi_1(\mathscr{M})$, choose a standard curve $p[z;c]$ that starts and ends at the base point p_0 for each equivalence class c in $\pi_1(\mathscr{M})$. For any two equivalence classes c_1 and c_2, define the 'product' $c_1 \times c_2$ as the equivalence class containing the curve $p[z,c_1,c_2]$ that starts at p_0, follows $p[z,c_1]$ back to p_0, and then follows $p[z,c_2]$ back again to p_0. Formally, we take

$$p[z,c_1,c_2] \equiv \begin{cases} p[2z,c_1] & 0 \leq z \leq \frac{1}{2} \\ p[2z-1,c_2] & \frac{1}{2} \leq z \leq 1 \end{cases}\,.$$

We must now show that multiplication defined in this way satisfies the conditions for a group. First, let us check that this multiplication is *associative*. For this purpose, note that $(c_1 \times c_2) \times c_3$ is the equivalence class containing a curve $p[z,c_1 \times c_2,c_3]$ that goes along the standard curve $p[z,c_1 \times c_2]$ from the base point and back again, and then along the standard curve $p[z,c_3]$ from the base point and back again, while $c_1 \times (c_2 \times c_3)$ is the equivalence class containing a curve $p[z,c_1,c_2 \times c_3]$ that goes along the standard curve $p[z,c_1]$ from the base point and back again, and then along the standard curve $p[z,c_2 \times c_3]$ from the base point and back again. By definition, the curve $p[z,c_1 \times c_2]$ may be deformed into a curve that goes along $p[z,c_1]$ from the base point and back again and then along $p[z,c_2]$ from the base point and back again, while $p[z,c_2 \times c_3]$ may be deformed into a curve that goes along $p[z,c_2]$ from the base point and back again and then along $p[z,c_3]$ from the base point and back again. Hence both curves $p[z,c_1 \times c_2,c_3]$ and $p[z,c_1,c_2 \times c_3]$ may be deformed into the curve that goes along $p[z,c_1]$ from the base point and back again, then along $p[z,c_2]$ from the base point and back again, and finally along $p[z,c_3]$ from the base point and back again, and hence they may be deformed into each other, showing that

$$(c_1 \times c_2) \times c_3 = c_1 \times (c_2 \times c_3).$$

The unit element e of $\pi_1(\mathscr{M})$ is defined as the equivalence class con-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

taining the curve $p[z,e] = p_0$ that stays at the base point. To check that $e \times c = c$, note that

$$p[z,e,c] = \begin{cases} p_0 & 0 \le z \le \frac{1}{2} \\ p[2z-1,c] & \frac{1}{2} \le z \le 1 \end{cases} .$$

But this curve can be continuously deformed into $p[z,c]$ by taking

$$p(z,t) = \begin{cases} p_0 & 0 \le z \le t/2 \\ p[(2z-t)/(2-t),c] & t/2 \le z \le 1 \end{cases} ,$$

which is the same as $p[z,e,c]$ for $t = 1$ and the same as $p[z,c]$ for $t = 0$. The product $e \times c$ is the equivalence class containing $p[z,e,c]$, which we now see is the same as the equivalence class containing $p[z,c]$, which is just c. The proof that $c \times e = c$ is similar.

The 'inverse' c^{-1} of the equivalence class c is the equivalence class that contains a curve $p^{-1}[z,c]$ which goes around the same path as the standard curve $p[z,c]$ but in the opposite direction; that is,

$$p^{-1}[z,c] \equiv p[1-z,c].$$

This is not necessarily the same as the 'standard' curve $p[z,c^{-1}]$, but by the definition of the equivalence class c^{-1}, the two curves may be deformed into each other. To check that $c^{-1} \times c = e$, note that by deforming $p[z,c^{-1}]$ into $p^{-1}[z,c]$, the curve $p[z,c^{-1},c]$ may be deformed into

$$p[z,c^{-1},c] \to \begin{cases} p[1-2z,c] & 0 \le z \le \frac{1}{2} \\ p[2z-1,c] & \frac{1}{2} \le z \le 1 \end{cases} .$$

But this curve can be continuously deformed into $p[z,e] = p_0$ by taking

$$p(z,t) = \begin{cases} p[1-2tz,c] & 0 \le z \le \frac{1}{2} \\ p[2tz+1-2t,c] & \frac{1}{2} \le z \le 1 \end{cases} ,$$

which is the same as $p[z,c^{-1},c]$ for $t = 1$ and the same as $p[z,e]$ for $t = 0$. The product $c^{-1} \times c$ is the equivalence class containing $p[z,c^{-1},c]$, which we now see is the same as the equivalence class containing $p[z,e]$, which is just e. The proof that $c \times c^{-1} = e$ is similar. The existence of a unit element and inverses shows that these equivalence classes form a *group*.

The classic example of a manifold $\mathcal{M}$ with a non-trivial first homotopy group is the circle itself, $\mathcal{M} = S_1$. This may be parameterized by an angle θ with $\theta = 0$ and $\theta = 2n\pi$ (with n any positive or negative integer) identified as the same point. The homotopy groups consist of classes of functions $\theta(z)$ for $0 \le z \le 1$ that begin at some base point, $\theta(0) = \theta_0$, and end at the same base point, $\theta(1) = \theta_0 + 2n\pi$. Two such functions may be continuously deformed into each other if and only if they have the same value of n, so $\pi_1(S_1)$ consists of a denumerably infinite number of classes c_n labelled by the positive or negative integer n. Furthermore, the 'product'

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

of two classes c_n and c_m consists of curves that go n times around the circle from the base point and back again, and then m times around the circle from the base point and back again, so here multiplication is just addition:

$$c_n \times c_m = c_{n+m} \tag{23.2.1}$$

and hence

$$\pi_1(S_1) = Z \ , \tag{23.2.2}$$

where Z is the group of addition of positive and negative integers. As an immediate physical application, we note that when a gauge group $SO(2)$ is completely spontaneously broken, the coset space is just $SO(2)$ itself, which has the topology of a circle, so in this case there is an infinite number of types of topologically stable vortex line, characterized by a positive or negative integer n. For instance, this is the case in a type II superconductor, where as we saw in Section 21.6 the $U(1)$ of electromagnetic gauge invariance is spontaneously broken to a discrete subgroup Z_2.

More generally, all spheres S_d with $d > 1$ are simply connected, which means that they have a trivial first homotopy group, a statement conventionally expressed as

$$\pi_1(S_d) = 0 \qquad \text{for } d > 1 \ . \tag{23.2.3}$$

Only a few of the more familiar Lie groups are multiply connected:

$$\pi_1(G) = \begin{cases} Z & G = U(k) & k \geq 1 \\ Z_2 & G = SO(k) & k \geq 3 \\ 0 & G = Spin(k) & k \geq 3 \\ 0 & G = SU(k) & k \geq 2 \\ 0 & G = USp(2k) & k \geq 1 \\ 0 & G = G_2, F_4, E_6, E_7, E_8 & \end{cases} \ . \tag{23.2.4}$$

Here Z_2 is the group with two elements 1 and -1, with group multiplication defined as ordinary multiplication, and $Spin(n)$ is the simply connected covering group of $SO(n)$. (As we saw in Section 2.7, $Spin(3)$ is the same as $SU(2)$.) Also, for a direct product of two manifolds $\mathcal{M}$ and $\mathcal{M}'$, the fundamental group is

$$\pi_1(\mathcal{M} \times \mathcal{M}') = \pi_1(\mathcal{M}) \times \pi_1(\mathcal{M}') \ . \tag{23.2.5}$$

We can appreciate the physical significance of the group structure of $\pi_1(\mathcal{M})$ by asking what happens when two distant parallel vortex lines in three dimensions are brought together. When the vortex lines are sufficiently far apart their fields do not interact, so the configuration can be described by specifying the classes c' and c'' in $\pi_1(G/H)$ to which

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

each belongs. The class c to which the whole configuration belongs is determined by the behavior of the fields on a very large circle surrounding both vortex lines. By a continuous deformation we can distort this very large circle into two large circles, one surrounding each vortex line, which intersect at a point midway between them. As we go around this closed curve in two-dimensional space we first trace out a curve in G/H which consists of one of the closed curves in the class c', and then one of the closed curves in the class c'', just as in the definition of the product of the classes. We conclude then that the whole configuration is in a class $c = c' \times c''$, and so the two vortex lines can only fuse together to form a vortex line of this class. In particular, they can only annihilate if $c'' = c'^{-1}$.

For instance, when $\pi_1(G/H) = Z$ (as in superconductivity, where $G/H = SO(2)/Z_2$), in three dimensions two vortex lines of classes c_n and c_m can fuse together to form a vortex line of class c_{n+m}, so they can only annihilate if $n = -m$. On the other hand, when $\pi_1(G/H) = Z_2$ (as when $G = SO(N)$ with $N \geq 3$ and H is trivial) there is only one kind of vortex line in three dimensions, corresponding to the element -1 of Z_2, and since $(-1)^2 = 1$, any two can annihilate.

Now let us consider the general homotopy group $\pi_k(\mathcal{M})$. This is much like $\pi_1(\mathcal{M})$, except that instead of considering mappings of the circle S_1 into a manifold $\mathcal{M}$, we consider mappings of the k-sphere S_k (the surface of a $(k+1)$-dimensional ball) into $\mathcal{M}$, again with one point of S_k always mapped into the same 'base point' p_0 of $\mathcal{M}$. Two such mappings are equivalent if one can be continuously deformed into the other, keeping the same point of S_k always mapped into the base point. The kth homotopy group $\pi_k(\mathcal{M})$ has elements consisting of equivalence classes of these mappings.

It is often convenient to picture the d-sphere S_d as the interior of a d-dimensional hypercube, with all points on the boundary identified as a single point. For instance, we have already seen that the circle S_1 can be treated as the interval $0 \leq \theta \leq 2\pi$, with points $\theta = 0$ and $\theta = 2\pi$ identified. Similarly we can make a map of an S_2 like the earth's surface by cutting out the south pole and spreading out the resulting sheet on the unit square $0 \leq z_1 \leq 1$, $0 \leq z_2 \leq 1$. Continuous mappings of this square into $\mathcal{M}$ must take all points on the boundary into the same point of $\mathcal{M}$, because all points on the boundary are the same point, the south pole. In general, two mappings $p(z_1, \cdots z_d)$ and $p'(z_1, \cdots z_d)$ of S_d into $\mathcal{M}$ are homotopically equivalent if one can be continuously deformed into the other, while keeping p on the boundary of the hypercube equal to the base point p_0.

As before, for each equivalence class c we choose a standard mapping $p(z_1, \cdots z_d; c)$. The product of c_1 and c_2 is defined as the equivalence class

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

that contains the mapping

$$p(z_1, z_2, \cdots z_d; c_1, c_2) = \begin{cases} p(2z_1, z_2, \cdots z_d; c_1) & 0 \le z_1 \le \frac{1}{2} \\ p(2z_1 - 1, z_2, \cdots z_d; c_2) & \frac{1}{2} \le z_1 \le 1 \end{cases} . \tag{23.2.6}$$

The unit element e is defined as the equivalence class that contains the mapping with $p = p_0$ for all z, and the inverse c^{-1} of c is defined as the equivalence class that contains the mapping with

$$p^{-1}[z_1, z_2, \cdots z_d; c] = p[1 - z_1, z_2, \cdots z_d; c] . \tag{23.2.7}$$

In the same way as for π_1, it can be shown that this multiplication is associative, and that $e \times c = c \times e = c$ and $c^{-1} \times c = c \times c^{-1} = e$. All $\pi_n(\mathcal{M})$ for $n \geq 2$ are Abelian. (There are manifolds $\mathcal{M}$ for which $\pi_1(\mathcal{M})$ is non-Abelian, such as the plane with two or more points removed.)

In any case in which $\pi_k(\mathcal{M}) = Z$, there must be a one-to-one mapping of the k-sphere S_k into a k-sphere S_k' in $\mathcal{M}$, which corresponds to the element 'one' of Z (not the unit element, which is zero). The element ν of Z with $\nu = 2, 3, \ldots$ corresponds to the mapping of S_k into the same k-sphere S_k' in $\mathcal{M}$, which covers S_k' ν times, with the Jacobian of the transformation $S_k \to S_k'$ positive. The element ν of Z with $\nu = -1, -2, \ldots$ corresponds to the mapping $S_k \to S_k'$, which covers S_k' $|\nu|$ times, with the Jacobian of the transformation $S_k \to S_k'$ negative.

For instance, we saw in the previous section that magnetic monopoles arise when a simply connected group G is broken to the $U(1)$ of electromagnetism. In this case the appendix shows that

$$\pi_2(G/U(1)) = \pi_1(U(1)) = Z , \tag{23.2.8}$$

so a magnetic monopole carries an integer-valued quantum number ν, which as shown in Section 23.3 is proportional to the magnetic charge. This quantum number gives the number of times that a two-sphere of large radius surrounding the monopole is mapped into a two-sphere in the manifold $G/U(1)$ of Goldstone boson fields (with the relative orientation of the two two-spheres being the same or opposite for ν positive or negative, respectively) and is therefore known as the *winding number*. The structure of the group Z shows that this quantum number is conserved, in the sense that a monopole of quantum number ν can fuse with a monopole of quantum number ν' only to form a monopole with quantum number $\nu + \nu'$.

If the unbroken subgroup is $SO(n)$ with $n \geq 3$, then according to Appendix B of this chapter

$$\pi_2(G/SO(n)) = \pi_1(SO(n)) = Z_2 . \tag{23.2.9}$$

In this case there is just one sort of 'monopole,' corresponding to the element -1 of Z_2, which can annihilate only in pairs. It is important to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

distinguish between this case and that in which $SO(n)$ is replaced with its simply connected covering group $Spin(n)$, for which there are no monoples. We will come back to this point at the end of the next section.

Another example: we saw in the previous section that the skyrmions in quantum chromodynamics with n light quarks correspond to elements of $\pi_3(SU(n))$, which according to the appendix is Z. Thus these skyrmions carry a conserved integer-valued quantum number ν, which perhaps may be identified with baryon number. Similarly, recall from the previous section that the instantons in a gauge theory based on a simple gauge group G correspond to the elements of $\pi_3(G)$, which according to the appendix is Z, so these instantons like skyrmions carry an integer-valued quantum number ν, also known as the winding number. In Section 23.5 we shall see how to express this quantum number as a local functional of the gauge field.

23.3 Monopoles

As a detailed example of a topologically non-trivial field configuration, we shall now consider the monopole of 't Hooft and Polyakov,[2] and its generalizations. We saw in Section 23.1 that when a simply connected gauge group G is spontaneously broken to the $U(1)$ of electromagnetism, the configurations of finite energy are classified according to the elements of the group $\pi_2(G/U(1)) = \pi_1(U(1)) = Z$. (The case of non-simply connected Lie groups is considered at the end of this section.) According to the physical interpretation of homotopy groups discussed in Section 23.2, this means that these configurations have a conserved additive quantum number. But we still need to show that any of these stationary configurations actually exist, and to give a physical interpretation of their topological quantum numbers.

As an illustrative example, consider a theory (like the Georgi–Glashow electroweak model[12]) in which an $SU(2)$ gauge group is spontaneously broken by the vacuum expectation value of an $SU(2)$ triplet of scalar fields ϕ_n. (It is explained at the end of this section why in this case we say that the gauge group is $SU(2)$ rather than $SO(3)$.) The Lagrangian density for the scalars and gauge fields in Minkowskian spacetime is taken as

$$\mathscr{L} = -\tfrac{1}{4}F_{n\mu\nu}F_n^{\mu\nu} - \tfrac{1}{2}D_\mu\phi_n D^\mu\phi_n - V(\phi_n\phi_n)\,, \tag{23.3.1}$$

where

$$F_{n\mu\nu} \equiv \partial_\mu A_{n\nu} - \partial_\nu A_{n\mu} + e\,\epsilon_{nml}A_{m\mu}A_{l\nu}\,, \tag{23.3.2}$$

$$D_\mu\phi_n \equiv \partial_\mu\phi_n + e\,\epsilon_{nml}A_{m\mu}\phi_l\,, \tag{23.3.3}$$

and the function $V(\phi_n\phi_n)$ is assumed to be positive, with the value zero

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

at a non-vanishing value $\langle\phi\rangle$ (taken positive) of $\sqrt{\phi_n\phi_n}$. (In much work on monopoles V is taken to be the quartic polynomial $\lambda(\phi_n\phi_n - \langle\phi\rangle^2)^2$, with $\lambda > 0$, but we shall not make this assumption here.) Eq. (23.3.1) describes a theory with a spacetime-independent vacuum solution with $A_{n\mu} = 0$, in which a vacuum expectation value ϕ_n with $\phi_n\phi_n = \langle\phi\rangle^2$ breaks $SU(2)$ to its $U(1)$ subgroup, which can be identified with the gauge group of electrodynamics. Instead we here shall seek topologically non-trivial inhomogeneous but time-independent classical solutions in temporal gauge, with $A_n^0 = 0$ but $A_n^i \neq 0$. The Lagrangian density in this case is the negative of the potential energy density $\mathscr{H}$, given by

$$\mathscr{H} = \tfrac{1}{4}F_{n\,ij}^2 + \tfrac{1}{2}(D_i\phi_n)^2 + V(\phi)\,, \tag{23.3.4}$$

with squares including obvious index contractions. Because each term in (23.3.4) is positive, the integral of each term must separately converge in a configuration of finite energy.

In particular, for the integral of $V(\phi_n\phi_n)$ to converge, the vector ϕ_n must have the fixed length $\langle\phi\rangle$ at infinity, so each configuration of finite energy defines a smooth mapping of a large two-sphere $\mathscr{S}$ surrounding the monopole configuration into the two-sphere of the ϕ_n with $\phi_n\phi_n = \langle\phi\rangle^2$. As $\hat{\mathbf{x}}$ runs over $\mathscr{S}$, ϕ_n may run over the sphere $\phi_n\phi_n = \langle\phi\rangle^2$ any integer number N of times, either with Jacobian $\mathrm{Det}(\partial x/\partial\phi)$ positive, in which case we say that the winding number is N, or with Jacobian negative, in which case the winding number is $-N$.

To see what the winding number has to do with the magnetic monopole moment, we must first consider what in this theory is observed as 'the' magnetic field. Whatever the field configuration, we can introduce a gauge in which the scalar field ϕ_n points in some definite direction, say the three-direction, in any given finite region, so that in this region the gauge field associated with the unbroken $U(1)$ subgroup of $SU(2)$ is A_{3i} . 't Hooft[2] found a gauge-invariant tensor $\mathscr{F}_{\mu\nu}$ which reduces to the usual electromagnetic field strength tensor $\partial_\mu A_{3\nu} - \partial_\nu A_{3\mu}$ in this gauge:

$$\mathscr{F}_{\mu\nu} \equiv F_{n\mu\nu}\hat{\phi}_n - \frac{1}{e}\epsilon_{nml}\,\hat{\phi}_n\,D_\mu\hat{\phi}_m\,D_\nu\hat{\phi}_l\,, \tag{23.3.5}$$

where $\hat{\phi}_n \equiv \phi_n/\sqrt{\phi_m\phi_m}$. To check that $\mathscr{F}_{\mu\nu}$ is the ordinary electromagnetic field strength tensor in a gauge with constant $\hat{\phi}_n$ (and for later purposes) we use Eqs. (23.3.2) and (23.3.3) and the identity $\epsilon_{abc}\epsilon_{ade} = \delta_{bd}\delta_{ce} - \delta_{be}\delta_{cd}$ to write $\mathscr{F}_{\mu\nu}$ in the form[13]

$$\mathscr{F}_{\mu\nu} = \partial_\mu(\hat{\phi}_n A_{n\nu}) - \partial_\nu(\hat{\phi}_n A_{n\mu}) - \frac{1}{e}\epsilon_{nml}\,\hat{\phi}_n\,\partial_\mu\hat{\phi}_m\,\partial_\nu\hat{\phi}_l\,. \tag{23.3.6}$$

Thus in a gauge in which $\hat{\phi}_n$ is a fixed unit vector in the three-direction,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

we have as promised

$$\mathscr{F}_{\mu\nu} = \partial_\mu A_{3\nu} - \partial_\nu A_{3\mu} \,.$$

The magnetic monopole moment g of any localized field configuration is defined as $1/4\pi$ times the magnetic flux through a large closed surface $\mathscr{S}$ around the configuration:

$$4\pi g \equiv \frac{1}{2}\epsilon_{ijk}\int_{\mathscr{S}} \mathscr{F}_{ij}\, d^2S_k \,. \tag{23.3.7}$$

The first two terms in Eq. (23.3.6) for $\mathscr{F}_{ij}$ are derivatives, and therefore do not contribute to the integral (23.3.7), so that

$$g = -\frac{1}{8\pi e}\epsilon_{ijk}\epsilon_{nml}\int_{\mathscr{S}} \hat{\phi}_n\, \partial_i\hat{\phi}_m\, \partial_j\hat{\phi}_l\, d^2S_k \,. \tag{23.3.8}$$

This has the important property of being a topological invariant: integrating by parts where necessary, we can see that the change in g under an infinitesimal variation $\delta\hat{\phi}_n$ in $\hat{\phi}_n$ is

$$\delta g = -\frac{3}{8\pi e}\epsilon_{ijk}\,\epsilon_{nml}\int_{\mathscr{S}} \delta\hat{\phi}_n\, \partial_i\hat{\phi}_m\, \partial_j\hat{\phi}_l\, d^2S_k \,.$$

But because $\hat{\phi}$ is a unit vector, $\delta\hat{\phi}$ as well as $\partial_i\hat{\phi}$ and $\partial_j\hat{\phi}$ are all in the plane perpendicular to $\hat{\phi}$, so

$$\epsilon_{nml}\,\delta\hat{\phi}_n\,\partial_i\hat{\phi}_m\,\partial_j\hat{\phi}_l = 0 \,,$$

and therefore $\delta g = 0$. The quantity (23.3.8) is related to the topological invariant known as the *Kronecker index*.

Because g is a surface integral, it is additive: for any two distant localized configurations, the surface $\mathscr{S}$ used to calculate g may be taken to be a pair of spheres, one surrounding each configuration, connected by a thin neck between them, so the value of g for the whole system will be the sum of the values of g_1, g_2 for the individual localized configurations. Furthermore, since g is a topological invariant, we will have $g = g_1 + g_2$ for any field configuration that is formed by a smooth fusion of two configurations with magnetic monopole moments g_1 and g_2. It follows that g must be proportional to the winding number. Arafune, Freund, and Goebel[13] verified this and calculated the coefficient of proportionality by using the formula (23.3.8) for general winding number. Here we will simply calculate the coefficient by studying the 't Hooft–Polyakov monopole,[2] in which the fields have unit winding number.

As we saw in the previous section, the 'identity' (as opposed to the unit) element $\mathscr{I}$ of $\pi_2(SU(2)/U(1))$, which corresponds to the element 'one' of Z, consists of configurations in which the two-sphere $\mathscr{S}$ at infinity is mapped once (with positive Jacobian) into the sphere described by $\hat{\phi}_n$. As a representative of this homotopy class, we may take a configuration

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

in which at infinity ϕ_n is in the same direction as $\mathbf{x}$. To construct this configuration let us impose a symmetry under joint rotations on $\boldsymbol{\phi} = \{\phi_1, \phi_2, \phi_3\}$ and $\mathbf{x}$, as well as parity conservation, and make the ansatz:

$$\phi_n = \hat{x}_n \langle \phi \rangle F(r) \,, \tag{23.3.9}$$

$$A_{ni} = \frac{\epsilon_{nil}\hat{x}_l}{er} G(r) \,. \tag{23.3.10}$$

There is an important similarity between this field configuration and that for a vortex line in a superconductor. The solution $\phi = \pm \ell\varphi/2e$ for the Goldstone boson field in a vortex line found in Section 21.6 shows that, although gauge invariance and rotational invariance are spontaneously broken, the vortex line solution is invariant under the combination of a global gauge transformation for which $\phi \to \phi + \Lambda$ and a rigid rotation $\varphi \to \varphi \pm 2e\Lambda/\ell$. Similarly, a monopole solution of the form (23.3.9)–(23.3.10) is not invariant under rotations or gauge transformations, but it is invariant under the combination of a rigid three-dimensional spatial rotation and an equal global $SO(3)$ gauge transformation.

As already mentioned, in order for the integral of $V(\phi_n\phi_n)$ to converge it is necessary for $\phi_n\phi_n$ to approach $\langle\phi\rangle^2$ as $r \to \infty$, so in this limit $F(r) \to 1$. To derive the limiting behavior of $G(r)$, note that the covariant derivative of the scalar field is

$$D_i\phi_n = \pm \langle \phi \rangle \left[\Big(1 - G(r)\Big)\Big(\delta_{ni} - \hat{x}_n\hat{x}_i\Big) \frac{F(r)}{r} + \hat{x}_n\hat{x}_i F'(r) \right] \,, \tag{23.3.11}$$

so the scalar term in the Hamiltonian density is

$$\tfrac{1}{2}(D_i\phi_n)^2 = \langle\phi\rangle^2 \left[\frac{F^2(1-G)^2}{r^2} + \frac{F'^2}{2} \right] \,. \tag{23.3.12}$$

For this to have a finite integral it is necessary also that $G(r) \to 1$ and $F'(r) \to 0$ as $r \to \infty$. Finally, the field strength (23.3.2) is

$$F_{nij} = \frac{\epsilon_{ijk}}{e} \left[-\frac{1}{r} G'(r)\Big(\delta_{nk} - \hat{x}_n\hat{x}_k\Big) - \frac{1}{r^2}\Big(2G(r) - G^2(r)\Big)\hat{x}_n\hat{x}_k \right] \,, \tag{23.3.13}$$

so the Yang–Mills term in the Hamiltonian density is

$$\tfrac{1}{4}(F_{nij})^2 = \frac{1}{e^2}\left[\frac{G'^2}{r^2} + \frac{(2G - G^2)^2}{2r^4} \right] \,. \tag{23.3.14}$$

This has an integral that converges at large distances as long as $G'(r)$ vanishes sufficiently rapidly as $r \to \infty$.

We can use these results to calculate the magnetic monopole moment of this configuration. Eq. (23.3.13), together with the limits $G(\infty) = 1$ and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$G'(\infty) = 0$, shows that for $r \to \infty$, we have

$$F_{nij} \to -\frac{\epsilon_{ijk}\hat{x}_k\hat{x}_n}{er^2} . \tag{23.3.15}$$

Since $D_i\phi_n$ vanishes rapidly as $r \to \infty$, the magnetic part of the field strength tensor in this gauge is given for $r \to \infty$ by the first term in Eq. (23.3.5), so at large distances the magnetic field becomes

$$B_i \equiv \tfrac{1}{2}\epsilon_{ijk}\mathcal{F}_{jk} \to \tfrac{1}{2}\epsilon_{ijk}\hat{\phi}_n F_{njk} \to -\frac{\hat{x}_i}{er^2} . \tag{23.3.16}$$

Thus *this configuration has magnetic monopole moment* $g = -1/e$. According to the general argument above, the magnetic moment of a configuration with winding number v, corresponding to the element v of Z, is then

$$g_v = -v/e . \tag{23.3.17}$$

It requires a detailed numerical calculation to find the stable configuration that minimizes the integral of the Hamiltonian density (23.3.4). However there is a limiting case where an analytic solution is available. To see this, it is useful first to derive a general lower bound due to Bogomol'nyi[10] on the monopole energy for a given magnetic monopole moment g. Note that (23.3.4) may be written

$$\mathcal{H} = \tfrac{1}{4}\Big(F_{nij} \mp \epsilon_{ijk}D_k\phi_n\Big)^2 \pm \tfrac{1}{2}\epsilon_{ijk}F_{nij}D_k\phi_n + V(\phi_n\phi_n) . \tag{23.3.18}$$

Using the Bianchi identity (15.3.9), the second term may be written

$$\pm \tfrac{1}{2}\epsilon_{ijk}F_{nij}D_k\phi_n = \pm \tfrac{1}{2}\epsilon_{ijk}D_k\Big(F_{nij}\phi_n\Big) = \pm \tfrac{1}{2}\epsilon_{ijk}\partial_k\Big(F_{nij}\phi_n\Big) ,$$

so its integral is given by the magnetic monopole moment g

$$\pm \tfrac{1}{2}\epsilon_{ijk}\int d^3x\, F_{nij}D_k\phi_n = \pm\langle\phi\rangle\int \mathbf{B}\cdot d\mathbf{A} = \pm 4\pi\langle\phi\rangle g .$$

Since every other term in $\mathcal{H}$ is positive, we have a general lower bound on the energy of a configuration with magnetic monopole moment g

$$E = \int d^3x\, \mathcal{H} \geq 4\pi\langle\phi\rangle|g| . \tag{23.3.19}$$

For $g = \pm 1/e$, this gives an energy $E \geq 4\pi\langle\phi\rangle/e$, which for small coupling constant e is much greater than the corrections due to quantum fluctuations, which are at most of order $\langle\phi\rangle$. This is why we can take such a classical configuration seriously as the leading term in a perturbation expansion.

Now, it is tempting to try to minimize the energy for a given magnetic monopole moment by setting the first term in Eq. (23.3.18) equal to zero, so that

$$F_{nij} = \pm\epsilon_{ijk}D_k\phi_n , \tag{23.3.20}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

but in general this does not lead to a configuration at which the energy is stationary. The condition that the energy should be stationary with respect to variations in the scalar field is the field equation

$$D_k D_k \phi_n = 2\phi_n V'(\phi_n \phi_n) , \tag{23.3.21}$$

while Eq. (23.3.20) together with the Bianchi identity (15.3.9) would imply that $D_k D_k \phi_n = 0$. This argument suggests that in the special case where $V(\phi_n \phi_n)$ is very small, it *will* be possible nearly to reach the lower bound (23.3.19), and hence minimize the energy for a given magnetic monopole moment, by imposing the condition (23.3.20). (Where V is the quartic polynomial $\lambda(\phi_n \phi_n - \langle\phi\rangle^2)^2$, the assumption that V is small means that $\lambda \ll e^2$, as in a type I superconductor.) Such stable configurations were studied in this way by Bogomol'nyi[10]. They had been found earlier by Prasad and Sommerfeld[14] without direct use of Eq. (23.3.20), and are usually called *BPS monopoles*.

The Bogomol'nyi condition (23.3.20) provides first-order differential equations for $F(r)$ and $G(r)$, which are much easier to solve than the second-order field equations derived directly from the condition that the energy should be stationary. Using Eqs. (23.3.11) and (23.3.13), the terms in Eq. (23.3.20) proportional to $\epsilon_{ijk}[\delta_{kn} - \hat{x}_k \hat{x}_n]$ and $\epsilon_{ijk}\hat{x}_k \hat{x}_n$ respectively yield the differential equations

$$e\langle\phi\rangle F(1 - G) = G' , \tag{23.3.22}$$

$$e\langle\phi\rangle r^2 F' = G(2 - G) . \tag{23.3.23}$$

With the boundary condition that $F(r) \to 1$ and $G(r) \to 1$ as $r \to \infty$, these equations have the solution

$$F = \coth \rho - \frac{1}{\rho} , \qquad G = 1 - \frac{\rho}{\sinh \rho} , \tag{23.3.24}$$

where $\rho \equiv e\langle\phi\rangle r$. Note in particular that the field ϕ_n given by Eqs. (23.3.9) and (23.3.24) vanishes for $r \to 0$, so as remarked in Section 23.1, the $SU(2)$ symmetry is restored at the center of the monopole.

Let's now return to the case of a potential V of arbitrary strength. The 't Hooft–Polyakov monopole is stable, because there are no configurations of smaller topological quantum number into which it could decay. Configurations of higher magnetic monopole moment are generally unstable.[10] There are also interesting configurations with both magnetic monopole moments and electric charge, known as *dyons*.[15]

There is another way of understanding the value $1/e$ of the 't Hooft–Polyakov magnetic monopole moment, which goes back to the original work of Dirac on magnetic monopoles.[16] As mentioned earlier, instead of the gauge we have been using, we can make a gauge transformation $\phi_n \to R_{nm}(\mathbf{x})\phi_m$ that rotates ϕ_n to point in a fixed direction $\hat{\mathbf{v}}$, for instance

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

the three-direction. Then the field strength $B_{nk} \equiv \frac{1}{2}\epsilon_{ijk}F_{nij}$ is transformed into $R_{nm}B_{mk}$, which approaches $\hat{\mathbf{v}}_n\hat{x}_k/er^2$ for $r \to \infty$, so here we do not have to project this on a local unbroken symmetry direction. The price we need to pay for these conveniences is that the gauge transformation is singular; the rotation that takes a vector in the $\hat{\mathbf{x}}$ direction into some fixed direction $\hat{\mathbf{v}}$ is

$$R(\hat{\mathbf{x}};\hat{\mathbf{v}}) = 1 - \Big(1 - \hat{\mathbf{v}}\cdot\hat{\mathbf{x}}\Big)\hat{\mathbf{v}}\hat{\mathbf{v}}^T + \hat{\mathbf{v}}\Big(\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}}\Big)^T + \Big(\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}}\Big)\hat{\mathbf{v}}^T - \frac{\Big(\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}}\Big)\Big(\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}}\Big)^T}{1 + \hat{\mathbf{x}}\cdot\hat{\mathbf{v}}}, \tag{23.3.25}$$

which is singular at $\hat{\mathbf{x}} = -\hat{\mathbf{v}}$. This R is not unique; for instance, we could perform the rotation $R(\hat{\mathbf{x}};-\hat{\mathbf{v}})$ that takes $\hat{\mathbf{x}}$ into the $-\hat{\mathbf{v}}$ direction, followed by a fixed rotation of 180° around some axis perpendicular to $\hat{\mathbf{v}}$, but this would become singular at $\hat{\mathbf{x}} = +\hat{\mathbf{v}}$. To avoid the singularities, we have to adopt different gauges in different regions; for instance, with $\hat{\mathbf{v}}$ in the three-direction, we can use a gauge for $0 < \theta < \theta_0$ that is singular at $\theta = \pi$ and a gauge for $\theta_0 < \theta < \pi$ that is singular at $\theta = 0$, where θ_0 is an arbitrary angle with $0 < \theta_0 < \pi$, often taken as $\pi/2$. Everywhere except at $\theta = 0$ and $\theta = \pi$ the magnetic field will be given at large distances by $B \to g\hat{\mathbf{x}}/r^2$, where g is the magnetic monopole strength. This can be written as the curl $\nabla \times \mathbf{A}$ of a vector potential whose only non-vanishing component is in the azimuthal φ direction. For $0 < \theta < \theta_0$ we must take $A_\varphi = g(1 - \cos\theta)/r\sin\theta$, which is only singular at $\theta = \pi$, while for $\theta_0 < \theta < \pi$ we must take $A_\varphi = -g(1 + \cos\theta)/r\sin\theta$, which is only singular at $\theta = 0$. The difference $\Delta\mathbf{A}$ between these two vector potentials is a gradient $\nabla\Lambda$ with $\Lambda = 2g\varphi$, which of course does not affect the magnetic field for $0 < \theta < \pi$, but could affect the dynamics of charged fields. A gauge transformation with $\Lambda = 2g\varphi$ will change a field of charge q by a factor $\exp(2iqg\varphi)$, which is not single-valued unless $2qg$ is an integer. This is the Dirac quantization condition; the existence of any magnetic monopole with monopole moment g would require all electric charges to be integer multiples of $(2g)^{-1}$. For the 't Hooft–Polyakov monopole this condition is automatically satisfied, because here $g = 1/e$, and all charges in the Georgi–Glashow model are integer multiples of $e/2$.

The Georgi–Glashow model was ruled out as a theory of weak and electromagnetic interactions by the discovery of neutral currents, but magnetic monopoles are expected to occur in other theories, where a simply connected group G is spontaneously broken not to $U(1)$, but to some subgroup $H' \times U(1)$, where H' is simply connected. (According to Appendix B of this chapter, for simply connected groups G we have $\pi_2(G/H) = \pi_1(H)$, which for $H = H' \times U(1)$ equals $\pi_1(H') \times \pi_1(U(1)) = \pi_1(U(1)) = Z$.) There are no monopoles produced in the spontaneous breaking of the gauge group

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$SU(2) \times U(1)$ of the standard electroweak theory, which is not simply connected. (About this, more below.) But we do find monopoles when the simply connected gauge group G of theories of unified strong and electroweak interactions, such as $SU(4) \times SU(4)$ or $SU(5)$ or $Spin(10)$, is spontaneously broken to the gauge group $SU(3) \times SU(2) \times U(1)$ of the standard model. (See Section 21.5.) The monopoles in this case are expected to have a mass larger by an inverse square gauge coupling constant than the vector boson masses $M \approx 10^{15}$–10^{16} GeV produced by this symmetry breaking. Such monopoles would have been produced when the universe underwent a phase transition in which G was spontaneously broken to $SU(3) \times SU(2) \times U(1)$, at a temperature T of order M.

This poses a problem for some cosmological models.[17] The scalar fields before this phase transition would have necessarily been uncorrelated at distances larger than the horizon distance, the furthest distance that light could have travelled since the initial singularity. At an early time t in standard cosmological theories[18] the horizon distance is of order $t \approx (G_N T^4)^{-1/2}$ (where $G_N \simeq (10^{19}\ \text{GeV})^{-2}$ is Newton's constant), so the number density of monopoles produced at this time would have been of order $t^{-3} \approx (G_N M^4)^{3/2}$, which is smaller than the photon density M^3 at $T \approx M$ by a factor of order $(G_N M^2)^{3/2}$. For $M \approx 10^{15}$ GeV this factor is of order 10^{-12}. If monopoles did not find each other to annihilate, then this ratio would remain roughly constant to the present, but with at least 10^9 microwave background photons per nucleon today, this would give at least 10^{-3} monopoles per nucleon, in gross disagreement with what is observed. This potential paradox was one of the factors leading to inflationary cosmological models,[19] in which there was a period of exponential expansion, which if it occurred before the monopoles were produced would have greatly extended the horizon, and if it occurred after the production of monopoles (but before a period of reheating) would have greatly diluted the monopole density.

The discovery of monopoles of any sort would create opportunities for the observation of remarkable phenomena, including the existence of fermion–monopole configurations of fractional fermion number,[20] and the violation of baryon conservation in fermion–monopole scattering.[21]

* * *

In the above discussion we have considered only monopoles associated with the spontaneous breakdown of a simply-connected gauge group G. This raises a question. For every Lie group G, whether simply connected or not, there is a simply-connected group $\overline{G}$ with the same Lie algebra, known as its *covering group*. (For examples, see Section 2.7.) Any non-simply connected group has fewer representations than its covering group (for instance, the doubly-connected groups $SO(n)$ have only scalar, vector,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

and tensor representations, while their covering groups $Spin(n)$ also have spinor representations). If a theory does not happen to involve fields that belong to the extra representations of the covering group, then we are free to consider the gauge group of the theory to be either the non-simply connected group G or its covering group $\overline{G}$? In particular, does the menu of possible monopoles depend on whether the theory contains only fields transforming as representations of a non-simply connected group G, or additional fields that would only furnish representations of its covering group $\overline{G}$? For instance, in the original Georgi–Glashow model[12] the only scalar fields belonged to a representation of $SO(3)$, the three-vector, but there were fermions that belonged to spinor representations of the covering group $Spin(3) = SU(2)$. Would the allowed values of the magnetic monopole moment change if we added scalar fields belonging to the spinor representations of $SU(2)$? Would it change if we removed the fermions?

The answer is that the menu of possible monopoles does not depend on whether we say that the gauge group is a non-simply connected group G or its covering group $\overline{G}$, and therefore is unaffected when we add or remove fields belonging to representations of $\overline{G}$ that are not representations of G. As we saw in Section 23.1, in general the topologically stable monopole-like configurations are classified according to the elements of $\pi_2(G/H)$. According to a result quoted in Appendix B of this chapter, this homotopy group consists of those elements of $\pi_1(H)$ that correspond to the trivial element of $\pi_1(G)$ when H is embedded in G. But if we replace G with its covering group $\overline{G}$, we also replace H with a different subgroup H', because some of the loops in H do not return to the base point when H is embedded in $\overline{G}$. These are just the loops that do not become trivial when H is embedded in G, so $\pi_2(G/H) = \pi_1(H')$, and thus as far as monopoles are concerned we could just as well say that the gauge group is $\overline{G}$ rather than G.

For instance, as far as the scalar fields are concerned, the gauge group of the Georgi–Glashow model might be considered to be the doubly-connected $SO(3)$ rather than its simply-connected covering group $SU(2)$. The unbroken subgroup is then $SO(2)$, in which we identify transformations that differ by a 360° rotation. Thus $\pi_1(SO(2))$ includes loops that extend from the unit element to a 360° rotation, which would not be loops when $SO(3)$ is embedded in $SU(2)$. But $\pi_2(SO(3)/SO(2))$ is not the same as $\pi_1(SO(2))$, but rather excludes the loops that are homotopically non-trivial when $SO(3)$ is embedded in $SU(2)$, which are just the loops that extend from the unit element to a 360° rotation, so $\pi_2(SO(3)/SO(2))$ is the $U(1)$ subgroup of $SU(2)$, just as if the gauge group were taken to be $SU(2)$ from the beginning.

It is convenient always to consider the gauge group G associated with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

any semi-simple gauge algebra to be the simply-connected covering group, so that we can use the simple result that $\pi_2(G/H) = \pi_1(H)$. As we have just seen, the connectivity properties of H are then fixed by its embedding in G, or more precisely, by the embedding of the Lie algebra of H in the Lie algebra of G. For instance, an $SU(3)$ gauge algebra might be spontaneously broken either into an $SU(2)$ subalgebra, under which the defining representation of $SU(3)$ transforms as a doublet plus a singlet, or into an $SO(3)$ subalgebra, under which the defining representation of $SU(3)$ transforms as a three-vector. In the first case we do not have the option of considering the unbroken subgroup to be $SO(3)$; since $\pi_1(SU(2)) = 0$, there are no monopoles here. In the second case the unbroken gauge group must be regarded as $SO(3)$, not $SU(2)$, so the theory does have monopole-like configurations, classified according to the elements of $\pi_1(SO(3)) = Z_2$. The nature of the unbroken subalgebra of H and its embedding in the gauge algebra G may be dynamically affected by the variety of field types that we introduce in the Lagrangian, but once the algebra of H and its embedding in the algebra of G are fixed, the menu of monopoles is otherwise entirely unaffected by the variety of fields in the theory.

In particular, the argument that led to the Dirac quantization condition shows that, in any theory in which a Lie algebra G is spontaneously broken to a subalgebra including the electric charge operator, the allowed magnetic monopole moments are integer multiples of the reciprocal of the smallest electric charge that appears in the representations of the covering group of G, whether or not there is actually any particle that carries that charge in the theory. If the algebra of G itself contains a $U(1)$ generator, then we must consider the covering group of this $U(1)$, which is the non-compact group of translations along the real line. If this $U(1)$ generator appears as a term in the electric charge operator, as in the standard electroweak theory, then there is no minimum electric charge in the representations of the covering group, and hence no monopoles.

23.4 The Cartan–Maurer Integral Invariant

In understanding the topology of various compact manifolds, it is a great help that there is often a topologically invariant quantity that can be written as an integral over the manifold. This will be important in our discussion of instantons in the next section, and it has already been used in studying Wess–Zumino–Witten terms in Sections 19.8 and 22.7.

Consider a mapping of an arbitrary compact manifold $\mathscr{S}$ of odd dimensionality d with coordinates $\theta^1, \theta^2, \cdots \theta^d$ into a manifold $\mathscr{M}$ of matrices $g(\theta^1, \theta^2, \cdots \theta^d)$ with $\mathrm{Det}\, g \neq 0$. (For the applications that concern us here,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$\mathscr{S}$ is usually a sphere S_d, and the g are the elements of a Lie group G in some representation.) We define a functional of $g(\theta)$, known as the *Cartan–Maurer form*:

$$\mathscr{I}[g] = \int d\theta^1\, d\theta^2 \cdots d\theta^d\, \epsilon^{i_1 i_2 \cdots i_d} \times \mathrm{Tr}\left\{ g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_1}} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_2}} \cdots g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} \right\}, \quad (23.4.1)$$

where $\epsilon^{i_1 i_2 \cdots i_d}$ is the totally antisymmetric quantity with $\epsilon^{12\cdots d} = 1$. From the fact that $\epsilon^{i_1 i_2 \cdots i_d} = -(-1)^d \epsilon^{i_2 \cdots i_d i_1}$, we see that $\mathscr{I}[g]$ vanishes where $\mathscr{S}$ is even-dimensional, so we will restrict ourselves here to the case where d is odd. The usefulness of this quantity arises from its several remarkable properties.

First, this integral is independent of the coordinate system used to parameterize the manifold $\mathscr{S}$. This follows rather obviously from the fact that $\epsilon^{i_1 i_2 \cdots i_d}$ is a contravariant tensor density, in the sense that

$$\epsilon^{i_1 i_2 \cdots i_d} \frac{\partial \theta'^{j_1}}{\partial \theta^{i_1}} \frac{\partial \theta'^{j_2}}{\partial \theta^{i_2}} \cdots \frac{\partial \theta'^{j_d}}{\partial \theta^{i_d}} = \mathrm{Det}\left(\frac{\partial \theta'}{\partial \theta} \right) \epsilon^{j_1 j_2 \cdots j_d} .$$

Second, the integral (23.4.1) is also invariant under small deformations of the mapping $\mathscr{S} \mapsto \mathscr{M}$. Using the properties of the trace, we see that under an infinitesimal change $g \to g + \delta g$ of the function $g(\theta)$, the change in each factor $g^{-1}\partial g/\partial \theta^i$ in (23.4.1) makes the same contribution to the change in $\mathscr{I}[g]$:

$$\delta\mathscr{I}[g] = d \int d\theta^1\, d\theta^2 \cdots d\theta^d\, \epsilon^{i_1 i_2 \cdots i_d} \times \mathrm{Tr}\left\{ g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_1}} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_2}} \cdots \delta\left(g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} \right) \right\} .$$

Now, the last factor in the trace is

$$\begin{aligned} \delta\left(g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} \right) &= -g^{-1}(\theta)\, \delta g(\theta)\, g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} + g^{-1}(\theta) \frac{\partial \delta g(\theta)}{\partial \theta^{i_d}} \\ &= g^{-1}(\theta) \frac{\partial}{\partial \theta^{i_d}} \left(\delta g(\theta) g^{-1}(\theta) \right) g(\theta) . \end{aligned}$$

When we integrate by parts, the derivative $\partial/\partial\theta^{i_d}$ gives no contribution when acting on the partial derivatives $\partial g(\theta)/\partial\theta^{i_n}$ because $\epsilon^{i_1 i_2 \cdots i_d}$ is antisymmetric. The remaining $d-1$ terms where $\partial/\partial\theta^{i_d}$ acts on $g^{-1}(\theta)$ are all equal except for an alternating sign, so since there are an even number of them they add up to zero.

Finally, let us specialize to the case where $\mathscr{S}$ is the sphere S_d. Because $\mathscr{I}[g]$ is invariant under small variations of $g(\theta)$, it can be regarded as a function $\mathscr{I}(c)$ only of the homotopy class c to which $g(\theta)$ belongs. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

integrals $\mathscr{I}(c)$ (or strictly speaking, $\exp\{\mathscr{I}(c)\}$) furnish a representation of the homotopy group $\pi_d(\mathscr{M})$, in the sense that

$$\mathscr{I}(c_a \times c_b) = \mathscr{I}(c_a) + \mathscr{I}(c_b) \,. \tag{23.4.2}$$

(If $g_a(\theta)$ and $g_b(\theta)$ are elements of the homotopy classes c_a and c_b respectively, then the homotopy class $c_a \times c_b$ consists of mappings homotopically equivalent to

$$g_{ab}(\theta) = \begin{cases} g_a(2\theta_1, \theta_2 \cdots \theta_d) & 0 \le \theta_1 \le \frac{1}{2} \\ g_b(2\theta_1 - 1, \theta_2 \cdots \theta_d) & \frac{1}{2} \le \theta_1 \le 1 \end{cases} .$$

The part of the integral $\mathscr{I}[g_{ab}]$ over the hemispheres with $0 \le \theta_1 \le 1/2$ and $1/2 \le \theta_1 \le 1$ can be done by changing variables to $\theta_1' = 2\theta_1$ and $\theta_1' = 2\theta_1 - 1$, respectively, yielding the terms $\mathscr{I}(c_a)$ and $\mathscr{I}(c_b)$ in (23.4.2).)

In particular, this tells us that the homotopy classes e, c, $c \times c$, etc., and c^{-1}, $c^{-1} \times c^{-1}$, etc., have

$$\mathscr{I}(c^n) = n\mathscr{I}(c) \,. \tag{23.4.3}$$

If $\mathscr{I}(c) \neq 0$ for some c, then these invariants are all different, so the classes c^n are all different, and therefore form a subgroup Z of $\pi_d(\mathscr{M})$. This goes far to explain the great difference between the sizes of the homotopy groups for odd and even dimensions shown in Appendix B of this chapter. For instance, $\pi_1(U(1)) = Z$, $\pi_3(G) = Z$ for all simple Lie groups G, and $\pi_5(SU(n)) = Z$ for all $n \geq 3$, while for all Lie groups G, $\pi_2(G) = 0$ and $\pi_4(G)$ is finite.

As a simple example where $\mathscr{I}(c) \neq 0$, consider the homotopy group $\pi_1(U(1))$, which is the same as the group $\pi_1(S_1)$ used as an example at the beginning of the previous section. Any mapping $S_1 \mapsto U(1)$ may be characterized by ν, the number of times that the phase of the $U(1)$ element goes counterclockwise around S_1 minus the number of times it goes around S_1 in the opposite direction, as the coordinate θ goes around S_1, two mappings being homotopically equivalent if and only if they have the same ν. The νth class contains the mapping $g_\nu(\theta) = \exp(2i\nu\pi\theta)$ with $0 \le \theta \le 1$, for which

$$\mathscr{I}[g_\nu] = \int_0^1 d\theta \, \exp(-2i\nu\pi\theta) \frac{d}{d\theta} \exp(2i\nu\pi\theta) = 2i\nu\pi \,,$$

thus verifying that $\pi_1(U(1)) = Z$.

As an aid to calculating $\mathscr{I}(g)$ in less simple cases, suppose that we can continuously deform the manifold $\mathscr{M}$ into a Lie group H of dimensionality d. The result of performing the H transformation with parameters θ followed by the H transformation with parameters φ is an H transformation, say with parameters $\theta'(\theta, \varphi)$. In terms of a matrix representation

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$g(\theta)$, this reads

$$g(\varphi)g(\theta) = g(\theta'(\theta, \varphi)) .$$

Differentiating with respect to θ' with φ fixed and multiplying on the left with the inverse of this equation gives

$$\frac{\partial \theta^i}{\partial \theta'^j} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^i} = g^{-1}(\theta') \frac{\partial g(\theta')}{\partial \theta'^j} .$$

The integrand of $\mathscr{I}[g]$ at a point θ' is therefore

$$\epsilon^{j_1 j_2 \cdots j_d} \operatorname{Tr} \left\{ g^{-1}(\theta') \frac{\partial g(\theta')}{\partial \theta'^{j_1}} g^{-1}(\theta') \frac{\partial g(\theta')}{\partial \theta'^{j_2}} \cdots g^{-1}(\theta') \frac{\partial g(\theta')}{\partial \theta'^{j_d}} \right\}$$
$$= \operatorname{Det} \left(\frac{\partial \theta}{\partial \theta'} \right) \epsilon^{i_1 i_2 \cdots i_d} \operatorname{Tr} \left\{ g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_1}} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_2}} \cdots g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} \right\} .$$

Now, every Lie group H has a metric $\gamma_{ij}(\theta)$ (not necessarily unique) which is form-invariant in the sense that

$$\gamma_{ij}(\theta') \equiv \frac{\partial \theta^k}{\partial \theta'^i} \frac{\partial \theta^\ell}{\partial \theta'^j} \gamma_{k\ell}(\theta) . \tag{23.4.4}$$

For instance, we can take

$$\gamma_{ij}(\theta) = -\frac{1}{2} \operatorname{Tr} \left\{ g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^i} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^j} \right\} . \tag{23.4.5}$$

For any choice of $\gamma_{ij}(\theta)$, the determinant of Eq. (23.4.4) gives

$$\operatorname{Det} \left(\frac{\partial \theta}{\partial \theta'} \right) = \sqrt{\frac{\operatorname{Det} \gamma(\theta')}{\operatorname{Det} \gamma(\theta)}} .$$

By replacing the coordinates θ in Eq. (23.4.1) with θ', we find

$$\mathscr{I}[g] = \epsilon^{i_1 i_2 \cdots i_d} \operatorname{Tr} \left\{ g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_1}} g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_2}} \cdots g^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta^{i_d}} \right\}$$
$$\times \frac{1}{\sqrt{\operatorname{Det} \gamma(\theta)}} \int d^d \theta' \sqrt{\operatorname{Det} \gamma(\theta')} . \tag{23.4.6}$$

Since the parameters φ of the second H transformation are arbitrary, we may regard θ and θ' as independent variables, and evaluate the right-hand side of (23.4.6) at any value of θ, say $\theta^i = 0$. It will be convenient to normalize the generators t_i and coordinates θ^i so that for $\theta \to 0$,

$$g(\theta) \to 1 + 2i\theta^i t_i . \tag{23.4.7}$$

In this case, Eq. (23.4.6) reads

$$\mathscr{I}[g] = (2i)^d \epsilon^{i_1 i_2 \cdots i_d} \operatorname{Tr} \{ t_{i_1} t_{i_2} \cdots t_{i_d} \} \frac{1}{\sqrt{\operatorname{Det} \gamma(0)}} \int d^d \theta' \sqrt{\operatorname{Det} \gamma(\theta')} . \tag{23.4.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

We will be especially interested in the case $d = 3$. Bott[22] has shown that for any simple Lie group G, all continuous mappings $S_3 \mapsto G$ may be continuously deformed into mappings of S_3 into a 'standard' $SU(2)$ subgroup of G. (Where $G = SU(n)$, this standard $SU(2)$ subgroup is the one that acts only on the first two components of the defining representation of $SU(n)$. Not all $SU(2)$ subgroups of $SU(n)$ are equivalent to this one.) As remarked in Section 2.7, in a 2×2 representation the general element of $SU(2)$ may be written

$$g(\theta) = \begin{pmatrix} \theta_4 + i\theta_3 & \theta_2 + i\theta_1 \\ -\theta_2 + i\theta_1 & \theta_4 - i\theta_3 \end{pmatrix} = \theta_4 + 2i\boldsymbol{\theta} \cdot \mathbf{t}\,, \tag{23.4.9}$$

where as usual

$$t_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad t_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad t_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and θ_4 and $\boldsymbol{\theta}$ are real, with $(\theta_4)^2 = 1 - \boldsymbol{\theta}^2$. (Note that Eq. (23.4.9) is consistent with the normalization convention (23.4.7).) A straightforward calculation gives the metric (23.4.5) as

$$\gamma_{ij}(\theta) = \delta_{ij} + \frac{\theta_i \theta_j}{1 - \boldsymbol{\theta}^2}\,, \tag{23.4.10}$$

so that

$$\text{Det}\ \gamma(\theta) = \frac{1}{1 - \boldsymbol{\theta}^2}\,. \tag{23.4.11}$$

Hence Eq. (23.4.8) here reads

$$\mathscr{I}[g] = -8i\epsilon^{ijk}\ \text{Tr}\ \{t_i t_j t_k\} \int \frac{d^3\theta}{\sqrt{1 - \boldsymbol{\theta}^2}}\,.$$

Using $4t_i t_j = \delta_{ij} + 2i\epsilon^{ij\ell} t_\ell$ and $\text{Tr}\,\{t_\ell t_k\} = \frac{1}{2}\delta_{\ell k}$, we see that

$$8\epsilon^{ijk}\ \text{Tr}\ \{t_i\, t_j t_k\} = 2i\epsilon^{ijk}\epsilon^{ijk} = 12i\,.$$

Also, for the 'identity' mapping g_1, the integral here runs twice over the interior of the unit ball (because θ_4 can be positive or negative), and gives

$$\int \frac{d^3\theta}{\sqrt{1 - \boldsymbol{\theta}^2}} = 2\int_0^1 \frac{4\pi r^2\, dr}{\sqrt{1 - r^2}} = 2\pi^2\,.$$

For the class c of mappings homotopic to g_1, we have then

$$\mathscr{I}(c) = 24\pi^2 \tag{23.4.12}$$

and so

$$\mathscr{I}(c^\nu) = 24\pi^2 \nu\,. \tag{23.4.13}$$

The integer ν is known as the *winding number*. This result is for a representation and normalization conventions for which the standard

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$SU(2)$ subalgebra has generators t_i with structure constants ϵ_{ijk} and $\mathrm{Tr}\,(t_i t_j) = \frac{1}{2}\delta_{ij}$. More generally, if $[t_i, t_j] = ig\epsilon_{ijk}t_k$ and $\mathrm{Tr}\,(t_i t_j) = \frac{1}{2}Ng^2\delta_{ij}$, then

$$\mathscr{I}(c^v) = 24\pi^2 N v\ . \tag{23.4.14}$$

The results (23.4.13) or (23.4.14) show incidentally that for every simple Lie group, $\pi_3(G)$ contains Z. As listed in Appendix B to this chapter, $\pi_3(G) = Z$ for all simple Lie groups. Thus the homotopy class of $g(\theta)$ for any simple Lie group is entirely determined by its homotopy class when the group is deformed into its standard $SU(2)$ subgroup.

23.5 Instantons

As we saw in Section 23.1, the topologically non-trivial solutions of a pure gauge theory with a simple gauge group G in $d = 4$ Euclidean spacetime dimensions correspond to the elements of the homotopy group $\pi_3(G) = Z$. These are four-dimensional field configurations, known as instantons, that (for reasons discussed in the next section) must be included along with their fluctuations in path integrals. After Belavin, Polyakov, Schwarz, and Tyupkin[1] demonstrated the existence of these solutions, 't Hooft[23] showed that the inclusion of these configurations in path integrals solved the $U(1)$ problem outlined in Section 19.10. Here we shall first discuss the instantons themselves, and then consider their role in path integrals.

According to Eq. (23.1.7), in order for a topologically non-trivial gauge field to have finite action, the gauge fields must approach a pure gauge at $r \to \infty$ (where here $r \equiv \sqrt{x_i x_i}$, with i summed over the values 1, 2, 3, 4):

$$iA_i(x) \to g^{-1}(\hat{x})\,\partial_i g(\hat{x})\ , \tag{23.5.1}$$

where $A_i \equiv t_\alpha A_{\alpha i}$ and $g(\hat{x})$ is a direction-dependent element of the gauge group G. The topological invariant discussed in the previous section may therefore be written in terms of the asymptotic behavior of the gauge field

$$\begin{aligned}\mathscr{I}[g] &\equiv \int d\theta^1\, d\theta^2\, d\theta^3\ \epsilon^{abc} \\ &\quad \times \mathrm{Tr}\left\{ g^{-1}(\theta)\frac{\partial g(\theta)}{\partial\theta^a} g^{-1}(\theta)\frac{\partial g(\theta)}{\partial\theta^b} g^{-1}(\theta)\frac{\partial g(\theta)}{\partial\theta^c}\right\} \\ &= -i\lim_{r\to\infty} r^3 \int d\theta^1\, d\theta^2\, d\theta^3\ \epsilon^{abc}\frac{\partial\hat{x}_i}{\partial\theta^a}\frac{\partial\hat{x}_j}{\partial\theta^b}\frac{\partial\hat{x}_k}{\partial\theta^c}\ \mathrm{Tr}\,\{A_iA_jA_k\}\ ,\end{aligned} \tag{23.5.2}$$

where θ^a with $a = 1, 2, 3$ are any three parameters used to specify the direction of the unit four-vector $\hat{x}$. This surface integral can be evaluated using Gauss's theorem. In analogy with the current (22.2.29) in Minkowski

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

spacetime, we can define a current in Euclidean spacetime

$$G_l \equiv \epsilon^E_{lijk}\left[A_{\gamma i}F_{\gamma jk} - \tfrac{1}{3}C_{\alpha\beta\gamma}A_{\alpha i}A_{\beta j}A_{\gamma k}\right], \tag{23.5.3}$$

whose divergence is

$$\partial_l G_l = \tfrac{1}{2}\epsilon^E_{ijkl}F_{\alpha ij}F_{\alpha kl}. \tag{23.5.4}$$

(Here ϵ^E_{ijkl} is the totally antisymmetric tensor with $\epsilon^E_{1234} \equiv 1$.) We are using a representation of the gauge group with totally antisymmetric structure constants, so that

$$\mathrm{Tr}\{t_\alpha t_\beta\} = \tfrac{1}{2}N\delta_{\alpha\beta}, \tag{23.5.5}$$

with N a constant that depends on the representation in which we calculate the trace in Eq. (23.5.2). Hence Eq. (23.5.3) may be written

$$G_l \equiv (2/N)\,\epsilon^E_{lijk}\,\mathrm{Tr}\left[A_iF_{jk} + (2i/3)A_iA_jA_k\right]. \tag{23.5.6}$$

For $r \to \infty$, the field strength F_{kl} vanishes, so

$$G_l \to (4i/3N)\,\epsilon^E_{lijk}\,\mathrm{Tr}\left[A_iA_jA_k\right]. \tag{23.5.7}$$

Hence Eq. (23.5.2) gives

$$\mathscr{I}[g] = -(3N/4)\int (d^4x)_E\,\partial_l G_l = -(3N/8)\,\epsilon^E_{ijkl}\int (d^4x)_E\,F_{\alpha ij}F_{\alpha kl}. \tag{23.5.8}$$

Thus in order to demonstrate the existence of topologically non-trivial field configurations, we have to show that there are configurations for which the Chern–Pontryagin density $\epsilon^E_{ijkl}F_{\alpha ij}F_{\alpha kl}$ has a non-vanishing integral.

For this purpose, it is very useful to take advantage of what is known as the *Bogomol'nyi inequality*.[10] From the fact that

$$0 \le \int \left(F_{\alpha ij} \mp \tfrac{1}{2}\epsilon^E_{ijkl}F_{\alpha kl}\right)^2 (d^4x)_E$$

(with squares indicating obvious index contractions), we have

$$S[A] \ge \tfrac{1}{8}\left|\epsilon^E_{ijkl}\int F_{\alpha ij}F_{\alpha kl}\,(d^4x)_E\right| = |\mathscr{I}[g]|/3N, \tag{23.5.9}$$

where $S[A]$ is (up to a factor) the Euclidean action

$$S[A] \equiv \tfrac{1}{4}\int F_{\alpha ij}F_{\alpha ij}\,(d^4x)_E. \tag{23.5.10}$$

The lower bound (23.5.9) is evidently reached if and only if the gauge field is self-dual or anti-self-dual, in the sense that

$$F_{\alpha ij} = \pm\tfrac{1}{2}\epsilon^E_{ijkl}F_{\alpha kl}. \tag{23.5.11}$$

Hence any solution of the first-order equation (23.5.11) is a minimum of $S[A]$ for gauge fields of winding number unity, and hence also a solution of the second-order Yang–Mills field equation.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

Belavin *et al.*[1] found a solution of Eq. (23.5.11) of the form

$$iA_i(x) = \left(\frac{r^2}{r^2+R^2}\right) g_1^{-1}(\hat{x})\partial_i g_1(\hat{x}) , \tag{23.5.12}$$

where R is an arbitrary scale factor, and $g_1(\hat{x})$ is an element of an $SU(2)$ subgroup of the gauge group, with

$$g_1(\hat{x}) = \left(\frac{x_4 + 2i\mathbf{x}\cdot\mathbf{t}}{r}\right) , \tag{23.5.13}$$

and

$$t_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad t_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad t_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{23.5.14}$$

It is clear that this solution has the asymptotic behavior (23.5.1), with $g(\hat{\mathbf{x}})$ the same as the 'identity' mapping (23.4.9), so this solution belongs to the homotopy class of the identity map, and therefore as we saw in the previous section has winding number $\nu = 1$. From Eqs. (23.4.14) and (23.5.9) (which is here an equality) we have then

$$S[A] = 8\pi^2 . \tag{23.5.15}$$

From Eqs. (23.5.10) and (23.5.11) (with positive sign), we also have

$$\epsilon^E_{ijkl}\int F_{\alpha ij}\, F_{\alpha kl}\,(d^4x)_E = 64\pi^2 . \tag{23.5.16}$$

This solution is not unique, because it can be translated or subjected to a gauge transformation, but aside from these degrees of freedom, there are no other solutions of the field equations with winding number unity.[24]

Because we have found field configurations with $\nu = 1$, we know that there are also field configurations with any integer ν. For instance, solutions with ν a positive integer $\mathcal{N}$ can be constructed by superimposing $\mathcal{N}$ solutions with $\nu = 1$ with centers so far apart that at these distances the non-linearities of the field equations become unimportant. A solution with $\nu = -1$ negative can be found by replacing g_1 in Eq. (23.5.12) with g_1^{-1}, and solutions with ν a negative integer $-\mathcal{N}$ can be found by superimposing $\mathcal{N}$ of these at large separations. For general winding number, Eqs. (23.5.15) and (23.5.16) become

$$S[A] = 8\pi^2|\nu| , \tag{23.5.17}$$

$$\epsilon^E_{ijkl}\int F_{\alpha ij}\, F_{\alpha kl}\,(d^4x)_E = 64\pi^2\nu . \tag{23.5.18}$$

These results are for a gauge field normalized as in Eqs. (23.5.11)–(23.5.13). With this normalization, the action $I[A]$ is not $-S[A]$, but

$$I[A] = -S[A]/g^2 = -8\pi^2|\nu|/g^2 , \tag{23.5.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

where g is the conventional coupling constant. If we had used our usual convention of including a factor g in the generators and structure constant, then the action $I[A]$ would be the same as $-S[A]$, but with $A_{\alpha\mu}$ and $F_{\alpha\mu\nu}$ carrying a factor $1/g$ we would have $S[A] = 8\pi^2/g^2$ instead of Eq. (23.5.15), so in this case again the action would be $-8\pi^2/g^2$, but now in place of Eq. (23.5.18) we would have

$$\epsilon^E_{ijkl} \int F_{\alpha ij} F_{\alpha kl} (d^4x)_E = 64\pi^2 \nu / g^2 . \tag{23.5.20}$$

We shall see in the next section that in path integrals we must sum the effects of instantons of all winding numbers. The contribution of configurations of winding number $\nu \neq 0$ to Euclidean path integrals is suppressed by a factor $\exp(I[A]) = \exp(-8|\nu|\pi^2/g^2)$. In Section 23.7 we will see that the coefficient of this exponential is a negative power $-n$ of g: in quantum chromodynamics, $n = 12$. The function $g^{-n}\exp(-8|\nu|\pi^2/g^2)$ and all its derivatives with respect to g vanish at $g = 0$, so such contributions are non-perturbative — they will never be encountered in any order of perturbation theory.

This does not necessarily mean that these contributions are small. As we saw in Chapter 18, in quantum chromodynamics the coupling g is not a fixed dimensionless parameter, but a function of a sliding energy scale, and becomes large at low energies. The effective energy scale to use in the coupling in Eq. (23.5.19) is determined by quantum fluctuations, the subject of Section 23.7, but on dimensional grounds it cannot be very different from $1/R$, where R is the instanton size in Eq. (23.5.12). The instanton size is not fixed, but must be integrated over, with some weight function that depends on the process under consideration. In quantum chromodynamics the running coupling constant g_μ is given for large μ by Eq. (18.7.7) as

$$g_\mu^2 = \frac{8\pi^2}{\beta_0 \ln(\mu/\Lambda)} ,$$

where $\beta_0 = 11 - 2n_f/3$ and $\Lambda \approx 250$ MeV is the quantum chromodynamic scale factor. The factor $\exp(-8\pi^2/g^2_{1/R})$ is therefore, for small instantons,

$$\exp\left(-8\pi^2/g^2_{1/R}\right) = \left(R\Lambda\right)^{\beta_0} .$$

We cannot calculate this factor for large instantons, with $R\Lambda \gg 1$, but it is clear that in this case there is no suppression of instanton effects.

Despite initial hopes, the discovery of instantons has not led to much improvement in our ability to do quantitative calculations in quantum chromodynamics. On the other hand, as we shall now see, it has produced spectacular *qualitative* changes in our understanding of quantum chromodynamics and other gauge theories.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

The mere fact that there are solutions of the field equations for which the integral (23.5.15) does not vanish is enough to provide a solution of the $U(1)$ problem discussed in Section 19.10. Under *global* $U(1)$ transformations $\psi \to \exp(i\gamma_5\alpha)\psi$, the measure for integration over quark fields undergoes a change given by Eq. (22.2.10):

$$[d\psi][d\bar{\psi}] \to \exp\left\{ i\alpha \int \mathscr{A}(x)\,(d^4x)_E \right\} [d\psi][d\bar{\psi}]\,, \tag{23.5.21}$$

where (with the matrix t here equal to unity) the anomaly function $\mathscr{A}(x)$ is given by Eq. (22.2.45):

$$\mathscr{A}(x) = \frac{1}{16\pi^2}\,\epsilon^E_{ijkl}\,F_{ij\alpha}\,F_{kl\beta}\,\mathrm{tr}\,t_\alpha t_\beta\,. \tag{23.5.22}$$

The existence of this anomaly would not in itself solve the $U(1)$ problem, because the quantity (23.5.22) is a total derivative, and so would have a zero integral for a non-singular gauge field that vanishes sufficiently rapidly at infinity. The instanton solution vanishes only as $1/r$, and yields a non-vanishing value for this integral, given by Eqs. (23.5.22), (23.5.18), and (23.5.5) as

$$\int (d^4x)_E\,\mathscr{A}(x) = 2Nv\,, \tag{23.5.23}$$

showing that the anomaly *does* violate the $U(1)$ chiral symmetry.

As we saw in Section 22.4, the currents of baryon and lepton number also contain anomalies due to the interaction of quarks and leptons with the $SU(2) \times U(1)$ gauge fields of the standard model. Instanton configurations of the $SU(2)$ gauge field therefore produce violations of baryon and lepton conservation.[25] As noted in Section 22.4, there are various currents whose conservation is not violated by anomalies or anything else, such as baryon number minus lepton number, and the differences of electron, muon, and tau lepton numbers, so these will be conserved in any baryon- and lepton-non-conserving process. For instance, the decay of a proton or deuteron is forbidden, but the decay $\mathrm{He}^3 \to e^+ + \mu^+ + \bar{\nu}_\tau$ is allowed. The amplitudes for these effects are suppressed by the same factor $\exp(-8|v|\pi^2/g^2)$ as before, but now with g the $SU(2)$ coupling constant $e/\sin\theta$, evaluated not at a sliding scale but at the natural scale for electroweak processes, roughly of order m_Z. Taking $e^2/4\pi = 1/129$ (see Section 18.2) and $\sin^2\theta = 0.23$, the suppression factor for $|v| = 1$ is $\exp(-373)$. He^3 decay into three antileptons is not likely to be observed.

* * *

There is another approach[26] that illuminates some aspects of instantons. In temporal gauge, with $A_{\alpha 4}(\mathbf{x}, x_4) = 0$, the gauge field for $x_4 \to \pm\infty$ is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

expected to approach time-independent pure gauges for $x_4 \to \pm\infty$:

$$iA(\mathbf{x}, x_4) \equiv it_\alpha A_\alpha(\mathbf{x}, x_4) \to g_\pm^{-1}(\mathbf{x})\nabla g_\pm(\mathbf{x}) , \tag{23.5.24}$$

where $g_\pm(\mathbf{x})$ are group elements in the representation generated by t_α. Assuming that $A_\alpha(\mathbf{x}, x_4)$ vanishes for $\mathbf{x} \to \infty$, the group elements $g_\pm(\mathbf{x})$ must approach constants $g_\pm$ for $\mathbf{x} \to \infty$, so the three-spaces at $x_4 \to \infty$ may be regarded as three-spheres, with the point at infinity regarded as an ordinary point. Following the same argument that led to Eq. (23.4.13), we then have

$$\epsilon_{ijk} \int d^3x \, \mathrm{Tr} \left\{ g_\pm^{-1}(\mathbf{x}) \partial_i g_\pm(\mathbf{x}) \, g_\pm^{-1}(\mathbf{x}) \partial_j g_\pm(\mathbf{x}) \, g_\pm^{-1}(\mathbf{x}) \partial_k g_\pm(\mathbf{x}) \right\} = 24\pi^2 n_\pm , \tag{23.5.25}$$

where $n_\pm$ are integers. The integral over the boundary of four-space in Eq. (23.5.2) may be regarded as the integral over the 'plane' $x_4 = +\infty$ minus the integral over the 'plane' $x_4 = -\infty$, so from (23.5.8) and (23.5.18) (with $N = 1$) we have

$$\nu = n_+ - n_- . \tag{23.5.26}$$

The exponential factor $\exp(-8|\nu|\pi^2/g^2)$ from Eq. (23.5.19) may therefore be regarded as the amplitude for a transition from a configuration with spatial winding number n_- at $x_4 \to -\infty$ to one of spatial winding number n_+ at $x_4 \to +\infty$. The exponential form of this factor for $\nu \neq 0$ reflects the fact that this is a *tunneling* process; no continuous sequence of pure gauge fields can take us from a configuration of one spatial winding number to a configuration with a different spatial winding number.

The interpretation of the factor $\exp(-8|\nu|\pi^2/g^2)$ as a tunneling amplitude suggests that baryon- and lepton-non-conserving processes may proceed rapidly at temperatures above about 1 TeV, where instead of having to tunnel through the barrier, thermal fluctuations can take the vacuum *over* the barrier.[27] This process may be of cosmological importance, but it is still subject to the selection rules mentioned above: thermal fluctuations will not change the density of baryon number minus lepton number, or the differences of the densities of the three varieties of lepton number.

23.6 The Theta Angle

We have seen that there are configurations of arbitrary integer winding number, but how do we know that these configurations must be included in the path integral? To keep an open mind, suppose we add up configurations with arbitrary weight factors $f(\nu)$ for each winding number, leaving

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

open the possibility that some or most of these weight factors might vanish. The expectation value of a local observable $\mathcal{O}$ located within a large Euclidean spacetime volume Ω is then

$$\langle \mathcal{O} \rangle_\Omega = \frac{\sum_\nu f(\nu) \int_\nu [d\phi] \exp\left(I_\Omega[\phi]\right) \mathcal{O}[\phi]}{\sum_\nu f(\nu) \int_\nu [d\phi] \exp\left(I_\Omega[\phi]\right)} , \tag{23.6.1}$$

where ϕ stands for all the fields of the theory; the subscript ν on the integrals over ϕ indicates that we are to include only configurations of winding number ν; and $I_\Omega[\phi]$ is the integral of the Lagrangian density over the spacetime volume Ω. Now suppose that Ω is divided into two very large volumes Ω_1 and Ω_2, with $\mathcal{O}$ in the volume Ω_1. The integral over all fields with winding number ν may be written as an integral over all fields with winding number ν_1 in volume Ω_1 and winding number ν_2 in volume Ω_2, with ν_1 and ν_2 summed over all values with $\nu_1 + \nu_2 = \nu$, and so to a good approximation Eq. (23.6.1) becomes

$$\langle \mathcal{O} \rangle_\Omega = \frac{\sum_{\nu_1,\nu_2} f(\nu_1+\nu_2) \int_{\nu_1} [d\phi] \exp\left(I_{\Omega_1}[\phi]\right) \mathcal{O}[\phi] \int_{\nu_2} [d\phi] \exp\left(I_{\Omega_2}[\phi]\right)}{\sum_{\nu_1,\nu_2} f(\nu_1+\nu_2) \int_{\nu_1} [d\phi] \exp\left(I_{\Omega_1}[\phi]\right) \int_{\nu_2} [d\phi] \exp\left(I_{\Omega_2}[\phi]\right)} . \tag{23.6.2}$$

But then for general weight factors the average is not the same as if we omitted the volume Ω_2, in contradiction with our general ideas about cluster decomposition. (See Chapter 4.) In order for the factors involving the volume Ω_2 to cancel in this ratio, we must have

$$f(\nu_1 + \nu_2) = f(\nu_1) f(\nu_2) .$$

This will be the case if and only if $f(\nu)$ is of the form

$$f(\nu) = \exp(i\theta\nu) , \tag{23.6.3}$$

where θ is a free parameter. Thus in particular we cannot arbitrarily discard all configurations with non-zero winding number, because then an instanton of winding number ν in one region would have to be balanced with an instanton of winding number $-\nu$ in some other region, making it impossible to calculate expectation values without considering what is happening far from the location of the operators being measured.

The factor $f(\nu)$ may be put in a more familiar form. According to Eq. (23.5.18), with gauge fields normalized so that in the standard $SU(2)$ subgroup the structure constants are ϵ_{ijk}, the winding number may be written as an integral:

$$\nu = \frac{1}{64\pi^2} \int (d^4x)_E \, \epsilon^E_{ijkl} \, F_{\alpha ij} \, F_{\alpha kl} . \tag{23.6.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

This can be expressed in terms of a Minkowskian path integral; since $(d^4x)_E = i\,d^4x$; $F_{\alpha 34} = -iF_{\alpha 30}$; and $\epsilon^{1230} = -1$, Eq. (23.6.4) may be written

$$\nu = -\frac{1}{64\pi^2}\int d^4x\, \epsilon^{\kappa\lambda\rho\sigma} F_{\alpha\kappa\lambda} F_{\alpha\rho\sigma}\ . \tag{23.6.5}$$

Thus inclusion of the weighting factor (23.6.3) is therefore equivalent to adding a term

$$\mathscr{L}_\theta = -\frac{\theta}{64\pi^2}\epsilon^{\kappa\lambda\rho\sigma} F_{\alpha\kappa\lambda} F_{\alpha\rho\sigma} \tag{23.6.6}$$

to the Lagrangian density. But as mentioned at the beginning of Section 15.2, we might have included such a term in the Lagrangian of any non-Abelian gauge theory anyway, with arbitrary real θ.

The inclusion of a term (23.6.6) in the Lagrangian density would violate P and CP conservation. We could of course simply set $\theta = 0$, but this would invalidate what in Section 18.7 was scored as one of the successes of quantum chromodynamics: it made it *automatic* for P and CP to be conserved by the strong interactions even though they are evidently violated by the weak interactions.

To assess the physical consequences of including the term (23.6.6) in the Lagrangian, consider the effect of redefinition of all the fermion fields

$$\psi_f \to \exp(i\gamma_5\alpha_f)\psi_f\ , \tag{23.6.7}$$

where f is a flavor index, and α_f are a set of real phases. According to Eqs. (22.2.10) and (22.2.24), the effect on the measure for path integrals over fermion fields is

$$[d\psi][d\bar\psi] \to \exp\left\{\frac{-i}{32\pi^2}\int d^4x\, \epsilon_{\mu\nu\rho\sigma} F_\alpha^{\mu\nu} F_\alpha^{\rho\sigma} \sum_f \alpha_f\right\}[d\psi][d\bar\psi]\ . \tag{23.6.8}$$

(We are using generators normalized so that $\mathrm{Tr}\,(t_\alpha t_\beta) = \delta_{\alpha\beta}/2$.) Comparing this with Eq. (23.6.6) shows that this is equivalent to shifting θ by

$$\theta \to \theta + 2\sum_f \alpha_f\ . \tag{23.6.9}$$

The redefinition of the fermion fields will also change the mass terms in the Lagrangian density. In order to take account of masses involving γ_5, let us write the fermion mass term in the Lagrangian as

$$\mathscr{L}_m = -\tfrac{1}{2}\sum_f \mathscr{M}_f\bar\psi_f(1+\gamma_5)\psi_f - \tfrac{1}{2}\sum_f \mathscr{M}_f^*\bar\psi_f(1-\gamma_5)\psi_f\ , \tag{23.6.10}$$

with mass parameters $\mathscr{M}_f$ which, if complex, would violate P and CP conservation. Then the redefinition (23.6.7) changes these parameters by

$$\mathscr{M}_f \to \exp(2i\alpha_f)\mathscr{M}_f\ . \tag{23.6.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

A mere change of path integration variables cannot have any physical effect, so observable quantities cannot depend separately on θ or the phases of the mass parameters $\mathscr{M}_f$, but only on the combination

$$\exp(-i\theta)\prod_f \mathscr{M}_f\ . \tag{23.6.12}$$

In particular, we can always define the fermion fields so that $\theta = 0$, but at the price of perhaps introducing P and CP violating phases in the mass parameters.

This discussion shows that if any of the quark masses were to vanish, then the theta angle would have no effect and there would be no P or CP non-conservation in quantum chromodynamics. The analysis of quark mass ratios in Section 19.7 indicates that the quarks all have non-zero masses, though it is sometimes suggested that the inclusion of terms of second order in m_s in this analysis might allow m_u to vanish.[28] But there is no question that the u and d quarks are quite light, which leads to some suppression of the effects of a non-zero theta angle. We saw in Section 19.4 that m_u and m_d are roughly of order m_π^2/m_N (where m_N is used here as a typical quantum chromodynamic mass scale), so we might expect the effects of the theta angle to be suppressed by four factors of m_π, but this is not quite correct. With P and CP not conserved there would be a non-zero amplitude for the π^0 to disappear into the vacuum, proportional to m_π^4, but the 'tadpole' graphs with pion lines ending in such vacuum transition vertices would be enhanced by a factor m_π^{-2} from the pion propagator, giving a net effect proportional to m_π^2, not m_π^4. In particular, if we define fermion fields so that all $\mathscr{M}_f$ are real, then a non-zero theta angle would produce a P- and T-non-conserving neutron electric dipole moment d_n proportional to $|\theta|$ and m_π^2, and hence on dimensional grounds of order[29]

$$d_n \approx |\theta| e m_\pi^2/m_N^3 \approx 10^{-16}|\theta|\ e\ \text{cm}\ . \tag{23.6.13}$$

The neutron electric dipole moment is known to be less than about $10^{-25}\ e$ cm, so $|\theta| < 10^{-9}$.

In order to explain in a natural way why θ is so small, Peccei and Quinn[30] proposed a theory in which θ became in effect a dynamical variable, which could relax to a minimum of the effective potential, at which P and CP would be conserved. Their idea was taken up by Wilczek and myself,[31] who noted that it would require the existence of a light spinless particle, the *axion*. The axion that appeared in the original Peccei–Quinn model has been ruled out by experiment, but there are more general possibilities,[32] in which the axion couples too weakly to ordinary matter to have been observed.

The common feature of all versions of the axion theory is that there

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

is some $U(1)$ symmetry that is spontaneously broken at energies much higher than those associated with quantum chromodynamics, and is also broken by an anomaly involving the gluon fields. According to the general formalism of Chapters 19 and 22, the low energy effective field theory will contain a Goldstone boson field ϕ, so that, under the symmetry transformation

$$\phi \to \phi + F_\phi \epsilon \,, \tag{23.6.14}$$

the effective Lagrangian undergoes the transformation

$$\delta \mathscr{L}_{\text{eff}} = -\frac{\epsilon A}{64\pi^2} \epsilon_{\mu\nu\rho\sigma} F_\alpha^{\mu\nu} F_\alpha^{\rho\sigma} \,, \tag{23.6.15}$$

where A is a dimensionless constant of order unity, characterizing the anomaly, and F_ϕ is a constant of the order of the energy scale at which the symmetry is spontaneously broken (with ϕ defined to be canonically normalized.) Then the terms in the effective Lagrangian involving ϕ are

$$\mathscr{L}_\phi = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{64\pi^2}\frac{\phi}{M}\epsilon_{\mu\nu\rho\sigma} F_\alpha^{\mu\nu} F_\alpha^{\rho\sigma} + \cdots \,, \tag{23.6.16}$$

where $M \equiv F_\phi/A$ and '$\cdots$' denotes possible interactions involving derivatives of ϕ. Comparing Eqs. (23.6.16) and (23.6.6), we see that, for a constant ϕ, all observables will be functions not of ϕ and θ separately, but only of $\theta + \phi/M$ (This is with fermion fields defined to make the mass parameters $\mathscr{M}_f$ all real; otherwise observables will depend on $\theta - \sum_f \text{Arg}\mathscr{M}_f + \phi/M$.) If everything in the theory but the theta term (23.6.6) and the ϕ interaction in Eq. (23.6.16) conserves P or CP, then the effective potential will be even in $\theta + \phi/M$, so it will have a stationary point at $\theta + \phi/M = 0$, preserving the conservation of both P and CP. In the real world P and CP are not exact, but the only observed violations are in the weak interactions, and would shift the expectation value of ϕ only a small way[33] from $-M\theta$.

Even without specifying the underlying theory, by using effective field theory techniques it is possible to say a fair amount about the general properties of the axion. The most general Lagrangian for the axion field ϕ, now including the theta term and all interactions with u and d quarks up to order $1/M$, is of the form

$$\begin{aligned}\mathscr{L}_\phi = &-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{64\pi^2}\left[\frac{\phi}{M} + \theta\right]\epsilon_{\mu\nu\rho\sigma} F_\alpha^{\mu\nu} F_\alpha^{\rho\sigma} \\ &-\frac{if_u}{M}\,\partial_\mu\phi\,\bar{u}\gamma_5\gamma^\mu u - \frac{if_d}{M}\,\partial_\mu\phi\,\bar{d}\gamma_5\gamma^\mu d \,,\end{aligned} \tag{23.6.17}$$

where f_u and f_d are dimensionless coupling constants, expected to be of order unity. As shown by Eq. (23.6.8), the redefinition (23.6.7) of the quark fields has the effect of subjecting the quantity $\phi(x)/M + \theta$ in the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

second term of the Lagrangian (23.6.17) to the replacement

$$\frac{\phi(x)}{M} + \theta \to \frac{\phi(x)}{M} + \theta + 2\Big[\alpha_u(x) + \alpha_d(x)\Big] \,. \tag{23.6.18}$$

By choosing $\alpha_f = -(\theta + \phi/M)c_f/2$ with constant coefficients c_f satisfying $c_u + c_d = 1$, we eliminate the term in the Lagrangian involving $\epsilon_{\mu\nu\rho\sigma}F_\alpha^{\mu\nu}F_\alpha^{\rho\sigma}$, and change the mass term in the low energy quantum chromodynamic Lagrangian density to

$$\mathscr{L}_m = -m_u\,\bar{u}\exp\Big[-ic_u(\theta+\phi/M)\gamma_5\Big]u - m_d\,\bar{d}\exp\Big[-ic_d(\theta+\phi/M)\gamma_5\Big]d \,. \tag{23.6.19}$$

In addition, from the kinematic part of the quark Lagrangian we pick up a derivative interaction term

$$\tfrac{1}{2}ic_u\,(\bar{u}\gamma^\mu\gamma_5)u\,\partial_\mu\phi/M + \tfrac{1}{2}ic_d\,(\bar{d}\gamma^\mu\gamma_5)d\,\partial_\mu\phi/M \,, \tag{23.6.20}$$

so that f_u and f_d in Eq. (23.6.17) are replaced with $f'_u = f_u - c_u/2$ and $f'_d = f_d - c_d/2$. To derive an effective Lagrangian for low-energy pions and axions, we follow the procedure of Section 19.5 (but now with u and d quarks in place of protons and neutrons), and make the replacements:

$$\begin{aligned}
&\bar{u}u \to v\cos(\pi^0/F_\pi)\,, && \bar{d}d \to v\cos(\pi^0/F_\pi)\,,\\
&\bar{u}\gamma_5 u \to iv\sin(\pi^0/F_\pi)\,, && \bar{d}\gamma_5 d \to -iv\sin(\pi^0/F_\pi)\,,\\
&i\bar{u}\gamma^\mu\gamma_5 \to \tfrac{1}{2}F_\pi\partial^\mu\pi^0 + \cdots\,, && i\bar{d}\gamma^\mu\gamma_5 \to \tfrac{1}{2}F_\pi\partial^\mu\pi^0 + \cdots\,,
\end{aligned} \tag{23.6.21}$$

where v is the constant $v = \langle\bar{u}u\rangle = \langle\bar{d}d\rangle$, and '$\cdots$' denotes terms that do not have the one-pion pole. From Eqs. (23.6.19) and (23.6.20) we find the effective pion–axion Lagrangian

$$\begin{aligned}
\mathscr{L}_{\text{eff}} = &-\frac{1}{2}\partial_\mu\pi^0\partial^\mu\pi^0 - \frac{1}{2}\partial_\mu\phi'\partial^\mu\phi' - \left(\frac{f'_u - f'_d}{2M}\right)F_\pi\partial_\mu\phi'\partial^\mu\pi^0\\
&-m_u v\cos\left(\frac{\pi^0}{F_\pi} - \frac{c_d\phi'}{M}\right) - m_d v\cos\left(\frac{\pi^0}{F_\pi} + \frac{c_d\phi'}{M}\right) \,,
\end{aligned} \tag{23.6.22}$$

where ϕ' is the difference between the axion field and its expectation value

$$\phi' \equiv \phi - \langle\phi\rangle = \phi + M\theta \,. \tag{23.6.23}$$

Since c_u and c_d are arbitrary except for the condition that $c_u + c_d = 1$, we are free to eliminate the $\partial_\mu\phi\partial^\mu\pi^0$ cross term by taking $c_u = \frac{1}{2} + f_u - f_d$ and $c_d = \frac{1}{2} + f_d - f_u$, so that $f'_u = f'_d$. The quadratic part of the Lagrangian (23.6.22) is

$$\mathscr{L}_{\text{quad}} = -\frac{1}{2}\partial_\mu\pi^0\partial^\mu\pi^0 - \frac{1}{2}\partial_\mu\phi'\partial^\mu\phi' - \frac{1}{2}\begin{pmatrix}\pi^0\\ \phi'\end{pmatrix}^T M_0^2\begin{pmatrix}\pi^0\\ \phi'\end{pmatrix} \,, \tag{23.6.24}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

where

$$M_0^2 = \begin{pmatrix} (m_u + m_d)\, v/F_\pi^2 & (-m_u c_u + m_d c_d)\, v/F_\pi M \\ (-m_u c_u + m_d c_d)\, v/F_\pi M & (m_u c_u^2 + m_d c_d^2)\, v/M^2 \end{pmatrix} . \quad (23.6.25)$$

For $M \gg F_\pi$ one eigenvalue of M_0^2 is $(m_u + m_d)\, v/F_\pi^2$, which according to Eq. (19.7.20) is the π^0 squared mass. The other eigenvalue is then the axion mass

$$m_a^2 = \frac{v}{M^2} \frac{m_d m_u}{m_d + m_u} = \frac{F_\pi^2}{M^2} \frac{m_d m_u}{(m_d + m_u)^2} m_\pi^2 . \quad (23.6.26)$$

For the quark mass ratio derived in Section 19.7, this gives $m_a = 13\ \text{MeV}/M(\text{GeV})$.

We can also use this formalism to say something about the interactions of axions with hadrons. The eigenvector of M_0^2 with eigenvalue m_a^2 has a component along the original π^0 direction equal to $(m_u c_u - m_d c_d) F_\pi/(m_u + m_d) M$. As mentioned earlier, because of the one-pion pole this is the dominant axion–hadron coupling. We see that the ratio of the axion and pion production interaction amplitudes will typically be of order F_π/M. The fact that axions are not observed in such collisions indicates that $M > 3$ TeV, in contradiction with the original expectation[30,31] that the anomalous $U(1)$ symmetry is spontaneously broken by the same scalar vacuum expectation values of order 0.3 TeV that break the electroweak $SU(2) \times U(1)$ symmetry. It is possible to explain why axions are not found in reactor or accelerator experiments by taking M as an independent parameter,[32] much larger than the electroweak breaking scale, but there are still astrophysical limitations. Limits on the rate of cooling of red giant stars give[34] $M > 10^7$ GeV, while observations of the supernova SN1987A indicate[35] that $M > 10^{10}$ GeV.* Cosmological arguments suggest[36] an *upper* bound $M < 10^{12}$ GeV, leaving an open but narrow window of allowed axion parameters.

* For $M > 10^7$ GeV the axion mass would be less than about 1 eV, so that stars are hot enough to produce axions. The ratio of the axion and π^0 decay rates into two photons is expected to be of order $(F_\pi/M)^2$ times a phase space ratio of order $(m_a/m_\pi)^3$, or

$$\frac{\Gamma(a \to \gamma + \gamma)}{\Gamma(\pi^0 \to \gamma + \gamma)} = \left(\frac{F_\pi}{M}\right)^2 \left(\frac{m_a}{m_\pi}\right)^3 \approx \left(\frac{F_\pi}{M}\right)^5 .$$

Hence for $M > 10^7$ GeV the axion lifetime is expected to be longer than about 10^{24} s, which is ample time for the axion to travel even cosmological distances before decaying.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

23.7 Quantum Fluctuations around Extended Field Configurations

Topologically non-trivial four-dimensional field configurations such as instantons provide only zeroth-order contributions to path integrals. Now we must consider the effect of quantum fluctuations around these configurations.

To start in a very general way, consider a set of fields, collectively called $\phi(x)$, whose dynamics is described by the Euclidean action $I[\phi]$, and suppose we have a set of configurations $\phi_{v,u}$ at which $I[\phi]$ is stationary, where v labels the topological type of the configuration, and u represents a set of continuous collective parameters on which the configuration depends. For instance, for instantons v is the winding number, an integer, and u includes the position and scale as well as its 'direction' in the gauge group. Euclidean path integrals may then be written

$$\int [d\phi] \exp\Big(I[\phi]\Big)\mathcal{O} = \sum_v \int du \int_u [d\phi'] \exp\Big(I[\phi_{v,u}+\phi']\Big)\mathcal{O}\,, \qquad (23.7.1)$$

where the subscript on the integral over the fluctuation ϕ' indicates that we are to integrate only over fluctuations that do not entail changes in the collective parameters, and $\mathcal{O}$ stands for any product of local functions of field operators. Because $I[\phi]$ is stationary at $\phi = \phi_{v,u}$, its expansion to second order in the fluctuations takes the form

$$I[\phi_{v,u}+\phi'] \simeq I_v - \tfrac{1}{2}\int d^4x\, d^4y\, K_{xl,ym}(v,u)\,\phi'_l(x)\,\phi'_m(y)\,, \qquad (23.7.2)$$

where l and m here include spin and species indices, and $I_v \equiv I[\phi_{v,u}]$ is a function of v alone because the action is supposed to be stationary at all field configurations $\phi_{v,u}$. The integral over the fluctuation ϕ' in Eq. (23.7.1) will then yield a sum of terms from contractions of the fields in $\mathcal{O}$, times (for real bosonic fields) an over-all factor $[\mathrm{Det}\, K(v,u)]^{-1/2}$, it being understood that K now acts only in the subspace of fluctuations that do not entail changes in the collective parameters u. This factor can be written as a product over the eigenvalues $\lambda_n(v,u)$ of the 'matrix' $K(v,u)$:

$$\Big(\mathrm{Det}\, K(v,u)\Big)^{-1/2} = {\prod_n}' \lambda_n^{-1/2}(v,u)\,, \qquad (23.7.3)$$

the prime here indicating that we are to exclude the 'zero modes', the zero eigenvalues of K, for which the eigenfunctions correspond to changes in the collective parameters.

These remarks allow us to refine the results given in Section 23.5 for the coupling-constant dependence of instanton contributions with different winding numbers. Suppose we define the fields of our theory in such

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

a way that the action takes the form $I[\phi, g] = g^{-2} I_1[\phi]$ where $I_1[\phi]$ is independent of coupling constants. (For instance, as discussed at the end of Section 15.2, in Yang–Mills theories we would use as a gauge field the canonically normalized gauge field times a factor g.) Then all eigenvalues of K are proportional to g^{-2}, and the factor (23.7.3) is proportional to g to a power equal to the number of non-zero eigenvalues of K. This power is of course infinite, but it can be written as the number of *all* eigenvalues of K, which is also infinite but is independent of v, minus the number $\mathcal{N}(v)$ of zero modes, equal to the number of collective parameters, which is finite and dependent on v. We conclude then that aside from factors that do not depend on v, the contribution of fluctuations around configurations of topological type v is a factor with a coupling-constant dependence that is given in terms of the number $\mathcal{N}(v)$ of collective parameters by

$$g^{-\mathcal{N}(v)} . \tag{23.7.4}$$

This estimate is based on the approximation (23.7.2), which corresponds to a one-loop approximation, so when higher-order terms are taken into account the factor (23.7.4) will be multiplied with a power series in g, whose details depend on the operators $\mathcal{O}$ appearing in the path integral.

Let's see how this applies to instantons. The configuration with $v = 0$ of course has no collective parameters, and so its contribution to path integrals is just a power series in g. The configuration with $v = 1$ has four collective parameters giving the spacetime location of the instanton, one collective parameter giving the scale of the instanton, and a number N_1 of collective parameters corresponding to the rotations and/or global gauge group transformations that do not leave the instanton invariant, so (now including the factor $\exp(I_v)$ given by Eq. (23.5.19)), the coupling-constant dependence for $v = 1$ instantons is $g^{-5-N_1} \exp(-8\pi^2/g^2)$. For the $v = 1$ instanton (23.5.12) in an $SU(2)$ Yang–Mills theory, there are three independent rotations and three independent $SU(2)$ transformations, but since the instanton is invariant under three combined rotations and $SU(2)$ transformations, we have $N_1 = 3$, so fluctuations around the $v = 1$ instanton yield a coupling-constant dependence $g^{-8} \exp(-8\pi^2/g^2)$. In an $SU(3)$ Yang–Mills theory we have three independent rotations and eight independent $SU(3)$ transformations, but again the instanton is invariant under three independent combined rotations and $SU(3)$ transformations in the standard $SU(2)$ subgroup (say, acting on the first two components of the defining representation of $SU(3)$), and now also invariant under one additional $SU(3)$ transformation that (like hypercharge) commutes with all generators of the standard $SU(2)$ subgroup. Hence $N_1 = 3+8-3-1 = 7$, and the $v = 1$ instanton yields a factor proportional to $g^{-12} \exp(-8\pi^2/g^2)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

Now suppose that the action I also contains a term

$$-\int d^4x \int d^4y \; \bar{\psi}_l(x) \mathscr{K}_{lx,my}(v,u) \psi_m(y) ,$$

involving independent fermion fields ψ and $\bar{\psi}$. If none of these fields appear in $\mathcal{O}$ then the integral over these fields yields a factor $\mathrm{Det}\,\mathscr{K}(v,u)$, which vanishes if $\mathscr{K}(v,u)$ has any zero modes. This result can be simply understood in terms of the rules for integration over fermionic parameters. By expanding $\psi(x)$ and $\bar{\psi}(x)$ in the eigenmodes of $\mathscr{K}$, we can write the integral over $\psi(x)$ and $\bar{\psi}(x)$ as an integral over the coefficients in these expansions. The coefficients of the zero modes do not appear in the quadratic approximation to the action, so for each such fermionic zero mode we have an integral over a fermionic parameter that does not appear in the integrand, which vanishes according to the general rules of Section 9.5. The only terms in an integral over fermionic variables that do not vanish are those for which the integrand contains a single factor of each integrated variable. Hence the integral over fermionic fields in Eq. (23.7.1) will not vanish only if there is a single fermionic field in $\mathcal{O}$ for each zero mode of $\mathscr{K}$. For instantons there are fermionic zero modes whose numbers and chiralities are governed by index theorems, like the Atiyah–Singer index theorem derived in Section 22.2, so only certain processes are allowed for a given winding number. It was on this basis that 't Hooft[25] showed that the baryon- and lepton-nonconserving effective interaction produced by $\nu = 1$ instantons in the electroweak standard model must involve just one of each lepton flavor.

23.8 Vacuum Decay

A vacuum state is stable if its scalar field expectation values are at a true minimum of the effective potential. But if the scalar field expectation values are at a local minimum that is higher than the true minimum then this vacuum will be metastable. A metastable 'false' vacuum state corresponding to a local minimum will decay into the stable 'true' vacuum corresponding to the true minimum by a process of barrier penetration, analogous to nuclear alpha decay or spontaneous fission. This is not a process that can be observed in our laboratories, but it has presumably occurred several times in the history of the universe as various symmetries have become spontaneously broken, so it is important to be able to calculate the rate of such false vacuum decay. As we shall now see, this calculation involves consideration of yet another extended field configuration.[6]

Let us concentrate on the component ϕ of the scalar field multiplet that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

acquires an expectation value $\langle\phi\rangle$ in the true vacuum. For instance, in the theory of broken chiral symmetry discussed in Section 19.5, ϕ would be the fourth component of a chiral four-vector. It will turn out that, where barrier penetration is strongly suppressed, the other scalar fields (including those of any Goldstone bosons) do not affect the dominant suppression factor in the decay rate. For definiteness, we will take the Lagrangian density in the form

$$\mathscr{L} = -\tfrac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi)\,. \tag{23.8.1}$$

We assume that the lowest-order effective potential $V(\phi)$ has a true minimum at $\phi = \langle\phi\rangle$ and a local minimum at $\phi = 0$, and we adjust an additive constant in the Lagrangian density so that $V(0) = 0$, in which case $V(\langle\phi\rangle) < 0$. We want to calculate the rate at which the false vacuum state with scalar field expectation value zero will decay into the true vacuum state with scalar field expectation value $\langle\phi\rangle$.

The results of Appendix A of this chapter (Eqs. (23.A.6), (23.A.21), and (23.A.23)) show that the energy E_0 of the false vacuum state in which the scalar field vacuum expectation value vanishes is given by

$$E_0 = -\lim_{T\to\infty}\frac{1}{T}\ln\left[\int\exp\Big(-S[\phi;T]\Big)\prod_{\mathbf{x},t}d\phi(\mathbf{x},t)\right], \tag{23.8.2}$$

where $S[\phi;T]$ is the Euclidean action derived from Eq. (23.8.1)

$$S[\phi;T] = \int d^3x\int_{-T/2}^{+T/2}dt\,\left[\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi)\right], \tag{23.8.3}$$

and the integral in Eq. (23.8.2) is over all fields $\phi(\mathbf{x},t)$ satisfying the conditions

$$\phi(\mathbf{x},T/2) = \phi(\mathbf{x},-T/2) = 0\,. \tag{23.8.4}$$

The energy (23.8.2) is complex; its imaginary part will give us the decay rate.

To calculate the functional integral in Eq. (23.8.2), we look for a stationary 'point' of the Euclidean action $S[\phi,T]$. The fields at which (23.8.3) is stationary satisfy the field equations

$$0 = \frac{\delta S}{\delta\phi} = -\frac{\partial^2\phi}{\partial t^2} - \nabla^2\phi + \frac{dV(\phi)}{d\phi}\,, \tag{23.8.5}$$

subject to the boundary conditions (23.8.4). Because of these boundary conditions, such a solution is known as a *bounce*.

We shall look for these bounce solutions by making the ansatz that $\phi(\mathbf{x},t)$ is invariant under rotations around a point $\mathbf{x}_0$, t_0 in four dimensions:

$$\phi(\mathbf{x},t) = \phi(\rho) \quad \text{where} \quad \rho \equiv \sqrt{(\mathbf{x}-\mathbf{x}_0)^2 + (t-t_0)^2}\,. \tag{23.8.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

There are solutions that are not rotationally invariant in four dimensions, but these other solutions have higher values[37] of S, and so become negligible for large T. Using Eq. (23.8.6) in Eq. (23.8.5) yields the ordinary differential equation

$$\frac{d^2\phi}{d\rho^2} + \frac{3}{\rho}\frac{d\phi}{d\rho} = V'(\phi) \,. \tag{23.8.7}$$

Strictly speaking, a solution of this form is consistent with the boundary conditions (23.8.4) only when T is very large compared with the characteristic time associated with $V(\rho)$, in which case we can take T to be infinite in Eq. (23.8.4), which then becomes the condition that $\phi(\rho)$ must vanish when $\rho \to \infty$. Also $\phi(\mathbf{x}, t)$ must be an analytic function of $\mathbf{x}$ near $\mathbf{x} = 0$ at all t including at $t = t_0$, so $\phi(\rho)$ is a power series in ρ^2 for $\rho \to 0$, and in particular $d\phi/d\rho = 0$ at $\rho = 0$. With these conditions, Eq. (23.8.7) is the equation of motion of a particle of unit mass with 'position' ϕ at 'time' ρ, moving under the influence of a potential $-V(\phi)$ and a viscous force $-(3/\rho)d\phi/d\rho$, that travels from rest at some finite initial value ϕ_0 of ϕ at $\rho = 0$, and just reaches $\phi = 0$ at $\rho \to \infty$, losing its initial 'energy' $-V(\phi_0) > 0$ to viscosity along the way. The Euclidean action (23.8.2) for such a solution is

$$B \equiv \int_0^\infty 2\pi^2 \rho^3 d\rho \, \left[\frac{1}{2} \left(\frac{d\phi}{d\rho} \right)^2 + V(\phi) \right] . \tag{23.8.8}$$

It is crucial to determine the sign of B. For this purpose,[38] we use the same device that was used to prove Derrick's theorem in Section 23.1, and consider the action (23.8.8) for a modified field $\phi_R(\rho) \equiv \phi(\rho/R)$. By rescaling the variable of integration, we find

$$S[\phi_R] = \int_0^\infty 2\pi^2 \rho^3 d\rho \, \left[\frac{R^2}{2} \left(\frac{d\phi}{d\rho} \right)^2 + R^4 V(\phi) \right] . \tag{23.8.9}$$

If $\phi(\rho)$ is the solution of Eq. (23.8.7) then the action must be stationary with respect to any variation in ϕ, so that $dS[\phi_R]/dR$ must vanish at $R = 1$, and therefore

$$\int_0^\infty \rho^3 d\rho \, \left(\frac{d\phi}{d\rho} \right)^2 = -4 \int_0^\infty \rho^3 d\rho \, V(\phi) \,. \tag{23.8.10}$$

It follows then that

$$B = \frac{\pi^2}{2} \int_0^\infty \rho^3 d\rho \, \left(\frac{d\phi}{d\rho} \right)^2 > 0 \,. \tag{23.8.11}$$

We will come back later to an explicit approximate solution for $\phi(\rho)$, but first let's consider how such solutions are to be used.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

We have here not just a single one-bounce configuration at which the Euclidean action is stationary, but a continuum, characterized by the collective coordinates $\mathbf{x}_0$ and t_0. According to the results of Section 23.7, we must integrate over these parameters, which, since B is independent of $\mathbf{x}_0$ and t_0, yields factors of $\mathcal{V}$ and T in a box of spatial volume $\mathcal{V}$. The contribution of all one-bounce configurations to the functional integral in Eq. (23.8.2) is given in one-loop order by

$$\mathcal{V} T A \exp(-B) . \tag{23.8.12}$$

The coefficient A is proportional* to a product $\prod_n' \lambda_n^{-1/2}$, where the λ_n are the eigenvalues of the kernel $\delta^2 S[\phi]/\delta\phi(\mathbf{x},t)\delta\phi(\mathbf{x}',t')$, and the prime indicates that we are to omit the zero eigenvalues, which correspond to changes in the collective coordinates $\mathbf{x}_0$ and t_0. Using Eq. (23.8.10), we can see that the second derivative of Eq. (23.8.9) with respect to R is negative at $R = 1$, so there is at least one (and in fact just one[39]) negative eigenvalue, and hence to this order A is imaginary. We will not attempt to calculate A here, but will just note that A unlike $\exp(-B)$ does not have a dramatic dependence on the parameters of the theory, so that we can estimate it on dimensional grounds to be roughly of order iM^{-4}, where M is some characteristic mass scale of the theory.

For large T and $\mathcal{V}$ we can find additional stationary configurations by superimposing any number N of these bounce configurations, yielding a contribution that is the Nth power of the quantity (23.8.12), divided by $N!$ to take account of the fact that in integrating over N of the $\mathbf{x}_0$ and t_0 we are summing over configurations that only differ by permutations of the N identical bounces. Summing over N then gives the exponential of the quantity (23.8.12), so the energy (23.8.2) is just the quantity (23.8.12) divided by $-T$

$$E_0 = -\mathcal{V} A \exp(-B) . \tag{23.8.13}$$

Since A is of order iM^{-4}, the decay rate per volume of the false vacuum is thus of order

$$\Gamma/\mathcal{V} \approx M^{-4} \exp(-B) . \tag{23.8.14}$$

Note that this is a decay rate per volume, because the decay does not occur by a change in the scalar field simultaneously everywhere in space, but by the occurrence of bubbles of true vacuum in a false vacuum background.

* Where a continuous symmetry G is spontaneously broken to a subgroup H, there are additional collective coordinates giving the orientation of H within G. The integration over these parameters yields an additional factor in A equal to the volume of the coset space G/H.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

The result (23.8.14) is chiefly useful in the case where B is large, so that barrier penetration is strongly suppressed, and we can estimate the suppression factor as simply $\exp(-B)$. Fortunately, the most natural circumstance in which B is large is one in which it is possible to calculate B in closed form. This is the case in which the energy $V(\langle\phi\rangle) \equiv -\epsilon$ of the true vacuum is only slightly below the zero energy of the false vacuum, but $V(\phi)$ is positive and not small between $\phi = 0$ and $\phi = \langle\phi\rangle$. To minimize the Euclidean action (23.8.3) in this case, we must take ϕ to be near $\langle\phi\rangle$ within a four-dimensional ball with a large radius R, at which ϕ drops to zero within a shell of thickness given by some length $L \approx M^{-1}$ characteristic of the potential in the limit $\epsilon \to 0$. (This is sometimes called the 'thin wall approximation,' but perhaps a better term would be the big bubble approximation.) The action (23.8.3) in this approximation is

$$S(R) \simeq -\tfrac{1}{2}\pi^2 R^4 \epsilon + 2\pi^2 R^3 \mathscr{S} , \tag{23.8.15}$$

where $\mathscr{S}$ is a surface tension, equal to the shell contribution to the action per area. The second term on the left-hand side of Eq. (23.8.7) becomes negligible for $\rho \approx R$, so this is now essentially a one-dimensional problem. We can therefore take the surface tension from Eq. (23.1.4), which for a solution of the field equations is an equality, and in present notation reads

$$\mathscr{S} = \int_0^{\langle\phi\rangle} \sqrt{2V(f)}\, df . \tag{23.8.16}$$

The action (23.8.15) is stationary at a radius

$$R \simeq 3\mathscr{S}/\epsilon , \tag{23.8.17}$$

so the action at its stationary point has the value

$$B \simeq \frac{27\pi^2 \mathscr{S}^4}{2\epsilon^3} . \tag{23.8.18}$$

Note that B is large for small ϵ, so in this case the decay rate of the false vacuum is strongly suppressed. After the barrier has been penetrated the bubble of true vacuum will grow at the speed of light, colliding eventually with other bubbles, until all space is in the state of lowest energy.

Appendix A Euclidean Path Integrals

This appendix will outline the use of Euclidean path integrals in quantum field theory. As mentioned in Section 9.1, it is possible to formulate quantum field theory in a four-dimensional Euclidean spacetime. Instead of going into the non-trivial analytic continuation needed to calculate S-matrix elements in this approach, here we shall illustrate the use of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

Euclidean path integrals by addressing a problem for which they are naturally suited.

We consider a set of Hermitian canonical variables Q_a and P_a, with commutation relations

$$[Q_a, P_b] = i\delta_{ab} \,, \tag{23.A.1}$$

$$[Q_a, Q_b] = [P_a, P_b] = 0 \,. \tag{23.A.2}$$

In quantum field theory the index a is understood, as in Section 9.1, to consist of a spatial position $\mathbf{x}$ and any discrete Lorentz and species indices m, and the Kronecker delta in Eq. (23.A.1) is understood as $\delta_{\mathbf{x}m,\mathbf{y}n} = \delta^3(\mathbf{x}-\mathbf{y})\delta_{mn}$. We define eigenstates of the Q_a

$$Q_a|q\rangle = q_a|q\rangle \,, \tag{23.A.3}$$

normalized so that

$$\langle q'|q\rangle = \delta(q'-q) \equiv \prod_a \delta(q'_a - q_a) \,, \tag{23.A.4}$$

and likewise for eigenstates $|p\rangle$ of the P_a.

The problem we consider is the calculation of the matrix element

$$F(q',q;T) \equiv \langle q'|\exp\Big(-H[Q,P]T\Big)|q\rangle \,, \tag{23.A.5}$$

where H is the Hamiltonian, and T is an arbitrary positive constant. One application is to the study of ground-state energies. If the smallest eigenvalue of H is E_0, with eigenvector $|0\rangle$, then for $T \to \infty$

$$F(q',q;T) \to \langle q'|0\rangle\langle 0|q\rangle \exp(-E_0 T) \,,$$

so

$$E_0 = -\lim_{T\to\infty}\left(\frac{\ln F(q',q;T)}{T}\right) \,. \tag{23.A.6}$$

Also, we can calculate the partition function of statistical mechanics from the trace:

$$Z(\beta) \equiv \mathrm{Tr}\,\exp(-\beta H) = \int \left[\prod_a dq_a\right] F(q,q;\beta) \,, \tag{23.A.7}$$

where $1/\beta$ is the temperature.

To derive a path integral formula for $F(q',q;T)$ we define Euclidean time dependent operators

$$Q_a(t) \equiv e^{Ht}Q_a e^{-Ht} \,, \qquad P_a(t) \equiv e^{Ht}P_a e^{-Ht} \,, \tag{23.A.8}$$

and corresponding left- and right-eigenstates

$$|q,t\rangle \equiv \exp(Ht)|q\rangle \,, \qquad \langle q,t| \equiv \langle q|\exp(-Ht) \,, \tag{23.A.9}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

$$|p,t\rangle \equiv \exp(Ht)|p\rangle\,, \qquad \langle p,t| \equiv \langle p|\exp(-Ht)\,, \tag{23.A.10}$$

such that

$$Q_a(t)|q,t\rangle = q_a|q,t\rangle\,, \qquad \langle q,t|Q_a(t) = q_a\langle q,t|\,, \tag{23.A.11}$$

and

$$P_a(t)|p,t\rangle = p_a|p,t\rangle\,, \qquad \langle p,t|P_a(t) = p_a\langle p,t|\,. \tag{23.A.12}$$

One difference between this and the usual Minkowskian formalism is that the 'time' evolution of the operators is governed by a non-unitary similarity transformation (23.A.8), so there is no simple relation between the right-eigenstates $\langle q,t|$ of $Q(t)$ and the Hermitian adjoint of the left-eigenstates $|q,t\rangle$, except at $t=0$.

In this language, the definition (23.A.5) of $F(q',q;T)$ may be rewritten

$$F(q',q;T) = \langle q',T/2|q,-T/2\rangle\,. \tag{23.A.13}$$

Let us first calculate $\langle q',t+dt|q,t\rangle$ for infinitesimal dt. Adopting the convention that $H(Q,P)$ is written with all Qs to the left of all Ps, we have

$$\langle q',t+dt|q,t\rangle = \langle q',t|\exp\Big(-H(q',P)dt\Big)|q,t\rangle\,.$$

Let us expand $|q,t\rangle$ in a complete set of eigenstates of the $P_a(t)$ operators. From Eq. (23.A.1) we have as usual

$$\langle q,t|p,t\rangle = \prod_a \frac{\exp(ip_aq_a)}{\sqrt{2\pi}}\,, \qquad \langle p,t|q,t\rangle = \prod_a \frac{\exp(-ip_aq_a)}{\sqrt{2\pi}}\,,$$

so

$$\langle q',t+dt|q,t\rangle = \int\left(\prod_a \frac{dp_a}{2\pi}\right)\exp\left(i\sum_a p_a(q'_a-q_a) - H(q',p)dt\right)\,.$$

(The summation convention is suspended here.)

As in Section 9.1, we divide the time interval from $-T/2$ to $T/2$ into a large number of very small intervals and insert a sum over Q eigenstates for each interval. Defining functions $q(t)$ and $p(t)$ that interpolate between the values of the Q and P eigenvalues at each interval, we obtain the path integral expression for F in its most general form

$$F(q',q;T) = \int_{q(-T/2)=q,\,q(T/2)=q'}\Big(\prod_{a,t} dq_a(t)\Big)\int\Big(\prod_{a,t}\frac{dp_a(t)}{2\pi}\Big)$$
$$\times\,\exp\left(\int_{-T/2}^{T/2} dt\,\Big[i\sum_a \dot{q}_a(t)p_a(t) - H\Big(q(t),\,p(t)\Big)\Big]\right)\,. \tag{23.A.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

To calculate the partition function (23.A.7) we would integrate over the ps and qs subject only to the condition that $q(t)$ is periodic with period equal to the inverse temperature β:

$$Z(\beta) = \int_{q(\beta/2)=q(-\beta/2)} \left(\prod_{a,t} dq_a(t)\right) \int \left(\prod_{a,t} \frac{dp_a(t)}{2\pi}\right) \times \exp\left(\int_{-\beta/2}^{\beta/2} dt \left[i\sum_a \dot{q}_a(t)p_a(t) - H\Big(q(t), p(t)\Big)\right]\right) . \quad (23.A.15)$$

Eqs. (23.A.14) and (23.A.15) look a little odd, with one term in the exponent real, and the other imaginary. These formulas begin to look more familiar after we do the path integral over the $p_a(t)$s. This integral is trivial in the important class of theories where $H(q,p)$ is quadratic in the Ps:

$$H(q,p) = \frac{1}{2}\sum_{a,b} A_{ab}(q)p_a p_b + \sum_a B_a(q)p_a + C(q) . \quad (23.A.16)$$

As shown in the appendix to Chapter 9, the integral over the ps in Eq. (23.A.14) yields

$$F(q',q;T) = \int_{q(-T/2)=q,\, q(T/2)=q'} \left(\prod_{a,t} dq_a(t)\right) \Big(\mathrm{Det}\left[2i\pi\mathscr{A}(q)\right]\Big)^{-1/2} \times \exp\left(\int_{-T/2}^{T/2} dt \left[i\sum_a \dot{q}_a(t)\bar{p}_a(t) - H\Big(q(t), \bar{p}(t)\Big)\right]\right), \quad (23.A.17)$$

where $\mathscr{A}(q)$ is the 'matrix'

$$\mathscr{A}(q)_{at',bt} = \delta(t'-t)A_{ab}(q(t)) , \quad (23.A.18)$$

and $\bar{p}(t)$ is the stationary 'point' of the argument of the exponential in Eq. (23.A.17) — that is, the solution of the equation

$$i\dot{q}_a(t) = \left.\frac{\delta H\Big(q(t),p\Big)}{\delta p_a}\right|_{p=\bar{p}(t)} . \quad (23.A.19)$$

The factor i here should not be surprising, because Eq. (23.A.19) is the same equation that is satisfied by the non-Hermitian operators (23.A.8):

$$i\dot{Q}_a(t) = i[H, Q_a(t)] = \frac{\delta H\Big(Q(t),P(t)\Big)}{\delta P_a(t)} .$$

For a Hamiltonian in the general quadratic form (23.A.16), the solution of Eq. (23.A.19) is

$$\bar{p}_a = \sum_b \Big[A^{-1}(q)\Big]_{ab}\Big(iq_b - B_b(q)\Big) , \quad (23.A.20)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

and so Eq. (23.A.17) takes the form

$$F(q',q;T) = \int_{q(-T/2)=q,\, q(T/2)=q'} \Big(\prod_{a,t} dq_a(t)\Big) \times \Big(\mathrm{Det}\left[2i\pi\mathscr{A}(q)\right]\Big)^{-1/2} \exp\Big(-S[q]\Big), \quad (23.A.21)$$

where $S[q]$ is the action

$$S[q] \equiv \int_{-T/2}^{T/2} dt \left[\frac{1}{2}\sum_{ab} A^{-1}_{ab}(q)\dot{q}_a\dot{q}_b + i\sum_{ab} A^{-1}_{ab}(q)B_a(q)\dot{q}_b - \frac{1}{2}\sum_{ab} A^{-1}_{ab}(q)B_a(q)B_b(q) + C(q)\right]. \quad (23.A.22)$$

In the special (but common) case where $B_a(q) = 0$, this simplifies to

$$S[q] \equiv \int_{-T/2}^{T/2} dt \left[\frac{1}{2}\sum_{ab} A^{-1}_{ab}(q)\dot{q}_a\dot{q}_b + C(q)\right]. \quad (23.A.23)$$

Thus in this case the 'Lagrangian' appearing in the path integral is equal to what the Hamiltonian would be in Minkowskian spacetime when the ps are expressed in terms of qs and $\dot{q}$s.

Appendix B A List of Homotopy Groups

This appendix presents a list of homotopy groups[40] for various manifolds. Here Z denotes the group of the integers, with group composition defined by addition, so that zero is the unit element. Also, Z_n is the group of the integers modulo n. The trivial group, consisting of the element 0, is denoted 0. Homotopy groups for direct products of manifolds may be obtained from the homotopy groups of the manifolds themselves by the product rule:

$$\pi_n(\mathscr{M}_1 \times \mathscr{M}_2) = \pi_n(\mathscr{M}_1) \times \pi_n(\mathscr{M}_2)\,.$$

Spheres

$$\begin{aligned}
\pi_n(S_m) &= 0 \ \text{ for } \ n < m \\
\pi_n(S_n) &= Z \\
\pi_{n+1}(S_n) &= Z_2 \ \text{ except } \ \pi_2(S_1) = 0;\ \ \pi_3(S_2) = Z \\
\pi_{n+2}(S_n) &= Z_2 \ \text{ except } \ \pi_3(S_1) = 0 \\
\pi_{n+3}(S_n) &= Z_{24} \text{ except } \ \pi_4(S_1) = 0;\ \pi_5(S_2) = Z_2;\ \pi_6(S_3) = Z_{12}; \\
&\qquad\qquad \pi_7(S_4) = Z \times Z_{12} \\
\pi_n(S_1) &= 0 \quad \text{except } \ \pi_1(S_1) = Z
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

Lie Group Manifolds

$$\pi_1(G) = \begin{cases} Z & G = U(1) \\ Z_2 & G = SO(n)\ (n \geq 3) \\ 0 & \text{other simple compact connected Lie groups} \end{cases}$$

$$\pi_2(G) = 0 \qquad G \text{ any compact connected Lie group}$$

$$\pi_3(G) = Z \qquad G \text{ any compact connected simple Lie group}$$

$$\pi_4(G) = \begin{cases} Z_2 \times Z_2 & G = SO(4),\ Spin(4) \\ Z_2 & G = USp(2n),\ SU(2),\ SO(3),\ Spin(5),\ SO(5) \\ 0 & G = SU(n)\ (n \geq 3),\ SO(n)\ (n \geq 6),\ G_2,\ F_4,\ E_n \end{cases}$$

$$\pi_{4n+2}(USp(2n)) = \begin{cases} Z_{(2n+2)!} & n \text{ even} \\ Z_{2(2n+2)!} & n \text{ odd} \end{cases} \qquad \pi_{2n}(SU(n)) = Z_{n!}$$

Bott Periodicity Theorems

For $n \geq (k-1)/4$, $k \geq 2$,

$$\pi_k(USp(2n)) = \begin{cases} Z & k = 3, 7\ (mod\ 8) \\ Z_2 & k = 4, 5\ (mod\ 8) \\ 0 & k = 0, 1, 2, 6\ (mod\ 8) \end{cases}$$

For $n \geq k+2$, $k \geq 2$,

$$\pi_k(SO(n)) = \begin{cases} Z & k = 3, 7\ (mod\ 8) \\ Z_2 & k = 0, 1\ (mod\ 8) \\ 0 & k = 2, 4, 5, 6\ (mod\ 8) \end{cases}$$

For $n \geq (k+1)/2$, $k \geq 2$,

$$\pi_k(SU(n)) = \begin{cases} Z & k \text{ odd} \\ 0 & k \text{ even} \end{cases}$$

Coset Spaces

For any Lie group G and any Lie subgroup $H \subset G$,

$$\pi_2(G/H) = \ker\{\pi_1(H) \mapsto \pi_1(G)\}\ .$$

That is, $\pi_2(G/H)$ is the subgroup of $\pi_1(H)$ that maps into the trivial element of $\pi_1(G)$ when H is embedded in G. As a special case,

$$\pi_2(G/H) = \pi_1(H) \qquad \text{for } \pi_1(G) = 0\ .$$

Problems

1. What sort of term would need to be added to the action for Goldstone

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

boson fields in four-dimensional Euclidean spacetime to make it possible to have topologically non-trivial field configurations at which the action is stationary?

2. Consider a theory of scalar fields in six space dimensions in which the chiral symmetry $SU(2) \times SU(2)$ of the Lagrangian is spontaneously broken to the $SU(2)$ of isospin. Suppose that enough higher derivative terms are added to the Lagrangian so that skyrmions are stabilized. What sort of conservation law do these skyrmions obey? (Hint: Note that, as shown in Section 2.7, $SU(2)$ is topologically the same as S_3.)

3. Show that all $\pi_n(\mathcal{M})$ for $n > 1$ and arbitrary manifolds $\mathcal{M}$ are Abelian.

4. What is the coupling-constant dependence of the contribution of instantons of unit winding number in an $SU(4)$ gauge theory?

5. Derive a formula for the axion mass in the case in which m_u, m_d, and m_s are all small.

6. Consider the theory of a real scalar field ϕ with Lagrangian density

$$\mathcal{L} = -\tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi + \tfrac{1}{2}m^2\phi^2 - \tfrac{1}{4}\lambda\phi^4 - g\phi\,,$$

where m^2 and λ are positive, and g is very small. Calculate the exponential suppression factor $\exp(-B)$ in the rate of decay of the metastable vacuum state in terms of m^2, λ, and g.

References

1. A. A. Belavin, A. M. Polyakov, A. S. Schwarz and Yu. S. Tyupkin, *Phys. Lett.* **59B**, 85 (1975).

2. G. 't Hooft, *Nucl. Phys.* **B79**, 276 (1974); A. M. Polyakov, *JETP. Letters* **20**, 194 (1974).

3. T. H. R. Skyrme, *Proc. Roy. Soc. London* **A260**, 127 (1961).

4. H. Nielsen and P. Oleson, *Nucl. Phys.* **B61**, 45 (1973).

5. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun', *Sov. Phys. JETP* **40**, 2 (1975).

6. S. Coleman, *Phys. Rev.* **D15**, 2929 (1977); C. G. Callan and S. Coleman, *Phys. Rev.* **D16**, 1762 (1977). The discussion in Section 23.8 is based on these references. For earlier work on the formation

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

of regions of broken symmetry, see T. D. Lee and G. C. Wick, *Phys. Rev.* **D9**, 2291 (1974); M. B. Voloshin, I. Yu. Kobzarev, and L. B. Okun', *Sov. J. Nucl. Phys.* **20**, 644 (1975).

7. S. Coleman, *Nucl. Phys.* **B262**, 263 (1985); T. D. Lee and Y. Pang, *Phys. Rep.* **221**, 251 (1992), and other references by Lee and collaborators quoted therein.

8. For an extremely lucid review of these topics and references to much of the original literature, see S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1985): Chapters 6 and 7. A good, more recent review is given by E. J. Weinberg, *Ann. Rev. Nucl. Part. Sci.* **42**, 177 (1992).

9. G. H. Derrick, *J. Math. Phys.* **5**, 1252 (1964).

10. E. B. Bogomol'nyi, *Sov. J. Nucl. Phys.* **24**, 449 (1976).

11. Homotopy theory has been used to study symmetry breaking in the early universe by T. W. B. Kibble, *J. Phys. A* **9**, 1387 (1976); A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge University Press, Cambridge, 1994), and references quoted therein.

12. H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **28**, 1494 (1972).

13. J. Arafune, P. G. O. Freund, and C. J. Goebel, *J. Math. Phys.* **16**, 433 (1975).

14. M. K. Prasad and C. M. Sommerfeld, *Phys. Rev. Lett.* **35**, 760 (1975).

15. B. Julia and A. Zee, *Phys. Rev.* **D11**, 276 (1974); **D13**, 819 (1976).

16. P. A. M. Dirac, *Proc. Roy. Soc. London* **A133**, 60 (1934); *Phys. Rev.* **74**, 817 (1948).

17. Ya. B. Zel'dovich and M. Yu. Khlopov, *Phys. Lett.* **79B**, 239 (1978); J. Preskill, *Phys. Rev. Lett.* **43**, 1365 (1979). For a review, see J. Preskill, *Annual Rev. Nucl. Part. Science* **34**, 461 (1984).

18. See, e. g., S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972): Chapter 15.

19. A. Guth, *Phys. Rev.* **D23**, 347 (1981); A. D. Linde, *Phys. Lett.* **108B**, 389 (1982); A. Albrecht and P. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).

20. R. Jackiw and C. Rebbi, *Phys. Rev.* **D13**, 3398 (1976).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

21. V. A. Rubakov, *JETP Lett.* **33**, 644 (1981); *Nucl. Phys.* **B212**, 391 (1982); C. G. Callan, *Phys. Rev.* **D25**, 2141 (1982); *Phys. Rev.* **D26**, 2058 (1982); *Nucl. Phys.* **B212**, 391 (1982).

22. R. Bott, *Bull. Soc. Math. France* **84**, 251 (1956).

23. G. 't Hooft, *Phys. Rev.* **D14**, 3432 (1976); *Phys. Rep.* **142**, 357 (1986).

24. M. Atiyah and R. Ward, *Commun. Math. Phys.* **55**, 117 (1977).

25. G. 't Hooft, *Phys. Rev. Lett.* **37**, 8 (1976). Instead of the identification of anomaly-free conserved currents used here, 't Hooft's derivation of the selection rules was based on an analysis of the zero modes of the Dirac operator (see Section 23.7), which gives the change in each quantum number for an instanton of a given winding number.

26. C. G. Callan, R. F. Dashen, and D. J. Gross, *Phys. Lett.* **63B**, 334 (1976); R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* **37**, 172 (1976).

27. V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *Phys. Lett.* **155B**, 36 (1985). The transition is dominated by field configurations known as sphalerons; see N. S. Manton, *Phys. Rev.* **D28**, 2019 (1983); F. R. Klinkhammer and N. S. Manton, *Phys. Rev.* **D30**, 2212 (1984); R. F. Dashen, B. Hasslacher, and A. Neveu, *Phys. Rev.* **D10**, 4138 (1974).

28. H. Georgi and I. McArthur, unpublished (1981); D. B. Kaplan and A. V. Manohar, *Phys. Rev. Lett.* **56**, 2004 (1986); K. Choi, *Nucl. Phys.* **B383**, 58 (1992).

29. For a more detailed calculation, see V. Baluni, *Phys. Rev.* **D19**, 2227 (1978); R. J. Crewther. P. Di Vecchia, G. Veneziano, and E. Witten, *Phys. Lett.* **88B**, 123 (1979).

30. R. D. Peccei and H. Quinn, *Phys. Rev. Lett.* **38**, 1440 (1977); *Phys. Rev.* **D16**, 1791 (1977).

31. S. Weinberg, *Phys. Rev. Lett.* **40**, 223 (1978); F. Wilczek, *Phys. Rev. Lett.* **40**, 279 (1978).

32. J. E. Kim, *Phys. Rev. Lett.* **43**, 103 (1979); M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Nucl. Phys.* **B166**, 493 (1980); A. R. Zhitnitsky, *Sov. J. Nucl. Phys.* **31**, 260 (1980); M. Dine, W. Fischler, and M. Srednicki, *Phys. Lett.* **104B**, 199 (1981).

33. S. Weinberg, in *Neutrinos '78*, ed. E. C. Fowler (Purdue, Lafayette, 1978).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

34. D. A. Dicus, E. W. Kolb, V. I. Teplitz, and R. V. Wagoner, *Phys. Rev.* **D18**, 1829 (1978); **D22**, 839 (1980).

35. For reviews, see M. S. Turner, *Phys. Rept.* **197**, 67 (1990); G. G. Raffelt, *Phys. Rept.* **198**, 1 (1990).

36. J. Preskill, M. B. Wise, and F. Wilczek, *Phys. Lett.* **120B**, 127 (1983); L. F. Abbott and P. Sikivie, *Phys. Lett.* **120B**, 127 (1983); M. Dine and W. Fischler, *Phys. Lett.* **120B**, 137 (1983). For reviews, see J. E. Kim, *Phys. Rep.* **150**, 1 (1987); M. S. Turner, *Phys. Rep.* **197**, 68 (1990).

37. S. Coleman, V. Glaser, and A. Martin, *Commun. Math. Phys.* **58**, 211 (1978).

38. S. Coleman, Ref. 8, Section 6.2.

39. S. Coleman, Ref. 8, Section 6.5.

40. A. Actor, *Rev. Mod. Phys.* **51**, 461 (1979); *Encyclopedic Dictionary of Mathematics* (MIT Press, Cambridge, 1980): Appendix A.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:50, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174.011

Author Index

Where page numbers are given in *italics*, they refer to publications cited in lists of references.

Abachi, S. *355*
Abbot, L. *110, 477*
Abe, F. *355, 356*
Abe, K. *162*
Abrikosov, A. A. *358*
Actor, A. *477*
Ademollo, M. *293*
Adler, S. *160, 248, 250,* 361, 380, *417*
Albrecht, A. *475*
Altarelli, G. *160, 162,* 280, *294*
Alvarez, O. *418*
Alvarez-Gaumè, L. *61, 417, 418*
Amaldi, U. *357*
Anderson, P. M. 332, *357*
Anselmi, D. *110*
Arafune, J. 438, *475*
Arnison, G. *356*
Atiyah, M. F. *417, 418, 476*
Aubert, J. J. *355*
Augustine, J. E. *355*

Baker, M. *160*
Balian, R. *110*
Baluni, V. *476*
Bamert, P. *356*
Banks, T. *417, 419*
Barbieri, R. *419*
Bardeen, W. A. *161*, 332, 342, *357*, 379, 380, *417, 418*
Barnich, G. *61, 110*, 406, *420*
Batalin, I. A. *61, 62*, 405
Baulieu, L. *61*
Baur, R. *248*
Beane, S. R. *249*
Becchi, C. 28, *60, 110*
Belavin, A. A. 425, 450, 452, *474*
Bell, J. S. 361, *417*
Bender, C. M. *78*
Benfatto, G. *358*
Benvenuti, A. *356*
Bernard, C. *293*
Bernstein, J. *248*
Bijnens, J. *248, 250*
Bismut, J. M. *418*
Bjorken, J. D. 273, *293*
Bleuler, K. *60*
Bloom, E. D. *161, 293*
Bludman, S. *355*
Bogoliubov, N. N. *159*
Bogomol'nyi, E. B. 338, *358*, 440, 441, *475*
Bonora, L. *61*
Borel, A. *251*
Bott, R. *251*, 449, *476*
Bouchiat, C. *418*
Branchina, V. *79*
Brandt, F. *61, 110*, 406, *419, 420*
Breidenbach, M. *161, 293*
Brézin, E. *160, 161*
Brout, R. *354*
Brown, D. E. *358*
Burgess, C. P. *356*

Cabibbo, N. *355*
Cahill, K. *79*
Callan, C. G. *159, 160, 250*, 275, *294, 474, 476*
Carlson, C. *357, 418*
Cartan, E. 11, 26, *62*
Castorina, P. *79*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Chaikin, P. M. *160*
Chanda, R. *419*
Choi, K. *476*
Chou, K.-C. *248*
Christ, N. *60, 294*
Christensen, S. M. *418*
Chu, C-S. 416, *420*
Cohen, T. *249*
Coleman, S. 69, *78*, 127, *160, 247, 250, 251, 419, 474, 475, 477*
Collins, J. C. *160*
Cooper, F. *78*
Cooper, L. N. 332, 342, *357*
Cornwall, J. M. *357*
Cotta-Ramusino, P. *61*
Crewther, R. J. *476*
Critchley, R. *418*
Curci, G. *60, 61*

D'Hoker, E. *251*
Das, T. *248*
Dashen, R. F. *248, 250, 476*
de Boer, W. *357*
de Wit, B. *61*
De Witt, B. S. 2, 28, *60, 78, 110, 418, 419*
De Witt, C. *78*
Delbouro, R. *418*
Derrick, G. H. *475*
Deser, S. *293, 417*
Di Vecchia, P. *476*
Dicus, D. A. *477*
Dimopoulos, S. *357, 419*
Dine, M. *476, 477*
Dirac, P. A. M. 14, *475*
Dixon, J. A. *419*
Domb, C. *160*
Donoghue, J. F. *248, 249, 250*, 272, *293*
Dragon, N. *419*
Dubois-Violette, M. *419*
Duff, M. J. *418*
Duncan, A. *293, 294*
Dyson, F. J. 283, *294*
Dzyaloshinskii, I. Ye. *358*

Ecker, G. *250*
Eguchi, T. *418*
Eichten, E. *161, 357*
Einstein, A. 163
Ellis, J. *161, 355, 356, 357*
Englert, F. *354*
Ericson, T. *248*

Faddeev, L. D. 2, 28, *60, 419*
Fanchotti, S. *356*
Farhi, E. *251, 357, 419*
Farrar, G. *419*
Fayyazuddin *293*
Feldman, J. *358*
Fernandez, R. *160*
Ferrara, S. *62*
Ferrari, R. *60, 61*
Feynman, R. P. 2, *60*, 274, *293*
Finkelstein, R. J. *417*
Fischler, W. *476, 477*
Fisher, M. E. 147, *160*
Fogli, G. L. *355, 356*
Fowler, E. C. *476*
Fradkin, E. S. *61*
Fradkina, T. E. *61*
Freed, D. S. *418*
Freedman, D. Z. *62, 356*
Freund, P. G. O. *418*, 438, *475*
Friar, J. *249*
Friedman, J. L. *161*, 273
Frishman, Y. *417, 419*
Fritzsch, H. *161*
Fröhlich, J. *160*
Fubini, S. *248*
Fujikawa, K. 301, *354*, 361, 362, *417*
Fujimoto, Y. *78*
Fukuda, H. *417*
Furstmann, H. *357*

Gaillard, M. K. *161*
Gallavotti, G. *358*
Gasser, J. 232, 233, *250*
Gamow, G. *355*
Gatto, R. *161*
Gell-Mann, M. *60*, 111, 126, 157, *159, 161*, 225, 226, 231, *248, 249, 250*, 349, *356*, 413
Georgi, H. 281, *294, 355, 357, 418, 419, 475, 476*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Gibbons, G. W. *419*
Gilkey, P. B. *418*
Gilman, F. J. *356*
Gilmore, R. *62*
Ginsparg, P. *417, 418*
Ginzburg, V. L. *358*
Glaser, V. *477*
Glashow, S. L. *60, 250, 251, 355, 357, 418, 475*
Glimm, J. *160*
Goebel, C. J. 438, *475*
Goldberger, M. L. *248*
Goldman, T. *249*
Goldstone, J. *78*, 167, 191, *247, 354*
Golowich, E. *249, 250*, 272, *293*
Gomis, J. *62, 110*
Gor'kov, L. P. 337, *358*
Green, M. *356*
Green, M. S. *160*
Greenberg, O. W. *161*
Greub, W. *251*
Gribov, V. *60*
Grisaru, M. *293, 417, 418*
Gross, D. J. *59*, 153, *160, 161*, 275, 281, *294, 418, 476*
Gupta, S. N. *60*
Guralnik, G. S. *248, 250, 354*
Guth, A. *475*

Hagberg, E. *355*
Hagen, C. R. *354*
Hall, L. J. *357*
Halperin, S. *251*
Han, M. Y. *161*
Hanson, A. J. *418*
Hanson, G. *161*
Harada, M. *110*
Hardy, G. N. *294*
Hardy, J. C. *355*
Harvey, J. *358*
Hasert, F. J. *355, 356*
Hasslacher, B. *294, 476*
Hawking, S. W. *160, 418, 419*
Haymaker, R. W. *78*
Heitler, W. *60*
Henneaux, M. *61, 62, 110*, 406, *419, 420*
Herb, S. W. *355*
Higgs, P. W. *354*
Hinchliffe, I. *162*
Hindmarsh, M. *79*
Ho, P-M. *420*
Hollik, W. *356*
Holstein, B. R. *248, 249, 250*
Hubbard, J. *358*

Iliopoulis, J. *78, 355, 418*
Isham, C. *110*
Israel, W. *160*
Itzykson, C. *78*
Iverson, G. *418*
Iyanaga, S. *251*

Jackiw, R. *60, 160, 251, 357*, 361, *417, 418, 419, 475, 476*
Jaffe, A. *61, 160*
Johnson, K. *160, 357*
Johnston, D. *79*
Jona-Lasinio, G. *78*, 191, 248
Josephson, B. D. 335, *357*
Jost, R. *160*
Julia, B. *475*

Kallosh, R. E. *62*
Kaplan, D. B. *249, 476*
Kawarabayashi, K. *293*
Kawada, Y. *251*
Kelley, S. *357*
Kemmer, N. *355*
Kendall, H. W. *161*, 273
Khlopov, M. Yu. *475*
Kibble, T. W. B. *354, 475*
Killing, W. 11
Kim, J. E. *476, 477*
Kimura, T. *61*
Kirk, W. T. *161*
Klein, O. *59*
Klinkhammer, F. R. *476*
Kleinknecht, K. *356*
Kniehl, B. *356*
Kobayashi, M. *355*
Kobzarev, I. Yu. *474, 475*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Kogut, J. *161*
Kolb, E. W. *477*
Koslowsky, V. T. *355*
Kotchan, D. *248*
Kreuzer, M. *419*
Kugo, T. *60, 110*
Kursunoglu, B. *60*
Kuzmin, V. A. *476*

Landau, L. D. 136, *160, 358*
Lane, K. *161, 357*
Langacker, P. *250, 356, 357*
Lautrup, B. *60, 294*
Lavrov, P. M. *61, 110*
Le Guillou, J. C. *160, 161, 294*
Lee, B. W. *110*, 301, *354*
Lee, C. Y. *249*
Lee, T. D. *60, 355, 356, 474, 475*
Lehmann, H. *61*
Leites-Lopes, J. *355*
Leutwyler, H. *161*, 232, 233, *250*
Lévy, M. *248, 249, 419*
Lipatov, L. N. 284, *294*
Lisi, E. *355, 356*
LoSecco, J. *293*
Low, F. E. 111, 126, 157, *159, 248*, 349
Lubensky, T. C. *160*
Lurie, D. *248*
Luttinger, J. M. *160*

Maiani, L. *250, 355, 419*
Maksymyk, I. *356*
Maltman, K. *248*
Manohar, A. V. *476*
Mañes, J. *61, 418, 419*
Manton, N. S. *476*
Martin, A. *78, 477*
Maskawa, K. *355*
Mathur, V. S. *248*
McArthur, I. *476*
Mehta, M. L. *418*
Meissner, U. G. *250*
Meyer, Ph. *418*
Miller, G. *249*
Mills, R. L. 1, 11, *59*
Min, D.-P. *249*
Mitter, P. K. *61, 419*
Miyamoto, Y. *417*
Miyazawa, H. *248*
Moshhin, P. Yu *61*
Mostow, G. D. *250*
Mueller, A. H. *294*
Musset, P. *356*
Muzinich, I. *60*

Nakanishi, N. *60*
Nambu, Y. *161*, 167, 191, *247, 248*
Nanopoulos, D. V. *357*
Ne'eman, Y. 225, 226, 231, *250*, 413
Neveu, A. *476*
Nickel, B. G. *161*
Nielsen, H. *474*
Nielsen, N. K. *418*
Norton, R. E. *356*
Novikov, V. A. *356*

O'Raifeartaigh, L. *78*
Oakes, R. J. *250*
Oehme, R. *248*
Ojima, I. *60, 61*
Okubo, S. 231, *250, 251*
Okun', L. *356, 474, 475*
Oleson, P. *474*
Ordóñez, C. *249*

Pagels, H. *250*
Palais, R. S. *250*
Pancheri, G. *250*
Pang, Y. *474*
Parravicini, G. *78*
París J. *62*
Parisi, G. 280, *294*
Park, T.-S. *249*
Pati, J. *357*
Paver, N. *250*
Peccei, R. D. 458, *476*
Pendleton, H. *293, 417*
Penrose, R. *110*
Perez-Mercader, J. *78*
Perl, M. L. *355*
Perlmutter, A. *60*
Pernice, S. *294*
Perry, M. J. *418*
Peskin, M. *161*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Peterman, A. *159*
Petronzio, R. *419*
Pich, A. *250*
Polchinski, J. *60, 358*
Politzer, H. D. 153, *161*, 281, *294*
Polonsky, N. *357*
Polyakov, A. M. 425, 436, 450, *474*
Pope, C. *418*
Popov, V. N. 2, 28, *60*
Prasad, M. K. 441, *475*
Preskill, J. *251, 419, 475, 477*

Quinn, H. R. *357*, 458, *476*

Rabi, I. I. *250*
Raby, S. *357, 419*
Raffelt, G. G. *477*
Ramond, P. *356*
Ray, L. *249*
Rebbi, C. *60, 475, 476*
Renk, B. *356*
Renner, B. *250*
Reshetnyak, A. A. *61*
Rho, M. *249*
Riazuddin *293*
Romer, R. *418*
Rosenberg, L. *417*
Rouet, A. 28, *60, 110*
Roy, P. *419*
Rubako, V. A. *476*

Sakata, S. *417*
Salam, A. *78*, 167, 191, *247, 354, 355*
Samuel, S. *62*
Sanda, A. 301, *354*
Sanford, J. *250*
Sarid, U. *357*
Savage, M. *249*
Savard, G. *355*
Schrieffer, J. R. 332, 342, *357*
Schur, I. *418*
Schwarz, A. S. *61*, 425, 450, *474*
Schwimmer, A. *417, 419*
Schwinger, J. *355, 417, 419*
Schwitters, R. F. *161*
Sciama, D. *110*
Scott, L. *60*
Seely, R. *251*
Shankar, R. *358*
Shaposhnikov, M. E. *476*
Shellard, E. P. S. *475*
Shifman, M. A. *162, 476*
Shirkov, D. V. *159*
Shrauner, E. *248*
Siegel, W. *61*
Sigg, D. *249*
Sikivie, P. *477*
Singer, I. M. *61, 417, 418*
Sirlin, A. *356*
Skyrme, T. H. R. *474*
Slansky, R. *356*
Slavnov, A. A. *79*
Sokal, A. D. *160*
Sommerfield, C. M. 441, *475*
Srednicki, M. *476*
Srivastava, P. K. *418*
Steinberger, J. 360, *417*
Steinhardt, P. *475*
Sterman, G. *161*
Stora, R. 28, *60, 61, 110, 418, 419*
Stratonovich, R. L. *358*
Stueckelberg, E. C. G. *159*
Susskind, L. *357, 419*
Sutherland, D. G. 360, *417*
Suzuki, M. *293*
Svartholm, N. *354*
Symanzik, K. *78, 159*

't Hooft, G. *61, 110*, 153, *161, 251, 294*, 300, *354*, 390, 394, 395, 408, *418*
Talon, M. *419*
Tanikawa, Y. *417*
Taylor, J. C. *79*, 273
Teitelboim, C. *62*
Teller, E. *355*
Teplitz, V. I. *477*
Thierry-Mieg, J. *61*
Thirring, W. *248*
Tinkham, M. 340
Titchmarsh, E. C. *247*
Tomozawa, Y. *249*
Towner, I. S. *248, 355*
Townsend, P. K. *419*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Treiman, S. *248, 294*
Troost, W. *420*
Trubowitz, E. *258*
Turner, M. S. *477*
Tyupkin, Yu. S. 425, 450, *474*
Tyutin, I. V. 28, *60, 62, 110*

Urech, R. *248*
Utiyama, R. *60*

Vafa, C. 239, 240, *251*, 395
Vainshtein, A. I. *358, 476*
van Holten, J. W. *62*
van Kolck, U. *249*
van Nieuwenhuizen, P. *62, 356, 418, 420*
Van Proeyen, A. *420*
Vanstone, R. *251*
Velo, G. *60, 110*
Veltman, M. *110, 354, 356*, 360, *417*
Veneziano, G. *293, 476*
Viallet, C. M. *419*
Vilenkin, A. *475*
Vilkovisky, G. A. *61, 62*, 405
Voloshin, M. B. *475*
Voronov, B. L. *62, 110*
Vysotsky, M. I. *356*

Wagoner, R. V. *477*
Ward, J. *355*
Ward, R. *476*
Weinberg, E. 69, *78*, 127, *475*
Weinberg, S. *60, 78, 110, 160, 161*, 167, 191, *247, 248, 249, 250, 251, 293, 354, 355, 356, 357, 358, 418, 419*, 458, *475, 476*
Weingarten, D. 238, *251*
Weisberger, W. I. *248*
Weise, W. *248*
Wentzel, G. *355*
Wess, J. 234, *250, 251*, 396, 397, 408, *419*
Wick, G. C. *474*
Wightman, A. S. *60, 110*
Wigner, E. P. *62*
Wilczek, F. 153, *161*, 281, *294, 357*, 458, *476, 477*
Wilson, K. G. 112, 147, *160, 161*, 252, *293*
Wise, M. B. *249, 477*
Witten, E. 234, 236, 239, 240, *251*, 383, 396, 415, *418, 419, 420, 476*
Wu, Y-S. *417*
Wyler, D. *248*

Yamawaki, K. *110*
Yanagida, T. *356*
Yang, C. N. 1, 11, *59, 355*
Yankielowicz, S. *417, 419*
Young, J. E. *248*

Zakharov, V. I. *476*
Zappalà, D. *79*
Zee, A. 153, *161, 357, 417, 419, 475*
Zel'dovich, Ya. B. *474, 475*
Zhitnitsky, A. *476*
Zichichi, A. *294*
Zimmerman, W. 252, *293*
Zinn-Justin, J. 42, *62*, 80, *110, 160, 161, 294, 354*
Zumino, B. *61*, 234, *250, 251*, 396, 397, 408, *417, 418, 419, 420*
Zweig, G. *161*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:49, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Subject Index

a_1 meson, 271
accidental symmetries, 155
  in quantum chromodynamics, 155
  in standard model, 317–18, 326
adjoint representation, defined, 3
Adler–Bell–Jackiw anomaly, *see* anomalies in symmetries
Adler zero, 175
Adler–Bardeen theorem, 380–1
Adler–Weisberger sum rule, 191, 210–11, 248
affine connection, 6–7
Altarelli–Parisi equations, 280–2
anomalies in symmetries, 42, 359–416, 454, 459
anomalous dimensions, 133
anomaly constant $D_{\alpha\beta\gamma}$, defined, 373
antibrackets, 45–7, 82; *also see* Batalin–Vilkovisky formalism, anticanonical transformations
anti-BRST transformations, 41, 62
anticanonical transformations, 47, 93
antifields, 42–3, 92, 405
antighosts, *see* ghost and antighost fields
antighost translation invariance, 88
asymptotic freedom, 133–6, 153
Atiyah–Singer index theorem, 370. 464
axial coupling constant g_A, 187, 203–4
axial gauge, 5, 15–18, 35, 41
axions, 458–61

b quark, 154, 313–15
background field gauge, 95–100
Bardeen formula, 379–80, 398, 414
Bardeen–Cooper–Schrieffer theory, 342–52
barrier penetration, *see* tunneling
baryon number, 239, 436, 443, 454–5, 464
baryons, 225–6; *also see* nucleons
base point, defined, 430
Batalin–Vilkovisky formalism, 42–50, 91–5, 404–7
beta-function
  defined, 120–1
  dependence on coupling definition, 138–9, 141
  for electrodynamics, 126–7, 150
  for minimum subtraction, 150
  for multiple couplings, 140–2
  for quantum chromodynamics, 152–3
  for scalar field theory, 121, 129
  also see renormalization group
Bianchi identity (for gauge fields), 13, 58
Bjorken scaling, 273–5, 280, 283
Bogomol'nyi inequalities, 425, 440, 451
Borel transform, defined, 283
Bott periodicity theorems, 473
bounce, 421–2, 465–468
bound states, 238–9, 389–96
BPS monopoles, 441
BRST quantization, 35–6, 41, 91
BRST symmetry, 27–41, 46, 81, 398–404
bubble formation, 467–8

c quark, 152, 312–15

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

Cabibbo angle, 185, 312–13
Callan–Gross relation, 275, 280
Callan–Symanzik equation, 121; *also see* renormalization group
canonical transformations (in Batalin–Vilkovisky formalism), see anticanonical transformations
Cartan decomposition, 215
Cartan–Maurer integral invariant, 445–450
charge conjugation invariance, 155, 240
Chern–Pontryagin density, 365, 451
Chern–Simons form, 366, 402
chiral symmetry, 182–5, 191; *also see* $SU(2) \times SU(2)$, $SU(3) \times SU(3)$
cluster decomposition principle, 167, 456
cohomology,
  antibracket, 94–5, 405–7
  anti-BRST, 61
  BRST, 32–3, 41, 399
  de Rham, 238
collective parameters, 462–4, 467
color, defined, 152
compact Lie algebras, defined, 9
compact Lie groups, 10
consistency conditions, *see* Wess–Zumino consistency conditions
constraints, 15
convexity of effective potential, 74
correlation length, 337–8
cosets, defined, 214
cosmic strings, 429
cosmology, 426, 443, 455, 461, 464
Coulomb gauge, 15
coupling constants
  electroweak, 305–10, 325–6, 329–30, 464
  for non-Abelian gauge fields, 11–12, 100, 108–9, 297
  for scalar fields, 120
covariant derivatives
  in gauge theories, 4
  in general relativity, 6
  in spontaneous symmetry breaking, 195, 219, 321
covering groups, 58, 443–4
CP non-conservation, 7–8, 315, 457–61
critical exponents, phenomena, 145–8
currents,
  axial-vector, 185–90, 199
  in gauge theories, 12–13
  in superconductors, 323, 336
  of anomalous symmetries, 365–7, 377–8, 380
  of broken symmetries, 169–70
  of semileptonic weak interactions, 311–13
  used to define effective action, 63–5
  vector (of isospin) 185
curvature, see Riemann–Christoffel curvature tensor
custodial symmetry, 326
cut-off, 108, 112, 364–5

d quarks, *see u* and *d* quarks
Debye frequency, 351
deep inelastic scattering, 153, 252, 272–83
Δ operator, defined, 49
Δ_+-function, defined, 170
Derrick's theorem, 424, 428–9, 466
descent equations, 402–4
De Witt notation, 37
De Witt–Faddeev–Popov method, *see* Faddeev–Popov–De Witt method
differential forms 36, 38–9, 400
dimensional regularization, 108–9, 122, 148–9; *also see* minimum subtraction
Dirac quantization conditions, 441–2, 445
direct sums, defined, 9, 52
domain boundaries, 421, 425–6

effective action, *see* quantum effective action
effective field theory, 145
  general broken global symmetry, 211–25, 425
  pions and nucleons, 192–211
  superconductivity, 342–52
  also see Wess–Zumino–Witten terms

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

effective potential, 68–74, 127–30, 168, 347–50
eightfold way, *see* $SU(3) \times SU(3)$
electric dipole moment of neutron, 458
electrodynamics, 1, 33–4, 108, 118–19, 125–7, 150, 157–8
electron, 305–11, 385
  $e^- - e^+$ annihilation, 127, 154, 272, 273, 287, 316
  also see deep-inelastic scattering, electrodynamics
electroweak mixing angle θ, 307–8, 310–11, 330
electroweak theory, 305–18, 325–7, 454, 464
electroweak–strong unification, 327–32
energy–momentum tensor, 13–14, 283, 391
η meson, 225–6, 229–31, 244–5, 361
Euclidean path integrals, 240, 362, 368–70, 451, 468–472
exceptional Lie groups and algebras, 11, 57, 384
extended technicolor, 327
exterior derivative, defined, 36, 400
external fields K_n, 80–1

F parameter for broken symmetry, defined, 172
Faddeev–Popov–De Witt method, 2, 19–24, 43
Fermi coupling constant, 185, 316
Fermi surface, 343
ferromagnetism, 147
Feynman gauge, 23
field strength tensor $F_{\mu\nu}$, defined, 5
fixed point, 133, 145–8
flavor, defined, 152
flux quantization, 335, 338, 340, 342, 429
free energy, 147
Fujikawa–Lee–Sanda gauge, 301–4
fundamental group, 430

gap function, 348–9
Gasser–Leutwyler parameters, 232–3
gauge fixing functional, 19, 43, 48, 301
gauge invariance, 2–7, 319, 332
Gell-Mann–Low function, *see* beta-function
general relativity, *see* gravitation
generalized Feynman (ξ) gauge, 23–4
generators (of symmetries), 14
Georgi–Glashow electroweak model, 305, 436, 442, 444
ghost and antighost fields, 24–7, 38, 302–4
ghost number, 25, 43, 88
ghosts of ghosts, 39
GIM mechanism, 312
Ginzburg–Landau theory, 337, 341
gluons, 154–6, 278
Goldberger–Treiman relation, 187, 190–1, 204, 360
Goldstone bosons, 167–91, 215–25, 295–6, 408–16; *also see* mesons, pions, skyrmions
grand unification, *see* electroweak–strong unification
gravitation, 1, 6–8, 13–14, 36, 43, 91, 386
Gribov ambiguity, 15

Haar measure, 22
hadrons, *see* mesons, baryons
Higgs boson, 316
Higgs mechanism, defined, 295
Hodge operator, 40, 62
homotopic equivalence, defined, 430–1
homotopy groups, 235, 423–5, 427–36, 444–7, 472–3
Hubbard–Stratonovich transformation, 345
hypercharge y, defined, 306

infrared divergences, 113–14, 154
infrared-safe, defined, 114
instantons, 286, 421, 426–7, 436, 450–64
irreducible, defined, 51–2
irrelevant interactions, 145, 148, 344
isotopic spin, 3, 163, 182–3, 243

J/ψ particle, 313
Jacobi identity, 3, 39, 45, 46

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

jets, 114, 154–5
Josephson junction, 335–6

K mesons, 228–31, 236, 312–13
Kobayashi–Maskawa matrix, 313
Kronecker index, 438
KSFR formula, 271

λ matrices, defined, 226
Λ parameter in quantum chromodynamics, 155–7
Landau (or Lorentz) gauge, 23, 301
Landau ghost, 132, 136
lattice approximation, 22, 37–8, 112, 136–9, 238, 239
Legendre transformation, 65, 81
lepton number, 306, 317–18, 464
Lie algebras, *see* compact Lie algebras, semisimple Lie algebras, simple Lie algebras, $U(1)$ Lie algebras
Lie groups, see compact Lie groups, exceptional Lie groups, orthogonal groups, symplectic groups, unitary groups

magnetic monopoles, 225, 421, 429, 435–45
Mandelstam variables, defined, 114
marginal interactions, 145
master equation, 43, 45, 82; *also see* quantum master equation
measure
  for integration over fermion fields, 361–70, 454
  for integration over gauge fields, 18
Meissner effect, 334
mesons, 225–34, 236, 243–6, 271; *also see* pions
metric
  for gauge groups, 7–12, 50–3
  for Goldstone boson fields, 423
  for Lie groups, 448
minimal variables, 44
minimum subtraction, 122, 146–51, 157
modified minimum subtraction, *see* minimum subtraction
Mott cross section, defined, 273
muon, 309–11, 389

Nakanishi–Lautrup field, 28
Nambu–Goldstone bosons, *see* Goldstone bosons
neutral currents, 311–13, 315, 327
neutrinos, 305–11, 385
  ν–e and $\bar{\nu}$–e reactions, 272, 311
  masses and oscillations, 317–18
neutron, *see* electric dipole moment of neutron, nucleon
non-Abelian gauge theories, 1–58, 86–91, 152–7, 426–30, 436; *also see* electroweak theory, quantum chromodynamics, electroweak–strong unification
non-linear realizations, 195, 215–24
nucleon
  as skyrmion, 436
  axial form factors, 186–7, 190
  interaction with pions, 202–9
  masses 209, 233–4

ω meson, 232
one-particle-irreducible graphs, 65–7
open gauge algebras, 42, 44, 48–50
operator product expansions, 252–83, 287–92
order parameters, 224–5, 336–7, 429
orthogonal groups, 11, 55–8, 384, 473
^{14}O decay, 185, 312

p-form fields, *see* differential forms
parity conservation, 155, 240, 315, 416, 457–61
partial conservation of axial current (PCAC), 191
partition function, 469, 471
parton model, 274–5
Peccei–Quinn symmetry, 458–61
penetration depth, 334, 337, 352
persistent currents, 335–6
persistent mass condition, 238–9, 395–6
phase transitions, 148
ϕ meson, 236

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

pions
as Goldstone bosons, 182–211
coupling to nucleon G_π, 187
$\pi^\pm$ decay, 185–6, 189
π^0 decay, 359–61, 368
mass, 189, 201
scattering on nucleons, 206–9
scattering on pions, 198–202, 249
tadpoles, 458
power counting, 197–8, 200, 205–6, 203–4, 323, 344
proton decay, 318,
pseudo-Goldstone bosons, 177–82, 324
pseudo-real representations, defined, 384
pure gauge fields, 6, 334, 412, 427, 428, 449, 455

quantum chromodynamics, 152–7, 181, 192, 225, 238, 239, 287–8, 421, 453; *also see* quarks, gluons, asymptotic freedom, deep inelastic scattering
quantum effective action, 63–8, 75–7, 164, 167, 346–7, 352
quantum electrodynamics, *see* electrodynamics
quantum master equation, 49, 92, 407
quarks, 152–7, 194, 225–7, 275, 312–15
masses, 231–2, 243, 457–61
also see anomalies

real and pseudoreal representations, 384
reducible gauge algebras, 38–9
relevant interactions, 145, 148
renormalization
at sliding scale, 111–12, 119–30
in general theories, 91–5, 141–2
in 'renormalizable' gauge theories, 82–91
of electron–electron potential, 349
of fields, 99–100, 118–19
of gauge coupling constants, 100, 108–9, 119
of general operators, 115–18, 123, 255, 260, 291
of masses, 99, 143–4
of pion self-interaction coupling constants, 199
of scalar coupling constants, 114–15
renormalization group, 111–58, 263–5, 329–32, 349–50, 453; *also see* anomalous dimensions, asymptotic freedom, asymptotic safety, beta function, critical phenomena, fixed point, Landau ghost
renormalons, 283–8
ρ mesons, 232, 271
Riemann–Christoffel curvature tensor, 6, 386
Riemann–Lebesgue theorem, 166
right cosets, defined, 214
running coupling constant, *see* coupling constants, beta-function

s quark, 152, 226, 244, 312–15
Schwinger terms, 403–4
second class currents, 186
σ model, 193–6
σ term, 209
simple Lie algebras, defined, 9
skyrmions, 421, 423–6
Slavnov operator, defined, 39
Slavnov–Taylor identities, 76–7, 81, 407
$SO(10)$ (or $Spin(10)$) symmetry, 328, 386
spectral function sum rules, 266–72
sphalerons, 476
$Spin(n)$ group, 436, 444
spontaneous symmetry breaking 63
approximate symmetries, 177–82
dynamical symmetry breaking, 318–27
global symmetries, 163–246, 265–6
in superconductivity, 332
local symmetries, 2, 295–352
standard model, 384–88; *also see* electroweak theory, quantum chromodynamics
standard $SU(2)$ subgroup, 449
Stora–Zumino descent equations, *see* descent equations
strangeness, 155, 239
string theories, 36, 37, 38, 39, 41

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

structure constants, 2–3, 8–11, 39, 50–3, 89
structure functions, 272–80
subalgebras, defined, 9
superconductivity, 225, 332–52
supergravity, 42
supersymmetry, 331
surface tension, 426, 468
$SU(2)\times U(1)$ symmetry, *see* electroweak theory
$SU(2) \times SU(2)$ symmetry, 182–5, 191, 271, 360–1
$SU(3)$ and $SU(3) \times SU(3)$ symmetry, 225–38, 243, 273
$SU(4) \times SU(4)$ symmetry, 328
$SU(5)$ symmetry, 328
symmetric space, defined, 215
symplectic groups, 9, 56–8, 384, 473

t quark, 152, 154, 313–16
τ lepton, 272, 313
technicolor, 239, 326–7
temperature effects, 146–7, 455, 469, 471
temporal gauge, 427–8, 437, 454
theta angle, 455–61
thin wall approximation, 468
threshold corrections, 358
time reversal invariance, 240, 315
topology, 422–30; *also see* homotopy, cohomology
totally reducible, defined, 52
trapping, 2, 153–5
tree graphs, defined, 66
trivial pairs, 44
triviality, 137–8
tunneling, 165, 455, 464, 468
twist, 277

u and d quarks, 152, 182, 226, 244, 275, 312–15, 367, 385
$U(1)$ Lie algebras, defined, 9
$U(1)$ problem, 243–6, 454
unbroken symmetries, 238–43
unitarity gauge, 295–301, 324, 352–3, 473
unitary groups and algebras, 11, 54–5, 58, 384, 433, 473
universality classes, 148

vacuum alignment, 179, 181, 188–9
vacuum decay, 129–30, 421–2, 464–8
vacuum degeneracy, 163–7
vortex lines, 225, 338–42, 421, 429

$W^{\pm}$ particles, 307–11, 316
weak interactions, 185
Wess–Zumino consistency conditions, 396–407, 409, 411
Wess–Zumino–Witten terms, 234–8, 413–15
Wilson–Fisher expansion, 147–8
winding number, 435–6, 449, 452–3, 455–7

Yang–Mills theory, *see* non-Abelian gauge theories
Yukawa interactions, 308–10

Z and Z_n groups, defined, 433, 472
Z^0 particle, 127, 157, 307–11, 316
zero mass singularities, *see* infrared divergences
zero modes, 426. 462, 464, 467, 476
Zinn-Justin equation, 42, 80–2, 93, 405–6

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:26:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644174

In this third volume of *The Quantum Theory of Fields*, Nobel Laureate Steven Weinberg continues his masterly exposition of quantum field theory. This volume presents a self-contained, up-to-date, and comprehensive introduction to supersymmetry, a highly active area of theoretical physics that is likely to be at the center of future progress in the physics of elementary particles and gravitation. The text introduces and explains a broad range of topics, including supersymmetric algebras, supersymmetric field theories, extended supersymmetry, supergraphs, non-perturbative results, theories of supersymmetry in higher dimensions, and supergravity. A thorough review is given of the phenomenological implications of supersymmetry, including theories of both gauge- and gravitationally mediated supersymmetry breaking. Also provided is an introduction to mathematical techniques, based on holomorphy and duality, that have proved so fruitful in recent developments. This book contains much material not found in other books on supersymmetry, including previously unpublished results. Problems are included.

From reviews of Volume I

'...an impressively lucid and thorough presentation of the subject... Weinberg manages to present difficult topics with richness of meaning and marvellous clarity. Full of valuable insights, his treatise is sure to become a classic, doing for quantum field theory what Dirac's *Quantum Mechanics* did for quantum mechanics. I eagerly await the publication of the second volume.'

S. S. Schweber, *Nature*

'For over twenty years there has been no good modern textbook on the subject. For all that time, Steven Weinberg has been promising to write one. That he has finally done it is cause for celebration among those who try to teach and try to learn the subject. Weinberg's book is for serious students of field theory... it is the first textbook to treat quantum field theory the way it is used by physicists today.'

Howard Georgi, *Science*

'Steven Weinberg, who contributed to the development of quantum chromodynamics and shared the Nobel Prize in Physics for his contributions to the electroweak theory, has written a definitive text on the physical foundations of quantum field theory. His book differs significantly from the long line of previous books on quantum field theory... To summarize, *Foundations* builds the structure of quantum field theory on the sure footing of physical insight. It is beautifully produced and meticulously edited... and it is a real bargain in price. If you want to learn quantum field theory, or have already learned it and want to have a definitive reference at hand, purchase this book.'

O. W. Greenberg, *Physics Today*

'In addition to a superb treatment of all the conventional topics there are numerous sections covering areas that are not normally emphasized, such as the subject of field redefinitions, higher-rank tensor fields and an unusually clear and thorough treatment of infrared effects... this latest book reinforces his high scholarly standards. It provides a unique exposition that will prove invaluable both to new research students as well as to experienced research workers. Together

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

with Volume II, this will become a classic text on a subject of central importance to a wide area of theoretical physics.'

M. B. Green, *CERN Courier*

'I believe that what readers will find particularly helpful in this volume is the consistency of the whole approach, and the emphasis on quantities and properties that are directly useful to particle physicists. This is particularly true for those who are interested in the more phenomenological aspects. The reader only needs limited background knowledge, and a clear line is followed throughout the book, making it easy to follow. The author presents extremely thorough but elementary discusssions of important physical questions, some of which seem to be an original way of addressing the subject.'

J. Zinn-Justin, *Physics World*

'This is a well-written book by one of the masters of the subject... it is certainly destined to become a standard text book and should find its way to the shelves of every physics library.'

J. Madore, *Classical and Quantum Gravity*

'The book starts out with an excellent historical introduction, not found anywhere else, giving citations to many by now classic papers... a valuable reference work as well as a textbook for graduate students.'

G. Roepstorff, *Zentralblatt für Mathematik*

From reviews of Volume II

'It is a majestic exposition. The two volumes are structured in a logical way. Everything is explained with incisive clarity. Weinberg always goes to the heart of any argument, and includes many things that cannot be found elsewhere in the literature. Often I find myself thinking: "Ah! Now I understand that properly."... I find it hard to imagine a better treatment of quantum field theory than Weinberg's. All serious students and researchers will want to have these volumes on their shelves.'

John C. Taylor, *Nature*

'Weinberg's *Modern Applications* goes to the boundaries of our present understanding of field theory. It is unmatched by any other book on quantum field theory for its depth, generality and definitive character, and it will be an essential reference for serious students and researchers in elementary particle physics.'

O. W. Greenberg, *Physics Today*

'... Steven Weinberg is one of our most gifted makers of theoretical tools as well as a virtuoso in their use. His new book conveys both the satisfaction of understanding nature and the feel of the atelier, for the "modern applications" of its subtitle include both the derivation of physical consequences and the development of new tools for understanding and applying field theory itself... *Modern Applications* is a splendid book, with abundant useful references to the original literature. It is a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

very interesting read from cover to cover, for the wholeness Weinberg's personal perspective gives to quantum field theory and particle physics.'

Chris Quigg, *Science*

'Experienced researchers and beginning graduate students will delight in the gems of wisdom to be found in these pages. This book combines exposition of technical detail with physical insight in a unique manner that confirms the promise of Volume I and I have no doubt that these two volumes will rapidly constitute the classic treatment of this important subject.'

M. B. Green, *CERN Courier*

'...a valued reference and a mine of useful information for professional field theorists.'

Tom Kibble, *New Scientist*

'...a clear presentation of the subject, explaining the underlying concepts in much depth and in an accessible style. I expect that these volumes will become the first source we turn to when trying to answer the challenging questions asked by bright postgraduates when they first encounter quantum field theory... I have no doubt that *The Quantum Theory of Fields* will soon be found on the bookshelves of most particle theorists, and that it will be one of the main sources used in the preparation of lectures on the subject for postgraduate students.'

C. T. C. Sachrajda, *The Times Higher Education Supplement*

'...Weinberg has produced a masterpiece that will be a standard reference on the field for a long time to come.'

B. E. Y. Svensson, *Elementa*

From reviews of Volume III

'...has produced a treatise that many of us had long awaited, perhaps without fully realizing it... with the publication of *The Quantum Theory of Fields*, Vol. III, has performed an analogous service for supersymmetry... Although this volume is the third in a trilogy, it is quite different from its two predecessors, and it stands on its own... May a new generation of students imbibe its content and spirit.'

Physics Today

'The third volume of *The Quantum Theory of Fields* is a self-contained introduction to the world of supersymmetry and supergravity. It will be useful both for experienced researchers in the field and for students who want to take the first steps towards learning about supersymmetry. Unlike other books in this field, it covers the wide spectrum of possible applications of supersymmetry in physics.'

Hans Peter Nilles, *Nature*

'Weinberg is of course one of the creators of modern quantum field theory, as well as of its physical culmination, the standard model of all (nongravitational) interactions. It is... very timely that this latest part of his monograph, devoted to supersymmetry and supergravity, has just appeared. As a text, it has been

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

pretested by Weinberg for a freestanding one-year graduate course; as a clear organizing reference to this extremely vast field, it will help the experts as well... Weinberg's style of presentation is as clear and meticulous as in his previous works.'

Stanley Deser, *Journal of General Relativity and Gravitation*

'Weinberg tries to be as elementary and clear as possible and steers clear of more sophisticated mathematical tools. Together with the previous volumes, this volume will serve as an invaluable reference to researchers and a textbook for graduate students.'

G. Roepstorff, *Zentralblatt MATH*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

The Quantum Theory of Fields

Volume III
Supersymmetry

Steven Weinberg
University of Texas at Austin

CAMBRIDGE
UNIVERSITY PRESS

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo,
Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge, CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521670555

© S.Weinberg 2000, 2005

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000
Paperback edition published 2005
8th printing 2012

Printed and bound by MPG Books Group, UK

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-67055-5 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables and other factual information given in this work are correct at the time of first printing but Cambridge University Press does not guarantee the accuracy of such information thereafter.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Contents

Sections marked with an asterisk are somewhat out of the book's main line of development and may be omitted in a first reading.

PREFACE TO VOLUME III xvi

NOTATION xx

24 HISTORICAL INTRODUCTION 1

24.1 Unconventional Symmetries and 'No-Go' Theorems 1

$SU(6)$ symmetry □ Elementary no-go theorem for unconventional semi-simple compact Lie algebras □ Role of relativity

24.2 The Birth of Supersymmetry 4

Bosonic string theory □ Fermionic coordinates □ Worldsheet supersymmetry □ Wess–Zumino model □ Precursors

Appendix A $SU(6)$ **Symmetry of Non-Relativistic Quark Models** 8

Appendix B The Coleman–Mandula Theorem 12

Problems 22

References 22

25 SUPERSYMMETRY ALGEBRAS 25

25.1 Graded Lie Algebras and Graded Parameters 25

Fermionic and bosonic generators □ Super-Jacobi identity □ Grassmann parameters □ Structure constants from supergroup multiplication rules □ Complex conjugates

25.2 Supersymmetry Algebras 29

Haag–Lopuszanski–Sohnius theorem □ Lorentz transformation of fermionic generators □ Central charges □ Other bosonic symmetries □ R-symmetry □ Simple

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

and extended supersymmetry □ Four-component notation □ Superconformal algebra

25.3 Space Inversion Properties of Supersymmetry Generators 40

Parity phases in simple supersymmetry □ Fermions have imaginary parity □ Parity matrices in extended supersymmetry □ Dirac notation

25.4 Massless Particle Supermultiplets 43

Known particles are massless for unbroken supersymmetry □ Helicity raising and lowering operators □ Simple supersymmetry doublets □ Squarks, sleptons, and gauginos □ Gravitino □ Extended supersymmetry multiplets □ Chirality problem for extended supersymmetry

25.5 Massive Particle Supermultiplets 48

Raising and lowering operators for spin 3-component □ General massive multiplets for simple supersymmetry □ Collapsed supermultiplet □ Mass bounds in extended supersymmetry □ BPS states and short supermultiplets

Problems 53

References 54

26 SUPERSYMMETRIC FIELD THEORIES 55

26.1 Direct Construction of Field Supermultiplets 55

Construction of simplest $N = 1$ field multiplet □ Auxiliary field □ Infinitesimal supersymmetry transformation rules □ Four-component notation □ Wess–Zumino supermultiplets regained

26.2 General Superfields 59

Superspace spinor coordinates □ Supersymmetry generators as superspace differential operators □ Supersymmetry transformations in superspace □ General superfields □ Multiplication rules □ Supersymmetric differential operators in superspace □ Supersymmetric actions for general superfields □ Parity of component fields □ Counting fermionic and bosonic components

26.3 Chiral and Linear Superfields 68

Chirality conditions on a general superfield □ Left- and right-chiral superfields □ Coordinates $x_{\pm}^{\mu}$ □ Differential constraints □ Product rules □ Supersymmetric $\mathscr{F}$-terms □ $\mathscr{F}$-terms equivalent to D-terms □ Superpotentials □ Kahler potentials □ Partial integration in superspace □ Space inversion of chiral superfields □ R-symmetry again □ Linear superfields

26.4 Renormalizable Theories of Chiral Superfields 75

Counting powers □ Kinematic Lagrangian □ $\mathscr{F}$-term of the superpotential □ Complete Lagrangian □ Elimination of auxiliary fields □ On-shell superalgebra □ Vacuum solutions □ Masses and couplings □ Wess–Zumino Lagrangian regained

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

26.5 Spontaneous Supersymmetry Breaking in the Tree Approximation 83

O'Raifeartaigh mechanism □ R-symmetry constraints □ Flat directions □ Goldstino

26.6 Superspace Integrals, Field Equations, and the Current Superfield 86

Berezin integration □ D- and $\mathscr{F}$-terms as superspace integrals □ Potential superfields □ Superspace field equations □ Conserved currents as components of linear superfields □ Conservation conditions in superspace

26.7 The Supercurrent 90

Supersymmetry current □ Superspace transformations generated by the supersymmetry current □ Local supersymmetry transformations □ Construction of the supercurrent □ Conservation of the supercurrent □ Energy-momentum tensor and R-current □ Scale invariance and R conservation □ Non-uniqueness of supercurrent

26.8 General Kahler Potentials* 102

Non-renormalizable non-derivative actions □ D-term of Kahler potential □ Kahler metric □ Lagrangian density □ Non-linear σ-models from spontaneous internal symmetry breaking □ Kahler manifolds □ Complexified coset spaces

Appendix Majorana Spinors 107

Problems 111

References 112

27 SUPERSYMMETRIC GAUGE THEORIES 113

27.1 Gauge-Invariant Actions for Chiral Superfields 113

Gauge transformation of chiral superfields □ Gauge superfield V □ Extended gauge invariance □ Wess–Zumino gauge □ Supersymmetric gauge-invariant kinematic terms for chiral superfields

27.2 Gauge-Invariant Action for Abelian Gauge Superfields 122

Field strength supermultiplet □ Kinematic Lagrangian density for Abelian gauge supermultiplet □ Fayet–Iliopoulos terms □ Abelian field-strength spinor superfield W_α □ Left- and right-chiral parts of W_α □ W_α as a superspace derivative of V □ Gauge invariance of W_α □ 'Bianchi' identities in superspace

27.3 Gauge-Invariant Action for General Gauge Superfields 127

Kinematic Lagrangian density for non-Abelian gauge supermultiplet □ Non-Abelian field-strength spinor superfield $W_{A\alpha}$ □ Left- and right-chiral parts of $W_{A\alpha}$ □ θ-term □ Complex coupling parameter τ

27.4 Renormalizable Gauge Theories with Chiral Superfields 132

Supersymmetric Lagrangian density □ Elimination of auxiliary fields □ Conditions for unbroken supersymmetry □ Counting independent conditions and field

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

variables □ Unitarity gauge □ Masses for spins 0, 1/2, and 1 □ Supersymmetry current □ Non-Abelian gauge theories with general Kahler potentials □ Gaugino mass

27.5 Supersymmetry Breaking in the Tree Approximation Resumed 144

Supersymmetry breaking in supersymmetric quantum electrodynamics □ General case: masses for spins 0, 1/2, and 1 □ Mass sum rule □ Goldstino component of gaugino and chiral fermion fields

27.6 Perturbative Non-Renormalization Theorems 148

Non-renormalization of Wilsonian superpotential □ One-loop renormalization of terms quadratic in gauge superfields □ Proof using holomorphy and new symmetries with external superfields □ Non-renormalization of Fayet–Iliopoulos constants ξ_A □ For $\xi_A = 0$, supersymmetry breaking depends only on superpotential □ Non-renormalizable theories

27.7 Soft Supersymmetry Breaking* 155

Limitation on supersymmetry-breaking radiative corrections □ Quadratic divergences in tadpole graphs

27.8 Another Approach: Gauge-Invariant Supersymmetry Transformations 157

De Wit–Freedman transformation rules □ Preserving Wess–Zumino gauge with combined supersymmetry and extended gauge transformations

27.9 Gauge Theories with Extended Supersymmetry* 160

$N = 2$ supersymmetry from $N = 1$ supersymmetry and R-symmetry □ Lagrangian for $N = 2$ supersymmetric gauge theory □ Eliminating auxiliary fields □ Supersymmetry currents □ Witten–Olive calculation of central charge □ Non-renormalization of masses □ BPS monopoles □ Adding hypermultiplets □ $N = 4$ supersymmetry □ Calculation of beta function □ $N = 4$ theory is finite □ Montonen–Olive duality

Problems 175

References 176

28 SUPERSYMMETRIC VERSIONS OF THE STANDARD MODEL 179

28.1 Superfields, Anomalies, and Conservation Laws 180

Quark, lepton, and gauge superfields □ At least two scalar doublet superfields □ $\mathscr{F}$-term Yukawa couplings □ Constraints from anomalies □ Unsuppressed violation of baryon and lepton numbers □ R-symmetry □ R parity □ μ-term □ Hierarchy problem □ Sparticle masses □ Cosmological constraints on lightest superparticle

28.2 Supersymmetry and Strong–Electroweak Unification 188

Renormalization group equations for running gauge couplings □ Effect of super-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

symmetry on beta functions □ Calculation of weak mixing angle and unification mass □ Just two scalar doublet superfields □ Coupling at unification scale

28.3 Where is Supersymmetry Broken? 192

Tree approximation supersymmetry breakdown ruled out □ Hierarchy from non-perturbative effects of asymptotically free gauge couplings □ Gauge and gravitational mediation of supersymmetry breaking □ Estimates of supersymmetry-breaking scale □ Gravitino mass □ Cosmological constraints

28.4 The Minimal Supersymmetric Standard Model 198

Supersymmetry breaking by superrenormalizable terms □ General Lagrangian □ Flavor changing processes □ Calculation of $K^0 \leftrightarrow \overline{K}^0$ □ Degenerate squarks and sleptons □ CP violation □ Calculation of quark chromoelectric dipole moment □ 'Naive dimensional analysis' □ Neutron electric dipole moment □ Constraints on masses and/or phases

28.5 The Sector of Zero Baryon and Lepton Number 209

D-term contribution to scalar potential □ μ-term contribution to scalar potential □ Soft supersymmetry breaking terms □ Vacuum stability constraint on parameters □ Finding a minimum of potential □ $B\mu \neq 0$ □ Masses of CP-odd neutral scalars □ Masses of CP-even neutral scalars □ Masses of charged scalars □ Bounds on masses □ Radiative corrections □ Conditions for electroweak symmetry breaking □ Charginos and neutralinos □ Lower bound on $|\mu|$

28.6 Gauge Mediation of Supersymmetry Breaking 220

Messenger superfields □ Supersymmetry breaking in gauge supermultiplet propagators □ Gaugino masses □ Squark and slepton masses □ Derivation from holomorphy □ Radiative corrections □ Numerical examples □ Higgs scalar masses □ μ problem □ A_{ij} and C_{ij} parameters □ Gravitino as lightest sparticle □ Next-to-lightest sparticle

28.7 Baryon and Lepton Non-Conservation 235

Dimensionality five interactions □ Gaugino exchange □ Gluino exchange suppressed □ Wino and bino exchange effects □ Estimate of proton lifetime □ Favored modes of proton decay

Problems 240

References 241

29 BEYOND PERTURBATION THEORY 248

29.1 General Aspects of Supersymmetry Breaking 248

Finite volume □ Vacuum energy and supersymmetry breaking □ Partially broken extended supersymmetry? □ Pairing of bosonic and fermionic states □ Pairing of vacuum and one-goldstino state □ Witten index □ Supersymmetry unbroken

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

in the Wess–Zumino model □ Models with unbroken supersymmetry and zero Witten index □ Large field values □ Weighted Witten indices

29.2 Supersymmetry Current Sum Rules 256

Sum rule for vacuum energy density □ One-goldstino contribution □ The supersymmetry-breaking parameter F □ Soft goldstino amplitudes □ Sum rule for supersymmetry current-fermion spectral functions □ One-goldstino contribution □ Vacuum energy density in terms of $\mathscr{F}$ and D vacuum values □ Vacuum energy sum rule for infinite volume

29.3 Non-Perturbative Corrections to the Superpotential 266

Non-perturbative effects break external field translation and R-conservation □ Remaining symmetry □ Example: generalized supersymmetric quantum chromodynamics □ Structure of induced superpotential for $C_1 > C_2$ □ Stabilizing the vacuum with a bare superpotential □ Vacuum moduli in generalized supersymmetric quantum chromodynamics for $N_c > N_f$ □ Induced superpotential is linear in bare superpotential parameters for $C_1 = C_2$ □ One-loop renormalization of $[W_\alpha W_\alpha]_{\mathscr{F}}$ term for all C_1, C_2

29.4 Supersymmetry Breaking in Gauge Theories 276

Witten index vanishes in supersymmetric quantum electrodynamics □ C-weighted Witten index □ Supersymmetry unbroken in supersymmetric quantum electrodynamics □ Counting zero-energy gauge field states in supersymmetric quantum electrodynamics □ Calculating Witten index for general supersymmetric pure gauge theories □ Counting zero-energy gauge field states for general supersymmetric pure gauge theories □ Weyl invariance □ Supersymmetry unbroken in general supersymmetric pure gauge theories □ Witten index and R anomalies □ Adding chiral scalars □ Model with spontaneously broken supersymmetry

29.5 The Seiberg–Witten Solution* 287

Underlying $N = 2$ supersymmetric Lagrangian □ Vacuum modulus □ Leading non-renormalizable terms in the effective Lagrangian □ Effective Lagrangian for component fields □ Kahler potential and gauge coupling from a function $h(\Phi)$ □ $SU(2)$ R-symmetry □ Prepotential □ Duality transformation □ $h(\Phi)$ translation □ $\mathbb{Z}_8$ R-symmetry □ $SL(2,\mathbb{Z})$-symmetry □ Central charge □ Charge and magnetic monopole moments □ Perturbative behavior for large $|a|$ □ Monodromy at infinity □ Singularities from dyons □ Monodromy at singularities □ Seiberg–Witten solution □ Uniqueness proof

Problems 305

References 305

30 SUPERGRAPHS 307

30.1 Potential Superfields 308

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Problem of chiral constraints □ Corresponding problem in quantum electrodynamics □ Path integrals over potential superfields

30.2 Superpropagators 310

A troublesome invariance □ Change of variables □ Defining property of superpropagator □ Analogy with quantum electrodynamics □ Propagator for potential superfields □ Propagator for chiral superfields

30.3 Calculations with Supergraphs 313

Superspace quantum effective action □ Locality in fermionic coordinates □ D-terms and $\mathscr{F}$-terms in effective action □ Counting superspace derivatives □ No renormalization of $\mathscr{F}$-terms

Problems 316

References 316

31 SUPERGRAVITY 318

31.1 The Metric Superfield 319

Vierbein formalism □ Transformation of gravitational field □ Transformation of gravitino field □ Generalized transformation of metric superfield H_μ □ Interaction of H_μ with supercurrent □ Invariance of interaction □ Generalized transformation of H_μ components □ Auxiliary fields □ Counting components □ Interaction of H_μ component fields □ Normalization of action

31.2 The Gravitational Action 326

Einstein superfield E_μ □ Component fields of E_μ □ Lagrangian for H_μ □ Value of κ □ Total Lagrangian □ Vacuum energy density □ Minimum vacuum energy □ De Sitter and anti-de Sitter spaces □ Why vacuum energy is negative □ Stability of flat space □ Weyl transformation

31.3 The Gravitino 333

Irreducibility conditions on gravitino field □ Gravitino propagator □ Gravitino kinematic Lagrangian □ Gravitino field equation □ Gravitino mass from broken supersymmetry □ Gravitino mass from s and p

31.4 Anomaly-Mediated Supersymmetry Breaking 337

First-order interaction with scale non-invariance superfield X □ General formula for X □ General first-order interaction □ Gaugino masses □ Gluino mass □ B parameter □ Wino and bino masses □ A parameters

31.5 Local Supersymmetry Transformations 341

Wess–Zumino gauge for metric superfield □ Local supersymmetry transformations □ Invariance of action

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

31.6 Supergravity to All Orders 343

Local supersymmetry transformation of vierbein, gravitino, and auxiliary fields □ Extended spin connection □ Local supersymmetry transformation of general scalar supermultiplet □ Product rules for general superfields □ Real matter superfields □ Chiral matter superfields □ Product rules for chiral superfields □ Cosmological constant and gravitino mass □ Lagrangian for supergravity and chiral fields with general Kahler potential and superpotential □ Elimination of auxiliary fields □ Kahler metric □ Weyl transformation □ Scalar field potential □ Conditions for flat space and unbroken supersymmetry □ Complete bosonic Lagrangian □ Canonical normalization □ Combining superpotential and Kahler potential □ No-scale models

31.7 Gravity-Mediated Supersymmetry Breaking 355

Early theories with hidden sectors □ Hidden sector gauge coupling strong at energy Λ □ First version: Observable and hidden sectors □ Separable bare superpotential □ General potential □ Terms of order $\kappa^4\Lambda^8 \approx m_g^4$ □ Λ estimated as $\approx 10^{11}$ GeV □ μ- and $B\mu$-terms □ Squark and slepton masses □ Gaugino masses □ A-parameters □ Second version: Observable, hidden, and modular sectors □ Dynamically induced superpotential for modular superfields □ Effective superpotential of observable sector □ μ-term □ Potential of observable sector scalars □ Terms of order $\kappa^8\Lambda^{12} \approx m_g^4$ □ Soft supersymmetry-breaking terms □ Λ estimated as $\approx 10^{13}$ GeV □ Shifts in modular fields □ Absence of C_{ij} terms □ Squark and slepton masses □ Gaugino masses

Appendix The Vierbein Formalism 375

Problems 378

References 379

32 SUPERSYMMETRY ALGEBRAS IN HIGHER DIMENSIONS 382

32.1 General Supersymmetry Algebras 382

Classification of fermionic generators □ Definition of weight □ Fermionic generators in fundamental spinor representation □ Fermionic generators commute with P_μ □ General form of anticommutation relations □ Central charges □ Anticommutation relations for odd dimensionality □ Anticommutation relations for even dimensionality □ R-symmetry groups

32.2 Massless Multiplets 393

Little group $O(d-2)$ □ Definition of 'spin' j □ Exclusion of $j > 2$ □ Missing fermionic generators □ Number of fermionic generators ≤ 32 □ $N = 1$ supersymmetry for $d = 11$ □ Three-form massless particle □ Types IIA, IIB and heterotic supersymmetry for $d = 10$

32.3 p-Branes 397

New conserved tensors □ Fermionic generators still in fundamental spinor repre-

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

sentation □ Fermionic generators still commute with P_μ □ Symmetry conditions on tensor central charges □ 2-form and 5-form central charges for $d = 11$

Appendix Spinors in Higher Dimensions 401

Problems 407

References 407

AUTHOR INDEX 411

SUBJECT INDEX 416

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Preface To Volume III

This volume deals with quantum field theories that are governed by supersymmetry, a symmetry that unites particles of integer and half-integer spin in common symmetry multiplets. These theories offer a possible way of solving the 'hierarchy problem,' the mystery of the enormous ratio of the Planck mass to the 300 GeV energy scale of electroweak symmetry breaking. Supersymmetry also has the quality of uniqueness that we search for in fundamental physical theories. There is an infinite number of Lie groups that can be used to combine particles of the same spin in ordinary symmetry multiplets, but there are only eight kinds of supersymmetry in four spacetime dimensions, of which only one, the simplest, could be directly relevant to observed particles.

These are reasons enough to devote this third volume of *The Quantum Theory of Fields* to supersymmetry. In addition, the quantum field theories based on supersymmetry have remarkable properties that are not found among other field theories: some supersymmetric theories have couplings that are not renormalized in any order of perturbation theory; other theories are finite; and some even allow exact solutions. Indeed, much of the most interesting work in quantum field theory over the past decade has been in the context of supersymmetry.

Unfortunately, after a quarter century there is no direct evidence for supersymmetry, as no pair of particles related by a supersymmetry transformation has yet been discovered. There is just one significant piece of indirect evidence for supersymmetry: the high-energy unification of the $SU(3)$, $SU(2)$, and $U(1)$ gauge couplings works better with the extra particles called for by supersymmetry than without them.

Nevertheless, because of the intrinsic attractiveness of supersymmetry and the possibility it offers of resolving the hierarchy problem, I and many other physicists are reasonably confident that supersymmetry will be found to be relevant to the real world, and perhaps soon. Supersymmetry is a primary target of experiments at high energy planned at existing accelerators, and at the Large Hadron Collider under construction at the CERN laboratory.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.001

After a historical introduction in Chapter 24, Chapters 25–27 present the essential machinery of supersymmetric field theories: the structure of the supersymmetry algebra and supersymmetry multiplets and the construction of supersymmetric Lagrangians in general, and in particular for theories of chiral and gauge superfields. Chapter 28 then uses this machinery to incorporate supersymmetry in the standard model of electroweak and strong interactions, and reviews experimental difficulties and opportunities. Chapters 29–32 deal with topics that are mathematically more advanced: non-perturbative results, supergraphs, supergravity, and supersymmetry in higher dimensions.

I have made the treatment of supersymmetry here as clear and self-contained as I could. Wherever possible I take the reader through calculations, rather than just reporting results from the literature. Where calculations are too lengthy or complicated to be included in a book of this sort, especially in Chapter 28, I have tried to present simpler versions that give the reader an idea of the physical issues involved.

I have made a point of including topics here that have generally not been covered in earlier books, some because they are too new. These include: the use of holomorphy to study perturbative and non-perturbative radiative corrections; the calculation of central charges; gauge-mediated and anomaly-mediated supersymmetry breaking; the Witten index; duality; the Seiberg–Witten calculation of the effective Lagrangian in $N = 2$ supersymmetric gauge theories; supersymmetry breaking by modular fields; and a first look at the rapidly developing topic of supersymmetry in higher dimensions, including theories with p-branes.

On the other hand, I have shortened the usual treatment of two topics that seemed to me to have been well covered in earlier books. One of these is the use of supergraphs. Many of the previous applications of the supergraph formalism in studying the general structure of radiative corrections can now be handled more easily by using the arguments of holomorphy described in Sections 27.6 and 29.3. The other is supergravity. In Sections 31.1–31.5 I have given a detailed and self-contained treatment of supergravity in the weak-field limit, which makes it clear why the ingredients of supergravity theories — the graviton, gravitino, and auxiliary fields — are what they are, and which allows us to derive some of the most important results of supergravity theory, including the formula for the gravitino mass and for the gaugino masses produced by anomaly-mediated supersymmetry breaking. In Section 31.6 I have outlined the calculations that generalize supergravity theory to gravitational fields of arbitrary strength, but these calculations are so lengthy and unlovely that I was content to quote other sources for the results. However, in Section 31.7 I have given a fuller than usual treatment of gravitationally mediated supersymmetry breaking. I regret that I have not been able to include exciting work of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.001

past decade on supersymmetry related to string theory, but string theory is beyond the scope of this book, and I did not want to report results for which I had not provided a basis of explanation.

I have given citations both to the classic papers on supersymmetry and to useful references on topics that are mentioned but not presented in detail in this book. I did not always know who was responsible for material presented here, and the mere absence of a citation should not be taken as a claim that the material presented here is original, but some of it is. I hope that I have improved on the original literature or standard textbook treatments in several places, as for instance in the proof of the Coleman–Mandula theorem; in the treatment of parity matrices in extended supersymmetry theories; in the inclusion of new soft supersymmetry-breaking terms in the minimum supersymmetric standard model; in the derivation of supercurrent sum rules; and in the proof of the uniqueness of the Seiberg–Witten solution.

I have also supplied problems for each chapter. Some of these problems aim simply at providing exercise in the use of techniques described in the chapter; others are intended to suggest extensions of the results of the chapter to a wider class of theories.

In teaching a course on supersymmetry, I have found that this book provides enough material for a one-year course for graduate students. I intended that this book should be accessible to students who are familiar with quantum field theory at the level it is presented in the first two volumes of this treatise. It is not assumed that the reader has gone through Volumes I and II, but for the convenience of those fortunate readers who have done so I use the same notation here, and give cross-references to material in Volumes I and II wherever appropriate.

* * *

I must acknowledge my special intellectual debt to colleagues at the University of Texas, notably Luis Boya, Phil Candelas, Bryce and Cecile De Witt, Willy Fischler, Daniel Freed, Joaquim Gomis, Vadim Kaplunovsky, and especially Jacques Distler. Also, Sally Dawson, Michael Dine, Michael Duff, Lawrence Hall, Hitoshi Murayama, Joe Polchinski, Edward Witten, and Bruno Zumino gave valuable help with special topics. Jonathan Evans read through the manuscript of this volume, and made many valuable suggestions. For pointing out various errors in the first printing of this book, I am greatly indebted to Stephen Adler, Jose Espinora, Tony Gherghetta, and San Fu Tuan. For corrections to the first printing of this volume I am indebted to several colleagues, especially Stephen Adler. Thanks are due to Alyce Wilson, who prepared the illustrations, to Terry

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.001

Riley for finding countless books and articles, and to Jan Duffy for many helps. I am grateful to Maureen Storey of Cambridge University Press for working to ready this book for publication, and especially to my editor, Rufus Neal, for his continued friendly good advice.

STEVEN WEINBERG

Austin, Texas
May, 1999

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.001

Notation

The big issue in choosing notation for a book on supersymmetry is whether to use a two-component or a four-component notation for spinors. The standard texts on supersymmetry have opted for the two-component Weyl notation. I have chosen instead to use the four-component Dirac notation except in the first stages of constructing the supersymmetry algebra and multiplets, because I think this will make the book more accessible to those physicists who work on particle phenomenology and model building. It would be a pity to see the growth of a separate enclave of supersymmetry specialists, who communicate well with each other but are cut off by their notation from the larger community of particle theorists.

There is no great difficulty anyway in converting expressions in four-component form into the two-component formalism. In the representation of the Dirac matrices used throughout this book, in which γ_5 is the diagonal matrix with elements $+1$, $+1$, -1, and -1 on the main diagonal, any four-component Majorana spinor ψ_α (such as the supersymmetry generator Q_α, the superspace coordinate θ_α, or the superderivative $\mathscr{D}_\alpha$) may be written in terms of a two-component spinor χ_a as

$$\psi = \begin{pmatrix} e\chi^* \\ \chi \end{pmatrix},$$

where e is the 2×2 antisymmetric matrix with $e_{12} = +1$. The two-component spinor χ_a is what in other books is often called $\psi_a^* = \bar{\psi}_{\dot{a}}$, while $(e\chi^*)_a$ would be called ψ^a. A summary of useful properties of four-component Majorana spinors is given in the appendix to Chapter 26.

Here are some other features of the notation used in this book:

Latin indices i, j, k, and so on generally run over the three spatial coordinate labels, usually taken as 1, 2, 3. Where specifically indicated, they run over values 1, 2, 3, 4, with $x^4 = it$.

Greek indices μ, ν, etc. from the middle of the Greek alphabet generally run over the four spacetime coordinate labels 1, 2, 3, 0, with x^0 the time coordinate. Where it is necessary to distinguish between spacetime

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.002

coordinates in a general coordinate system and in a locally inertial system, indices μ, ν, etc. are used for the former and a, b, etc. for the latter.

Greek indices α, β, etc. from the beginning of the Greek alphabet generally (except in Chapter 24) run over the components of four-component spinors. To avoid confusion, I depart here from the conventions of Volume II, and use upper-case letters A, B, etc. to label the generators of a symmetry algebra. Components of two-component spinors are labelled with indices a, b, etc. In particular, four-component supersymmetry generators are denoted Q_α, while two-component generators (the bottom two components of Q_α) are called Q_a.

Repeated indices are generally summed, unless otherwise indicated.

The spacetime metric $\eta_{\mu\nu}$ is diagonal, with elements $\eta_{11} = \eta_{22} = \eta_{33} = 1$, $\eta_{00} = -1$.

The d'Alembertian is defined as $\Box \equiv \eta^{\mu\nu}\partial^2/\partial x^\mu \partial x^\nu = \nabla^2 - \partial^2/\partial t^2$, where ∇^2 is the Laplacian $\partial^2/\partial x^i \partial x^i$.

The 'Levi–Civita tensor' $\epsilon^{\mu\nu\rho\sigma}$ is defined as the totally antisymmetric quantity with $\epsilon^{0123} = +1$.

Dirac matrices γ_μ are defined so that $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\eta_{\mu\nu}$. Also, $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$, and $\beta = i\gamma^0 = \gamma_4$. Where explicit matrices are needed, they are given by the block matrices

$$\gamma^0 = -i\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{\gamma} = -i\begin{bmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{bmatrix},$$

where $\mathbf{1}$ is the unit 2×2 matrix, $\mathbf{0}$ is the 2×2 matrix with elements zero, and the components of $\boldsymbol{\sigma}$ are the usual Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We also frequently make use of the 4×4 block matrices

$$\gamma_5 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \epsilon = \begin{bmatrix} e & 0 \\ 0 & e \end{bmatrix},$$

where e is again the antisymmetric 2×2 matrix $i\sigma_2$. For instance, our phase convention for four-component Majorana spinors s may be expressed as $s^* = -\beta\,\gamma_5\,\epsilon\,s$.

The step function $\theta(s)$ has the value $+1$ for $s > 0$ and 0 for $s < 0$.

The complex conjugate, transpose, and Hermitian adjoint of a matrix or vector A are denoted A^*, A^{T}, and $A^\dagger = A^{*\mathrm{T}}$, respectively. We use

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.002

an asterisk * for the Hermitian adjoint of an operator or the complex conjugate of a number, except where a dagger † is used for the transpose of the matrix formed from the Hermitian adjoints of operators or complex conjugates of numbers. +H.c. or +c.c. at the end of an expression indicates the addition of the Hermitian adjoint or complex conjugate of the foregoing terms. A bar on a four-component spinor u is defined by $\bar{u} = u^\dagger \beta$.

Units are used with $\hbar$ and the speed of light taken to be unity. Throughout $-e$ is the rationalized charge of the electron, so that the fine structure constant is $\alpha = e^2/4\pi \simeq 1/137$. Temperatures are in energy units, with the Boltzmann constant taken equal to unity.

Numbers in parenthesis at the end of quoted numerical data give the uncertainty in the last digits of the quoted figure. Where not otherwise indicated, experimental data are taken from 'Review of Particle Physics,' The Particle Data Group, *European Physics Journal C* **3**, 1 (1998).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.002

24
Historical Introduction

The history of supersymmetry is as peculiar as anything in the history of science. Suggested in the early 1970s, supersymmetry has been elaborated since then into a beautiful mathematical formalism that unites particles of different spin into symmetry multiplets and has profound implications for fundamental physics. Yet there is so far not a shred of direct experimental evidence and only a few bits of indirect evidence that supersymmetry has anything to do with the real world. If (as I expect) supersymmetry does turn out to be relevant to nature, it will represent a striking success of purely theoretical insight.

Chapter 25 will begin the construction of supersymmetry theories from first principles. In the present chapter we shall introduce supersymmetry along chronological rather than logical lines.

24.1 Unconventional Symmetries and 'No-Go' Theorems

In the early 1960s the symmetry $SU(3)$ of Gell-Mann and Ne'eman (discussed in Section 19.7) successfully explained the relations between various strongly interacting particles of different charge and strangeness but of the same spin. The idea then grew up that perhaps $SU(3)$ is part of a larger symmetry, which has the unconventional effect of uniting $SU(3)$ multiplets of different spin.[1] There is such an approximate symmetry in the non-relativistic quark model, under $SU(6)$ transformations on quark spins and flavors, analogous to an earlier $SU(4)$ symmetry of nuclear physics that had been introduced in 1937 by Wigner.[2] As described in detail in Appendix A of this chapter, this $SU(6)$ symmetry unites the pseudoscalar meson octet π, K, $\bar{K}$, and η, the vector meson octet ρ, K^*, $\bar{K}^*$, and ω, and the vector meson singlet ϕ in a single **35** multiplet, and also unites the spin 1/2 baryon octet N, Σ, Λ, and Ξ with the spin 3/2 baryon decuplet Δ, $\Sigma(1385)$, $\Xi(1530)$, and Ω in a single **56** multiplet. The $SU(6)$ symmetry scored a number of successes, but it is actually nothing but a consequence of the approximate spin and flavor independence of forces in the quark

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

model; the $SU(6)$ symmetry is somewhat weaker than the assumption of spin and flavor independence, but, as shown in Appendix A, there is no evidence that the predictions of $SU(6)$ symmetry are any more accurate than those of complete spin and flavor independence.

Nevertheless, there were various attempts to generalize the $SU(6)$ symmetry of the non-relativistic quark model to a fully relativistic quantum theory.[3] These attempts all failed, and a number of authors showed under various restrictive assumptions that this is in fact impossible.[4] The most far-reaching theorem of this sort was proved in 1967 by Coleman and Mandula.[5] They adopted reasonable assumptions about the finiteness of the number of particle types below any given mass, the existence of scattering at almost all energies, and the analyticity of the S-matrix, and used them to show that the most general Lie algebra of symmetry operators that commute with the S-matrix, that take single-particle states into single-particle states, and that act on multiparticle states as the direct sum of their action on single-particle states consists of the generators P_μ and $J_{\mu\nu}$ of the Poincaré group, plus ordinary internal symmetry generators that act on one-particle states with matrices that are diagonal in and independent of both momentum and spin. We will use this theorem as an essential ingredient in our analysis of all possible supersymmetry algebras in four spacetime dimensions in Chapter 25, and in higher spacetime dimensions in Chapter 32. In Section 32.3 we will consider supersymmetry algebras in theories that involve extended objects, for which the Coleman–Mandula theorem does not apply.

Coleman and Mandula's proof is ingenious and complicated. A version is presented in Appendix B to this chapter. In the present section we shall give a very simple purely kinematic proof of one piece of this theorem, but this piece is enough to show clearly why an unconventional symmetry like $SU(6)$ is possible in non-relativistic but not in relativistic theories. We shall use Lorentz invariance to show that if the Lie algebra of all symmetry operators B_α that commute with the momentum generators P_μ consists of the P_μ themselves plus the Hermitian generators B_A of some finite-parameter semi-simple compact* Lie subalgebra $\mathscr{A}$, then the B_A must be the generators of an ordinary internal symmetry, in the sense that they act on single-particle states with matrices that are diagonal in and independent of both momentum and spin. No use is made in this theorem of the properties of the S-matrix, of the finiteness of the particle spectrum, or of assumptions about how the symmetry generators act on physical states. The Lie algebra of $SU(6)$ is of course both semi-simple

* For the definition of semi-simple and compact Lie algebras, see the footnote in Section 15.2.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

and compact, so this theorem rules out the use of any such symmetry in relativistic theories to derive relations among particles of different spin.

Here is the proof. Let all symmetry generators that commute with the four-momentum P_μ form a Lie algebra spanned by the generators B_α. Consider the effect on these generators of a proper Lorentz transformation $x^\mu \to \Lambda^\mu{}_\nu x^\nu$, which is represented on Hilbert space by the unitary operator $U(\Lambda)$. It is easy to see that the operator $U(\Lambda)B_\alpha U^{-1}(\Lambda)$ is a Hermitian symmetry generator that commutes with $\Lambda_\mu{}^\nu P_\nu$, so since $\Lambda_\mu{}^\nu$ is non-singular, this operator must commute with P_μ, and therefore must be a linear combination of the B_α:

$$U(\Lambda)B_\alpha U^{-1}(\Lambda) = \sum_\beta D^\beta{}_\alpha(\Lambda)\, B_\beta \,, \tag{24.1.1}$$

with $D^\beta{}_\alpha(\Lambda)$ a set of real coefficients that furnish a representation of the homogeneous Lorentz group

$$D(\Lambda_1)\, D(\Lambda_2) = D(\Lambda_1\, \Lambda_2) \,. \tag{24.1.2}$$

Further, the $U(\Lambda)B_\alpha U^{-1}(\Lambda)$ satisfy the same commutation relations as the B_α, so the structure constants $C^\gamma_{\alpha\beta}$ of this Lie algebra are invariant tensors in the sense that

$$C^\gamma_{\alpha\beta} = \sum_{\alpha'\beta'\gamma'} D^{\alpha'}{}_\alpha(\Lambda)\, D^{\beta'}{}_\beta(\Lambda)\, D^\gamma{}_{\gamma'}(\Lambda^{-1})\, C^{\gamma'}_{\alpha'\beta'} \,. \tag{24.1.3}$$

Contracting this with the corresponding equation for $C^\alpha_{\gamma\delta}$, we find that

$$g_{\beta\delta} = \sum_{\beta'\delta'} D^{\beta'}{}_\beta(\Lambda)\, D^{\delta'}{}_\delta(\Lambda)\, g_{\beta'\delta'} \,, \tag{24.1.4}$$

where $g_{\beta\delta}$ is the Lie algebra metric

$$g_{\beta\delta} \equiv \sum_{\alpha\gamma} C^\gamma_{\alpha\beta}\, C^\alpha_{\gamma\delta} \,. \tag{24.1.5}$$

Because all of these generators commute with P_μ, we have $C^\alpha_{\mu\beta} = -C^\alpha_{\beta\mu} = 0$, so $g_{\mu\alpha} = g_{\alpha\mu} = 0$.

We will distinguish the symmetry generators other than the P_μ by using subscripts A, B, etc. in place of α, β, etc. Using the vanishing of $C^A_{\mu B} = -C^A_{B\mu}$ in Eq. (24.1.5) gives $g_{AB} = \sum_{CD} C^D_{AC}\, C^C_{BD}$. We assumed that the generators B_A span a compact semi-simple Lie algebra, so the matrix g_{AB} is positive-definite. Eqs. (24.1.4) and (24.1.2) show that the matrices $g^{1/2}D(\Lambda)g^{-1/2}$ furnish a real orthogonal and hence unitary finite-dimensional representation of the homogeneous Lorentz group. But because the Lorentz group is non-compact, *the only such representation is the trivial one*, for which $D(\Lambda) = 1$. (This is the place where relativity makes all the difference; the semi-simple part of the Galilean group is the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

compact group $SU(2)$, which of course has an infinite number of unitary finite-dimensional representations.) With $D(\Lambda) = 1$, the generators B_A commute with $U(\Lambda)$ for all Lorentz transformations $\Lambda^{\mu}{}_{\nu}$.

Acting on the state $|p,n\rangle$ of a single stable particle with momentum p^{μ} and spin and species labelled by a discrete index n, an operator like B_A that commutes with P_{μ} can only yield a linear combination of such states

$$B_A|p,n\rangle = \sum_m \Big(b_A(p)\Big)_{mn}|p,m\rangle \ . \tag{24.1.6}$$

The fact that the B_A commute with what we called 'boosts' in Section 2.5 implies that the $b_A(p)$ are independent of momentum, and the fact that the B_A commute with rotations implies that the $b_A(p)$ act as unit matrices on spin indices, so the B_A are the generators of an ordinary internal symmetry, as was to be proved.

24.2 The Birth of Supersymmetry

If theoretical physics followed logic in its evolution, then after the proof of the Coleman–Mandula theorem someone seeking exceptions to this theorem would have noticed that it deals only with transformations that take bosons into bosons and fermions into fermions and are therefore generated by operators that satisfy commutation relations rather than anti-commutation relations. This would have raised the question of whether a relativistic theory can have symmetries acting non-trivially on particle spins that take fermions and bosons into each other, and that therefore satisfy anticommutation relations rather than commutation relations. Exploring the possible structure of such a superalgebra along the lines described in the following chapter, supersymmetry would have emerged as the only possibility.

This is not what happened. Instead, supersymmetry arose in a series of articles on string theory and independently in a pair of little-noticed papers, about which more later, none of which show any sign that the authors were at all concerned with the Coleman–Mandula theorem.

Starting in the late 1960s, the effort to construct S-matrix elements for strong-interaction processes that would satisfy various theoretical requirements led to a new picture of various types of hadrons, as different modes of vibration of a string.[6] A point on a string labelled by a parameter σ will at time τ on some fixed clock have spacetime coordinates $X^{\mu}(\sigma,\tau)$, so the theory of a string's motion in d spacetime dimensions may be regarded

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

as a two-dimensional field theory with d bosonic fields, with action

$$I[X] = \frac{T}{2}\int d\sigma \int d\tau\, \eta_{\mu\nu}\left[\frac{\partial X^\mu}{\partial\tau}\frac{\partial X^\nu}{\partial\tau} - \frac{\partial X^\mu}{\partial\sigma}\frac{\partial X^\nu}{\partial\sigma}\right]$$
$$= T\int d\sigma^+ \int d\sigma^-\, \eta_{\mu\nu}\frac{\partial X^\mu}{\partial\sigma^+}\frac{\partial X^\nu}{\partial\sigma^-}\,, \qquad (24.2.1)$$

where T is a constant known as the string tension; $\mu = 0, 1, \ldots, d-1$; and $\sigma^\pm$ are the two-dimensional 'light-cone' coordinates $\sigma^\pm \equiv \tau \pm \sigma$. This action can be derived from a more general version, with complete invariance* under transformations of a pair of 'worldsheet coordinates' σ_k

$$I[X] = -\frac{T}{2}\int d^2\sigma\, \eta_{\mu\nu}\sqrt{\mathrm{Det}\, g}\, g^{kl}\frac{\partial X^\mu}{\partial\sigma^k}\frac{\partial X^\nu}{\partial\sigma^l}\,, \qquad (24.2.2)$$

by passing to a special coordinate system in which the worldsheet metric g_{kl} satisfies the condition

$$\sqrt{\mathrm{Det}\, g}\, g^{kl} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (24.2.3)$$

In much the same way that in electrodynamics the problems introduced by the negative sign of the action for timelike photons are eliminated by the gauge invariance of the theory, here the problems introduced by the negative sign of $\eta_{\mu\nu}$ in Eqs. (24.2.1) and (24.2.2) for $\mu = \nu = 0$ are eliminated by the invariance of the action (24.2.2) (for appropriate boundary conditions) under general transformations of the worldsheet coordinates. In the special coordinate system in which the action takes the form (24.2.1), there is an important remnant of invariance under general worldsheet coordinate transformations: invariance under the global *conformal transformations*:

$$\sigma^\pm \to f^\pm(\sigma^\pm)\,, \qquad (24.2.4)$$

with $f^\pm$ a pair of independent arbitrary functions.

The particles described by this string theory do not match those seen in the real world. Ramond[7] and Neveu and Schwarz[8] in 1971, aiming respectively to introduce particles with half-integer spin or with the quantum numbers of pions, suggested the addition of d fermionic field doublets $\psi_1^\mu(\sigma,\tau)$ and $\psi_2^\mu(\sigma,\tau)$. Shortly after, Gervais and Sakita[9] introduced an action for this theory:

$$I[X,\psi] = \int d\sigma^+ \int d\sigma^- \left[T\frac{\partial X^\mu}{\partial\sigma^+}\frac{\partial X_\mu}{\partial\sigma^-} + i\psi_2^\mu\frac{\partial}{\partial\sigma^+}\psi_{2\mu} + i\psi_1^\mu\frac{\partial}{\partial\sigma^-}\psi_{1\mu}\right], \qquad (24.2.5)$$

* This symmetry is violated by quantum anomalies, like those discussed in Chapter 22, except in $d = 26$ spacetime dimensions for the purely bosonic theory, or $d = 10$ dimensions after the introduction of fermions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

and noted that conformal invariance could be maintained by extending the conformal transformations (24.2.4) to act also on the fermion fields

$$\psi_1^\mu \to \left(\frac{df^+}{d\sigma^+}\right)^{-1/2} \psi_1^\mu , \qquad \psi_2^\mu \to \left(\frac{df^-}{d\sigma^-}\right)^{-1/2} \psi_2^\mu . \tag{24.2.6}$$

Gervais and Sakita pointed out that, in addition to two-dimensional conformal invariance and d-dimensional Lorentz invariance, for appropriate boundary conditions this theory has a symmetry under infinitesimal transformations that interchange the bosonic field X^μ with the fermionic fields ψ_r^μ:

$$\begin{aligned}
\delta\psi_2^\mu(\sigma^+,\sigma^-) &= iT\,\alpha_2(\sigma^-)\,\frac{\partial}{\partial\sigma^-}X^\mu(\sigma^+,\sigma^-)\,,\\
\delta\psi_1^\mu(\sigma^+,\sigma^-) &= iT\,\alpha_1(\sigma^+)\,\frac{\partial}{\partial\sigma^+}X^\mu(\sigma^+,\sigma^-)\,,\\
\delta X^\mu(\sigma^+,\sigma^-) &= \alpha_2(\sigma^-)\,\psi_2^\mu(\sigma^+,\sigma^-) + \alpha_1(\sigma^+)\,\psi_1^\mu(\sigma^+,\sigma^-)\,,
\end{aligned} \tag{24.2.7}$$

where α_1 and α_2 are a pair of infinitesimal *fermionic* functions of σ^+ and σ^-, respectively, like the Grassmann variables introduced in Section 9.5. This was an example of what has subsequently come to be called supersymmetry, a symmetry connecting bosons and fermions, but thus far it was only a symmetry of a two-dimensional field theory, not of a physical theory in four spacetime dimensions.

A few years later Wess and Zumino[10] referred back to the example of supersymmetry that had been provided by References 7–9, and commented that it would be natural to try to extend the idea of supersymmetry to quantum field theories in *four* spacetime dimensions. They constructed several supersymmetric models. The simplest involved a single Majorana (self-charge-conjugate Dirac) field ψ, a pair of real scalar and pseudoscalar bosonic fields A and B, and a pair of real scalar and pseudoscalar bosonic auxiliary fields F and G, with invariance under the infinitesimal transformation**

$$\begin{aligned}
\delta A &= \left(\bar\alpha\,\psi\right), \qquad \delta B = -i\left(\bar\alpha\,\gamma_5\,\psi\right),\\
\delta\psi &= \partial_\mu(A + i\gamma_5 B)\gamma^\mu\alpha + (F - i\gamma_5 G)\alpha\,,\\
\delta F &= \left(\bar\alpha\,\gamma^\mu\,\partial_\mu\psi\right), \qquad \delta G = -i\left(\bar\alpha\,\gamma_5\gamma^\mu\,\partial_\mu\psi\right),
\end{aligned} \tag{24.2.8}$$

where α is an arbitrary constant infinitesimal Majorana fermion c-number

** The notation for Dirac matrices used here is explained in the Preface and in Section 5.4. The γ_5 used here (which satisfies $\gamma_5^2 = 1$) is a factor of i times that used by Wess and Zumino, and the covariant conjugate $\bar\psi$ of any spinor ψ is defined here as i times that of Wess and Zumino. For this reason, some of the phases in Eqs. (24.2.8)–(24.2.10) are different from those in Reference 10.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

parameter. If we require invariance of the action under these transformations, then the most general real, Lorentz-invariant, parity-conserving, renormalizable Lagrangian density built out of these ingredients is

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{2}\partial_\mu A\,\partial^\mu A - \tfrac{1}{2}\partial_\mu B\,\partial^\mu B - \tfrac{1}{2}\bar\psi\gamma^\mu\partial_\mu\psi \\ &+\tfrac{1}{2}(F^2+G^2) + m\left[FA + GB - \tfrac{1}{2}\bar\psi\psi\right] \\ &+ g\left[F(A^2+B^2) + 2GAB - \bar\psi(A+i\gamma_5 B)\psi\right]. \qquad (24.2.9)\end{aligned}$$

Since the auxiliary fields F and G enter quadratically, we can derive an equivalent Lagrangian by setting them equal to the values given by the field equations

$$F = -mA - g(A^2+B^2), \qquad G = -mB - 2gAB. \qquad (24.2.10)$$

The Lagrangian density then becomes

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{2}\partial_\mu A\,\partial^\mu A - \tfrac{1}{2}\partial_\mu B\,\partial^\mu B - \tfrac{1}{2}\bar\psi\gamma^\mu\partial_\mu\psi \\ &-\tfrac{1}{2}m^2[A^2+B^2] - \tfrac{1}{2}m\bar\psi\psi \\ &- gmA(A^2+B^2) - \tfrac{1}{2}g^2(A^2+B^2)^2 - g\bar\psi(A+i\gamma_5 B)\psi. \qquad (24.2.11)\end{aligned}$$

This Lagrangian density exhibits relations not only between scalar and fermion masses, but also between Yukawa interactions and scalar self-couplings, which are characteristic of supersymmetric theories. Wess and Zumino also described supersymmetry transformations and gave a Lagrangian for a supermultiplet containing a vector field. (We shall go into all this in more detail in Chapter 26.) Finally, in a second paper, Wess and Zumino[11] recalled the Coleman–Mandula theorem and traced the apparent violation of this theorem to the fact that the symmetry generators here satisfy anticommutation rather than commutation relations. It was a few more years before Gliozzi, Scherk and Olive[11a] showed that it was possible to construct a superstring theory with spacetime as well as worldsheet supersymmetry by imposing suitable periodicity conditions on the fields of the Ramond–Neveu–Schwarz model.

Unknown to Wess and Zumino, at the time of their first papers on supersymmetry in four spacetime dimensions this symmetry had already appeared in a pair of papers published in the Soviet Union. In 1971 Gol'fand and Likhtman[12] had extended the algebra of the Poincaré group discussed in Section 2.4 to a superalgebra and used the requirement of invariance under this superalgebra to construct supersymmetric field theories in four spacetime dimensions. Their paper though prophetic gave few details, and was generally ignored until much later. Independently Volkov and Akulov[13] in 1973 discovered what today would be called spontaneously broken supersymmetry, but they used their formalism to identify the Goldstone fermion associated with supersymmetry breaking

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

with the neutrino, an idea that met with no success. For most theorists, especially outside the Soviet Union, supersymmetry as a possible symmetry of nature in four spacetime dimensions began with the 1974 papers of Wess and Zumino.

Appendix A $SU(6)$ Symmetry of Non-Relativistic Quark Models

This appendix will describe the way that an $SU(6)$ symmetry that relates particles of different spin arises in non-relativistic quark models. This has nothing directly to do with supersymmetry, but it provides the historical background for the Coleman–Mandula theorem, which is an essential input to the construction of general supersymmetry algebras in Sections 25.1 and 31.1.

In general, the Hamiltonian of the non-relativistic quark model could depend not only on positions and momenta but also on the spin and flavor operators $\sigma_i^{(n)}$ and $\lambda_A^{(n)}$, where the $\sigma_i^{(n)}$ (with $i = 1, 2, 3$) act on the spin indices of the nth quark as the Pauli matrices σ_i defined by Eq. (5.4.18), while the $\lambda_A^{(n)}$ (with $A = 1, 2, \ldots, 8$) act on the flavor indices of the nth quark as the Gell-Mann $SU(3)$ matrices λ_A defined by Eq. (19.7.2). (Where n refers to an antiquark, $\sigma_i^{(n)}$ and $\lambda_A^{(n)}$ act as the matrices $-\sigma_i^T$ and $-\lambda_A^T$ of the contragredient representations.) If we were to assume only that there is no spin-orbit coupling, so that the total orbital angular momentum L_i is separately conserved, then we could conclude only that the Hamiltonian commutes with the total spin and unitary spin

$$S_i \equiv \tfrac{1}{2}\sum_n \sigma_i^{(n)}\,, \qquad T_A \equiv \tfrac{1}{2}\sum_n \lambda_A^{(n)}\,, \tag{24.A.1}$$

as well as L_i. On the other hand, if we were to suppose that the Hamiltonian depends only on quark positions and momenta, and not at all on spins or quark flavors, then such a Hamiltonian would commute not only with the total orbital angular momentum $\mathbf{L}$, but also with *each* of the operators $\sigma_i^{(n)}$ and $\lambda_A^{(n)}$. In between these two extremes there is the interesting possibility that in addition to commuting with the L_i, S_i, and T_A, the Hamiltonian also commutes with the operators

$$R_{iA} \equiv \tfrac{1}{2}\sum_n \pm\sigma_i^{(n)}\,\lambda_A^{(n)}\,, \tag{24.A.2}$$

where the sign is + or − for quarks and antiquarks, respectively.* The S_i, T_A, and R_{iA} form the Lie algebra of the group $SU(6)$, with commutation

* The minus sign for antiquarks arises because the terms in R_{iA} for antiquarks must act on spin and flavor indices as the matrix $-(\sigma_i\,\lambda_A)^T = -(-\sigma_i^T)(-\lambda_A^T)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

relations

$$[S_i\,,\,S_j] = i\sum_k \epsilon_{ijk}S_k\,, \qquad [T_A\,,\,T_B] = i\sum_C f_{ABC}T_C\,, \qquad [S_i\,,\,T_A] = 0\,,$$
$$[S_i\,,\,R_{jA}] = i\sum_k \epsilon_{ijk}R_{kA}\,, \qquad [T_A\,,\,R_{iB}] = i\sum_C f_{ABC}R_{iC}\,, \tag{24.A.3}$$
$$[R_{Ai}\,,\,R_{Bj}] = i\delta_{ij}\sum_C f_{ABC}\,T_C + \tfrac{2}{3}i\delta_{AB}\sum_k \epsilon_{ijk}\,S_k + i\sum_{kC}\epsilon_{ijk}\,d_{ABC}R_{kC}\,.$$

Here f_{ABC} and d_{ABC} are respectively totally antisymmetric and totally symmetric numerical coefficients,[14] with independent non-vanishing values given by

$$\begin{aligned} &f_{123} = 1\,, \qquad f_{458} = f_{678} = \sqrt{3}/2\,, \\ &f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = 1/2\,, \end{aligned} \tag{24.A.4}$$

and

$$\begin{aligned} &d_{146} = d_{157} = -d_{247} = d_{256} = d_{344} = d_{355} = -d_{366} = -d_{377} = 1/2\,, \\ &d_{118} = d_{228} = d_{338} = -d_{888} = 1/\sqrt{3}\,, \\ &d_{448} = d_{558} = d_{668} = d_{778} = -1/(2\sqrt{3})\,. \end{aligned} \tag{24.A.5}$$

This is the symmetry that remains if we include a spin- and flavor-dependent two-body interaction in the Hamiltonian that commutes with R_{iA} as well as with the S_i and T_A. There are such interactions, given by linear combinations of two-body operators of the form

$$H^{(nm)} \propto \left[1 \pm \sum_i \sigma_i^{(n)}\sigma_i^{(m)}\right]\left[\frac{2}{3} \pm \sum_A \lambda_A^{(n)}\,\lambda_A^{(m)}\right]\,, \tag{24.A.6}$$

where the sign $\pm$ is negative if one of the particles n, m is a quark and the other an antiquark and positive if they are both quarks or both antiquarks.

Of course, even in the non-relativistic quark model the $SU(6)$ symmetry is at best approximate. It is broken by spin–orbit and spin–spin forces, and also by the mass of the s quark, which reduces the flavor $SU(3)$ symmetry to the $SU(2)$ and $U(1)$ of isospin and hypercharge conservation. If we avoid the effects of this quark mass difference by restricting ourselves to hadrons built up from the light u and d quarks and antiquarks, then the only non-vanishing λ_A matrices are the λ_a with $a = 1, 2, 3$ (which for the u and d quarks are given by the Pauli matrices (5.4.18) (that in this context are conventionally called τ_a) and λ_8 (which is just the number $1/\sqrt{3}$ for the u and d quarks, and $-1/\sqrt{3}$ for the $\bar{u}$ and $\bar{d}$ antiquarks). The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

interaction (24.A.6) thus becomes

$$H^{(nm)} \propto \left[1 \pm \sum_i \sigma_i^{(n)} \sigma_i^{(m)}\right] \left[1 \pm \sum_A \tau_A^{(n)} \tau_A^{(m)}\right] . \tag{24.A.7}$$

Aside from quark number conservation, the remaining symmetry is then $SU(4)$, with generators S_i, T_a, and R_{ia} which commute with (24.A.7). This was proposed by Wigner[2] in 1937 as a symmetry of nuclear forces, though of course with protons and neutrons in place of u and d quarks. The interaction (24.A.7) is known in nuclear theory as a *Majorana potential* to distinguish it from an interaction that does not depend on spin or isospin, called a *Wigner potential*, or an interaction proportional to just the spin-dependent or just the isospin-dependent factor in (24.A.7), known respectively as a *Bartlett potential* and a *Heisenberg potential.*

It is amusing to note that, although in non-relativistic theories there is no theoretical barrier to symmetries like $SU(6)$ that act on spin as well as particle type, *there never was any experimental evidence for such an assumed $SU(6)$ symmetry of the non-relativistic quark model that is any better satisfied than the assumption of complete spin and flavor independence.* These assumptions are not the same; if the Hamiltonian of a system of N non-relativistic quarks and/or antiquarks is completely independent of spin and flavor, then its symmetry is $SU(6)^N$, not $SU(6)$. For instance, a two-particle interaction like (24.A.6) and various other multiparticle interactions break $SU(6)^N$ to $SU(6)$. Of course, all these symmetries are only approximate anyway. The question is whether $SU(6)$ is less badly broken than $SU(6)^N$?

This cannot be answered by studying the multiplet that contains the baryon octet, which consists of the nucleon and hyperons Λ, Σ, and Ξ. In the non-relativistic quark model these particles are interpreted as bound states of three quarks with zero orbital angular momentum. Because these states are color neutral, the wave function is completely antisymmetric in the suppressed color indices, and therefore it is completely symmetric under the combined interchange of spin and flavor. The baryon octet would therefore have to be put in the symmetric third-rank tensor representation **56** of $SU(6)$, which besides the baryon octet contains a spin 3/2 decuplet, which may be identified as the one consisting of the famous '3–3' resonance Δ and the $\Sigma(1385)$, $\Xi(1530)$, and Ω particles. (Numbers in parentheses give masses in MeV, where these are needed to distinguish the particles from others of the same isospin and strangeness but lower mass.) The $SU(6)$ symmetry leads to good predictions for the baryon magnetic moments: The quark charge operator is $q = e(\lambda_3/2 + \lambda_8/2\sqrt{3})$, so if quarks have the magnetic moments $3q/2m_N$ of Dirac particles of this

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

charge and mass $m_N/3$, then the magnetic moment operator is

$$\mu_i = 3\mu_N \left[\frac{1}{2}R_{i3} + \frac{1}{2\sqrt{3}}R_{i8}\right] ,$$

where $\mu_N \equiv e/2m_N$ is the nuclear magneton, and R_{iA} is defined by Eq. (24.A.2). It is straightforward to calculate the matrix elements of this symmetry generator between the members of the **56** multiplet, with the result that the magnetic moments for p, n, Λ, Σ^+, Σ^-, Ξ^-, and Ξ^0 in units of μ_N are respectively $+3$, -2, -1, $+3$, -1, -1, and -2, which may be compared with the corresponding experimental values $+2.79$, -1.91, -0.61, $+2.46$, -1.16, -0.65, and -1.25. The agreement is fair, and somewhat better (except for Σ^-) if we take the quark magnetic moment to be a little less than $3\mu_N$. Because of the symmetry of the three-quark wave function, nothing new is learned here if we assume that the Hamiltonian is completely independent of spin and flavor; the states of zero angular momentum would still have to fall in a multiplet consisting of $6 \times 7 \times 8/6! = 56$ members. In particular, the operator (24.A.6) has the same value 4 for any state of two quarks that is symmetric under simultaneous interchange of their spins and flavors, whether it is symmetric both under interchange of spins and interchange of flavors, or antisymmetric under both interchanges.

In order to decide whether $SU(6)$ is any better than $SU(6)^N$, it is more useful to study the mesons, which in the non-relativistic quark model are interpreted as bound states of a quark and antiquark. If the Hamiltonian of these states is completely independent of spin and flavor then its symmetry is $SU(6)^2$, and the meson states fall into its 36-dimensional $(6, \bar{6})$ representation, while for $SU(6)$ symmetry we could only say that the mesons belong to either of the two representations of $SU(6)$ contained in $\mathbf{6} \times \bar{\mathbf{6}}$: the adjoint representation **35** or the singlet representation. To be more specific, the **35** consists of an $SU(3)$ singlet with spin $S = 1$, an $SU(3)$ octet with $S = 0$, and an $SU(3)$ octet with $S = 1$, corresponding to the $SU(6)$ generators S_i, T_A, and R_{iA}, which is split from the $SU(3)$ singlet state with $S = 0$ by the interaction (24.A.6). Since all these assumptions are approximate anyway, the question in deciding whether $SU(6)$ symmetry is more accurate than complete spin and flavor independence is whether the splitting of the $SU(3)$ singlet $S = 0$ state from the other 35 states of the same orbital angular momentum is any greater than the splittings within the **35** supermultiplet.

For orbital angular momentum $L = 0$ the quark–antiquark states have negative parity P, and positive or negative charge-conjugation quantum number C (for self-charge-conjugate states) according to whether the total spin S is zero or one, respectively. (For an explanation, see Section 5.5.) The **35** therefore consists of a singlet with $J^{PC} = 1^{--}$, a 0^{-+} octet, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

a 1^{--} octet, which may be identified respectively as: the $\phi(1020)$; the pseudoscalar octet π, η, K, and $\bar{K}$; and the vector octet ρ, ω, K^*, and $\bar{K}^*$. There is also a 0^{-+} $SU(3)$ singlet η' at 958 MeV, which can be regarded as the $SU(6)$ singlet. The splitting of this singlet from the particles in the **35** multiplet is not distinctly greater than the splittings within the **35** multiplet.

It may be argued that the $L = 0$ mesons do not provide a good test of the symmetries of the non-relativistic quark model, because they include the Goldstone bosons π, η, K, and $\bar{K}$, which become massless for zero u and d quark masses and are therefore not well described by this model. Therefore let us consider the quark–antiquark states with $L = 1$. These states have P positive and C positive or negative according to whether $S = 1$ or $S = 0$, so the p-wave **35** consists of: $SU(3)$ singlets with $S = 1$ and $J^{PC} = 0^{++}$, 1^{++}, 2^{++}, which may be identified as the $f_0(1370)$, the $f_1(1285)$, and the $f_2(1270)$; an $S = 0$ 1^{+-} octet identified as $h_1(1170)$, $b_1(1235)$, $K_1(1400)$, and $\bar{K}_1(1400)$; and $S = 1$ octets: a 0^{++} octet consisting of $f_0(980)$, $a_0(980)$, $K_0^+(1950)$, and $\bar{K}_0^+(1950)$; a 1^{++} octet consisting of $f_1(1420)$, $a_1(1260)$, $K_1^+(1650)$, and $\bar{K}_1^+(1650)$; and a 2^{++} octet consisting of $f_2(1430)$, $a_2(1320)$, $K_2^+(1980)$, and $\bar{K}_2^+(1980)$. In addition to these 35×3 states, there is a another particle with the right quantum numbers to be the p-wave $SU(6)$ singlet: the 1^{+-} isoscalar $h_1(1380)$. Of course, we could interchange the identifications of $h_1(1170)$ and $h_1(1380)$, or identify the $SU(3)$ singlet and octet isoscalar 1^{+-} states as orthogonal linear combinations of $h_1(1170)$ and $h_1(1380)$. The important point is that there are *two* of these 1^{+-} isoscalars, with no indication that one of them, belonging to an $SU(6)$ singlet, is more strongly split from the particles of the **35** than the particles of the **35** are split from each other. Here too, then, there is no evidence that $SU(6)$ symmetry is any more accurate than the stronger assumption of complete spin and flavor independence.

Appendix B The Coleman–Mandula Theorem

This appendix provides a proof of the celebrated theorem of Coleman and Mandula[5] that the only possible Lie algebra (as opposed to super-algebra) of symmetry generators consists of the generators P_μ and $J_{\mu\nu}$ of translations and homogeneous Lorentz transformations, together with possible internal symmetry generators, which commute with P_μ and $J_{\mu\nu}$ and act on physical states by multiplying them with spin-independent,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

momentum-independent Hermitian matrices.* By 'symmetry generators' here is meant any Hermitian operators: that commute with the S-matrix; whose commutators are also symmetry generators; which take one-particle states into one-particle states; and whose action on multiparticle states is the direct sum of their action on one-particle states (as in Eq. (24.B.1)). A further technical requirement will be added when needed later. Apart from the general principles of relativistic quantum mechanics described in Chapters 2 and 3, the only other assumptions needed in this proof are:

Assumption 1 *For any M there are only a finite number of particle types with mass less than M.*

Assumption 2 *Any two-particle state undergoes some reaction at almost all energies (that is, at all energies except perhaps an isolated set).*

Assumption 3 *The amplitudes for elastic two-body scattering are analytic functions of the scattering angle at almost all energies and angles.***

It is not necessary to assume that the S-matrix is governed by a local quantum field theory. The proof presented here is somewhat rearranged and streamlined, and it spells out some steps that Coleman and Mandula left to the reader.

It is convenient to start by proving this theorem for the subalgebra consisting of those symmetry generators B_α that commute with the four-momentum operator P_μ. (This part of the theorem is of some interest in itself; it rules out symmetries in relativistic theories that act like the $SU(6)$ symmetry of the non-relativistic quark model.) The action of such

* As we shall see, in theories with only massless particles there is also the possibility that in addition to the generators P_μ and $J_{\mu\nu}$ there are additional generators D and K_μ that fill out the Lie algebra of the conformal group.[15]

** Strictly speaking, this assumption is not satisfied in theories with infrared divergences such as quantum electrodynamics, where, as shown in Section 13.3, the S-matrix element for any one scattering process involving charged particles actually vanishes, except for elastic forward scattering. In Abelian gauge theories like electrodynamics this problem can be avoided by applying the Coleman–Mandula theorem to the theory with a fictitious gauge boson mass, and then working only with 'infrared-safe' quantities like masses and suitably integrated cross-sections that are finite in the limit of zero gauge boson mass. There is no problem in non-Abelian gauge theories like quantum chromodynamics, in which all massless particles are trapped — symmetries if unbroken would only govern S-matrix elements for gauge-neutral bound states, like the mesons and baryons in quantum chromodynamics. As far as I know, the Coleman–Mandula theorem has not been proved for non-Abelian gauge theories with untrapped massless particles, like quantum chromodynamics with many quark flavors.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

symmetry generators on multiparticle states is given by

$$B_\alpha |p\,m,\ q\,n,\ \ldots\rangle = \sum_{m'} \Big(b_\alpha(p)\Big)_{m'm} |p\,m',\ q\,n,\ \ldots\rangle + \sum_{n'} \Big(b_\alpha(q)\Big)_{n'n} |p\,m,\ q\,n',\ \ldots\rangle + \cdots , \tag{24.B.1}$$

where m, n, etc. are discrete indices labelling spin z-components and particle type for particles of a definite mass $\sqrt{-p_\mu\,p^\mu}$, and the $b_\alpha(p)$ are finite Hermitian matrices, which define the action of the B_α on one-particle states.

Now, we can see from Eq. (24.B.1) that the mapping that takes the B_α into $b_\alpha(p)$ for some fixed p is a homomorphism in the sense that the commutation relations

$$[B_\alpha, B_\beta] = i\sum_\gamma C^\gamma_{\alpha\beta} B_\gamma \tag{24.B.2}$$

are also satisfied by the Hermitian matrices $b_\alpha(p)$:

$$[b_\alpha(p), b_\beta(p)] = i\sum_\gamma C^\gamma_{\alpha\beta}\, b_\gamma(p)\,. \tag{24.B.3}$$

A well-known theorem proved in Section 15.2 tells us that any Lie algebra of finite Hermitian matrices like $b_\alpha(p)$ must be a direct sum of a compact semi-simple Lie algebra and $U(1)$ algebras. However, we cannot immediately apply this result to the operator algebras B_α because the homomorphism between the operators B_α and matrices $b_\alpha(p)$ is not necessarily an isomorphism. For it to be an isomorphism would require also that whenever $\sum_\alpha c^\alpha b_\alpha(p) = 0$ for some coefficients c^α and momentum p, then $\sum_\alpha c^\alpha b_\alpha(k) = 0$ for all momenta k, which is equivalent to the condition $\sum_\alpha c^\alpha B_\alpha = 0$.

Instead of considering the homomorphism that maps the B_α into the one-particle matrices $b_\alpha(p)$, Coleman and Mandula considered the homomorphism that maps the B_α into the matrices that define the action of B_α on *two-particle states* with fixed four-momenta p and q:

$$\Big(b_\alpha(p,q)\Big)_{m'n',mn} = \Big(b_\alpha(p)\Big)_{m'm}\delta_{n'n} + \Big(b_\alpha(q)\Big)_{n'n}\delta_{m'm}\,. \tag{24.B.4}$$

The invariance of the S-matrix for the elastic or quasi-elastic scattering of two particles with four-momenta p and q into two particles with momenta p' and q', with masses $\sqrt{-p'_\mu\,p'^\mu} = \sqrt{-p_\mu\,p^\mu}$ and $\sqrt{-q'_\mu\,q'^\mu} = \sqrt{-q_\mu\,q^\mu}$, yields the condition

$$b_\alpha(p',q')\,S(p',q';p,q) = S(p',q';p,q)\,b_\alpha(p,q)\,. \tag{24.B.5}$$

Here $S(p',q';p,q)$ is a matrix of the same dimensionality as $b(p,q)$ and $b(p',q')$, defined in terms of the connected S-matrix elements $S(pm, qn \to$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

$p'm', q'n'$) by

$$S(p\,m, q\,n \to p'm', q'n') \equiv \delta^4(p'+q'-p-q)\left(S(p', q'; p, q)\right)_{m'n', mn} . \quad (24.B.6)$$

According to Assumption 2 and the optical theorem (see Section 3.6), for almost any choice of p and q the elastic scattering amplitude is non-vanishing in the forward direction, and Assumption 3 then tells us that the matrix $S(p', q'; p, q)$ is non-singular for almost all p' and q' on the same mass shells and satisfying the conservation condition $p'+q' = p+q$, so for almost all such four-momenta Eq. (24.B.5) is a *similarity transformation.*

It follows then that if $\sum_\alpha c^\alpha b_\alpha(p, q) = 0$ for almost any fixed four-momenta p and q, then $\sum_\alpha c^\alpha b_\alpha(p', q') = 0$ for almost any four-momenta p' and q' on the same mass shells that satisfy $p'+q' = p+q$. Unfortunately, this does not tell us that $\sum_\alpha c^\alpha b_\alpha(p')$ and $\sum_\alpha c^\alpha b_\alpha(q')$ vanish, but only that these matrices are proportional to the unit matrix (with opposite coefficients). To do better, it is necessary to consider not the $b_\alpha(p)$ or $b_\alpha(p, q)$, but their traceless parts.

One immediate consequence of Eq. (24.B.5) is that

$$\mathrm{Tr}\, b_\alpha(p', q') = \mathrm{Tr}\, b_\alpha(p, q) . \quad (24.B.7)$$

With Eq. (24.B.4), this tells us that

$$\begin{aligned} &N(\sqrt{-q_\mu q^\mu})\,\mathrm{tr}\, b_\alpha(p') + N(\sqrt{-p_\mu p^\mu})\,\mathrm{tr}\, b_\alpha(q') \\ &= N(\sqrt{-q_\mu q^\mu})\,\mathrm{tr}\, b_\alpha(p) + N(\sqrt{-p_\mu p^\mu})\,\mathrm{tr}\, b_\alpha(q) , \end{aligned} \quad (24.B.8)$$

where $N(m)$ is the multiplicity[†] of particle types with mass m, and the lower case t in 'tr' indicates a sum over one-particle rather than two-particle labels. In order for this to be satisfied for almost all mass-shell four-momenta for which $p' + q' = p + q$, it is necessary that the function $\mathrm{tr}\, b_\alpha(p)/N(\sqrt{-p_\mu p^\mu})$ be linear[††] in p:

$$\frac{\mathrm{tr}\, b_\alpha(p)}{N(\sqrt{-p_\mu p^\mu})} = a^\mu_\alpha\, p_\mu , \quad (24.B.9)$$

with a^μ_α independent of p (and of everything else but the displayed indices.) We may define new symmetry generators by subtracting terms linear in

[†] These multiplicity factors were not shown explicitly by Coleman and Mandula. They are needed in justifying a step that Coleman and Mandula made without explanation, that of defining the symmetry generators $B^\sharp_\alpha$ with traceless kernels.

[††] A constant term is easily seen to be ruled out in Eq. (24.B.9) by the existence of processes in which the number of particles is not conserved, processes which are inevitable in any relativistic quantum theory satisfying the cluster decomposition principle. Even if we considered only two-particle processes and did not use this argument, a constant term in Eq. (24.B.9) would only amount to a change in the action of internal symmetries on physical states.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

the momentum operator P_μ:

$$B_\alpha^\# \equiv B_\alpha - a_\alpha^\mu P_\mu \,, \tag{24.B.10}$$

which according to Eq. (24.B.9) are represented on one-particle states by the traceless matrices

$$\left(b_\alpha^\#(p)\right)_{n'n} = \left(b_\alpha(p)\right)_{n'n} - \frac{\operatorname{tr} b_\alpha(p)}{N(\sqrt{-p_\mu p^\mu})}\,\delta_{n'n} \,. \tag{24.B.11}$$

Because P_μ commutes with B_α and the unit matrix commutes with everything, the commutators of the $B_\alpha^\#$ are the same as those of the B_α, and the commutators of the $b_\alpha^\#(p)$ are the same as those of the $b_\alpha(p)$:

$$[B_\alpha^\#, B_\beta^\#] = i\sum_\gamma C_{\alpha\beta}^\gamma B_\gamma = i\sum_\gamma C_{\alpha\beta}^\gamma \left[B_\gamma^\# + a_\gamma^\mu P_\mu\right] , \tag{24.B.12}$$

$$[b_\alpha^\#(p), b_\beta^\#(p)] = i\sum_\gamma C_{\alpha\beta}^\gamma b_\gamma(p) = i\sum_\gamma C_{\alpha\beta}^\gamma \left[b_\gamma^\#(p) + a_\gamma^\mu p_\mu\right] . \tag{24.B.13}$$

Also, Eq. (24.B.13) and the fact that commutators of the finite matrices $b_\alpha^\#(p)$ have zero trace[‡] imply that $\sum_\gamma C_{\alpha\beta}^\gamma a_\gamma^\mu = 0$, and using this in Eq. (24.B.12) then shows that the $B_\alpha^\#$ satisfy the same commutation relations as the B_α:

$$[B_\alpha^\#, B_\beta^\#] = i\sum_\gamma C_{\alpha\beta}^\gamma B_\gamma^\# \,. \tag{24.B.14}$$

Because $B_\alpha^\#$ is a symmetry generator, the scattering amplitude satisfies

$$b_\alpha^\#(p',q')\,S(p',q';p,q) = S(p',q';p,q)\,b_\alpha^\#(p,q) \,, \tag{24.B.15}$$

where the $b_\alpha^\#(p,q)$ are the matrices representing the $B_\alpha^\#$ on the two-particle states

$$\left(b_\alpha^\#(p,q)\right)_{m'n',mn} = \left(b_\alpha^\#(p)\right)_{m'm}\delta_{n'n} + \left(b_\alpha^\#(q)\right)_{n'n}\delta_{m'm} \tag{24.B.16}$$

and satisfy the same commutation relations as the $B_\alpha^\#$:

$$[b_\alpha^\#(p,q), b_\beta^\#(p,q)] = i\sum_\gamma C_{\alpha\beta}^\gamma\, b_\gamma^\#(p,q) \,. \tag{24.B.17}$$

The advantage of dealing with these two-particle matrices is that, since $S(p',q';p,q)$ is a non-singular matrix, it follows that if $\sum_\alpha c^\alpha b_\alpha^\#(p,q) = 0$ for some fixed mass-shell four-momenta p and q, then $\sum_\alpha c^\alpha b_\alpha^\#(p',q') = 0$ for

[‡] This is one of the places where we use Assumption 1, without which commutators need not have zero trace. Also, at this point it is crucial that we are dealing with commutation rather than anticommutation relations, since the unit matrix does not anticommute with other matrices, and anticommutators of finite matrices do not necessarily have zero trace.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

almost all p' and q' on the same respective mass shells with $p'+q' = p+q$. Because we are dealing now with traceless matrices, this tells us that

$$\sum_\alpha c^\alpha b_\alpha^\#(p') = \sum_\alpha c^\alpha b_\alpha^\#(q') = 0\,. \tag{24.B.18}$$

We would like to conclude from this that $\sum_\alpha c^\alpha b_\alpha^\#(k) = 0$ for *all* mass-shell four-momenta k, but so far we have proved only that $\sum_\alpha c^\alpha b_\alpha^\#(p') = 0$ for almost all of those p' for which $q' = p+q-p'$ as well as p' is on the mass shell (and correspondingly for q'.) To get around this limitation, we can use a trick of Coleman and Mandula, noting that if $\sum_\alpha c^\alpha b_\alpha^\#(p, q) = 0$ then (24.B.18) and (24.B.16) together yield

$$\sum_\alpha c^\alpha b_\alpha^\#(p, q') = 0\,,$$

so that, according to Eq. (24.B.15),

$$\sum_\alpha c^\alpha b_\alpha^\#(k\,, p+q'-k) = 0\,,$$

and therefore

$$\sum_\alpha c^\alpha b_\alpha^\#(k) = 0\,, \tag{24.B.19}$$

for almost all mass-shell four-momenta k for which $p+q'-k$ is also on the mass shell. Now, the conditions that q' and $p+q-q'$ be on the mass shell leave two parameters free in q', so that we have enough freedom in choosing q' that the condition that $p+q'-k$ is on the mass shell leaves us free to choose $\mathbf{k}$ to be anything we like, at least within a finite volume of momentum space. This volume can be adjusted to be as large as we like by taking $\mathbf{p}$ and $\mathbf{q}$ sufficiently large, so if $\sum_\alpha c^\alpha b_\alpha^\#(p,q) = 0$ for some fixed mass-shell four-momenta p and q then $\sum_\alpha c^\alpha b_\alpha^\#(k) = 0$ for almost all mass-shell four-momenta k. But then if $\sum_\alpha c^\alpha b_\alpha^\#(k_0) \neq 0$ for some particular mass-shell four-momentum k_0, a scattering process in which particles with four-momenta k_0 and k scatter into particles with four-momenta k' and k'' will be forbidden by the symmetry generated by $\sum_\alpha c^\alpha B_\alpha^\#$ for almost all k, k', and k'', in contradiction with our assumption about the analyticity of the scattering amplitude. We conclude then that if $\sum_\alpha c^\alpha b_\alpha^\#(p,q) = 0$ for some fixed mass-shell four-momenta p and q then $\sum_\alpha c^\alpha b_\alpha^\#(k) = 0$ for *all* k, and therefore $\sum_\alpha c^\alpha B_\alpha^\# = 0$, so that the mapping that takes B_α into $b_\alpha^\#(p,q)$ is an isomorphism.

One immediate consequence is that, since the number of independent matrices $b_\alpha^\#(p,q)$ cannot exceed $N(\sqrt{-p_\mu p^\mu})\,N(\sqrt{-q_\mu q^\mu})$, there can be at most a finite number of independent symmetry generators B_α. As emphasized by Coleman and Mandula, in proving their theorem it is not necessary to make an independent assumption that the symmetry algebra is finite-dimensional.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

The theorems of Section 15.2 tell us further that a Lie algebra of finite Hermitian matrices like $b_\alpha^\sharp(p,q)$ for fixed p and q is at most the direct sum of a semi-simple compact Lie algebra and some number of $U(1)$ Lie algebras. We have seen that this Lie algebra is isomorphic to that of the symmetry generators $B_\alpha^\sharp$, so the $B_\alpha^\sharp$ must also span the direct sum of at most a compact semi-simple Lie algebra and $U(1)$ Lie algebras.

Let us first dispose of the $U(1)$ Lie algebras. For any pair of mass-shell momenta p and q we can find a Lorentz generator J that leaves both p and q invariant. (If p and q are lightlike and parallel then take J to generate rotations around the common direction of $\mathbf{p}$ and $\mathbf{q}$. Otherwise, $p+q$ will be timelike, so we can take J to generate rotations around the common direction of $\mathbf{p}$ and $\mathbf{q}$ in the center-of-mass frame in which $\mathbf{p}=-\mathbf{q}$.) We can choose the basis of two-particle states to diagonalize J, so that

$$J|pm,qn\rangle = \sigma(m,n)|pm,qn\rangle\ . \tag{24.B.20}$$

Now, P_μ commutes with all $B_\alpha^\sharp$, and $[J,P_\mu]$ is a linear combination of components of P_μ, so P_μ commutes with all $[J,B_\alpha^\sharp]$, and therefore the symmetry generator $[J,B_\alpha^\sharp]$ must be a linear combination of the B_β, which by definition form a complete set of symmetry generators that commute with P_μ. More specifically, since the matrices representing a commutator of symmetry generators are necessarily traceless, $[J,B_\alpha^\sharp]$ has to be a linear combination of the $B_\beta^\sharp$. But any $U(1)$ generator $B_i^\sharp$ (taken Hermitian) in the algebra of the $B_\beta^\sharp$ would have to commute with all the $B_\beta^\sharp$, and hence in particular must commute with $[J,B_i^\sharp]$:

$$[B_i^\sharp,\,[J,\,B_i^\sharp]] = 0\ .$$

Taking the expectation value of this double commutator in the two-particle basis in which J is diagonal, we have

$$0 = \sum_{m',n'} \Big(\sigma(m',n') - \sigma(m,n)\Big)\ \Big|\Big(b_i^\sharp(p,q)\Big)_{m'n',mn}\Big|^2\ , \tag{24.B.21}$$

for any m and n. The indices run over a finite range, so if there were any σ for which there existed an m and n with $\sigma(m,n)=\sigma$ and an m' and n' with $\sigma(m',n')\neq\sigma$, with $(b_i^\sharp(p,q))_{m'n',mn}\neq 0$, then there would have to be a *smallest* such σ, in which case the right-hand side of Eq. (24.B.21) for this m and n would be positive-definite, contradicting Eq. (24.B.21). We conclude that $(b_i^\sharp(p,q))_{m'n',mn}$ must vanish for all m, n, m', and n' for which $\sigma(m',n')\neq\sigma(m,n)$. Because the algebra of the $b_i^\sharp(p,q)$ is isomorphic to that of the $B_i^\sharp$, this means that each of the $U(1)$ generators $B_i^\sharp$ commutes with J. Since we can choose $p+q$ to be in any timelike direction, it follows that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

each of the $U(1)$ generators $B_i^\sharp$ commutes with all the generators $J_{\mu\nu}$ of the homogeneous Lorentz group. The fact that they commute with what we called 'boosts' in Section 2.5 implies that the $(b_i^\sharp(p))_{n'n}$ are independent of three-momentum, and the fact that they commute with rotations implies that the $(b_i^\sharp(p))_{n'n}$ act as unit matrices on spin indices, so these generators are the generators of an ordinary internal symmetry.

This leaves the $B_\alpha^\sharp$ that generate a semi-simple compact Lie algebra. The argument of Section 24.1 (a somewhat more explicit version of the reasoning given by Coleman and Mandula) tells us that the generators of the semi-simple compact part of the Lie algebra commute with Lorentz transformations and, as shown for the $U(1)$ generators, this means that they too are the generators of internal symmetries. We have thus shown that the symmetry generators B_α that commute with P_μ are either internal symmetry generators or linear combinations of the components of P_μ itself.

Next, we must take up the possibility of symmetry generators that do not commute with the momentum operator. The action of a general symmetry generator A_α on a one-particle state $|p\,n\rangle$ of four-momentum p would be

$$A_\alpha\,|p,n\rangle = \sum_{n'} \int d^4p' \left(\mathscr{A}_\alpha(p',p)\right)_{n'n} |p',n'\rangle\,, \tag{24.B.22}$$

where n and n' are again discrete indices labelling both spin z-components and particle types. Of course, the kernel $\mathscr{A}_\alpha(p',p)$ must vanish unless both p and p' are on the mass shell. We shall show first that the $\mathscr{A}_\alpha(p',p)$ vanishes for any $p' \neq p$.

For this purpose, note that if A_α is a symmetry generator, then so is

$$A_\alpha^f \equiv \int d^4x\ \exp(i\,P\cdot x)\,A_\alpha\ \exp(-i\,P\cdot x)\,f(x)\,, \tag{24.B.23}$$

where P_μ is the four-momentum operator, and $f(x)$ is a function that can be chosen as we like. Acting on a one-particle state, this gives

$$A_\alpha^f\,|p,n\rangle = \sum_{n'} \int d^4p'\ \tilde{f}(p'-p) \left(\mathscr{A}_\alpha(p',p)\right)_{n'n} |p',n'\rangle\,, \tag{24.B.24}$$

where $\tilde{f}$ is the Fourier transform

$$\tilde{f}(k) \equiv \int d^4x\ \exp(ix\cdot k)\,f(x)\,. \tag{24.B.25}$$

Suppose that there is some pair of mass-shell four-momenta p and $p+\Delta$ with $\Delta \neq 0$ such that $\mathscr{A}(p+\Delta,p) \neq 0$. Generic mass-shell four-momenta q, p', and q' with $p'+q' = p+q$ will not have $q+\Delta$ or $p'+\Delta$ or $q'+\Delta$ on the mass shell. If we take $\tilde{f}(k)$ to vanish outside a sufficiently small

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

region around Δ, then A_α^f will annihilate all one-particle states with four-momenta q, p', and q', but not the one-particle states with four-momentum p, so such a symmetry would forbid any scattering process in which any particles of momenta p and q go into any particles of momenta p' and q', in contradiction with the consequence of Assumptions 2 and 3 that there is some scattering at almost all energies and angles.

This result does not mean that any symmetry generator A_α must commute with P_μ, because the kernels $\mathscr{A}_\alpha(p',p)$ may include terms proportional to derivatives of $\delta^4(p'-p)$ as well as terms proportional to $\delta^4(p'-p)$ itself. To deal with this possibility, Coleman and Mandula made the 'ugly technical assumption' that the kernels $\mathscr{A}_\alpha(p',p)$ are *distributions*, which means that each can contain at most a finite number D_α of derivatives of $\delta^4(p'-p)$. To put this another way, each symmetry generator A_α is assumed to act on one-particle states as a polynomial of order D_α in the derivatives $\partial/\partial p_\mu$, with matrix coefficients that at this point are allowed to depend on momentum and spin. To use the above results for symmetry generators that commute with the momentum operators, Coleman and Mandula considered the D_α-fold commutator of momentum operators with A_α

$$B_\alpha^{\mu_1\cdots\mu_{D_\alpha}} \equiv [P^{\mu_1},[P^{\mu_2},\ldots[P^{\mu_{D_\alpha}},A_\alpha]\,]\ldots]\,. \tag{24.B.26}$$

The matrix elements of the commutators of $B_\alpha^{\mu_1\cdots\mu_{D_\alpha}}$ with P^μ between states with four-momenta p' and p are proportional to $D_\alpha+1$ factors of $p'-p$, times a polynomial of order D_α in momentum derivatives acting on $\delta^4(p'-p)$, and therefore vanish. Since the generators $B_\alpha^{\mu_1\cdots\mu_{D_\alpha}}$ commute with the momentum operators, according to the results obtained so far they act on one-particle states with matrices of the form

$$b_\alpha^{\mu_1\cdots\mu_{D_\alpha}}(p) = b_\alpha^{\sharp\mu_1\cdots\mu_{D_\alpha}} + a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}} p_\mu\, 1\,, \tag{24.B.27}$$

where the $b_\alpha^{\sharp\mu_1\cdots\mu_{D_\alpha}}$ are momentum-independent traceless Hermitian matrices generating an ordinary internal symmetry algebra, and the $a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}}$ are momentum-independent numerical constants, with both $b_\alpha^{\sharp\mu_1\cdots\mu_{D_\alpha}}$ and $a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}}$ symmetric in the indices $\mu_1\cdots\mu_{D_\alpha}$. Also, even though the A_α do not necessarily commute with P_μ, they cannot take one-particle states off the mass shell, so since Assumption 1 requires that the mass-squared operator $-P_\mu P^\mu$ has only discrete eigenvalues, the A_α must commute with $-P_\mu P^\mu$. It follows in particular that for $D \geq 1$

$$0 = [P^{\mu_1}P_{\mu_1},[P^{\mu_2},\ldots[P^{\mu_{D_\alpha}},A_\alpha]\,]\ldots] = 2P_{\mu_1}B_\alpha^{\mu_1\cdots\mu_{D_\alpha}}\,,$$

so that

$$0 = p_{\mu_1} b_\alpha^{\mu_1\cdots\mu_{D_\alpha}}(p)\,. \tag{24.B.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

As long as the theory contains massive particles this must be satisfied for p in any timelike direction, so for $D_\alpha \geq 1$

$$b_\alpha^{\sharp\mu_1\cdots\mu_{D_\alpha}} = 0\,, \tag{24.B.29}$$

and

$$a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}} = -a_\alpha^{\mu_1\mu\cdots\mu_{D_\alpha}}\,. \tag{24.B.30}$$

But for $D_\alpha \geq 2$ Eq. (24.B.30) together with the symmetry of $a_\alpha^{\mu\mu_1\cdots\mu_D}$ in the indices $\mu_1\cdots\mu_{D_\alpha}$ would require that $a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}} = 0$. (For then $a_\alpha^{\mu\mu_1\mu_2\cdots\mu_{D_\alpha}} = a_\alpha^{\mu\mu_2\mu_1\cdots\mu_{D_\alpha}} = -a_\alpha^{\mu_2\mu\mu_1\cdots\mu_{D_\alpha}} = -a_\alpha^{\mu_2\mu_1\mu\cdots\mu_{D_\alpha}} = a_\alpha^{\mu_1\mu_2\mu\cdots\mu_{D_\alpha}} = a_\alpha^{\mu_1\mu\mu_2\cdots\mu_{D_\alpha}} = -a_\alpha^{\mu\mu_1\mu_2\cdots\mu_{D_\alpha}}$.) We are left with at most two kinds of non-vanishing symmetry generator: those with $D_\alpha = 0$, for which the generator A_α commutes with P_μ and therefore must be either an internal symmetry generator or some linear combination of the P_μ; and those with $D_\alpha = 1$, in which case

$$[P^\nu, A_\alpha] = a_\alpha^{\mu\nu} P_\mu\,, \tag{24.B.31}$$

with $a_\alpha^{\mu\nu}$ some numerical constants antisymmetric in μ and ν. Eq. (24.B.31) requires that

$$A_\alpha = -\tfrac{1}{2} i\, a_\alpha^{\mu\nu} J_{\mu\nu} + B_\alpha\,, \tag{24.B.32}$$

where $J_{\mu\nu}$ is the generator of proper Lorentz transformations, which according to Eq. (2.4.13) satisfies $[P^\nu, J^{\rho\sigma}] = -i\eta^{\nu\rho}P^\sigma + i\eta^{\nu\sigma}P^\rho$, and B_α commutes with P_μ. Since A_α and $J_{\mu\nu}$ are symmetry generators, so is B_α, which must therefore be a linear combination of internal symmetry generators and/or the components of P_μ. Eq. (24.B.32) therefore completes the proof of the Coleman–Mandula theorem.

* * *

In theories with only massless particles Eq. (24.B.30) does not necessarily follow from Eq. (24.B.28); since $p_\mu p^\mu = 0$, we can also have

$$a_\alpha^{\mu\mu_1\cdots\mu_{D_\alpha}} + a_\alpha^{\mu_1\mu\cdots\mu_{D_\alpha}} \propto \eta^{\mu\mu_1}\,. \tag{24.B.33}$$

In this case the symmetry algebra consists of internal symmetries plus the algebra of the conformal group, which is spanned by generators K^μ and D together with the generators $J^{\mu\nu}$ and P^μ of the Poincaré group. The commutation relations are

$$\begin{aligned}
&[P^\mu, D] = iP^\mu\,, \qquad [K^\mu, D] = -iK^\mu\,,\\
&[P^\mu, K^\nu] = 2i\eta^{\mu\nu}D + 2iJ^{\mu\nu}\,, \qquad [K^\mu, K^\nu] = 0\,,\\
&[J^{\rho\sigma}, K^\mu] = i\eta^{\mu\rho}K^\sigma - i\eta^{\mu\sigma}K^\rho\,, \qquad [J^{\rho\sigma}, D] = 0\,,
\end{aligned} \tag{24.B.34}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

together with the commutation relations (2.4.12)–(2.4.14) of the Poincaré algebra

$$\begin{aligned} i[J^{\mu\nu}, J^{\rho\sigma}] &= \eta^{\nu\rho} J^{\mu\sigma} - \eta^{\mu\rho} J^{\nu\sigma} - \eta^{\sigma\mu} J^{\rho\nu} + \eta^{\sigma\nu} J^{\rho\mu} , \\ i[P^{\mu}, J^{\rho\sigma}] &= \eta^{\mu\rho} P^{\sigma} - \eta^{\mu\sigma} P^{\rho} , \\ [P^{\mu}, P^{\rho}] &= 0 . \end{aligned} \tag{24.B.35}$$

The infinitesimal group element

$$U(1+\omega, \epsilon, \lambda, \rho) = 1 + (i/2) J_{\mu\nu} \omega^{\mu\nu} + i P_{\mu} \epsilon^{\mu} + i\lambda D + i K_{\mu} \rho^{\mu} \tag{24.B.36}$$

induces the infinitesimal spacetime transformation

$$x^{\mu} \to x^{\mu} + \omega^{\mu\nu} x_{\nu} + \epsilon^{\mu} + \lambda x^{\mu} + \rho^{\mu} x^{\nu} x_{\nu} - 2 x^{\mu} \rho^{\nu} x_{\nu} . \tag{24.B.37}$$

These are the most general infinitesimal spacetime transformations that leave the light-cone invariant.

Problems

1. Show that the most general symmetry algebra allowed under the assumptions of the Coleman–Mandula theorem in the case where all particles are massless consists of internal symmetry generators plus either the Poincaré algebra or the conformal algebra (24.B.34), (24.B.35).

2. Show that the Gervais–Sakita action (24.2.5) is invariant under the worldsheet supersymmetry transformation (24.2.7).

3. Calculate the change in the Wess–Zumino Lagrangian density (24.2.9) under the spacetime supersymmetry transformation (24.2.8).

References

1. B. Sakita, *Phys. Rev.* **136**, B 1756 (1964); F. Gursey and L. A. Radicati, *Phys. Rev. Lett.* **13**, 173 (1964); A. Pais, *Phys. Rev. Lett.* **13**, 175 (1964); F. Gursey, A. Pais, and L. A. Radicati, *Phys. Rev. Lett.* **13**, 299 (1964). These articles are reprinted in *Symmetry Groups in Nuclear and Particle Physics*, F. J. Dyson, ed. (W. A. Benjamin, New York, 1966), along with a helpful set of Dyson's lecture notes on this subject.

2. E. P. Wigner, *Phys. Rev.* **51**. 106 (1937). This is reproduced in *Symmetry Groups in Nuclear and Particle Physics*, Ref. 1.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

3. A. Salam, R. Delbourgo, and J. Strathdee, *Proc. Roy. Soc. (London)* **A 284**, 146 (1965); M. A. Beg and A. Pais, *Phys. Rev. Lett.* **14**, 267 (1965); B. Sakita and K. C. Wali, *Phys. Rev.* **139**, B 1355 (1965). These are reproduced in *Symmetry Groups in Nuclear and Particle Physics*, Ref. 1.
4. W. D. McGlinn, *Phys. Rev. Lett.* **12**, 467 (1964); O. W. Greenberg, *Phys. Rev.* **135**, B 1447 (1964); L. Michel, *Phys. Rev.* **137**, B 405 (1964); L. Michel and B. Sakita, *Ann. Inst. Henri-Poincaré* **2**, 167 (1965); M. A. B. Beg and A. Pais, *Phys. Rev. Lett.* **14**, 509, 577 (1965); S. Coleman, *Phys. Rev.* **138**, B 1262 (1965); S. Weinberg, *Phys. Rev.* **139**, B 597 (1965); L. O'Raifeartaigh, *Phys. Rev.* **139**, B 1052 (1965). These are reproduced in *Symmetry Groups in Nuclear and Particle Physics*, Ref. 1.
5. S. Coleman and J. Mandula, *Phys. Rev.* **159**, 1251 (1967).
6. For an introduction with references to the original literature, see M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge University Press, Cambridge, 1987); J. Polchinski, *String Theory* (Cambridge University Press, Cambridge, 1998).
7. P. Ramond, *Phys. Rev.* **D3**, 2415 (1971). This article is reprinted in *Superstrings — The First 15 Years of Superstring Theory*, J. H. Schwarz, ed. (World Scientific, Singapore, 1985).
8. A. Neveu and J. H. Schwarz, *Nucl. Phys.* **B31**, 86 (1971); *Phys. Rev.* **D4**, 1109 (1971). These articles are reprinted in *Superstrings — The First 15 Years of Superstring Theory*, Ref. 7. Also see Y. Aharonov, A. Casher, and L. Susskind, *Phys. Rev.* **D5**, 988 (1972).
9. J.-L. Gervais and B. Sakita, *Nucl. Phys.* **B34**, 632 (1971). This article is reprinted in *Superstrings — The First 15 Years of Superstring Theory*, Ref. 7.
10. J. Wess and B. Zumino, *Nucl. Phys.* **B70**, 39 (1974). This article is reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987).
11. J. Wess and B. Zumino, *Phys. Lett.* **49B**, 52 (1974). This article is reprinted in *Supersymmetry*, Ref. 10.

11*a*. F. Gliozzi, J. Scherk, and D. Olive, *Nucl. Phys.* **B122**, 253 (1977).

12. Yu. A. Gol'fand and E. P. Likhtman, *JETP Letters* **13**, 323 (1971). This article is reprinted in *Supersymmetry*, Ref. 10.
13. D. V. Volkov and V. P. Akulov, *Phys. Lett.* **46B**, 109 (1973). This article is reprinted in *Supersymmetry*, Ref. 10.
14. M. Gell-Mann, Cal. Tech. Synchotron Laboratory Report CTSL–20 (1961), unpublished. This is reproduced along with other articles on

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

$SU(3)$ symmetry in M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).

15. R. Haag, J. T. Lopuszanski, and M. Sohnius, *Nucl. Phys.* **B88**, 257 (1975). This article is reprinted in *Supersymmetry*, Ref. 10.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.003

25

Supersymmetry Algebras

This chapter will develop the form of the supersymmetry algebra from first principles, following the treatment of Haag, Lopuszanski, and Sohnius.[1] As we shall see, under conditions in which the Coleman–Mandula theorem applies, this structure is almost uniquely fixed by the requirements of Lorentz invariance. The supermultiplet structure of one-particle states will then be deduced directly from the supersymmetry algebra.

25.1 Graded Lie Algebras and Graded Parameters

We saw in Section 2.2 how to express any continuous symmetry transformation in terms of a Lie algebra of linearly independent symmetry generators t_a that satisfy commutation relations $[t_a, t_b] = i\sum_c C^c_{ab} t_c$. In much the same way, supersymmetry is expressed in terms of symmetry generators t_a that form a *graded* Lie algebra,[2] embodied in commutation *and* anticommutation relations of the form

$$t_a t_b - (-1)^{\eta_a \eta_b} t_b t_a = i \sum_c C^c_{ab} t_c \ . \tag{25.1.1}$$

(The summation convention is suspended in this section.) Here η_a for each a is either $+1$ or 0 and is known as the *grading* of the generator t_a, and the C^c_{ab} are a set of numerical structure constants. Generators t_a for which $\eta_a = 1$ are called *fermionic*; the others, for which $\eta_a = 0$, are called *bosonic*. Eq. (25.1.1) provides commutation relations for bosonic operators with each other and with fermionic operators, but anticommutation relations for fermionic operators with each other. We will come back soon to the motivation for Eq. (25.1.1); for the moment we will just take a look at its consequences for the structure constants.

According to Eq. (25.1.1), the structure constants must satisfy the conditions

$$C^c_{ba} = -(-1)^{\eta_a \eta_b} C^c_{ab} \ . \tag{25.1.2}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

For any operator formed as a functional of field operators, the products of two bosonic or two fermionic operators are bosonic, and the products of a fermionic with a bosonic operator are fermionic, so that

$$C^c_{ab} = 0 \text{ unless } \eta^c = \eta^a + \eta^b \pmod 2 . \tag{25.1.3}$$

Also, for any operator formed in this way, the Hermitian adjoint of a bosonic or fermionic operator is, respectively, bosonic or fermionic. If the t_a are Hermitian operators, then the structure constants satisfy a reality condition

$$C^{c\,*}_{ab} = -C^c_{ba} . \tag{25.1.4}$$

The structure constants also satisfy a non-linear constraint, which follows from a super-Jacobi identity

$$(-1)^{\eta_c\eta_a}[[t_a,t_b\},t_c\} + (-1)^{\eta_a\eta_b}[[t_b,t_c\},t_a\} + (-1)^{\eta_b\eta_c}[[t_c,t_a\},t_b\} = 0 . \tag{25.1.5}$$

Here '[...}' denotes a commutator/anticommutator like that appearing on the left-hand side of Eq. (25.1.1), but here extended to any graded operators O, O', etc.

$$[O,O'\} \equiv OO' - (-1)^{\eta(O)\eta(O')}O'O = -(-1)^{\eta(O)\eta(O')}[O',O\} , \tag{25.1.6}$$

it now being understood that any product $O = t_at_bt_c\cdots$ of generators is given a grading $\eta(O) \equiv \eta_a + \eta_b + \eta_c + \cdots \pmod 2$. (To prove Eq. (25.1.5), it is sufficient to prove that the coefficients of $t_at_bt_c$ and $t_at_ct_b$ vanish, for then the symmetry of the left-hand side of Eq. (25.1.5) under the cyclic permutations $abc \to bca \to cab$ will ensure that the coefficients of all other products of generators also vanish. The coefficient of $t_at_bt_c$ in Eq. (25.1.5) is

$$(-1)^{\eta_c\eta_a} - (-1)^{\eta_a\eta_b}(-1)^{\eta_a(\eta_b+\eta_c)} = 0 ,$$

while the coefficient of $t_at_ct_b$ is

$$(-1)^{\eta_a\eta_b}(-1)^{\eta_b\eta_c}(-1)^{\eta_a(\eta_b+\eta_c)} - (-1)^{\eta_b\eta_c}(-1)^{\eta_c\eta_a} = 0 ,$$

completing the proof.) By inserting Eq. (25.1.1) in Eq. (25.1.5), we find the constraint

$$\sum_d (-1)^{\eta_c\eta_a}C^d_{ab}C^e_{dc} + \sum_d (-1)^{\eta_a\eta_b}C^d_{bc}C^e_{da} + \sum_d (-1)^{\eta_b\eta_c}C^d_{ca}C^e_{db} = 0 . \tag{25.1.7}$$

Of course, in the case where all generators are bosonic Eq. (25.1.5) is the usual Jacobi identity, and Eq. (25.1.7) is the usual non-linear condition (2.2.22) on the structure constants.

Eq. (25.1.1) can be taken as our starting point, but it can also be given a motivation like that given for ordinary Lie algebras in Section 2.2, as a necessary feature of finite continuous symmetry transformations. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

difference is that now these transformations depend on continuous *graded* parameters. A set of graded c-number parameters can be thought of as 'numbers,' including Grassmann parameters (see Section 9.5) as well as ordinary numbers, which satisfy the associative and distributive rules of arithmetic, but which instead of simply commuting satisfy relations

$$\alpha^a \beta^b = (-1)^{\eta_a \eta_b} \beta^b \alpha^a \,, \tag{25.1.8}$$

where α^a, β^a, ... are used to distinguish different values of the ath parameter, in the same way that in vector algebra we might let v^a and u^a denote the a-components of different values of some real vector. Again the ath graded parameter is given a grading η_a equal to $+1$ or 0 when α^a is a fermionic or bosonic parameter, respectively. That is, these parameters commute if either is bosonic, and anticommute if both are fermionic. The product $\alpha^a \beta^b \gamma^c \cdots$ of a set of graded parameters is given the grading $\eta_a + \eta_b + \eta_c + \cdots$ (mod 2); that is, such a product is fermionic if it involves an odd number of fermionic parameters, and is otherwise bosonic. With this grading, it is easy to see that products of graded parameters satisfy a commutation or anticommutation rule just like Eq. (25.1.8).

Consider a continuous transformation $T(\alpha)$, given by a formal power series in the graded parameters α^a:

$$T(\alpha) = 1 + \sum_a \alpha^a t_a + \sum_{ab} \alpha^a \alpha^b t_{ab} + \cdots \,, \tag{25.1.9}$$

where t_a, t_{ab}, etc. are a set of α-independent operator coefficients, not yet assumed to satisfy any algebraic relations like Eq. (25.1.1). Because the parameters α^a satisfy Eq. (25.1.8), the coefficients $t_{ab\cdots}$ must satisfy symmetry/antisymmetry conditions, such as

$$t_{ab} = (-1)^{\eta_a \eta_b} t_{ba} \,. \tag{25.1.10}$$

It is convenient also to assume that the transformation $T(\beta)$ commutes with any value α^a of any graded parameter, in which case the operator coefficients in (25.1.9) satisfy the conditions

$$\alpha^a t_b = (-1)^{\eta_a \eta_b} t_b \alpha^a \,, \tag{25.1.11}$$

$$\alpha^a t_{bc} = (-1)^{\eta_a (\eta_b + \eta_c)} t_{bc} \alpha^a \,. \tag{25.1.12}$$

That is, t_b and t_{bc} commute or anticommute with graded parameters as if they were graded parameters themselves, with grading η_b and $\eta_b + \eta_c$ (mod 2), respectively.

The other constraints on these operators follow from the requirement that the $T(\alpha)$ form a semi-group; that is, that the product of the operators for different values α and β of the graded parameters is itself a T-operator

$$T(\alpha) T(\beta) = T(f(\alpha, \beta)) \,, \tag{25.1.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

where $f^c(\alpha,\beta)$ is itself a formal power series in the graded parameters. Because $T(0)T(\beta) = T(\beta)$ and $T(\alpha)T(0) = T(\alpha)$, we must have

$$f^c(0,\beta) = \beta^c\,, \qquad f^c(\alpha,0) = \alpha^c\,, \tag{25.1.14}$$

and therefore the power series expansion for $f(\alpha,\beta)$ must take the form

$$f^c(\alpha,\beta) = \alpha^c + \beta^c + \sum_{ab} f^c_{ab}\,\alpha^a\,\beta^b + \cdots\,, \tag{25.1.15}$$

where f^c_{ab} is a set of ordinary (that is, bosonic) constants, and '$\cdots$' denotes terms of third or higher order in the graded parameters. In order for $f^c(\alpha,\beta)$ to be a graded parameter, it is necessary for each term in Eq. (25.1.15) to have the same grading, which implies that

$$f^c_{ab} = 0 \;\text{ unless }\; \eta^c = \eta^a + \eta^b \;(\text{mod } 2)\,. \tag{25.1.16}$$

Inserting the power series (25.1.9) and (25.1.15) into the product rule (25.1.13) gives

$$\begin{aligned}&\Big[1 + \sum_a \alpha^a t_a + \sum_{ab}\alpha^a\alpha^b t_{ab} + \cdots\Big]\Big[1 + \sum_a \beta^a t_a + \sum_{ab}\beta^a\beta^b t_{ab} + \cdots\Big]\\ &\quad = 1 + \sum_c \Big(\alpha^c + \beta^c + \sum_{ab} f^c_{ab}\alpha^a\beta^b + \cdots\Big)t_c\\ &\qquad + \sum_{cd}\Big(\alpha^c + \beta^c + \cdots\Big)\Big(\alpha^d + \beta^d + \cdots\Big)t_{cd} + \cdots\qquad.\end{aligned}$$

The coefficients of 1, α^a, β^a, $\alpha^a\alpha^b$, and $\beta^a\beta^b$ match on both sides of this equation, but the condition that the coefficients of $\alpha^a\beta^b$ are the same yields the non-trivial relation

$$(-1)^{\eta_a\eta_b}t_at_b = \sum_c f^c_{ab}t_c + t_{ab} + (-1)^{\eta_a\eta_b}t_{ba} = \sum_c f^c_{ab}t_c + 2t_{ab}\,. \tag{25.1.17}$$

(The sign factor on the left-hand side arises from the interchange of t_a and β^b.) Together with higher-order relations of the same sort, this allows us to calculate the whole function (25.1.9) if we know the generators t_a and the group-composition function $f^a(\alpha,\beta)$. But for this to be possible, t_a must satisfy a constraint. Using Eq. (25.1.10), the difference or sum of Eq. (25.1.17) and the same equation with a and b interchanged yields the Lie superalgebra relations Eq. (25.1.1), with structure constants given by

$$i\,C^c_{ab} = (-1)^{\eta_a\eta_b}f^c_{ab} - f^c_{ba}\,. \tag{25.1.18}$$

Also, Eq. (25.1.3) follows immediately from Eqs. (25.1.16) and (25.1.18).

The complex conjugate α^* of an anticommuting c-number α is defined so that the Hermitian adjoint of the product of α and an arbitrary operator $\mathcal{O}$ is

$$(\alpha\mathcal{O})^* = \mathcal{O}^*\alpha^*\,. \tag{25.1.19}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

It follows that products of c-numbers behave the same under complex conjugation as operators do under Hermitian conjugation:

$$(\alpha\beta)^* = \beta^*\alpha^* , \tag{25.1.20}$$

and that α^* has the same grading as α.

The graded Lie algebras of importance to physics are severely restricted by spacetime symmetries. We now turn to a consideration of these restrictions.

25.2 Supersymmetry Algebras

Consider a general graded Lie algebra of symmetry generators that commute with the S-matrix. If Q is any of the fermionic symmetry generators, then so will be $U^{-1}(\Lambda)\,Q\,U(\Lambda)$, where $U(\Lambda)$ is the quantum mechanical operator corresponding to an arbitrary homogeneous Lorentz transformation $\Lambda^{\mu}{}_{\nu}$. Therefore $U^{-1}(\Lambda)\,Q\,U(\Lambda)$ is a linear combination of the complete set of fermionic symmetry generators, and hence this set of generators must furnish a representation of the homogeneous Lorentz group. The individual generators may therefore be classified according to the irreducible representation of the homogeneous Lorentz group to which they belong.

As described in Section 5.6, the representations of the homogeneous Lorentz group furnished by any set of operators can be specified by giving their commutation relations with generators **A** and **B**, defined by

$$\mathbf{A} \equiv \tfrac{1}{2}\left(\mathbf{J} + i\mathbf{K}\right) , \qquad \mathbf{B} \equiv \tfrac{1}{2}\left(\mathbf{J} - i\mathbf{K}\right) , \tag{25.2.1}$$

where **J** and **K** are the Hermitian generators of rotations and boosts, respectively. These satisfy the commutation relations

$$[A_i, A_j] = \sum_k \epsilon_{ijk} A_k , \quad [B_i, B_j] = \sum_k \epsilon_{ijk} B_k , \quad [A_i, B_j] = 0 , \tag{25.2.2}$$

where i, j, and k run over the values 1, 2, 3, and ϵ_{ijk} is totally antisymmetric, with $\epsilon_{123} = +1$. Thus representations of the homogeneous Lorentz group are labelled, like states with two independent spins, by a pair of integers or half-integers A and B, with elements of the representation labelled by a pair of indices a and b, which run by unit steps from $-A$ to $+A$ and from $-B$ to $+B$, respectively. More specifically, a set of $(2A+1)(2B+1)$ operators Q^{AB}_{ab} that form an (A, B) representation of the homogeneous Lorentz group satisfies the commutation relations

$$[\mathbf{A}, Q^{AB}_{ab}] = -\sum_{a'} \mathbf{J}^{(A)}_{aa'} Q^{AB}_{a'b} , \qquad [\mathbf{B}, Q^{AB}_{ab}] = -\sum_{b'} \mathbf{J}^{(B)}_{bb'} Q^{AB}_{ab'} , \tag{25.2.3}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

where $\mathbf{J}^{(j)}$ is the spin three-vector matrix for angular momentum j:

$$\left(J_1^{(j)} \pm iJ_2^{(j)}\right)_{\sigma'\sigma} = \delta_{\sigma',\sigma\pm1}\sqrt{(j \mp \sigma)(j \pm \sigma + 1)}\,, \qquad \left(J_3^{(j)}\right)_{\sigma'\sigma} = \delta_{\sigma'\sigma}\sigma\,. \tag{25.2.4}$$

From Eq. (25.2.4), it follows that*

$$-\left(\mathbf{J}^{(j)}\right)^*_{\sigma',\sigma} = (-1)^{\sigma'-\sigma}\left(\mathbf{J}^{(j)}\right)_{-\sigma',-\sigma}\,. \tag{25.2.5}$$

Thus if Q^j_σ are a set of operators that transform according to the spin j representation of the rotation group, then so are $(-1)^{j-\sigma}Q^{j*}_{-\sigma}$. Also, Eq. (25.2.1) shows that $\mathbf{A}^* = \mathbf{B}$. By taking Hermitian adjoints in Eq. (25.2.3), we see that the Hermitian adjoint Q^{AB*}_{ab} of operators that transform according to the (A,B) representation of the homogeneous Lorentz group are related by a similarity transformation to operators $\bar{Q}^{BA}_{ba}$ that transform according to the (B,A) representation:

$$Q^{AB*}_{ab} = (-1)^{A-a}(-1)^{B-b}\,\bar{Q}^{BA}_{-b,-a}\,. \tag{25.2.6}$$

The Haag–Lopuszanski–Sohnius theorem[1] states in part that the fermion symmetry generators can only belong to the $(0,1/2)$ and $(1/2,0)$ representations. As we have seen, the Hermitian adjoint of a $(0,1/2)$ or $(1/2,0)$ operator is a linear combination respectively of $(1/2,0)$ or $(0,1/2)$ operators, so the complete set of fermionic symmetry operators may be divided into $(0,1/2)$ generators Q_{ar} (with the superscript $0\,\frac{1}{2}$ omitted) and their $(1/2,0)$ Hermitian adjoints Q^*_{ar}, where a is a spinor index running over the values $\pm1/2$, and r is used to distinguish different two-component generators with the same Lorentz transformation properties.** The theorem further states that the fermionic generators may be defined so as to satisfy the anticommutation relations

$$\{\mathrm{Q}_{ar}\,,\,\mathrm{Q}^*_{bs}\} = 2\delta_{rs}\,\sigma^\mu_{ab}\,P_\mu\,, \tag{25.2.7}$$

$$\{\mathrm{Q}_{ar}\,,\,\mathrm{Q}_{bs}\} = e_{ab}\,Z_{rs}\,, \tag{25.2.8}$$

where P_μ is the four-momentum operator, the $Z_{rs} = -Z_{sr}$ are bosonic symmetry generators, and σ_μ and e are the 2×2 matrices (with rows and

* We use an asterisk for the Hermitian adjoint of an operator or the complex conjugate of a number. The dagger † will be used for the transpose of the matrix formed from the Hermitian adjoints of operators or complex conjugates of numbers.

** We are using roman instead of italic letters for the two-component Weyl spinors Q_{ar}, to distinguish them from four-component Dirac spinors that will be introduced later in this section. There is a notation due to van der Waerden, according to which a $(0,1/2)$ operator like Q is written with dotted indices, as $\mathrm{Q}_{\dot{a}}$, while $(1/2,0)$ operators are written with undotted indices. We will not use this notation here, but will instead explicitly indicate which two-component spinors transform according to the $(0,1/2)$ or $(1/2,0)$ representations of the homogeneous Lorentz group.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

columns labelled $+1/2$, $-1/2$):

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad e = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{25.2.9}$$

Finally, the fermionic generators commute with energy and momentum:

$$[P_\mu, Q_{ar}] = [P_\mu, Q^*_{ar}] = 0, \tag{25.2.10}$$

and the Z_{rs} and Z^*_{rs} are a set of central charges of this algebra, in the sense that

$$\begin{aligned} 0 &= [Z_{rs}, Q_{at}] = [Z_{rs}, Q^*_{at}] = [Z_{rs}, Z_{tu}] = [Z_{rs}, Z^*_{tu}] \\ &= [Z^*_{rs}, Q_{at}] = [Z^*_{rs}, Q^*_{at}] = [Z^*_{rs}, Z^*_{tu}]. \end{aligned} \tag{25.2.11}$$

To prove these results, let us start by considering non-vanishing fermionic symmetry generators that belong to some (A, B) irreducible representation of the homogeneous Lorentz group, and that therefore may be labelled Q^{AB}_{ab}, where a and b run by unit steps from $-A$ to $+A$ and from $-B$ to $+B$, respectively. As already mentioned, the Hermitian adjoint is related to operators belonging to the (B, A) representation by Eq. (25.2.6), so the anticommutator of these operators must take the form

$$\{Q^{AB}_{ab}, Q^{AB*}_{a'b'}\} = (-1)^{A-a'}(-1)^{B-b'} \sum_{C=|A-B|}^{A+B} \sum_{D=|A-B|}^{A+B} \sum_{c=-C}^{C} \sum_{d=-D}^{D} \times C_{AB}(Cc; a, -b')\, C_{AB}(Dd; -a'b)\, X^{CD}_{cd}, \tag{25.2.12}$$

where $C_{AB}(j\sigma; ab)$ is the usual Clebsch–Gordan coefficient for coupling spins A and B to form spin j, and X^{CD}_{cd} is the (c, d)-component of an operator that transforms according to the (C, D) representation of the homogeneous Lorentz group. Using the well-known unitarity properties of the Clebsch–Gordan coefficients, we may express the operator X^{CD}_{cd} in terms of these anticommutators:

$$X^{CD}_{cd} = \sum_{a=-A}^{A} \sum_{b=-B}^{B} \sum_{a'=-A}^{A} \sum_{b'=-B}^{B} (-1)^{A-a'}(-1)^{B-b'} \times C_{AB}(Cc; a - b')\, C_{AB}(Dd; -a'b)\, \{Q^{AB}_{ab}, Q^{AB*}_{a'b'}\}. \tag{25.2.13}$$

Not all of these operators are necessarily non-zero. But the only non-zero Clebsch–Gordan coefficients $C_{AB}(j\sigma, ab)$ for $j = \sigma = A + B$ and $j = -\sigma = A + B$ are the ones for $a = A$, $b = B$ and $a = -A$, $b = -B$, respectively, which both have the value unity, so by taking $C = D = c = -d = A + B$ in Eq. (25.2.13), we find that

$$X^{A+B,A+B}_{A+B,-A-B} = (-1)^{2B} \{Q^{AB}_{A,-B}, Q^{AB*}_{A,-B}\}. \tag{25.2.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

This cannot vanish unless $Q^{AB}_{A,-B} = 0$, which would imply (taking commutators with the 'lowering' operators $A_1 - iA_2$ and 'raising' operators $B_1 + iB_2$) that all the Q^{AB}_{ab} vanish. Hence if there are any non-vanishing (A, B) fermionic generators, then their anticommutators with their adjoints must at least involve non-zero bosonic symmetry generators belonging to the $(A + B, A + B)$ representation.

Now, the Coleman–Mandula theorem tells us that the bosonic symmetry generators consist of the $(1/2, 1/2)$ generators P_μ of translations, the $(1,0) + (0,1)$ generators $J_{\mu\nu}$ of proper Lorentz transformations, and perhaps the $(0,0)$ generators T_A of various internal symmetries. (Recall that symmetric traceless tensors of rank N transform according to the representation $(N/2, N/2)$, antisymmetric tensors of rank 2 transform according to the representation $(1,0) + (0,1)$, while Dirac fields transform according to the representation $(1/2,0) + (0,1/2)$.) Hence the fermionic symmetry generators can only belong to representations (A, B) with $A + B \leq 1/2$. These operators turn bosons into fermions and vice-versa, so they cannot be scalars, leaving only the $(1/2,0)$ and $(0,1/2)$ representations, as was to be shown. Labelling the linearly independent $(0,1/2)$ fermionic generators as Q_{ar}, the anticommutator $\{Q_{ar}, Q^*_{bs}\}$ belongs to the representation $(0,1/2) \times (1/2,0) = (1/2,1/2)$, and therefore must be proportional to the only $(1/2,1/2)$ bosonic symmetry generator, the momentum four-vector P_μ. Lorentz invariance dictates that the form of this relation must be

$$\{Q_{ar}, Q^*_{bs}\} = 2N_{rs}\,\sigma^\mu_{ab}\,P_\mu\,, \tag{25.2.15}$$

where N_{rs} is a numerical matrix.

To see this, we use the isomorphism of the Lorentz group (or more properly its covering group) with the group $SL(2,C)$ of two-dimensional unimodular complex matrices λ, discussed in Section 2.7. The effect of a Lorentz transformation $\Lambda^\mu{}_\nu$ on the $(0,1/2)$ fermionic generators is

$$U^{-1}(\Lambda)\,Q_{ar}\,U(\Lambda) = \sum_b \lambda_{ab}\,Q_{br}\,, \tag{25.2.16}$$

where Λ is the Lorentz transformation defined by

$$\lambda\,\sigma_\mu\,\lambda^\dagger = \Lambda^\nu{}_\mu \sigma_\nu\,. \tag{25.2.17}$$

We can check that Eq. (25.2.16) applies for $(0,1/2)$ operators by noting that for an infinitesimal Lorentz transformation $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \omega^\mu{}_\nu$ with $\omega_{\mu\nu} = -\omega_{\nu\mu}$, Eq. (25.2.17) is satisfied for

$$\lambda = 1 + \tfrac{1}{2}\Big[\tfrac{1}{2}\,i\,\epsilon_{ijk}\omega_{ij} + \omega_{k0}\Big]\sigma_k\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

while[†]

$$U(\Lambda) = 1 + \tfrac{1}{2} i \omega_{\mu\nu} J^{\mu\nu} = 1 + \tfrac{1}{2} i \epsilon_{ijk} \omega_{ij} J_k - i \omega_{i0} K_i \,.$$

(Repeated latin indices i, j, k are summed over values 1, 2, 3.) In this case, by equating the coefficients of ω_{ij} and ω_{i0} in Eq. (25.2.16), we find that

$$[\mathbf{J}, Q_a] = -\tfrac{1}{2} \sum_b \boldsymbol{\sigma}_{ab} Q_b \,, \qquad [\mathbf{K}, Q_a] = -\tfrac{1}{2} i \sum_b \boldsymbol{\sigma}_{ab} Q_b \,,$$

or, equivalently,

$$[\mathbf{B}, Q_a] = -\tfrac{1}{2} \sum_b \boldsymbol{\sigma}_{ab} Q_b \,, \qquad [\mathbf{A}, Q_a] = 0 \,,$$

which shows that an operator satisfying Eq. (25.2.16) belongs to the $(0, 1/2)$ representation. Now, the σ_μ form a complete set of 2×2 matrices, so we can put the anticommutator $\{Q_{ar}, Q^*_{bs}\}$ in the form $N^\mu_{rs} (\sigma_\mu)_{ab}$, where N^μ is some matrix of operators. Eqs. (25.2.16) and (25.2.17) show that these operators are four-vectors, in the sense that $U^{-1}(\Lambda) N^\mu U(\Lambda) = \Lambda^\mu{}_\nu N^\nu$, and so by the Coleman–Mandula theorem they must be proportional to P^μ, the only four-vector of bosonic symmetry operators. Setting $N^\mu_{rs} = 2P^\mu N_{rs}$ then gives Eq. (25.2.15).

Now we will apply a linear transformation to the Q_{ar}, to put their anticommutators in the form (25.2.7). For this purpose, we need to establish that the matrix N_{rs} is Hermitian and positive-definite. That it is Hermitian follows immediately by taking the Hermitian adjoint of Eq. (25.2.15). To see that it is positive-definite, recall that the Q_{ar} are taken linearly independent, so for any non-zero linear combination $Q \equiv \sum_r d_a c_r Q_{ar}$ there must be some state $|\Psi\rangle$ that Q does not annihilate. Taking the expectation value of Eq. (25.2.15) in this state implies that

$$2\langle\Psi| \sum_{ab} \sigma^\mu_{ab} P_\mu d_a d^*_b |\Psi\rangle \sum_{rs} c_r c^*_s N_{rs} = \langle\Psi|\{Q, Q^*\}|\Psi\rangle > 0 \,.$$

This shows immediately that $\sum_{rs} c_r c^*_s N_{rs}$ cannot vanish for any c_r that are not all zero, so N_{rs} is either positive-definite or negative-definite. The operator $\sum_{ab} (\sigma_\mu)_{ab} P^\mu d_a d^*_b$ is positive on the space of physical states with $-P^\mu P_\mu \geq 0$ and $P^0 > 0$, so the matrix N_{rs} must be positive-definite.[††]

[†] Here K_i is defined as J_{i0}. There was a mistake in the first two printings of Volume I: K_i was defined as J^{i0} in Sections 2.4, 3.3, and 3.5, but as J_{i0} in Sections 5.6 and 5.9, with **A** and **B** throughout given by Eq. (25.2.1).

[††] This argument may be turned around. Assuming a supersymmetry with N_{rs} positive-definite, as in Eq. (25.2.7), we can deduce that $P^0 > 0$ for all states.[2] However, this conclusion is not valid when gravitation is taken into account, and unless gravitation is taken into account a shift in the energy of all states by the same amount would have no physical effects.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

We may now define new fermionic generators

$$Q'_{ar} \equiv \sum_s N_{rs}^{-1/2} Q_{as} ,$$

for which the anticommutators take the form

$$\{Q'_{ar}, Q'^{*}_{bs}\} = 2\delta_{rs}\, \sigma^{\mu}_{ab}\, P_{\mu} .$$

From now on we shall assume that the fermionic generators are defined in this way, and drop the primes, so that Eq. (25.2.7) is satisfied.

Next we must show that the Q_{ar} commute with the momentum four-vector P_μ. The commutator of a $(1/2, 1/2)$ operator like P_μ with a $(0, 1/2)$ operator like Q can only be a $(1/2, 0)$ or $(1/2, 1)$ operator, but we have seen that there are no $(1/2, 1)$ symmetry generators, so the commutator of P_μ with Q can only be proportional to the $(1/2, 0)$ symmetry generator Q^*. Lorentz invariance requires this relation to take the form

$$[\mathcal{M}_{ab}, Q_{cr}] = \sum_s e_{ac}\, K_{rs}\, Q^{*}_{bs} , \tag{25.2.18}$$

where K is a numerical matrix, and $\mathcal{M}$ is the matrix of operators

$$\mathcal{M} \equiv \sigma_\mu P^\mu . \tag{25.2.19}$$

(The matrix e_{ac} is the Clebsch–Gordan coefficient for coupling two spins $1/2$ to give zero spin.) It is straightforward then to calculate that

$$[\mathcal{M}_{-\frac{1}{2}-\frac{1}{2}}, [\mathcal{M}_{-\frac{1}{2}-\frac{1}{2}}, \{Q_{\frac{1}{2}r}, Q^{*}_{\frac{1}{2}s}\}]] = -4(\mathcal{M})_{-\frac{1}{2}-\frac{1}{2}}(KK^{\dagger})_{rs} . \tag{25.2.20}$$

Using Eq. (25.2.7), the left-hand side is a linear combination of multiple commutators $[P_\mu, [P_\nu, P_\lambda]]$, all of which vanish, while $(\mathcal{M})_{-1/2-1/2}$ is non-zero for generic momenta, so $KK^\dagger = 0$, and therefore $K = 0$, which with Eq. (25.2.18) shows that $[P_\mu, Q_{ar}] = 0$. The complex conjugate gives $[P_\mu, Q^*_{ar}] = 0$.

Now we can take up the anticommutator of two Qs. The anticommutator of two $(0, 1/2)$ symmetry generators must be a linear combination of $(0, 1)$ and $(0, 0)$ symmetry generators. The Coleman–Mandula theorem tells us that the only $(0, 1)$ symmetry generators are linear combinations of the generators $J_{\nu\lambda}$ of proper homogeneous Lorentz transformations, but since the Qs commute with P_μ so must their anticommutators, and Eq. (2.4.13) shows that no linear combination of the $J_{\nu\lambda}$ commutes with P_μ. This leaves just $(0, 0)$ operators, which commute with both P_μ and $J_{\mu\nu}$. Lorentz invariance then requires the anticommutators of the Qs with each other to take the form of Eq. (25.2.8). The internal symmetry generators Z_{rs} are antisymmetric in r and s because the whole expression must be symmetric under interchange of r and a with s and b, and the matrix e_{ab} is antisymmetric in a and b.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

It now only remains to show that the Zs are central charges. It follows immediately from Eqs. (25.2.8) and (25.2.10) that

$$[P_\mu, Z_{rs}] = 0 . \tag{25.2.21}$$

Next consider the generalized Jacobi identity (25.1.5) involving two Qs and a Q^*:

$$0 = [\{Q_{ar}, Q_{bs}\}, Q^*_{ct}] + [\{Q_{bs}, Q^*_{ct}\}, Q_{ar}] + [\{Q^*_{ct}, Q_{ar}\}, Q_{bs}] .$$

Eqs. (25.2.7) and (25.2.10) show that the second and third terms vanish, so

$$[Z_{rs}, Q^*_{ct}] = 0 . \tag{25.2.22}$$

Finally, consider the generalized Jacobi identity for a Z, a Q, and a Q^*:

$$0 = -[Z_{rs}, \{Q_{at}, Q^*_{bu}\}] + \{Q^*_{bu}, [Z_{rs}, Q_{at}]\} - \{Q_{at}, [Q^*_{bu}, Z_{rs}]\} .$$

The first and third terms vanish because of Eqs. (25.2.21) and (25.2.22), respectively, so we are left with the second term

$$\{Q^*_{bu}, [Z_{rs}, Q_{at}]\} = 0 . \tag{25.2.23}$$

Now, $[Z_{rs}, Q_{at}]$ is a $(0, 1/2)$ symmetry generator, so it must be a linear combination of the Qs:

$$[Z_{rs}, Q_{at}] = \sum_u M_{rstu} Q_{au} . \tag{25.2.24}$$

Eq. (25.2.23) then reads

$$\sigma^\mu_{ab} P_\mu M_{rstu} = 0 ,$$

for all a, b, r, s, t, and u. Since the operator $\sigma^\mu_{ab} P_\mu$ is not zero, we conclude that $M_{rstu} = 0$, so that

$$[Z_{rs}, Q_{at}] = 0 . \tag{25.2.25}$$

Using the anticommutation relation (25.2.8) and its adjoint together with the commutation relations (25.2.22) and (25.2.25) and their adjoints then gives

$$[Z_{rs}, Z_{tu}] = [Z_{rs}, Z^*_{tu}] = [Z^*_{rs}, Z^*_{tu}] = 0 , \tag{25.2.26}$$

which finishes the proof of Eq. (25.2.11), and with it the proof of the Haag–Lopuszanski–Sohnius theorem.

Of course, the fact that the Z_{rs} are central charges of the supersymmetry algebra does not rule out the possibility that there may be *other* Abelian or non-Abelian internal symmetries. Let T_A span the complete Lie algebra of bosonic internal symmetries. Then $[T_A, Q_{ar}]$ is a $(0, 1/2)$ symmetry generator, so it must be a linear combination of the Qs:

$$[T_A, Q_{ar}] = -\sum_s (t_A)_{rs} Q_{as} . \tag{25.2.27}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

From the Jacobi identity for two Ts and a Q, we learn that the t_A matrices furnish a representation of the internal symmetry algebra

$$[t_A, t_B] = i \sum_C C_{AB}^C t_C , \tag{25.2.28}$$

where the coefficients C_{AB}^C are the structure constants of the internal symmetry algebra

$$[T_A, T_B] = i \sum_C C_{AB}^C T_C . \tag{25.2.29}$$

The Z_{rs} will then be central charges not only of the superalgebra consisting of the Qs, Q*s, P_μ, Zs, and Z^*s, but also of the larger symmetry superalgebra that in addition contains all the T_A. To see this, note from Eqs. (25.2.27) and (25.2.8) that

$$[T_A, Z_{rs}] = -\sum_{r'} (t_A)_{rr'} Z_{r's} - \sum_{s'} (t_A)_{ss'} Z_{rs'} ,$$

so the Z_{rs} form an *invariant* Abelian subalgebra of the whole bosonic symmetry algebra. But recall that in proving the Coleman–Mandula theorem we found that the complete Lie algebra of internal bosonic symmetries, which in our case is spanned by the T_A, is isomorphic to a direct sum of a compact semi-simple Lie algebra and several $U(1)$ algebras. The only invariant Abelian subalgebras of such a Lie algebra are spanned by $U(1)$ generators, so the Z_{rs} must be $U(1)$ generators, and hence commute with all the T_A.

Even though the Zs commute with all symmetry generators, they are not just numbers; they are quantum operators, whose value may vary from state to state. In fact, the Zs must obviously take the value zero for a supersymmetric vacuum state, which is annihilated by all supersymmetry generators, but they need not vanish in general. In Section 27.9 we will see how the Zs may be calculated in gauge theories with extended supersymmetry.

In the absence of central charges, the supersymmetry algebra (25.2.7), (25.2.8) is invariant under a group $U(N)$ of internal symmetries

$$Q_{ar} \to \sum_s V_{rs} Q_{as} , \tag{25.2.30}$$

with V_{rs} an $N \times N$ unitary (not necessarily unimodular) matrix. This is known as *R-symmetry*. This symmetry may or may not be a good symmetry of the action, and if it is then it may be violated by anomalies, or it may be spontaneously broken, or it may be a good symmetry of nature.

A supersymmetry algebra where r, s, etc. run over $N > 1$ values is known as an *N-extended supersymmetry*. Where there is just one Q, the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

condition $Z_{rs} = -Z_{sr}$ tells us that the Zs vanish, yielding a simpler form of the anticommutation relations

$$\{Q_a, Q_b^*\} = 2\sigma_{ab}^{\mu} P_{\mu} , \tag{25.2.31}$$

$$\{Q_a, Q_b\} = 0 . \tag{25.2.32}$$

This is known as the case of *simple supersymmetry*, or $N = 1$ supersymmetry. In this case the R-symmetry transformations are $U(1)$ phase transformations

$$Q_a \to \exp(i\varphi) Q_a , \tag{25.2.33}$$

with φ a real phase.

For several purposes it is convenient to combine the $(0, 1/2)$ operators Q_{ar} together with $(1/2, 0)$ operators, which according to Eq. (25.2.6) may be taken as $e_{ab}Q_{br}^*$, into four-component Majorana spinor generators $Q_{\alpha r}$, defined as

$$Q_r \equiv \begin{pmatrix} eQ_r^* \\ Q_r \end{pmatrix} , \tag{25.2.34}$$

or more explicitly

$$Q_{1r} = Q_{-\frac{1}{2}r}^* , \quad Q_{2r} = -Q_{\frac{1}{2}r}^* , \quad Q_{3r} = Q_{\frac{1}{2}r} , \quad Q_{4r} = Q_{-\frac{1}{2}r} .$$

This is a Majorana spinor in the sense that

$$Q_r = -\beta\epsilon\gamma_5 Q_r^* ,$$

where β, ϵ, and γ_5 are 4×4 matrices that may be written as the 2×2 block matrices:

$$\beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \epsilon = \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix} \qquad \gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

(The properties of Majorana spinors are reviewed in the appendix to Chapter 26.) The form (25.2.34) is chosen in accordance with the usual notation for the four-component Dirac representation of the homogeneous Lorentz group, in which according to Eq. (5.4.4) the rotation and boost generators are represented according to Eqs. (5.4.19) and (5.4.20) by

$$\mathscr{J}_i = \frac{1}{2} \begin{bmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{bmatrix} , \qquad \mathscr{K}_i = -\frac{i}{2} \begin{bmatrix} \sigma_i & 0 \\ 0 & -\sigma_i \end{bmatrix} . \tag{25.2.35}$$

With Eq. (25.2.1), this shows that the operators **A** and **B** act only on the top two and bottom two components of the Dirac spinor, respectively, which is why we use the $(0, 1/2)$ operators Q_{ar} as the bottom rather than the top components in Eq. (25.2.34).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

In this four-component notation, the fundamental anticommutation relations (25.2.31) and (25.2.32) for simple supersymmetry read

$$\{Q\,,\,\overline{Q}\} = 2\begin{pmatrix} 0 & -e\,(\sigma_\mu P^\mu)^{\mathrm{T}}\,e \\ \sigma_\mu P^\mu & 0 \end{pmatrix} = -2i\,P_\mu\,\gamma^\mu\,. \tag{25.2.36}$$

Our notation for Dirac matrices is reviewed in the Preface to this volume; here we need only recall that

$$\gamma^0 = -i\beta = -i\begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}\,, \qquad \boldsymbol{\gamma} = -i\begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}\,, \tag{25.2.37}$$

while $e\,\boldsymbol{\sigma}^{\mathrm{T}}e = \boldsymbol{\sigma}$, $e\sigma_0 e = -\sigma_0$, and as usual $\overline{Q} \equiv Q^\dagger\beta$. The presence of central charges will change this formula in the case of extended supersymmetry; instead of Eq. (25.2.36), we have

$$\{Q_r\,,\,\overline{Q}_s\} = -2i\,P_\mu\,\gamma^\mu\delta_{rs} + \left(\frac{1+\gamma_5}{2}\right)Z^*_{sr} + \left(\frac{1-\gamma_5}{2}\right)Z_{rs}\,. \tag{25.2.38}$$

The analysis presented here for the case of four spacetime dimensions will be repeated, in a somewhat less explicit form, for general spacetime dimensions in Chapter 32. As we will see there, the supersymmetry generators always belong to the fundamental spinor representation of the higher-dimensional Lorentz group, even in theories with extended objects that allow the construction of bosonic symmetry generators other than those allowed by the Coleman–Mandula theorem.

* * *

In theories of massless particles that are invariant under the conformal symmetry algebra (24.B.34)–(24.B.35) there are two additional bosonic symmetry generators D and K_μ that can appear on the right-hand side of the supersymmetry anticommutation relations. These new generators have the Lorentz transformation properties of a scalar and a vector, respectively, just like Z_{rs} and P_μ, so once again the fermionic generators must belong to the fundamental $(1/2,0)$ spinor representation of the Lorentz algebra, and its Hermitian adjoint, the $(0,1/2)$ representation. It is convenient to classify all generators also according to their commutators with the dilation generator D; an operator X is said to have dimensionality a if

$$[X\,,\,D] = iaX\,. \tag{25.2.39}$$

Inspection of Eq. (24.B.34) shows that the bosonic symmetry generators $J^{\mu\nu}$, P^μ, K^μ, and D have dimensionalities 0, +1, −1, and 0, respectively. Also, the generators of any Lie group of internal symmetries have dimensionality 0. The anticommutator of a fermionic generator of dimensionality a with its adjoint is a positive-definite bosonic operator of dimensionality $2a$, so since the only positive-definite bosonic symmetry generators are linear combinations of the components of P_μ and K_μ, the only fermionic

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

symmetry generators have dimensionalities $+1/2$ and $-1/2$. The $(0,1/2)$ fermionic symmetry generators of dimensionality $1/2$ and their adjoints may again be assembled into Majorana spinors $Q_{r\alpha}$, with

$$\{Q_{r\alpha}\,,\,\bar{Q}_{s\beta}\} = -2iP_\mu(\gamma^\mu)_{\alpha\beta}\delta_{rs}\,, \tag{25.2.40}$$

$$[P_\mu\,,\,Q_{r\alpha}] = 0\,, \tag{25.2.41}$$

$$[D\,,\,Q_{r\alpha}] = -\tfrac{1}{2}iQ_{r\alpha}\,. \tag{25.2.42}$$

(Note that central charges are not allowed here, because their dimensionality would be 0, not $+1$.) The commutators of the K_μ with the $Q_{r\alpha}$ are linear combinations of Majorana fermionic symmetry generators $Q^\#_{r\alpha}$, which Lorentz invariance allows us to write in the form

$$[K^\mu\,,\,Q_{r\alpha}] = i(\gamma^\mu)_{\alpha\beta}Q^\#_{r\beta}\,. \tag{25.2.43}$$

(An arbitrary factor on the right-hand side has been absorbed into the normalization of $Q^\#_{r\beta}$. The phase of the right-hand side is chosen so that $Q^\#_{r\beta}$ will satisfy the standard reality condition (26.A.2) for Majorana spinors.) The $Q^\#_{r\beta}$ have dimensionality $+1/2-1=-1/2$, so

$$[D\,,\,Q^\#_{r\alpha}] = +\tfrac{1}{2}iQ^\#_{r\alpha}\,. \tag{25.2.44}$$

Taking the commutator of Eq. (25.2.43) with P^ν and using the commutator of K^μ with P^ν given by Eq. (24.B.34) gives

$$[P^\nu\,,\,Q^\#_{r\alpha}] = -i(\gamma^\nu)_{\alpha\beta}Q_{r\beta}\,. \tag{25.2.45}$$

We see that the Qs and $Q^\#$s are paired. By taking the commutator of the anticommutation relation (25.2.40) with K_μ we find the anticommutators of the $Q^\#$s and Qs:

$$\{Q^\#_{r\alpha}\,,\,\bar{Q}_{s\beta}\} = 2iD\delta_{rs}\delta_{\alpha\beta} + 2J_{\mu\nu}\delta_{rs}\mathscr{J}^{\mu\nu}_{\alpha\beta} + O_{rs}\delta_{\alpha\beta} + O'_{rs}(\gamma_5)_{\alpha\beta}\,, \tag{25.2.46}$$

where $\mathscr{J}^{\mu\nu} \equiv -i[\gamma^\mu\,,\,\gamma^\nu]/4$, and O_{rs} and O'_{rs} are Lorentz invariant operators of zero dimensionality with

$$O_{rs} = -O_{sr}\,, \qquad\qquad O'_{rs} = +O'_{sr}\,. \tag{25.2.47}$$

Taking the commutator of Eq. (25.2.43) with K_ν and using the fact that $[K_\nu, K_\mu] = 0$, we find that $(\gamma^\mu)_{\alpha\beta}[K^\nu, Q_{r\beta}]$ is symmetric in μ and ν, which with a little algebra tells us that

$$[K^\nu, Q_{r\beta}] = 0\,. \tag{25.2.48}$$

Also, taking the commutator of Eq. (25.2.46) with K_ν gives

$$\{Q^\#_{r\alpha}\,,\,\overline{Q^\#}_{s\beta}\} = +2iK_\mu(\gamma^\mu)_{\alpha\beta}\delta_{rs}\,. \tag{25.2.49}$$

Finally, taking the commutator of Eq. (25.2.46) with $Q_{t\gamma}$ shows that O_{rs} and O'_{rs} act as the generators of the R-symmetry group $U(N)$, with the left- and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

right-handed parts of $Q_{r\alpha}$ transforming according to the representations $\mathbf{N}$ and $\bar{\mathbf{N}}$, respectively, while the P_μ, K_μ, and D are $U(N)$-invariant. The $U(N)$ commutation relations of these generators with each other and with other generators together with Eqs. (24.B.34), (24.B.35), (25.2.40)–(25.2.49) and the commutators of $J_{\mu\nu}$ and D with the various generators constitute the *superconformal algebra.* One of the outstanding differences between this algebra and that of ordinary simple or N-extended supersymmetry is that the $U(N)$ symmetry is not merely an outer automorphism of the algebra that may or may not be a symmetry of the action — it is a part of the superconformal algebra, which therefore must be a symmetry of any action that is supersymmetric and conformally invariant.

25.3 Space Inversion Properties of Supersymmetry Generators

In theories that respect the conservation of parity, the result $\mathsf{P}^{-1}\mathrm{Q}_{ar}\mathsf{P}$ of acting on the fermionic symmetry generator Q_{ar} with the parity operator P must also be a fermionic symmetry generator. Since J_i and K_i are respectively even and odd under space inversion, Eq. (25.2.1) shows that the effect of acting on A_i with the parity operator is

$$\mathsf{P}^{-1}A_i\mathsf{P} = B_i\,. \tag{25.3.1}$$

According to Eq. (25.2.3), the definition of Q_{ar} as a $(0,1/2)$ operator means that

$$[B_i\,,\,\mathrm{Q}_{ar}] = -\tfrac{1}{2}\sum_b \left(\sigma_i\right)_{ab}\mathrm{Q}_{br}\,, \qquad [A_i\,,\,\mathrm{Q}_{ar}] = 0\,. \tag{25.3.2}$$

Applying the parity operator yields

$$[A_i\,,\,\mathsf{P}^{-1}\mathrm{Q}_{ar}\mathsf{P}] = -\tfrac{1}{2}\sum_b \left(\sigma_i\right)_{ab}\mathsf{P}^{-1}\mathrm{Q}_{br}\mathsf{P}\,, \qquad [B_i\,,\,\mathsf{P}^{-1}\mathrm{Q}_{ar}\mathsf{P}] = 0\,, \tag{25.3.3}$$

so $\mathsf{P}^{-1}\mathrm{Q}_{ar}\mathsf{P}$ is a $(1/2,0)$ symmetry generator, and therefore must be a linear combination of the Q^*_{ar}. According to Eq. (25.2.6), Lorentz invariance dictates the form of this relation as

$$\mathsf{P}^{-1}\mathrm{Q}_{ar}\mathsf{P} = \sum_{bs} \mathscr{P}_{rs}\, e_{ab}\mathrm{Q}^*_{bs}\,, \tag{25.3.4}$$

where $\mathscr{P}$ is a numerical matrix, and the matrix e is given by Eq. (25.2.9).

We can learn something about the properties of the matrix $\mathscr{P}$ by requiring that Eq. (25.3.4) be consistent with the fundamental anticommutation relation (25.2.7). Eq. (25.3.4) and its adjoint yield

$$\mathsf{P}^{-1}\{\mathrm{Q}_{ar}\,,\,\mathrm{Q}^*_{bs}\}\mathsf{P} = \sum_{cdtu} \mathscr{P}_{rt}\, e_{ac}\, \mathscr{P}^*_{su}\, e_{bd}\, \{\mathrm{Q}^*_{ct}\,,\,\mathrm{Q}_{du}\}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

Inserting Eq. (25.2.7), this becomes

$$\delta_{rs}\sigma^{\mu}_{ab}\,\mathsf{P}^{-1}P_{\mu}\mathsf{P} = \sum_{cdtu}\mathscr{P}_{rt}\,e_{ac}\,\mathscr{P}^{*}_{su}\,e_{bd}\,\delta_{tu}\,\sigma^{\mu}_{dc}\,P_{\mu}\,.$$

But $e\sigma_i^{\mathrm{T}}e^{-1} = -\sigma_i$ and $e\sigma_0^{\mathrm{T}}e^{-1} = +\sigma_0$, while $\mathsf{P}^{-1}P_i\mathsf{P} = -P_i$ and $\mathsf{P}^{-1}P_0\mathsf{P} = P_0$, so this reduces to the statement that $\mathscr{P}$ is unitary

$$\mathscr{P}\,\mathscr{P}^{\dagger} = 1\,. \tag{25.3.5}$$

The matrix $\mathscr{P}$ is to some extent arbitrary, because for any set of fermionic generators Q_{ar} that satisfies Eqs. (25.3.2) and (25.2.7), we construct another set Q'_{ar} that also satisfies Eqs. (25.3.2) and (25.2.7) by a unitary transformation

$$\mathrm{Q}'_{ar} = \sum_{s}\mathscr{U}_{rs}\,\mathrm{Q}_{as}\,, \qquad \mathscr{U}^{\dagger} = \mathscr{U}^{-1}\,, \tag{25.3.6}$$

so that the parity transformation rule (25.3.4) becomes

$$\mathsf{P}^{-1}\mathrm{Q}'_{ar}\mathsf{P} = \sum_{bs}\mathscr{P}'_{rs}\,e_{ab}\mathrm{Q}'^{*}_{bs}\,, \tag{25.3.7}$$

where

$$\mathscr{P}' = \mathscr{U}\,\mathscr{P}\,\mathscr{U}^{-1*} = \mathscr{U}\,\mathscr{P}\,\mathscr{U}^{\mathrm{T}}\,. \tag{25.3.8}$$

For simple supersymmetry $\mathscr{P}$ is just a 1×1 phase factor, and Eq. (25.3.4) reads

$$\mathsf{P}^{-1}\mathrm{Q}_a\mathsf{P} = \mathscr{P}\sum_{b}e_{ab}\mathrm{Q}^{*}_{b}\,. \tag{25.3.9}$$

Combining this with its adjoint gives

$$\mathsf{P}^{-2}\mathrm{Q}_a\,\mathsf{P}^2 = -\,\mathrm{Q}_a\,, \tag{25.3.10}$$

independently of the value chosen for the phase factor $\mathscr{P}$. This has the striking consequence that if a boson in a supermultiplet of particles has a real intrinsic parity, then the fermions obtained by acting on this boson state with Q_a have *imaginary* intrinsic parities.

Because for simple supersymmetry $\mathscr{U}$ and $\mathscr{P}$ are just phase factors, it is obvious from Eq. (25.3.8) that by a suitable choice of $\mathscr{U}$, the phase factor $\mathscr{P}'$ can be made anything we like. It will be convenient to choose $\mathscr{P}' = +i$, so that Eq. (25.3.7) takes the simple form (now dropping primes)

$$\mathsf{P}^{-1}\mathrm{Q}_a\mathsf{P} = i\sum_{b}e_{ab}\,\mathrm{Q}^{*}_{b}\,. \tag{25.3.11}$$

Just as for spinor field operators, the representation of space inversion is simpler if we combine the $(0,1/2)$ operators Q_a and the $(1/2,0)$ operators $\sum_b e_{ab}\mathrm{Q}^{*}_{b}$ into the four-component Dirac spinor generator Q_{α} defined by

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

Eq. (25.2.34). In these terms, Eq. (25.3.11) and its adjoint read

$$\mathsf{P}^{-1} Q \mathsf{P} = i\beta Q \,. \tag{25.3.12}$$

(We are using the notation for Dirac matrices given in Section 5.4 and in the Preface to this volume, which gives

$$\beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ,$$

where 1 and 0 are understood as 2×2 submatrices.)

For extended supersymmetry it is not always possible to choose $\mathscr{U}$ so that $\mathscr{P}'$ is diagonal. However, a theorem in matrix algebra (proved in Appendix C of Chapter 2) shows that it is possible to choose $\mathscr{U}$ so that $\mathscr{P}'$ is block diagonal, in general with some of the matrices on the diagonal equal to 1×1 submatrices that can be chosen equal to i (or any other phase factors we like), and other submatrices on the diagonal given by 2×2 matrices that can be chosen to have the form

$$\begin{pmatrix} 0 & \exp(i\phi) \\ \exp(-i\phi) & 0 \end{pmatrix} ,$$

with various phases ϕ. Correspondingly, with this choice of $\mathscr{U}$ (and now dropping primes) the two-component Qs are of two types. The Qs of the first type also satisfy Eq. (25.3.11):

$$\mathsf{P}^{-1} \mathrm{Q}_{ar} \mathsf{P} = i \sum_b e_{ab} \, \mathrm{Q}^*_{br} \,. \tag{25.3.13}$$

The two-component Qs of the second type come in pairs, which we will call $\mathrm{Q}_{a\,s1}$ and $\mathrm{Q}_{a\,s2}$, with the sth pair having the parity transformation rule

$$\mathsf{P}^{-1} \, \mathrm{Q}_{a\,s1} \, \mathsf{P} = e^{i\phi_s} \sum_b e_{ab} \, \mathrm{Q}^*_{b\,s2} \,, \quad \mathsf{P}^{-1} \, \mathrm{Q}_{a\,s2} \, \mathsf{P} = e^{-i\phi_s} \sum_b e_{ab} \, \mathrm{Q}^*_{b\,s1} \,. \tag{25.3.14}$$

In particular, we now have

$$\mathsf{P}^{-2} \mathrm{Q}_{a\,s1} \mathsf{P}^2 = -e^{2i\phi_s} \mathrm{Q}_{a\,s1} \,, \qquad \mathsf{P}^{-2} \mathrm{Q}_{a\,s2} \mathsf{P}^2 = -e^{-2i\phi_s} \mathrm{Q}_{a\,s2} \,. \tag{25.3.15}$$

This shows that, unless $\phi_s = 0 \pmod{\pi}$, it is not possible to form supersymmetry generators of the first type out of linear combinations of extended supersymmetry generators of the second type.

In terms of the four-component spinors (25.2.34), the effect of the parity operator on extended supersymmetry generators of the first type is

$$\mathsf{P}^{-1} Q_r \, \mathsf{P} = i\beta \, Q_r \,, \tag{25.3.16}$$

while for generators of the second type

$$\mathsf{P}^{-1} \, Q_{s1} \, \mathsf{P} = \beta \, \gamma_5 \, \exp(i\gamma_5 \phi_s) \, Q_{s2} \,, \qquad \mathsf{P}^{-1} \, Q_{s2} \, \mathsf{P} = \beta \, \gamma_5 \, \exp(-i\gamma_5 \phi_s) \, Q_{s1} \,. \tag{25.3.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

25.4 Massless Particle Supermultiplets

Supersymmetry requires that the known particles be accompanied in irreducible representations of the supersymmetry algebra by 'sparticles': bosonic 'squarks' and 'sleptons' accompanying quarks and leptons, and fermionic 'gauginos' accompanying gauge bosons. None of these sparticles have been observed, so supersymmetry is certainly broken, with the masses of the sparticles almost certainly much larger than the quark, lepton, and gauge boson masses produced by the spontaneous breakdown of the electroweak $SU(2) \times U(1)$ gauge group, and hence of the same order of magnitude as the splittings within supermultiplets. Thus it is very likely that at energy scales that are large enough so that we can neglect supersymmetry breaking and these mass splittings, we can also treat the known quarks, leptons, and gauge bosons and their superpartners as massless. Therefore we will be specially interested in supermultiplets of massless particles.

Consider a state containing a single massless particle belonging to some supermultiplet. We obtain the other states in the same supermultiplet by applying operators Q_{ar} and/or Q^*_{ar} to this state. Since Q_{ar} and Q^*_{ar} commute with P_μ, all these states have the same value of the four-momentum. We will work in a Lorentz frame in which the four-momentum of these states is $p^1 = p^2 = 0$ and $p^3 = p^0 = E$. With this choice of four-momentum, we have

$$\sigma_\mu p^\mu = E(\sigma_0 + \sigma_3) = 2E \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} , \tag{25.4.1}$$

which aside from the factor $2E$ is the projection matrix onto the subspace with helicity $+1/2$. The anticommutation relation (25.2.7) therefore shows that $\{Q_{(-1/2)r}, Q^*_{(-1/2)r}\}$ gives zero when acting on any state in a supermultiplet with this momentum, and thus so do $Q_{(-1/2)r}$ and $Q^*_{(-1/2)r}$. We must therefore construct the states of the supermultiplet by acting only with $Q_{(1/2)r}$ and $Q^*_{(1/2)r}$. Furthermore, we are labelling the Qs with their J_3 values, in the sense that

$$[J_3, Q_{ar}] = -a\, Q_{ar} , \tag{25.4.2}$$

so $Q_{(1/2)r}$ and $Q^*_{(1/2)r}$ respectively lower and raise the helicity by $1/2$.

We will first consider the case of simple supersymmetry. Consider a supermultiplet with maximum helicity $\lambda_{\max}$, and let $|\lambda_{\max}\rangle$ be any one-particle state with this helicity and four-momentum p^μ. Then

$$Q^*_{\frac{1}{2}} |\lambda_{\max}\rangle = 0 , \tag{25.4.3}$$

while acting on this state with $Q_{1/2}$ gives a state $|\lambda_{\max} - 1/2\rangle$ with helicity

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

$\lambda_{\max} - 1/2$. We will define this state as

$$|\lambda_{\max} - 1/2\rangle \equiv (4E)^{-1/2} Q_{\frac{1}{2}} |\lambda_{\max}\rangle \,. \tag{25.4.4}$$

The fundamental anticommutation relation (25.2.7) together with Eqs. (25.4.1) and (25.4.3) shows that this state is normalized in the same way as $|\lambda_{\max}\rangle$

$$\langle \lambda_{\max} - 1/2 | \lambda_{\max} - 1/2\rangle = \langle \lambda_{\max} | \lambda_{\max}\rangle \,, \tag{25.4.5}$$

and in particular this state cannot vanish. Eq. (25.2.32) shows that $Q_{1/2}^2 = 0$, so acting with $Q_{1/2}$ on $|\lambda_{\max} - 1/2\rangle$ gives zero:

$$Q_{\frac{1}{2}} |\lambda_{\max} - 1/2\rangle = (4E)^{-1/2} Q_{\frac{1}{2}}^2 |\lambda_{\max}\rangle = 0 \,. \tag{25.4.6}$$

On the other hand, acting with $Q_{1/2}^*$ on this state gives the state with which we started. That is,

$$Q_{\frac{1}{2}}^* |\lambda_{\max} - 1/2\rangle = (4E)^{-1/2} Q_{\frac{1}{2}}^* Q_{\frac{1}{2}} |\lambda_{\max}\rangle = (4E)^{-1/2} \{Q_{\frac{1}{2}}^* , Q_{\frac{1}{2}}\} |\lambda_{\max}\rangle \,,$$

so that Eq. (25.4.1) and the anticommutation relation (25.2.31) yield

$$Q_{\frac{1}{2}}^* |\lambda_{\max} - 1/2\rangle = (4E)^{1/2} |\lambda_{\max}\rangle \,. \tag{25.4.7}$$

Thus the supermultiplet consists of just two states, with helicities $\lambda_{\max}$ and $\lambda_{\max} - 1/2$. In the basis provided by these two states, the operators $Q_{1/2}$ and $Q_{1/2}^*$ are represented by the matrices

$$q_{\frac{1}{2}} = \sqrt{4E} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \,, \qquad q_{\frac{1}{2}}^\dagger = \sqrt{4E} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \,, \tag{25.4.8}$$

while the operators $Q_{-1/2}$ and $Q_{-1/2}^*$ are represented by zero.

It is worth emphasizing that this is the *only* kind of massless supermultiplet in theories with simple supersymmetry. There are no massless particles that are not accompanied with a superpartner, and none that have more than one superpartner. Of course, CPT invariance implies that for every supermultiplet of massless particles of helicities λ and $\lambda - 1/2$ there must be an antimultiplet with helicities $-\lambda + 1/2$ and $-\lambda$. In particular, a massless particle and antiparticle with helicities $+1/2$ and $-1/2$ must be accompanied by a massless particle and antiparticle either with helicities $+1$ and -1 or with helicities both zero.

How might the known quarks, leptons, and gauge bosons fit into this picture? We will assume that the supersymmetry generator commutes with the generators of the $SU(3) \times SU(2) \times U(1)$ gauge group.* The quarks

* In simple supersymmetry the generator Q_α must in any case commute with the $SU(3) \times SU(2)$ generators, because semi-simple algebras like $SU(3) \times SU(2)$ have no non-trivial one-dimensional representations.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

and leptons belong to different representations of the gauge group from the gauge bosons, so they cannot be in the same supermultiplets. We have to conclude then that in the limit of high energy where $SU(2) \times U(1)$ symmetry breaking may be neglected, the massless quarks and leptons of each color and flavor are in supermultiplets with *pairs* of massless squarks and sleptons of zero helicity and the same color and flavor, while the massless gauge bosons are accompanied by massless gauginos of helicity $\pm 1/2$ comprising an adjoint representation of $SU(3) \times SU(2) \times U(1)$.

Because gravity exists we know that in addition to the particles of the standard model there must also exist a massless particle of helicity ± 2, the *graviton*. Massless particles with helicity λ having $|\lambda| > 1/2$ must couple at low momentum to conserved quantities.** Soft massless particles of helicity ± 1 can couple to various internal symmetry generators, soft massless particles of helicity $\pm 3/2$ can couple to the supersymmetry generators Q_a, and a soft particle of helicity ± 2 can couple to a single conserved quantity, the momentum four-vector P_μ, but there are no conserved quantities to which a soft massless particle with $|\lambda| > 2$ could couple. We conclude that the graviton cannot be in a supermultiplet with particles of helicity $\pm 5/2$, so it must be in a supermultiplet with a massless particle of helicity $\pm 3/2$, known as a *gravitino*, coupled to the supersymmetry generators themselves. The field theory of this supermultiplet is known as *supergravity*, and will be discussed in Chapter 31.

Now let us consider the case of extended supersymmetry, with N supersymmetry generators. We first note that because the $Q_{(-1/2)r}$ all give zero when acting on the states of a supermultiplet (including a state obtained by letting $Q_{(1/2)s}$ act on any other state of the multiplet), the central charges Z_{rs} must also annihilate any state of the multiplet. With central charges out of the picture, the supersymmetry generators $Q_{(1/2)r}$ all anticommute when acting on a massless particle supermultiplet, so applying n of them to a one-particle state of maximum helicity $\lambda_{\max}$ and four-momentum p^μ gives $N!/n!(N-n)!$ one-particle states of the same four-momentum and helicity $\lambda_{\max} - n/2$, forming a rank n antisymmetric tensor representation of the $SU(N)$ R-symmetry† (25.2.30). The maximum value of n that gives a non-zero state is $n = N$, so the minimum helicity in a supermultiplet is given by

$$\lambda_{\min} = \lambda_{\max} - N/2 \,. \tag{25.4.9}$$

If we wish to exclude massless particle helicities λ with $|\lambda| > 2$, then

** This is discussed for the case of integer helicity in Section 13.1. The argument for half-integer helicity was given by Grisaru and Pendleton.[3]

† The $U(1)$ part of $U(N)$ R-symmetry is often violated by quantum mechanical anomalies.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

$\lambda_{\max} - \lambda_{\min} \leq 4$, so extended supersymmetries are allowed only with $N \leq 8$.

For $N = 8$, with helicities with $|\lambda| > 2$ excluded, there is only one possible supermultiplet, consisting of: 1 graviton with each helicity ± 2; 8 gravitinos with each helicity $\pm 3/2$; 28 gauge bosons with each helicity ± 1; 56 fermions with each helicity $\pm 1/2$; and 70 bosons with helicity zero.

Compare this with the case $N = 7$, again excluding helicities with $|\lambda| > 2$. Here there are two supermultiplets. One supermultiplet contains: 1 graviton with helicity $+2$; 7 gravitinos with helicity $+3/2$; 21 gauge bosons with helicity $+1$; 35 fermions with helicity $+1/2$; 35 bosons with helicity zero; 21 fermions with helicity $-1/2$; 7 gauge bosons with helicity -1; and 1 gravitino with helicity $-3/2$. The other is the CPT-conjugate supermultiplet, with all helicities reversed. Adding the numbers of particles in these two supermultiplets, we have 1 graviton with each helicity ± 2; $7 + 1 = 8$ gravitinos with each helicity $\pm 3/2$; $21 + 7 = 28$ gauge bosons with each helicity ± 1; $35 + 21 = 56$ fermions with each helicity $\pm 1/2$; and $35 + 35 = 70$ bosons with helicity zero. Extended supergravity theories with $N = 8$ and $N = 7$ thus have precisely the same particle content and are in fact identical.

On the other hand, extended supergravity theories with $N \leq 6$ have just N gravitinos of each helicity $\pm 3/2$ and are therefore all distinct.

For $N \leq 4$ there is also the possibility of *global* supersymmetry theories, theories with supermultiplets that do not include gravitons or gravitinos. For global $N = 4$ supersymmetry there is just one supermultiplet, containing: 1 gauge boson of each helicity ± 1; 4 fermions of each helicity $\pm 1/2$; and 6 bosons of helicity zero. This is equivalent to the global supersymmetry theory with $N = 3$, which has two supermultiplets: 1 supermultiplet with one gauge boson of helicity $+1$, 3 fermions with helicity $+1/2$; 3 bosons with helicity zero; and 1 fermion with helicity $-1/2$; and the other the CPT conjugate supermultiplet with opposite helicities. Adding the numbers of particles of each helicity in these two $N = 3$ supermultiplets gives the same particle content as for $N = 4$ global supersymmetry. The gauge field theory with $N = 4$ supersymmetry has remarkable properties, which will be discussed in Section 27.9.

For $N = 2$ extended global supersymmetry there are supermultiplets of two different types, apart from those related by CPT. There are *gauge* supermultiplets, each containing one gauge boson of helicity $+1$, two fermions of helicity $+1/2$ forming a doublet under the $SU(2)$ R-symmetry (25.2.30), and one boson of helicity zero, together with their CPT-conjugate supermultiplets, with helicities reversed. Together each gauge supermultiplet and its antimultiplet contain one gauge boson with each helicity ± 1, an $SU(2)$ doublet of fermions of each helicity $\pm 1/2$, and two $SU(2)$ singlet bosons of helicity zero. Then there are *hypermultiplets*,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

which contain one fermion of each helicity $\pm 1/2$ and an $SU(2)$ doublet of bosons of helicity zero, together with its CPT-conjugate. (In quantum field theory a hypermultiplet cannot be its own antimultiplet, because then the helicity zero particles would be described by just two *real* scalar fields, which cannot form an $SU(2)$ doublet.) Of course, in the real world there would also have to be a graviton supermultiplet, containing a graviton of helicity $+2$, an $SU(2)$ doublet of gravitinos with helicity $+3/2$, and one gauge boson with helicity $+1$, together with their CPT-conjugates, with opposite helicity. Gauge theories with $N = 2$ supersymmetry are constructed in Section 27.9, and explored non-perturbatively in Section 29.5.

The particle content of these supermultiplets reveals a difficulty in incorporating extended supersymmetry in realistic theories of particles at accessible energies. In all cases but one, the helicity $+1/2$ fermions belong to supermultiplets along with helicity $+1$ gauge bosons. Gauge bosons belong to the adjoint representation of the gauge group, so if the supersymmetry generators are invariant under the gauge group then the helicity $+1/2$ fermions must also belong to the adjoint representation, which is real. This is in conflict with the fact that the known quarks and leptons belong to a representation of $SU(3) \times SU(2) \times U(1)$ which is *chiral* — that is, for which the helicity $+1/2$ fermions belong to a complex representation, which is then necessarily different from the representation furnished by their CPT-conjugates, the helicity $-1/2$ fermions. The one exception, where helicity $+1/2$ fermions are not in a supermultiplet with gauge bosons, is the $N = 2$ hypermultiplet discussed above. But in this case particles of both helicities $+1/2$ and $-1/2$ are in the same supermultiplet, and therefore must transform the same under any gauge transformations that leave the supersymmetry generators invariant. They may belong to a complex representation of this gauge group, but then the CPT-conjugate of this hypermultiplet belongs to the complex-conjugate representation, and the fermions of each helicity then belong to the sum of the two representations, which is real, again in conflict with the chiral nature of the known quarks and leptons.

In contrast, for simple supersymmetry there are supermultiplets containing just helicity $+1/2$ and helicity zero, which may be in a complex representation of the gauge group, distinct from the representation furnished by the CPT-conjugate supermultiplet. Here there is no conflict with chirality. For this reason, most discussions of supersymmetry as a symmetry that remains unbroken at accessible energies have focused on simple rather than extended supersymmetry.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

25.5 Massive Particle Supermultiplets

Although the known quarks, leptons, and gauge bosons and their superpartners may probably be treated as massless at energies where supersymmetry breaking is negligible, this is not necessarily true for other particles, including the extra gauge bosons of large mass required by theories that unify the strong and electroweak interactions. Also, ever since the Wess–Zumino model, massive particle theories have been useful test cases for studying supersymmetry theories. It will therefore be worthwhile for us briefly to consider the implications of unbroken supersymmetry for massive particles.

As in the previous section, we obtain the various one-particle states in a supermultiplet by acting on any one of them with the operators Q_{ar} and $Q^\dagger_{ar}$, and all of these states have the same four-momentum. Unlike the case of zero mass, for mass $M > 0$ we may now take this to be the four-momentum of a particle at rest, with $p^i = 0$ for $i = 1, 2, 3$ and $p^0 = M$. In this frame of reference, we have

$$\sigma_\mu p^\mu = M\sigma_0 = M \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{25.5.1}$$

Thus, acting on any state $|\ \rangle$ in a supermultiplet with this four-momentum, the anticommutation relation (25.2.7) yields

$$\{Q_{ar}\,,\, Q^\dagger_{bs}\}|\ \rangle = 2M\,\delta_{ab}\,\delta_{rs}\,|\ \rangle\,. \tag{25.5.2}$$

In contrast with the case of zero mass, here no component of Q_{ar} or $Q^\dagger_{ar}$ can vanish on the whole multiplet, so we have two sets of raising and lowering operators: both $Q_{(1/2)r}$ and $Q^\dagger_{(-1/2)r}$ lower the spin 3-component by 1/2, while both $Q_{(-1/2)r}$ and $Q^\dagger_{(1/2)r}$ raise the spin 3-component by 1/2. However, as we shall see, for extended supersymmetry it is possible for certain linear combinations of the Qs and $Q^\dagger$s to vanish.

We will first consider the case of simple supersymmetry. By using the supersymmetry algebra (25.2.31), (25.2.32), we shall show that the general massive supermultiplet consists of a particle of spin $j + 1/2$, a *pair* of particles of spin j, and a particle of spin $j - 1/2$. Where parity is conserved, the particles of spin $j \pm 1/2$ have equal intrinsic parity, given by some phase η, while the two particles of spin j have parities $+i\eta$ and $-i\eta$. Here j is any integer or half-integer greater than zero. There is also a collapsed supermultiplet, consisting of two particles of spin zero and a particle of spin 1/2. When parity is conserved the particles of spin zero have parities $i\eta$ and $-i\eta$, where η is the parity of the particle of spin 1/2.

Here is the proof. We first show that any supermultiplet will contain at least one spin multiplet of states $|j, \sigma\rangle$ with spin 3-component σ running by unit steps from $-j$ to $+j$, having the special property that, for all such

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

σ and for $a = \pm 1/2$,

$$Q_a |j, \sigma\rangle = 0. \tag{25.5.3}$$

Starting with any non-zero state $|\psi\rangle$ in the supermultiplet, we can define non-zero states

$$|\psi'\rangle \equiv \begin{cases} (2M)^{-1/2} Q_{1/2} |\psi\rangle & Q_{1/2} |\psi\rangle \neq 0 \\ |\psi\rangle & Q_{1/2}|\psi\rangle = 0 \end{cases},$$

and

$$|\psi''\rangle \equiv \begin{cases} (2M)^{-1/2} Q_{-1/2} |\psi'\rangle & Q_{-1/2} |\psi'\rangle \neq 0 \\ |\psi'\rangle & Q_{-1/2} |\psi'\rangle = 0 \end{cases}.$$

Because the Q_a anticommute, $Q_{1/2}|\psi'\rangle = 0$, and so $Q_a|\psi''\rangle = 0$ for $a = \pm 1/2$. If any state $|\psi''\rangle$ satisfies the condition that $Q_a|\psi''\rangle = 0$, then so does $U(R)|\psi''\rangle$, where $U(R)$ is the unitary operator representing an arbitrary spatial rotation. It follows that the states that satisfy this condition may be decomposed into complete spin multiplets $|j, \sigma\rangle$, satisfying condition (25.5.3).

Now focus on any one of these spin multiplets satisfying Eq. (25.5.3), normalized so that

$$\langle j, \sigma' | j, \sigma\rangle = \delta_{\sigma'\sigma} . \tag{25.5.4}$$

For $j > 0$, by applying the spin 1/2 operators* Q_a^* to these states we can construct states of spin $j \pm 1/2$:

$$|j \pm 1/2, \sigma\rangle = \frac{1}{\sqrt{2M}} \sum_a C_{\frac{1}{2}\, j}\Big(j \pm 1/2\,,\, \sigma\,;\, a\,,\, \sigma - a\Big)\, Q_a^* |j, \sigma - a\rangle\,, \tag{25.5.5}$$

where $C_{jj'}(j'', \sigma''; \sigma, \sigma')$ is the conventional Clebsch–Gordan coefficient for coupling spins j and j' with 3-components σ and σ' to make spin j'' with 3-component σ''. Using Eqs. (25.5.2)–(25.5.5) and the orthonormality properties of the Clebsch–Gordan coefficients, we can show that these states are properly normalized:

$$\langle j \pm 1/2, \sigma | j \pm 1/2, \sigma'\rangle = \delta_{\sigma\sigma'}\,, \qquad \langle j \pm 1/2, \sigma | j \mp 1/2, \sigma'\rangle = 0\,, \tag{25.5.6}$$

so none of the states $|j \pm 1/2, \sigma\rangle$ can vanish. The only exception is for $j = 0$, in which case of course there is no state $|j - 1/2, \sigma\rangle$. We can also construct other states by applying *two* Q*s to $|j, \sigma\rangle$. Since each Q_a^*

* Since Q_a transforms under rotations like a field that destroys a particle of spin 1/2 and spin 3-component a, it is Q_a^* that transforms like a field that creates such a particle, and hence transforms like the particle itself. To put this formally, $[J_i\,,\, Q_a] = -\sum_b \frac{1}{2}(\sigma_i)_{ab}\, Q_b$, so $[J_i\,,\, Q_a^*] = \sum_b \frac{1}{2}(\sigma_i)_{ba}\, Q_b^*$, which may be compared with the transformation property of a spin 1/2 particle, $J_i |a\rangle = \sum_b \frac{1}{2}(\sigma_i)_{ba} |b\rangle$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

anticommutes with itself, the only such non-zero states are formed by applying the operator $Q^*_{1/2}Q^*_{-1/2} = -Q^*_{-1/2}Q^*_{1/2}$. This operator may be written as $\frac{1}{2}e_{ab}Q^*_aQ^*_b$, which shows that it is rotationally invariant, so this gives a second spin multiplet with spin j:

$$|j,\sigma\rangle^\flat = \frac{1}{2M}\, Q^*_{1/2}\, Q^*_{-1/2}|j,\sigma\rangle\,, \tag{25.5.7}$$

which is distinguished from $|j,\sigma\rangle$ by the fact that instead of Eq. (25.5.3), we have

$$Q^*_a|j,\sigma\rangle^\flat = 0\,. \tag{25.5.8}$$

Again using Eqs. (25.5.2)–(25.5.4), we find that these are also normalized states:

$${}^\flat\langle j,\sigma'|j,\sigma\rangle^\flat = \delta_{\sigma'\sigma}\,, \qquad \langle j,\sigma'|j,\sigma\rangle^\flat = 0\,. \tag{25.5.9}$$

It is then easy to show that the states constructed so far form a complete representation of the supersymmetry algebra. The orthonormality property of the Clebsch–Gordan coefficients allows us to rewrite Eq. (25.5.5) as

$$Q^*_a\,|j,\sigma\rangle = \sqrt{2M}\sum_{\pm} C_{1/2j}\Big(j\pm 1/2,\sigma+a;a,\sigma\Big)\,|j\pm 1/2,\sigma+a\rangle\,. \tag{25.5.10}$$

Also, Eq. (25.5.2) shows that, for any state $|\,\rangle$ in the supermultiplet,

$$\Big[Q_a\,,\, Q^*_{\frac{1}{2}}Q^*_{-\frac{1}{2}}\Big]\,|\,\rangle = 2M\sum_b e_{ab}\,Q^*_b\,|\,\rangle\,, \tag{25.5.11}$$

so Eqs. (25.5.7) and (25.5.3) give

$$\begin{aligned} Q_a\,|j,\sigma\rangle^\flat &= \sum_b e_{ab}\,Q^*_b\,|j,\sigma\rangle \\ &= \sqrt{2M}\sum_b e_{ab}\sum_{\pm} C_{\frac{1}{2}j}\Big(j\pm 1/2,\sigma+b;b,\sigma\Big)\,|j\pm 1/2,\sigma+b\rangle\,. \end{aligned} \tag{25.5.12}$$

From Eqs. (25.5.2), (25.5.3), and (25.5.5) we have

$$Q_a\,|j\pm 1/2,\sigma\rangle = \sqrt{2M}C_{\frac{1}{2}\,j}\Big(j\pm 1/2\,,\,\sigma\,;\,a\,,\,\sigma-a\Big)\,|j,\sigma-a\rangle\,, \tag{25.5.13}$$

while Eqs. (25.5.5), (25.2.31), and (25.5.7) yield

$$Q^*_a\,|j\pm 1/2,\sigma\rangle = \sqrt{2M}\sum_b e_{ab}\,C_{\frac{1}{2}\,j}\Big(j\pm 1/2,\sigma;b,\sigma-b\Big)\,|j,\sigma-b\rangle^\flat\,. \tag{25.5.14}$$

Eqs. (25.5.3), (25.5.8), (25.5.10), and (25.5.12)–(25.5.14) give the action of the Qs and Q*s on all the states of the supermultiplet.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

For $j = 0$ we have the collapsed supermultiplet: Eqs. (25.5.3), (25.5.8), (25.5.10), and (25.5.12)–(25.5.14) become

$$\begin{aligned} &Q_a|0,0\rangle = 0\,, && Q_a^*|0,0\rangle^\flat = 0\,,\\ &Q_a^*|0,0\rangle = \sqrt{2M}\,|1/2,a\rangle\,, && Q_a|0,0\rangle^\flat = \sqrt{2M}\textstyle\sum_b e_{ab}\,|1/2,b\rangle\,,\\ &Q_a|1/2,b\rangle = \sqrt{2M}\delta_{ab}|0,0\rangle\,, && Q_a^*|1/2,b\rangle = \sqrt{2M}e_{ab}\,|0,0\rangle^\flat\,. \end{aligned} \tag{25.5.15}$$

Now suppose that parity is conserved. Recall that the phase of the supersymmetry generator may be chosen so that the action of the parity operator on these generators is given by Eq. (25.3.13). Then Q_a^* acting on $\mathsf{P}|j,\sigma\rangle$ is a linear combination of the states $\mathsf{P}Q_a|j,\sigma\rangle$, which vanish, and since $\mathsf{P}|j,\sigma\rangle$ has the same rotation properties as $|j,\sigma\rangle^\flat$, it must simply be proportional to it

$$\mathsf{P}|j,\sigma\rangle = -\eta|j,\sigma\rangle^\flat\,. \tag{25.5.16}$$

Since P is unitary, η is a phase factor, with $|\eta| = 1$. A corresponding argument shows that $\mathsf{P}|j,\sigma\rangle^\flat$ is proportional to $|j,\sigma\rangle$. To find the proportionality coefficient, we note that

$$\begin{aligned} \mathsf{P}|j,\sigma\rangle^\flat &= (2M)^{-1}\mathsf{P}\,Q_{\frac12}^*\,Q_{-\frac12}^*\,|j,\sigma\rangle = -\eta(2M)^{-1}\,Q_{-\frac12}\,Q_{\frac12}\,|j,\sigma\rangle^\flat\\ &= -\eta(2M)^{-2}\,Q_{-\frac12}\,Q_{\frac12}\,Q_{\frac12}^*\,Q_{-\frac12}^*\,|j,\sigma\rangle = -\eta|j,\sigma\rangle\,. \end{aligned}$$

We can then define states of spin j

$$|j,\sigma\rangle^\pm \equiv \frac{1}{\sqrt2}\Big(|j,\sigma\rangle \pm i\,|j,\sigma\rangle^\flat\Big)\,, \tag{25.5.17}$$

with definite parity

$$\mathsf{P}|j,\sigma\rangle^\pm = \pm i\eta\,|j,\sigma\rangle^\pm\,. \tag{25.5.18}$$

Finally, applying the parity operator to Eq. (25.5.5) and using Eqs. (25.3.13) and (25.5.16) gives

$$\mathsf{P}\,|j\pm1/2,\sigma\rangle = -\frac{\eta}{\sqrt{2M}}\sum_a C_{\frac12\,j}\Big(j\pm1/2\,,\,\sigma\,;\,a\,,\,\sigma-a\Big)\sum_b e_{ab}Q_b\,|j,\sigma-a\rangle^\flat\,.$$

Eq. (25.5.12) and the orthonormality property of the Clebsch–Gordan coefficients then yield

$$\mathsf{P}\,|j\pm1/2,\sigma\rangle = \eta\,|j\pm1/2,\sigma\rangle\,, \tag{25.5.19}$$

as was to be shown.

We now turn briefly to the case of extended supersymmetry, with N supersymmetry generators. As mentioned in the previous section, there can be no massless particle with a non-vanishing eigenvalue for any central charge. We can go further and show that the eigenvalues of the central charge operators set a lower bound on the mass of any supermultiplet.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

Because the central charges Z_{rs} and Z^*_{rs} commute with each other and with P_μ, one-particle states can be chosen to be eigenstates of all the central charges as well as of P_μ, and because the central charges commute with the Q_{ar} and Q^*_{ar}, all states in a supermultiplet have the same eigenvalues.

To derive an inequality relating the mass M of a supermultiplet and the eigenvalues of the central charges on this multiplet, we use the anti-commutation relations (25.2.7) and (25.2.8) to write

$$\sum_{ar} \left\{ \left(Q_{ar} - \sum_{bs} e_{ab} U_{rs} Q^*_{bs} \right), \left(Q^*_{ar} - \sum_{ct} e_{ac} U^*_{rt} Q_{ct} \right) \right\} = 8NP^0 - 2\mathrm{Tr}\left(ZU^\dagger + UZ^\dagger \right), \tag{25.5.20}$$

where U_{rs} is an arbitrary $N \times N$ unitary matrix. The left-hand side is a positive-definite operator, so by letting this act on the states of the supermultiplet at rest, we find

$$M \geq \frac{1}{4N} \mathrm{Tr}\left(ZU^\dagger + UZ^\dagger \right), \tag{25.5.21}$$

where now Z_{rs} denotes the values of the central charges for the supermultiplet of mass M. The polar decomposition theorem tells us that any square matrix Z may be written as HV, where H is a positive Hermitian matrix and V is unitary. We can obtain a useful inequality (which is in fact optimal) by setting $U = V$, in which case Eq. (25.5.21) becomes

$$M \geq \frac{1}{2N} \mathrm{Tr}\, H = \frac{1}{2N} \mathrm{Tr} \sqrt{Z^\dagger Z}\,. \tag{25.5.22}$$

States for which M equals the minimum value allowed by this inequality are known as *BPS* states, by analogy with the Bogomol'nyi–Prasad–Sommerfeld magnetic monopole configurations discussed in Section 23.3, whose mass also equals a lower bound on general monopole masses. In fact, this is more than an analogy; we will see in Section 27.9 that the lower bound on monopole masses in gauge theories with extended supersymmetry is a special case of the lower bound (25.5.22).

As we can see from this derivation of Eq. (25.5.22), for BPS supermultiplets the operator $Q_{ar} - \sum_{bs} e_{ab} U_{rs} Q^*_{bs}$ gives zero when acting on any state of a supermultiplet, so there are only N independent helicity-lowering operators $Q_{(1/2)r}$ and only N independent helicity-raising operators $Q_{(-1/2)r}$, just as in the case of massless supermultiplets. This leads to smaller supermultiplets than would be found in the general case.

For instance, for $N = 2$ supersymmetry the central charge is given by a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

single complex number**

$$Z = \begin{pmatrix} 0 & Z_{12} \\ -Z_{12} & 0 \end{pmatrix} . \tag{25.5.23}$$

The inequality (25.5.22) here reads

$$M \geq |Z_{12}|/2 . \tag{25.5.24}$$

Where $M = |Z_{12}|/2$, the helicity contents of the massive particle supermultiplets are the same as for those of zero mass: there are gauge supermultiplets, consisting of one particle of spin 1, an $SU(2)$ R-symmetry doublet of spin $1/2$, and one particle of spin 0 (the other helicity zero state belonging to the spin one particle), and hypermultiplets, consisting of one particle of spin $1/2$ and an $SU(2)$ R-symmetry doublet of spin 0. These are sometimes called 'short' supermultiplets, to distinguish them from the larger supermultiplets encountered when $M > |Z_{12}|/2$.

Problems

1. Find a set of 2×2 matrices that form a graded Lie algebra containing fermionic as well as bosonic generators.

2. Following the approach of Haag, Lopuszanski, and Sohnius, derive the form of the most general symmetry superalgebra in $2 + 1$ spacetime dimensions. (Hint: With the generators of the Lorentz group in $2 + 1$ spacetime dimensions labelled $A_1 = -iJ_{10}$, $A_2 = -iJ_{20}$, $A_3 = J_{12}$, the commutation relations of the Poincaré algebra are $[A_i, , A_j] = i\sum_k \epsilon_{ijk} A_k$, so the representations of the homogeneous Lorentz group in $2 + 1$ spacetime dimensions are labelled with a *single* positive integer or half-integer A.) Assume that the conditions for the Coleman–Mandula theorem are satisfied here.

3. Suppose that there were no massless particles with helicity greater than $+3/2$ or less than $-3/2$. Find the most general massless particle multiplets for $N = 6$ extended supersymmetry and (using CPT symmetry) $N = 5$ extended supersymmetry. What does the comparison between the particle content you find suggest about these two extended supersymmetries?

4. What are the possible parities of the particles in the short supermultiplets of extended $N = 2$ supersymmetry?

** In some articles on $N = 2$ supersymmetry the central charge Z is what we would call $Z/2\sqrt{2}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

References

1. R. Haag, J. T. Lopuszanski, and M. Sohnius, *Nucl. Phys.* **B88**, 257 (1975). This article is reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987).
2. B. Zumino, *Nucl. Phys.* **B89**, 535 (1975). This article is reprinted in *Supersymmetry*, Ref. 1.
3. M. T. Grisaru and H. N. Pendleton, *Phys. Lett.* **67B**, 323 (1977).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.004

26

Supersymmetric Field Theories

Now we know the structure of the most general supersymmetry algebras, and we have seen how to work out the implications of this symmetry for the particle spectrum. In order to learn what supersymmetry has to say about particle interactions, we need to see how to construct supersymmetric field theories.

Originally the construction of field supermultiplets was done directly, by a repeated use of the Jacobi identities, much as in the construction of supermultiplets of one-particle states in Sections 25.4 and 25.5. Section 26.1 presents one example of this technique, used here to construct supermultiplets containing only scalar and Dirac fields. Fortunately there is an easier technique, invented by Salam and Strathdee,[1] in which supermultiplets of fields are gathered into 'superfields,' which depend on fermionic coordinates as well as on the usual four coordinates of spacetime. Superfields are introduced in Section 26.2, and used to construct supersymmetric field theories and to study some of their consequences in Sections 26.3–26.8. This chapter will be concerned only with $N = 1$ supersymmetry, where the superfield formalism has been chiefly useful. At the end of the next chapter we will construct theories with N-extended supersymmetry by imposing the $U(N)$ R-symmetry on theories of $N = 1$ superfields.

26.1 Direct Construction of Field Supermultiplets

To illustrate the direct construction of a field supermultiplet, we will consider fields that can destroy the particles belonging to the simplest supermultiplet of arbitrary mass discussed in Section 25.5: two spinless particles and one particle of spin 1/2. We saw in Eq. (25.5.15) that the one-particle zero spin state $|0,0\rangle$ is annihilated by the supersymmetry generator Q_a but not by $Q_a^\dagger$, so we would expect the scalar field $\phi(x)$ that creates this particle from the vacuum (which is assumed to be annihilated by all supersymmetry generators) to commute with Q_a but not with $Q_a^\dagger$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

That is,

$$[Q_a, \phi(x)] = 0, \tag{26.1.1}$$

$$-i\sum_b e_{ab}[Q_b^\dagger, \phi(x)] \equiv \zeta_a(x) \neq 0. \tag{26.1.2}$$

The antisymmetric 2×2 matrix e_{ab} (with $e_{1/2,-1/2} \equiv +1$) is introduced here because it is $\sum_b e_{ab}Q_b^\dagger$ that transforms under the homogeneous Lorentz group according to the $(1/2,0)$ representation. It follows that $\zeta_a(x)$ is a two-component spinor field that also belongs to the $(1/2,0)$ representation of the homogeneous Lorentz group.*

From Eqs. (26.1.1)–(26.1.2) and the anticommutation relation (25.2.31), we find

$$\{Q_b, \zeta_a\} = -i\sum_c e_{ac}[\{Q_b, Q_c^\dagger\}, \phi(x)] = 2i(\sigma^\mu e)_{ba}[P_\mu, \phi],$$

and so

$$\{Q_b, \zeta_a(x)\} = -2(\sigma^\mu e)_{ba}\partial_\mu\phi(x). \tag{26.1.3}$$

On the other hand, Eq. (26.1.2) and the anticommutation relation (25.2.32) gives

$$-i\sum_c e_{ac}\{Q_b^\dagger, \zeta_c\} = \{Q_b^\dagger, [Q_a^\dagger, \phi]\} = -\{Q_a^\dagger, [Q_b^\dagger, \phi]\} = i\sum_c e_{bc}\{Q_a^\dagger, \zeta_c\},$$

so $\sum_c e_{ac}\{Q_b^\dagger, \zeta_c\}$ is antisymmetric, and therefore proportional to the antisymmetric 2×2 matrix e_{ab}:

$$i\{Q_b^\dagger, \zeta_a(x)\} = 2\delta_{ab}\mathcal{F}(x). \tag{26.1.4}$$

Lorentz invariance requires that the coefficient $\mathcal{F}(x)$ is a scalar field.

We must now go one step further, and calculate the commutators of the supersymmetry generators with $\mathcal{F}(x)$. Using Eqs. (26.1.4), (26.1.2), and (25.2.32), we have

$$\delta_{ab}[Q_c^\dagger, \mathcal{F}] = \tfrac{1}{2}i[Q_c^\dagger, \{Q_b^\dagger, \zeta_a\}] = \tfrac{1}{2}i[\{Q_c^\dagger, \zeta_a\}, Q_b^\dagger] = -\delta_{ac}[Q_b^\dagger, \mathcal{F}].$$

By taking $a = b \neq c$, we find that this commutator vanishes:

$$[Q_c^\dagger, \mathcal{F}(x)] = 0. \tag{26.1.5}$$

* At this point we are not assuming anything about the masses or interactions of the particles described by these fields, but it may be noted that, as explained in Section 5.9, a $(1/2,0)$ free field can create massless particles only of helicity $+1/2$, in agreement with the result (25.5.15) that the massless spinless one-particle state $|0,0\rangle$ that is annihilated by Q_a is in a supermultiplet with a state of helicity $+1/2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Finally, using Eqs. (26.1.4), (25.2.31), and (26.1.3), we have

$$\begin{aligned}\delta_{ab}\,[\mathrm{Q}_c\,,\,\mathscr{F}] &= \tfrac{1}{2}i[\mathrm{Q}_c\,,\,\{\mathrm{Q}_b^*\,,\,\zeta_a\}] = \tfrac{1}{2}i[\{\mathrm{Q}_c\,,\,\mathrm{Q}_b^*\}\,,\,\zeta_a] - \tfrac{1}{2}i[\mathrm{Q}_b^*\,,\,\{\mathrm{Q}_c\,,\,\zeta_a\}]\\ &= -\sigma^\mu_{cb}\,\partial_\mu\zeta_a + i\,(\sigma^\mu e)_{ca}\,[\mathrm{Q}_b^*\,,\,\partial_\mu\phi]\\ &= -\sigma^\mu_{cb}\,\partial_\mu\zeta_a + \sum_d e_{bd}\,(\sigma^\mu e)_{ca}\,\partial_\mu\zeta_d\,.\end{aligned}$$

Contracting with δ_{ab}, this becomes

$$[\mathrm{Q}_c\,,\,\mathscr{F}(x)] = -\sum_a \sigma^\mu_{ca}\,\partial_\mu\zeta_a(x)\,. \tag{26.1.6}$$

Eqs. (26.1.1)–(26.1.6) show that the fields $\phi(x)$, $\zeta_a(x)$, and $\mathscr{F}(x)$ furnish a complete representation of the supersymmetry algebra. These fields are not Hermitian, so their complex conjugates furnish another supermultiplet:

$$[\mathrm{Q}_a^*\,,\,\phi^*(x)] = 0\,, \tag{26.1.7}$$

$$-i\sum_b e_{ab}[\mathrm{Q}_b\,,\,\phi^*(x)] = \zeta_a^*(x)\,, \tag{26.1.8}$$

$$\{\mathrm{Q}_b^*\,,\,\zeta_a^*(x)\} = 2(e\sigma^\mu)_{ab}\partial_\mu\phi^*(x)\,, \tag{26.1.9}$$

$$-i\{\mathrm{Q}_b\,,\,\zeta_a^*(x)\} = 2\delta_{ab}\,\mathscr{F}^*(x)\,, \tag{26.1.10}$$

$$[\mathrm{Q}_c\,,\,\mathscr{F}^*(x)] = 0\,, \tag{26.1.11}$$

$$[\mathrm{Q}_c^*\,,\,\mathscr{F}^*(x)] = \sum_a \sigma^\mu_{ac}\,\partial_\mu\zeta_a^*(x)\,. \tag{26.1.12}$$

We can express these commutation and anticommutation relations as transformation rules under a supersymmetry transformation, which shifts any bosonic or fermionic field operator $\mathscr{O}(x)$ by the infinitesimal amount

$$\delta\mathscr{O}(x) \equiv \left[\sum_a \left(\epsilon_a^*\mathrm{Q}_a + \epsilon_a\mathrm{Q}_a^*\right)\,,\,\mathscr{O}(x)\right]\,, \tag{26.1.13}$$

where ϵ_a is an infinitesimal fermionic c-number spinor. (Because ϵ_a and ϵ_a^* anticommute with Q_a and Q_a^*, the quantity $\epsilon_a^*\mathrm{Q}_a + \epsilon_a\mathrm{Q}_a^*$ is *anti-Hermitian*, so Eq. (26.1.13) gives $(\delta\mathscr{O})^* = \delta\mathscr{O}^*$.) The commutation and anticommutation rules (26.1.1)–(26.1.6) are equivalent to the transformation rules

$$\delta\phi(x) = -i\sum_{ab}\epsilon_a\,e_{ab}\,\zeta_b(x)\,, \tag{26.1.14}$$

$$\delta\zeta_a(x) = -2\sum_b \epsilon_b^*\,(\sigma^\mu e)_{ba}\,\partial_\mu\phi(x) - 2i\epsilon_a\,\mathscr{F}(x)\,, \tag{26.1.15}$$

$$\delta\mathscr{F}(x) = -\sum_{ab}\epsilon_b^*\,\sigma^\mu_{ba}\,\partial_\mu\zeta_a(x)\,. \tag{26.1.16}$$

This may be put into a four-component Dirac notation by introducing

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

an infinitesimal, Majorana,** four-component, spinor transformation parameter

$$\alpha \equiv -i \begin{pmatrix} \epsilon_a \\ \sum_b e_{ab} \epsilon_b^* \end{pmatrix} , \tag{26.1.17}$$

so that Eq. (26.1.13) becomes

$$i\,\delta \mathcal{O}(x) \equiv [\bar{\alpha} Q , \mathcal{O}(x)] . \tag{26.1.18}$$

The transformation rules (26.1.14)–(26.1.16) and their complex conjugates may be put into a convenient covariant form by introducing a set of real bosonic fields A, B, F, and G, defined by

$$\frac{A + iB}{\sqrt{2}} \equiv \phi , \qquad \frac{F - iG}{\sqrt{2}} \equiv \mathscr{F} , \tag{26.1.19}$$

and a four-component Majorana spinor ψ, defined by

$$\psi \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} \zeta_a \\ -\sum_b e_{ab} \zeta_b^* \end{pmatrix} . \tag{26.1.20}$$

Let's also recall the relation between the 4×4 Dirac matrices and the 2×2 matrices σ_μ:

$$\gamma_\mu = \begin{pmatrix} 0 & -i\, e \sigma_\mu^{\mathrm{T}} e \\ i\, \sigma_\mu & 0 \end{pmatrix} .$$

The transformation rules now take the form

$$\begin{aligned} &\delta A = \bar{\alpha}\, \psi , \qquad \delta B = -i\, \bar{\alpha}\, \gamma_5\, \psi , \\ &\delta \psi = \partial_\mu (A + i\gamma_5 B) \gamma^\mu \alpha + (F - i\gamma_5 G)\alpha , \\ &\delta F = \bar{\alpha}\, \gamma^\mu\, \partial_\mu \psi , \qquad \delta G = -i\, \bar{\alpha}\, \gamma_5 \gamma^\mu\, \partial_\mu \psi . \end{aligned} \tag{26.1.21}$$

A direct but tedious calculation shows that this transformation leaves

** With the phase convention we will be using here, a Majorana four-component spinor is formed from a $(1/2, 0)$ two-component spinor u_a as

$$\begin{pmatrix} u \\ -e\, u^* \end{pmatrix} .$$

Eq. (26.1.17) fits this definition, with $u = -i\epsilon$. Equivalently, a Majorana spinor can be formed from a two-component $(0, 1/2)$ spinor v_a as

$$\begin{pmatrix} e\, v^* \\ v \end{pmatrix} .$$

An example is provided by Eq. (25.2.34). Properties of Majorana spinors are considered in detail in the appendix to this chapter.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

invariant the action

$$I = \int d^4x \left\{ -\tfrac{1}{2}\partial_\mu A\,\partial^\mu A - \tfrac{1}{2}\partial_\mu B\,\partial^\mu B - \tfrac{1}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi \right.$$
$$+ \tfrac{1}{2}(F^2+G^2) + m\left[FA+GB-\tfrac{1}{2}\bar{\psi}\psi\right]$$
$$\left. + g\left[F(A^2+B^2)+2GAB-\bar{\psi}(A+i\gamma_5 B)\psi\right]\right\}. \qquad (26.1.22)$$

Eqs. (26.1.21) and (26.1.22) agree with the transformation rules (24.2.8) and Lagrangian density (24.2.9) found in the original work of Wess and Zumino. In the next three sections we will explore a convenient technique for checking the supersymmetry of Eq. (26.1.22) and for deriving more general supersymmetric actions.

Where the fermion field $\psi(x)$ satisfies the free-field Dirac equation $(\gamma^\mu\partial_\mu+m)\psi=0$, these transformation rules show that $F+mA$ and $G+mB$ are invariant, and therefore commute with Q_a and $Q_a^\dagger$, and hence also with P_μ. This does not prove that $F=-mA$ and $G=-mB$, but without changing any of the commutation and anticommutation rules (26.1.1)–(26.1.6) or transformation rules (26.1.21), we can redefine the fields F and G by subtracting the constants $F+mA$ and $B+mG$, respectively, so that the new fields F and G are given by $F=-mA$ and $G=-mB$, and therefore $\mathscr{F}=-m\phi^*$. This is not true in the presence of interactions, but even in the interacting case $\mathscr{F}(x)$, $F(x)$, and $G(x)$ are typically auxiliary fields that can be expressed in terms of the other fields of the supermultiplet, as is the case for the action (26.1.22).

26.2 General Superfields

It is straightforward to construct supermultiplets of fields by the direct technique illustrated in the previous section, but in order to construct supersymmetric actions we also need to know how to multiply field supermultiplets to make other supermultiplets. A great deal of work can be saved by using a formalism invented by Salam and Strathdee,[1] in which the fields in any supermultiplet are assembled into a single superfield.

Just as the four-momentum operators P_μ are defined as the generators of translations of the ordinary spacetime coordinates x^μ, the four supersymmetry generators Q_a and $Q_a^\dagger$ may be regarded as the generators of translations of four fermionic c-number superspace coordinates, which anticommute with each other and with fermionic fields but commute with the x^μ and all bosonic fields. We aim at constructing Lorentz-invariant Lagrangian densities, so it will be convenient to adopt the four-component Dirac formalism described in Section 25.2. The supersymmetry generators are gathered into a four-component Majorana spinor Q_α, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

correspondingly, the superspace coordinates are gathered into another four-component Majorana spinor θ_α. (Various properties of Majorana spinors are outlined in the appendix to this chapter.) The supersymmetry generators have non-vanishing anticommutators, so we cannot take them as simply proportional to the supercoordinate translation operators $\partial/\partial\theta_\alpha$. Instead, Salam and Strathdee found that the supersymmetry algebra would be satisfied if we suppose that the commutator or anticommutator of the supersymmetry generator Q with any bosonic or fermionic superfield $S(x,\theta)$ is

$$[Q\,,\,S\} = i\mathscr{Q}\,S\,, \tag{26.2.1}$$

where $\mathscr{Q}$ is the superspace differential operator

$$\mathscr{Q} \equiv -\frac{\partial}{\partial\bar{\theta}} + \gamma^\mu\,\theta\,\frac{\partial}{\partial x^\mu}\,. \tag{26.2.2}$$

(As usual $\bar{\theta} \equiv \theta^\dagger\beta$. All derivatives with respect to fermionic c-number variables should be understood as *left*-derivatives, calculated by moving the variable to the left of any expression before differentiating with respect to it.) For Majorana spinors $\bar{\theta} = \theta^{\mathrm{T}}\gamma_5\epsilon$, with the 4×4 matrix ϵ given by Eq. (26.A.3), so Eq. (26.2.1) may be expressed more explicitly as

$$\mathscr{Q}_\alpha = \sum_\gamma (\gamma_5\epsilon)_{\alpha\gamma}\,\frac{\partial}{\partial\theta_\gamma} + \sum_\gamma \gamma^\mu_{\alpha\gamma}\theta_\gamma\,\frac{\partial}{\partial x^\mu}\,. \tag{26.2.3}$$

Likewise,

$$\overline{\mathscr{Q}}_\beta = \sum_\gamma \mathscr{Q}_\gamma\,(\gamma_5\,\epsilon)_{\gamma\beta} = \frac{\partial}{\partial\theta_\beta} - \sum_\gamma (\gamma_5\,\epsilon\,\gamma^\mu)_{\beta\gamma}\,\theta_\gamma\,\frac{\partial}{\partial x^\mu}\,. \tag{26.2.4}$$

It is straightforward to calculate that

$$\left\{\mathscr{Q}_\alpha\,,\,\overline{\mathscr{Q}}_\beta\right\} = (\gamma_5\,\epsilon\,\gamma^\mu\,\gamma_5\,\epsilon)_{\beta\alpha}\,\frac{\partial}{\partial x^\mu} + \gamma^\mu_{\alpha\beta}\,\frac{\partial}{\partial x^\mu}\,. \tag{26.2.5}$$

But Eq. (5.4.35) shows that $\gamma_\mu^{\mathrm{T}} = -\mathscr{C}\,\gamma_\mu\,\mathscr{C}^{-1}$, where $\mathscr{C}$ is the matrix $\mathscr{C} = -\gamma_5\epsilon$, so both terms on the right-hand side of Eq. (26.2.5) are equal, and therefore

$$\left\{\mathscr{Q}_\alpha\,,\,\overline{\mathscr{Q}}_\beta\right\} = 2\gamma^\mu_{\alpha\beta}\,\frac{\partial}{\partial x^\mu}\,. \tag{26.2.6}$$

Eqs. (26.2.6) and (26.2.1) together with the generalized Jacobi identities (25.1.5) show that

$$[\{Q_\alpha\,,\,\overline{Q}_\beta\}\,,\,S] = \{\mathscr{Q}_\alpha\,,\,\overline{\mathscr{Q}}_\beta\}\,S = 2\gamma^\mu_{\alpha\beta}\partial_\mu S = -2i\gamma^\mu_{\alpha\beta}[P_\mu\,,\,S]\,, \tag{26.2.7}$$

in agreement with the anticommutation relation (25.2.36).

It is often more convenient to express the commutation and anticommutation relations (26.2.1) as transformation rules under infinitesimal

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

supersymmetry transformations. Combining Eqs. (26.1.18), (26.2.1), and (26.2.2) shows that a supersymmetry transformation with infinitesimal Majorana spinor parameter α changes a superfield $S(x,\theta)$ by an amount

$$\delta S = (\bar{\alpha}\,\mathcal{Q})\,S = -\left(\bar{\alpha}\,\frac{\partial S}{\partial\bar{\theta}}\right) + (\bar{\alpha}\,\gamma^{\mu}\,\theta)\,\frac{\partial S}{\partial x^{\mu}}\,. \tag{26.2.8}$$

Recall that $\partial/\partial\bar{\theta}$ here acts on the left of any expression. In particular, where M is any linear combination of the matrices 1, $\gamma_5\gamma_\mu$, and γ_5 for which $\bar{\theta}M\theta$ does not vanish, we have $\bar{\theta}'M\,\theta'' = \bar{\theta}''M\,\theta'$, so

$$\frac{\partial}{\partial\bar{\theta}}\,(\bar{\theta}M\theta) = 2M\theta\,. \tag{26.2.9}$$

The components of θ anticommute, so any product of their components vanishes if two of them are the same component. But θ has only four components, so any function of θ has a power series that terminates with its quartic term. Furthermore, as shown in the appendix to this chapter, a product of two θs is proportional to a linear combination of $(\bar{\theta}\theta)$, $(\bar{\theta}\gamma_\mu\gamma_5\theta)$, and $(\bar{\theta}\gamma_5\theta)$; a product of three θs is proportional to $(\bar{\theta}\gamma_5\theta)\theta$; and a product of four θs is proportional to $(\bar{\theta}\gamma_5\theta)^2$. The most general function of x^μ and θ may therefore be expressed as

$$\begin{aligned} S(x,\theta) = {} & C(x) - i\big(\bar{\theta}\,\gamma_5\,\omega(x)\big) - \frac{i}{2}\big(\bar{\theta}\,\gamma_5\,\theta\big)M(x) - \frac{1}{2}\big(\bar{\theta}\,\theta\big)N(x) \\ & + \frac{i}{2}\big(\bar{\theta}\,\gamma_5\,\gamma_\mu\,\theta\big)V^{\mu}(x) - i\big(\bar{\theta}\,\gamma_5\,\theta\big)\left(\bar{\theta}\left[\lambda(x) + \frac{1}{2}\,\partial\!\!\!/\omega(x)\right]\right) \\ & - \frac{1}{4}\big(\bar{\theta}\,\gamma_5\,\theta\big)^2\left(D(x) + \frac{1}{2}\Box C(x)\right). \end{aligned} \tag{26.2.10}$$

(The terms $\frac{1}{2}\partial\!\!\!/\omega$ and $\frac{1}{2}\Box C(x)$ are separated from $\lambda(x)$ and $D(x)$, respectively, for later convenience.) If $S(x,\theta)$ is a scalar then $C(x)$, $M(x)$, $N(x)$, and $D(x)$ are scalar (or pseudoscalar) fields; $\omega(x)$ and $\lambda(x)$ are four-component spinor fields; and $V^\mu(x)$ is a vector field. Also, using the reality properties of bilinear products of Majorana fields given in the appendix to this chapter, we see that if $S(x,\theta)$ is real, then $C(x)$, $M(x)$, $N(x)$, $V^\mu(x)$, and $D(x)$ are all real, while $\omega(x)$ and $\lambda(x)$ are Majorana spinors satisfying the phase convention $s^* = -\beta\epsilon\gamma_5 s$.

Now we must work out the supersymmetry transformation properties of the component fields in Eq. (26.2.10). Applying Eqs. (26.2.8) and (26.2.9)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

to the expansion (26.2.10) gives

$$\begin{aligned}\delta S = {} & (\bar\alpha\gamma^\mu\theta)\frac{\partial C}{\partial x^\mu} \\ & + i(\bar\alpha\gamma_5\omega) - i(\bar\alpha\gamma^\mu\theta)\left(\bar\theta\gamma_5\frac{\partial\omega}{\partial x^\mu}\right) \\ & + i(\bar\alpha\gamma_5\theta)M - \frac{i}{2}(\bar\alpha\gamma^\mu\theta)\left(\bar\theta\gamma_5\theta\right)\frac{\partial M}{\partial x^\mu} \\ & + (\bar\alpha\theta)N - \frac{1}{2}(\bar\alpha\gamma^\mu\theta)\left(\bar\theta\theta\right)\frac{\partial N}{\partial x^\mu} \\ & - i(\bar\alpha\gamma_5\gamma_\nu\theta)V^\nu + \frac{i}{2}(\bar\alpha\gamma^\mu\theta)\left(\bar\theta\gamma_5\gamma_\nu\theta\right)\frac{\partial V^\nu}{\partial x^\mu} \\ & + 2i(\bar\alpha\gamma_5\theta)\left(\bar\theta[\lambda + \tfrac{1}{2}\not\partial\omega]\right) + i\left(\bar\theta\gamma_5\theta\right)(\bar\alpha[\lambda + \tfrac{1}{2}\not\partial\omega]) \\ & - i(\bar\alpha\gamma^\mu\theta)\left(\bar\theta\gamma_5\theta\right)\left(\bar\theta\partial_\mu[\lambda + \tfrac{1}{2}\not\partial\omega]\right) + \left(\bar\theta\gamma_5\theta\right)(\bar\alpha\gamma_5\theta)[D + \tfrac{1}{2}\Box C]\,.\end{aligned}$$

We need to put each term in the standard form of Eq. (26.2.10). For this purpose, we note that: Eq. (26.A.9) gives

$$(\bar\alpha\gamma^\mu\theta)(\bar\theta\gamma_5\partial_\mu\omega) = -\tfrac{1}{4}(\bar\theta\theta)(\bar\alpha\not\partial\gamma_5\omega) - \tfrac{1}{4}(\bar\theta\gamma_5\gamma^\nu\theta)(\bar\alpha\not\partial\gamma_\nu\omega) - \tfrac{1}{4}(\bar\theta\gamma_5\theta)(\bar\alpha\not\partial\omega)\ ;$$

Eq. (26.A.16) gives

$$(\bar\alpha\gamma^\mu\theta)(\bar\theta\theta) = -(\bar\alpha\gamma^\mu\gamma_5\theta)(\bar\theta\gamma_5\theta)\ ;$$

Eq. (26.A.17) gives

$$(\bar\alpha\gamma^\mu\theta)(\bar\theta\gamma_5\gamma_\nu\theta) = -(\bar\alpha\gamma^\mu\gamma_\nu\theta)(\bar\theta\gamma_5\theta)\ ;$$

Eq. (26.A.9) gives

$$\begin{aligned}(\bar\alpha\gamma_5\theta)(\bar\theta[\lambda + \tfrac{1}{2}\not\partial\omega]) = {} & -\tfrac{1}{4}(\bar\theta\theta)(\bar\alpha\gamma_5[\lambda + \tfrac{1}{2}\not\partial\omega]) + \tfrac{1}{4}(\bar\theta\gamma_5\gamma^\mu\theta)(\bar\alpha\gamma_\mu[\lambda + \tfrac{1}{2}\not\partial\omega]) \\ & - \tfrac{1}{4}(\bar\theta\gamma_5\theta)(\bar\alpha[\lambda + \tfrac{1}{2}\not\partial\omega])\ ;\end{aligned}$$

and Eq. (26.A.19) gives

$$(\bar\alpha\gamma^\mu\theta)(\bar\theta\gamma_5\theta)(\bar\theta\partial_\mu[\lambda + \tfrac{1}{2}\not\partial\omega]) = -\tfrac{1}{4}(\bar\alpha\not\partial\gamma_5[\lambda + \tfrac{1}{2}\not\partial\omega])(\bar\theta\gamma_5\theta)^2\,.$$

Using these relations and rearranging terms in order of increasing numbers of θ factors, we have

$$\begin{aligned}\delta S = {} & i(\bar\alpha\gamma_5\omega) + (\bar\alpha[\not\partial C + i\gamma_5 M + N - i\gamma_5\not V]\theta) \\ & - \tfrac{1}{2}i\left(\bar\theta\theta\right)(\bar\alpha\gamma_5[\lambda + \not\partial\omega]) + \tfrac{1}{2}i\left(\bar\theta\gamma_5\theta\right)(\bar\alpha[\lambda + \not\partial\omega]) \\ & + \tfrac{1}{2}i\left(\bar\theta\gamma_5\gamma^\mu\theta\right)(\bar\alpha\gamma_\mu\lambda) + \tfrac{1}{2}i\left(\bar\theta\gamma_5\gamma^\nu\theta\right)(\bar\alpha\partial_\nu\omega) \\ & + \tfrac{1}{2}\left(\bar\theta\gamma_5\theta\right)(\bar\alpha[-i\not\partial M - \gamma_5\not\partial N - i\not\partial\not V + \gamma_5(D + \tfrac{1}{2}\Box C)]\theta) \\ & - \tfrac{1}{4}i\left(\bar\theta\gamma_5\theta\right)^2(\bar\alpha\gamma_5[\not\partial\lambda + \tfrac{1}{2}\Box\omega])\ ,\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

or, using the symmetry property (26.A.7),

$$\begin{aligned}\delta S = \; & i(\bar{\alpha}\gamma_5\omega) + \left(\bar{\theta}[-\not{\partial}C + i\gamma_5 M + N - i\gamma_5 \not{V}]\alpha\right) \\ & - \tfrac{1}{2}i\left(\bar{\theta}\theta\right)(\bar{\alpha}\gamma_5[\lambda + \not{\partial}\omega]) + \tfrac{1}{2}i\left(\bar{\theta}\gamma_5\theta\right)(\bar{\alpha}[\lambda + \not{\partial}\omega]) \\ & + \tfrac{1}{2}i\left(\bar{\theta}\gamma_5\gamma^\mu\theta\right)(\bar{\alpha}\gamma_\mu\lambda) + \tfrac{1}{2}i\left(\bar{\theta}\gamma_5\gamma^\nu\theta\right)(\bar{\alpha}\partial_\nu\omega) \\ & + \tfrac{1}{2}\left(\bar{\theta}\gamma_5\theta\right)\left(\bar{\theta}[i\not{\partial}M - \gamma_5\not{\partial}N - i\partial_\mu \not{V}\gamma^\mu + \gamma_5(D + \tfrac{1}{2}\Box C)]\alpha\right) \\ & - \tfrac{1}{4}i\left(\bar{\theta}\gamma_5\theta\right)^2(\bar{\alpha}\gamma_5[\not{\partial}\lambda + \tfrac{1}{2}\Box\omega]) \; .\end{aligned}$$

If we now compare this with the terms up to second order in θ in the expansion (26.2.10), we find the transformation rules:

$$\delta C = i\left(\bar{\alpha}\gamma_5\omega\right), \tag{26.2.11}$$

$$\delta\omega = \left(-i\gamma_5\not{\partial}C - M + i\gamma_5 N + \not{V}\right)\alpha, \tag{26.2.12}$$

$$\delta M = -\left(\bar{\alpha}\left[\lambda + \not{\partial}\omega\right]\right), \tag{26.2.13}$$

$$\delta N = i\left(\bar{\alpha}\gamma_5\left[\lambda + \not{\partial}\omega\right]\right), \tag{26.2.14}$$

$$\delta V_\mu = \left(\bar{\alpha}\gamma_\mu\lambda\right) + \left(\bar{\alpha}\partial_\mu\omega\right). \tag{26.2.15}$$

The terms of third and fourth order in θ give

$$\begin{aligned}\delta[\lambda + \tfrac{1}{2}\not{\partial}\omega] &= \tfrac{1}{2}\left[-\not{\partial}M - i\gamma_5\not{\partial}N + \partial_\mu \not{V}\gamma^\mu + i\gamma_5\left(D + \tfrac{1}{2}\Box C\right)\right]\alpha, \\ \delta[D + \tfrac{1}{2}\Box C] &= i\left(\bar{\alpha}\gamma_5[\not{\partial}\lambda + \tfrac{1}{2}\Box\omega]\right).\end{aligned}$$

Combining the latter two transformation rules with the transformation rules (26.2.11) and (26.2.12) for C and ω yields much simpler transformation rules for λ and D:

$$\delta\lambda = \left(\tfrac{1}{2}\left[\partial_\mu \not{V}, \gamma^\mu\right] + i\gamma_5 D\right)\alpha, \tag{26.2.16}$$

$$\delta D = i\left(\bar{\alpha}\gamma_5\not{\partial}\lambda\right). \tag{26.2.17}$$

It is to achieve this simplification that the terms $\frac{1}{2}\not{\partial}\omega$ and $\frac{1}{2}\Box C$ were separated from λ and D in the expansion (26.2.10).

The whole point of the superfield formalism is to simplify the task of making supermultiplets out of other supermultiplets. Given two superfields S_1 and S_2 that both satisfy the transformation rule (26.2.8), their product $S \equiv S_1 S_2$ satisfies

$$\begin{aligned}\delta S \equiv [(\bar{\alpha}Q), S_1 S_2] &= (\delta S_1)S_2 + S_1(\delta S_2) \\ &= \left((\bar{\alpha}\not{2})S_1\right)S_2 + S_1\left((\bar{\alpha}\not{2})\right)S_2 = (\bar{\alpha}\not{2})S,\end{aligned} \tag{26.2.18}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

and is therefore also a superfield. A straightforward calculation using Eqs. (26.A.7), (26.A.16), (26.A.18), and (26.A.19) gives its components as

$$C = C_1 C_2 \,, \tag{26.2.19}$$

$$\omega = C_1\omega_2 + C_2\omega_1 \,, \tag{26.2.20}$$

$$M = C_1 M_2 + C_2 M_1 + \tfrac{1}{2} i \left(\overline{\omega_1}\,\gamma_5\,\omega_2\right) , \tag{26.2.21}$$

$$N = C_1 N_2 + C_2 N_1 - \tfrac{1}{2} \left(\overline{\omega_1}\,\omega_2\right) , \tag{26.2.22}$$

$$V^\mu = C_1 V_2^\mu + C_2 V_1^\mu - \tfrac{1}{2} i \left(\overline{\omega_1}\,\gamma_5 \gamma^\mu \omega_2\right) , \tag{26.2.23}$$

$$\begin{aligned}\lambda = {} & C_1\lambda_2 + C_2\lambda_1 - \tfrac{1}{2}\gamma^\mu\omega_1\partial_\mu C_2 - \tfrac{1}{2}\gamma^\mu\omega_2\partial_\mu C_1 + \tfrac{1}{2} i \not{V}_1\gamma_5\,\omega_2 + \tfrac{1}{2} i \not{V}_2\gamma_5\,\omega_1 \\ & + \tfrac{1}{2}(N_1 - i\gamma_5 M_1)\,\omega_2 + \tfrac{1}{2}(N_2 - i\gamma_5 M_2)\,\omega_1 \,,\end{aligned} \tag{26.2.24}$$

$$\begin{aligned}D = {} & -\partial_\mu C_1\,\partial^\mu C_2 + C_1 D_2 + C_2 D_1 + M_1 M_2 + N_1 N_2 \\ & - \left(\overline{\omega_1}\,[\lambda_2 + \tfrac{1}{2}\not{\partial}\omega_2]\right) - \left(\overline{\omega_2}\,[\lambda_1 + \tfrac{1}{2}\not{\partial}\omega_1]\right) - V_{1\mu}V_2^\mu \,.\end{aligned} \tag{26.2.25}$$

It is trivial that linear combinations of superfields are superfields, in the same sense, and that spacetime derivatives and complex conjugates of superfields are superfields. But multiplying a superfield by some function of θ or differentiating it with respect to θ does not in general yield a superfield. (For instance, θ itself is evidently not a superfield, because θ is a fermionic c-number and therefore commutes with $\bar{\alpha}Q$, while $\mathscr{Q}\theta \neq 0$.) There is, however, a way of combining a derivative of a superfield with respect to θ and multiplying it by a factor θ which does yield another superfield.

Consider the superspace differential operator $\mathscr{D}_\alpha$ defined by

$$\mathscr{D} \equiv -\frac{\partial}{\partial\bar{\theta}} - \gamma^\mu\,\theta\,\frac{\partial}{\partial x^\mu} \,, \tag{26.2.26}$$

or more explicitly

$$\mathscr{D}_\alpha = \sum_\gamma (\gamma_5\epsilon)_{\alpha\gamma}\,\frac{\partial}{\partial\theta_\gamma} - \sum_\gamma \gamma^\mu_{\alpha\gamma}\theta_\gamma\,\frac{\partial}{\partial x^\mu} \,. \tag{26.2.27}$$

The only difference between the definitions of $\mathscr{D}$ and $\mathscr{Q}$ is a change of sign of the terms involving the spacetime derivative. In consequence of this change of sign, in the anticommutator of $\mathscr{D}_\beta$ with $\mathscr{Q}_\alpha$, instead of getting two equal terms like those in Eq. (26.2.5) with the same sign, we get them with opposite signs, so that they cancel:

$$\{\mathscr{D}_\beta\,,\,\mathscr{Q}_\alpha\} = 0 \,. \tag{26.2.28}$$

Since α is fermionic, it follows that $(\bar{\alpha}\mathscr{Q})$ commutes with $\mathscr{D}_\beta$, so if $S(x,\theta)$ is a superfield, then

$$\delta\mathscr{D}_\beta S \equiv -i[(\bar{\alpha}Q)\,,\,\mathscr{D}_\beta S] = -i\mathscr{D}_\beta[(\bar{\alpha}Q)\,,\,S] = \mathscr{D}_\beta(\bar{\alpha}\mathscr{Q})S = (\bar{\alpha}\mathscr{Q})\mathscr{D}_\beta S \,, \tag{26.2.29}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

so that $\mathscr{D}_\beta S$ is also a superfield. *Thus an arbitrary polynomial function of superfields S and their superderivatives* $\mathscr{D}_\beta S$, $\mathscr{D}_\beta \mathscr{D}_\gamma S$, *etc. is also a superfield.*

It is not necessary to add here that in constructing a superfield out of other superfields we can also include their spacetime derivatives, because these can be obtained from second superderivatives. Since the only difference between $\mathscr{D}_\beta$ and $\mathscr{Q}_\beta$ is in the sign of the term involving ∂_μ, the anticommutators of the $\mathscr{D}$s are the same as those of the $\mathscr{Q}$s, except for a change of sign:

$$\left\{\mathscr{D}_\alpha, \overline{\mathscr{D}}_\beta\right\} = -2\gamma^\mu_{\alpha\beta} \frac{\partial}{\partial x^\mu} . \tag{26.2.30}$$

Now let us consider how to construct a supersymmetric action out of superfields. There is no such thing as a supersymmetric Lagrangian density, because the anticommutation relation (26.2.6) shows that if $\delta\mathscr{L} = 0$ then $\mathscr{L}$ must be a constant. Even if the Lagrangian density is not supersymmetric, the action will still be supersymmetric if $\delta\mathscr{L}(x)$ is a derivative, which would not contribute to $\delta \int \mathscr{L}\, d^4x$. In general, the Lagrangian density $\mathscr{L}$ can be written as a sum of terms, each of which is some component of a superfield that is constructed out of elementary superfields and their superderivatives. Inspection of the transformation rules (26.2.11)–(26.2.17) for the individual components shows that in the absence of any special conditions on a general superfield, the only component of such a superfield whose variation is a derivative is the D-component. Also, for the D-component of any superfield to be a scalar, the superfield itself must be a scalar. Therefore, unless there are special conditions on the individual superfields from which the Lagrangian density is constructed, a supersymmetric action can only be the integral of the D-term of a scalar superfield Λ:

$$I = \int d^4x\, [\Lambda]_D . \tag{26.2.31}$$

But in fact no action of this sort would be physically satisfactory without special conditions on the superfields from which it is constructed. For a general superfield $S(x,\theta)$, the only sort of supersymmetric kinematic action I_0 that is bilinear in S and S^* and involves no more than two derivatives of the component fields is of the form

$$I_0 \propto \int d^4x \left[S^* S\right]_D . \tag{26.2.32}$$

From Eq. (26.2.25), we see that $S^* S$ has the D-component

$$\begin{aligned}\left[S^* S\right]_D = & -\partial_\mu C^* \partial^\mu C - \tfrac{1}{2}\left(\bar{\omega}\gamma^\mu \partial_\mu \omega\right) + \tfrac{1}{2}\left((\partial_\mu \bar{\omega})\gamma^\mu \omega\right) \\ & + C^* D + D^* C - \left(\bar{\omega}\lambda\right) - \left(\bar{\lambda}\omega\right) \\ & + M^* M + N^* N - V^*_\mu V^\mu . \end{aligned} \tag{26.2.33}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

The terms quadratic in C or ω look promising as the kinematic Lagrangians for massless fields of spins zero and 1/2; the final three terms are harmless; but the terms involving D or λ have the disastrous effect in path integrals of constraining C and ω to vanish. Fortunately, as we shall see in the following section, there are constrained superfields from which we *can* construct physically sensible actions. The introduction of these constrained superfields will also open up ways of constructing supersymmetric terms in the action that are not the D-components of functions of superfields.

If parity is conserved, then the space inversion properties of the component fields of a superfield will be related by supersymmetry. To work out this relation, we apply the parity operator P to the commutation/anticommutation relation (26.2.1) and use the transformation property (25.3.16) of supersymmetry generators, which gives

$$i\beta \left[Q, \mathsf{P}^{-1} S(x,\theta) \mathsf{P} \right\} = \not{2}\, \mathsf{P}^{-1} S(x,\theta) \mathsf{P} \,. \tag{26.2.34}$$

The solution of Eq. (26.2.34) for a scalar superfield is of the form

$$\mathsf{P}^{-1} S(x,\theta) \mathsf{P} = \eta\, S(\Lambda_P x, -i\beta\theta) \,, \tag{26.2.35}$$

where η is some phase (the intrinsic parity of the superfield) and $\Lambda_P x \equiv (-\mathbf{x}, +x^0)$. (To check that Eq. (26.2.35) satisfies Eq. (26.2.34), note that Eq. (26.2.35) gives the left-hand side of Eq. (26.2.34) as

$$i\eta\beta \left(-\frac{\partial}{\partial(-i\beta\theta)} + \gamma^\mu (-i\beta\theta) \frac{\partial}{\partial(\Lambda_P x)^\mu} \right) S(\Lambda_P x, \theta) = \eta\, \not{2}\, S(\Lambda_P x, -i\beta\theta) \,,$$

in agreement with what Eq. (26.2.35) gives for the right-hand side of Eq. (26.2.34).) Using the expansion (26.2.10) in Eq. (26.2.35) then gives the space inversion properties of the component fields:

$$\begin{aligned}
&\mathsf{P}^{-1} C(x) \mathsf{P} = \eta\, C(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} \omega(x) \mathsf{P} = -i\eta\, \beta\, \omega(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} M(x) \mathsf{P} = -\eta\, M(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} N(x) \mathsf{P} = \eta\, N(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} V^\mu(x) \mathsf{P} = -\eta\, (\Lambda_P)^\mu{}_\nu V^\nu(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} \lambda(x) \mathsf{P} = i\eta\, \beta\, \lambda(\Lambda_P x) \,, \\
&\mathsf{P}^{-1} D(x) \mathsf{P} = \eta\, D(\Lambda_P x) \,.
\end{aligned} \tag{26.2.36}$$

* * *

The general real superfield S involves four real spinless fields C, M, N, and D, plus one real four-vector field V_μ, for a total of eight independent bosonic field components. For comparison, there are two four-component Majorana spinor fields ω and λ, also for a total of eight independent

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

field components. The equality of the numbers of independent bosonic and fermionic field components holds in general not only for the unconstrained general superfield studied in this section, but also for all superfields obtained from the general superfield by imposing supersymmetric constraints, such as the chiral and other constrained superfields discussed in the next section.

To see this in general, suppose that we have a representation of the supersymmetry algebra provided by N_B linearly independent real bosonic field operators $b_n(x)$ and N_F linearly independent fermionic field operators $f_k(x)$. We will assume that these fields satisfy only non-trivial field equations, so that no linear combination of the b_n or f_k with non-vanishing coefficients satisfies a homogeneous linear field equation. Consider a real supersymmetry generator $Q(u)$, defined as

$$Q(u) \equiv \left(\bar{u}\, Q\right) = \left(\bar{Q}\, u\right), \tag{26.2.37}$$

where u is some ordinary numerical Majorana spinor (*not* an anticommuting c-number). (For extended supersymmetry, in place of Q_α we could use any one of the $Q_{r\alpha}$, say $Q_{1\alpha}$.) In order for the b_n and f_k to furnish a representation of the supersymmetry algebra, we must have

$$[\,Q(u)\,,\, b_n\,] = i\sum_k q_{nk}(\partial)\, f_k\,, \tag{26.2.38}$$

$$\{\,Q(u)\,,\, f_k\,\} = \sum_n p_{kn}(\partial)\, b_n\,, \tag{26.2.39}$$

for some matrix differential operators $q(\partial)$ and $p(\partial)$. Taking the anticommutator of Eq. (26.2.38) and the commutator of Eq. (26.2.39) with $Q(u)$ gives

$$[\,Q^2(u)\,,\, b_n\,] = i\sum_m \Big(q(\partial)\, p(\partial)\Big)_{nm}\, b_m\,, \tag{26.2.40}$$

$$[\,Q^2(u)\,,\, f_k\,] = i\sum_\ell \Big(p(\partial)\, q(\partial)\Big)_{k\ell}\, f_\ell\,. \tag{26.2.41}$$

The anticommutation relation (25.2.36) or (25.2.38) gives the square of $Q(u)$ as $Q^2(u) = -iP_\mu\left(\bar{u}\gamma^\mu u\right)$. Hence the square matrices $p(\partial)q(\partial)$ and $q(\partial)p(\partial)$ must both be non-singular, because if there were any non-vanishing coefficients $c_n(\partial)$ or $d_k(\partial)$ for which $\sum_n c_n(\partial)(q(\partial)p(\partial))_{nm} = 0$ or $\sum_k d_k(\partial)(p(\partial)q(\partial))_{k\ell} = 0$ then b_n or f_k would satisfy the homogeneous linear field equations

$$\left(\bar{u}\gamma^\mu u\right)\partial_\mu \sum_n c_n(\partial) b_n = 0 \quad \text{or} \quad \left(\bar{u}\gamma^\mu u\right)\partial_\mu \sum_k d_k(\partial) f_k = 0\,,$$

in contradiction with our assumption that the fields do not satisfy such field equations. In order for qp to be non-singular we must have $N_F \geq N_B$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

and in order for pq to be non-singular we must have $N_B \geq N_F$, so we can conclude that $N_B = N_F$. Also the square matrices q and p must both be non-singular, so the complex conjugate of Eq. (26.2.38) tells us that $f^* = q^{*-1}qf$, so that the number of independent fermion fields is N_F rather than $2N_F$, and is therefore equal to the number N_B of independent boson fields, as was to be shown.

26.3 Chiral and Linear Superfields

In the previous section we found that the presence of D and λ components in a general superfield stood in the way of using such superfields in a physically satisfactory Lagrangian density. Suppose then that we consider a superfield with

$$\lambda = D = 0 \,. \tag{26.3.1}$$

Are these conditions preserved by supersymmetry transformations? According to Eqs. (26.2.17) and (26.2.16), the condition $D = 0$ is invariant if $\lambda = 0$, but the condition $\lambda = 0$ is only invariant if we also impose the condition that $\partial_\mu V_\nu - \partial_\nu V_\mu = 0$, which requires that V_μ be a pure gauge:

$$V_\mu(x) = \partial_\mu Z(x) \,. \tag{26.3.2}$$

Eq. (26.2.15) shows that, with $\lambda = 0$, this condition is preserved by a supersymmetry transformation. We have thus arrived at a reduced superfield, subject to the constraints (26.3.1) and (26.3.2), with component fields having the transformation properties

$$\delta C = i\left(\bar{\alpha}\gamma_5\omega\right), \tag{26.3.3}$$

$$\delta\omega = \left(-i\gamma_5\not{\partial}C - M + i\gamma_5 N + \not{\partial}Z\right)\alpha, \tag{26.3.4}$$

$$\delta M = -\left(\bar{\alpha}\not{\partial}\omega\right), \tag{26.3.5}$$

$$\delta N = i\left(\bar{\alpha}\gamma_5\not{\partial}\omega\right), \tag{26.3.6}$$

$$\delta Z = \left(\bar{\alpha}\omega\right). \tag{26.3.7}$$

Comparing with Eq. (26.1.21), we see that this is the same as the supermultiplet constructed by direct methods in Section 26.1, with the identifications

$$C = A, \quad \omega = -i\gamma_5\psi, \quad M = G, \quad N = -F, \quad Z = B \,. \tag{26.3.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

A superfield satisfying the conditions (26.3.1) and (26.3.2) is said to be *chiral.*[*]

To distinguish a chiral superfield $X(x,\theta)$ from the general superfield $S(x,\theta)$ of the previous section, we will use A, B, F, G, and ψ for its components, instead of C, M, N, Z, and ω. By using Eqs. (26.3.1), (26.3.2), and (26.3.8) in Eq. (26.2.10), we find the form of a general chiral superfield to be

$$\begin{aligned} X(x,\theta) = {} & A(x) - \left(\bar{\theta}\,\psi(x)\right) + \frac{1}{2}\left(\bar{\theta}\,\theta\right)F(x) - \frac{i}{2}\left(\bar{\theta}\,\gamma_5\,\theta\right)G(x) \\ & + \frac{i}{2}\left(\bar{\theta}\,\gamma_5\,\gamma_\mu\,\theta\right)\partial^\mu B(x) + \frac{1}{2}\left(\bar{\theta}\,\gamma_5\,\theta\right)\left(\bar{\theta}\gamma_5\,\not{\partial}\psi(x)\right) \\ & - \frac{1}{8}\left(\bar{\theta}\,\gamma_5\,\theta\right)^2\Box A(x)\,. \end{aligned} \tag{26.3.9}$$

(We could just as well have taken $C = -B$, $\omega = \psi$, $M = -F$, $N = -G$, and $Z = A$. We make the identifications (26.3.8) because, as we see here, for a scalar superfield they are consistent with the usual convention that A and F are scalars while B and G are pseudoscalars.)

The chiral superfield (26.3.9) may be further decomposed, as

$$X(x,\theta) = \frac{1}{\sqrt{2}}\left[\Phi(x,\theta) + \tilde{\Phi}(x,\theta)\right], \tag{26.3.10}$$

where

$$\begin{aligned} \Phi(x,\theta) = {} & \phi(x) - \sqrt{2}\left(\bar{\theta}\psi_L(x)\right) + \mathscr{F}(x)\left(\bar{\theta}\left(\frac{1+\gamma_5}{2}\right)\theta\right) + \frac{1}{2}\left(\bar{\theta}\gamma_5\gamma_\mu\theta\right)\partial^\mu\phi(x) \\ & - \frac{1}{\sqrt{2}}\left(\bar{\theta}\gamma_5\theta\right)\left(\bar{\theta}\,\not{\partial}\psi_L(x)\right) - \frac{1}{8}\left(\bar{\theta}\gamma_5\theta\right)^2\Box\phi(x)\,, \end{aligned} \tag{26.3.11}$$

$$\begin{aligned} \tilde{\Phi}(x,\theta) = {} & \tilde{\phi}(x) - \sqrt{2}\left(\bar{\theta}\psi_R(x)\right) + \tilde{\mathscr{F}}(x)\left(\bar{\theta}\left(\frac{1-\gamma_5}{2}\right)\theta\right) - \frac{1}{2}\left(\bar{\theta}\gamma_5\gamma_\mu\theta\right)\partial^\mu\tilde{\phi}(x) \\ & + \frac{1}{\sqrt{2}}\left(\bar{\theta}\gamma_5\theta\right)\left(\bar{\theta}\,\not{\partial}\psi_R(x)\right) - \frac{1}{8}\left(\bar{\theta}\gamma_5\theta\right)^2\Box\tilde{\phi}(x)\,, \end{aligned} \tag{26.3.12}$$

with component fields defined by

$$\phi \equiv \frac{A+iB}{\sqrt{2}}\,, \qquad \psi_L \equiv \left(\frac{1+\gamma_5}{2}\right)\psi\,, \qquad \mathscr{F} \equiv \frac{F-iG}{\sqrt{2}}\,, \tag{26.3.13}$$

[*] Some authors use the term 'chiral' to describe a special case of such superfields, introduced below, which are here called left-chiral or right-chiral. Our use here of the term chiral may at first seem strange, because it has no counterpart for Dirac spinors. Any Dirac spinor is the sum of Dirac spinors that are left-chiral and right-chiral, in the sense that they are respectively proportional to $1+\gamma_5$ and $1-\gamma_5$, so no special term is needed for such sums of Dirac spinors. In contrast, it is only superfields satisfying Eqs. (26.3.1) and (26.3.2) that can be expressed as the sum of a left-chiral and a right-chiral superfield.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$$\tilde{\phi} \equiv \frac{A - iB}{\sqrt{2}}\,, \qquad \psi_R \equiv \left(\frac{1-\gamma_5}{2}\right)\psi\,, \qquad \tilde{\mathscr{F}} \equiv \frac{F + iG}{\sqrt{2}}\,. \tag{26.3.14}$$

The component fields of either Φ or $\tilde{\Phi}$ furnish complete representations of the supersymmetry algebra:

$$\delta\psi_L = \sqrt{2}\partial_\mu\phi\,\gamma^\mu\,\alpha_R + \sqrt{2}\mathscr{F}\,\alpha_L\,, \tag{26.3.15}$$

$$\delta\mathscr{F} = \sqrt{2}\left(\overline{\alpha_L}\,\partial\!\!\!/\psi_L\right), \tag{26.3.16}$$

$$\delta\phi = \sqrt{2}\left(\overline{\alpha_R}\psi_L\right), \tag{26.3.17}$$

$$\delta\psi_R = \sqrt{2}\partial_\mu\tilde{\phi}\,\gamma^\mu\,\alpha_L + \sqrt{2}\tilde{\mathscr{F}}\,\alpha_R\,, \tag{26.3.18}$$

$$\delta\tilde{\mathscr{F}} = \sqrt{2}\left(\overline{\alpha_R}\,\partial\!\!\!/\psi_R\right), \tag{26.3.19}$$

$$\delta\tilde{\phi} = \sqrt{2}\left(\overline{\alpha_L}\psi_R\right), \tag{26.3.20}$$

where as usual

$$\alpha_L = \left(\frac{1+\gamma_5}{2}\right)\alpha\,, \qquad \alpha_R = \left(\frac{1-\gamma_5}{2}\right)\alpha\,,$$

and likewise for θ. A superfield of the form (26.3.11) or (26.3.12) is known as *left-chiral* or *right-chiral*, respectively. In the special case where a chiral superfield $X(x,\theta)$ is *real*, its left-chiral and right-chiral parts Φ and $\tilde{\Phi}$ are complex conjugates, so that $\tilde{\phi} = \phi^*$, $\tilde{\mathscr{F}} = \mathscr{F}^*$, and ψ is a Majorana field. However, if we do not require $X(x,\theta)$ to be real then in general there is no relation between Φ and $\tilde{\Phi}$; it is even possible that one of the two vanishes.

The component fields of the superfield Φ include two complex bosonic components ϕ and $\mathscr{F}$, or four independent real bosonic components, and one Majorana fermion field ψ, which has four independent fermionic components. This is another example of the general result, derived at the end of the previous section, that any set of fields that furnish a representation of the supersymmetry algebra must have an equal number of independent bosonic and fermionic components.

We can use Eqs. (26.A.5), (26.A.17), and (26.A.18) to rewrite Eqs. (26.3.11) and (26.3.12) in a form that clarifies the way that these superfields depend on θ_L and θ_R:

$$\Phi(x,\theta) = \phi(x_+) - \sqrt{2}\left(\theta_L^{\mathrm{T}}\epsilon\,\psi_L(x_+)\right) + \mathscr{F}(x_+)\left(\theta_L^{\mathrm{T}}\epsilon\,\theta_L\right), \tag{26.3.21}$$

$$\tilde{\Phi}(x,\theta) = \tilde{\phi}(x_-) + \sqrt{2}\left(\theta_R^{\mathrm{T}}\epsilon\,\psi_R(x_-)\right) - \tilde{\mathscr{F}}(x_-)\left(\theta_R^{\mathrm{T}}\epsilon\,\theta_R\right), \tag{26.3.22}$$

where

$$x_\pm^\mu \equiv x^\mu \pm \tfrac{1}{2}\left(\bar{\theta}\gamma_5\gamma^\mu\theta\right) = x^\mu \pm \left(\theta_R^{\mathrm{T}}\epsilon\gamma^\mu\theta_L\right). \tag{26.3.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

The expansions of $\phi(x_+)$ and $\tilde{\phi}(x_-)$ in powers of $x^\mu - x_\pm^\mu$ terminate with quadratic terms, the expansions of $\psi_{L,R}(x_\pm)$ terminate with linear terms, while the expansions of $\mathscr{F}(x_+)$ and $\tilde{\mathscr{F}}(x_-)$ terminate at zeroth order, because all higher terms make contributions in Eqs. (26.3.21) and (26.3.22) that contain three or more factors of θ_L or θ_R and therefore vanish. For the same reason, it is easy to see that any superfield that depends only on θ_L and x_+^μ but not otherwise on θ_R must take the form (26.3.21), and any superfield that depends only on θ_R and x_-^μ but not otherwise on θ_L must take the form (26.3.22).

We have seen that for a superfield to be left-chiral or right-chiral is entirely a matter of what the superfield is allowed to depend on. It follows immediately that *any function of left-chiral superfields (or of right-chiral superfields), but not their complex conjugates or spacetime derivatives, is a left-(or right-)chiral superfield.* This can also be shown in a more formal way. Because $\Phi(x,\theta)$ depends on θ_R only through its dependence on x_+, and $\tilde{\Phi}(x,\theta)$ depends on θ_L only through its dependence on x_-, they satisfy the conditions

$$\mathscr{D}_{R\alpha}\Phi = \mathscr{D}_{L\alpha}\tilde{\Phi} = 0\,, \tag{26.3.24}$$

where $\mathscr{D}_R$ and $\mathscr{D}_L$ are the right- and left-handed parts of the superderivative (26.2.26):

$$\mathscr{D}_{R\alpha} \equiv \left[\left(\frac{1-\gamma_5}{2}\right)\mathscr{D}\right]_\alpha = -\sum_\beta \epsilon_{\alpha\beta}\frac{\partial}{\partial\theta_{R\beta}} - (\gamma^\mu\theta_L)_\alpha\frac{\partial}{\partial x^\mu}\,, \tag{26.3.25}$$

$$\mathscr{D}_{L\alpha} \equiv \left[\left(\frac{1+\gamma_5}{2}\right)\mathscr{D}\right]_\alpha = +\sum_\beta \epsilon_{\alpha\beta}\frac{\partial}{\partial\theta_{L\beta}} - (\gamma^\mu\theta_R)_\alpha\frac{\partial}{\partial x^\mu}\,, \tag{26.3.26}$$

for which

$$\mathscr{D}_{R\alpha}x_+^\mu = \mathscr{D}_{L\alpha}x_-^\mu = 0\,.$$

Conversely, if a superfield Φ satisfies $\mathscr{D}_R\Phi = 0$ it is left-chiral, and if it satisfies $\mathscr{D}_L\Phi = 0$ it is right-chiral. Any function $f(\Phi)$ of superfields Φ_n that all satisfy $\mathscr{D}_R\Phi_n = 0$ or all satisfy $\mathscr{D}_L\Phi_n = 0$ will satisfy $\mathscr{D}_R f(\Phi) = 0$ or $\mathscr{D}_L f(\Phi) = 0$, and hence be left-chiral or right-chiral, respectively. But a function of left-chiral *and* right-chiral superfields is in general not chiral at all.

Using the representation (26.3.21) for left-chiral superfields makes it easy to work out their multiplication properties. For instance, if Φ_1 and Φ_2 are two left-chiral superfields, then their product $\Phi = \Phi_1\Phi_2$ is a left-chiral superfield, with components

$$\phi = \phi_1\phi_2\,, \tag{26.3.27}$$

$$\psi_L = \phi_1\psi_{2L} + \phi_2\psi_{1L}\,, \tag{26.3.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$$\mathscr{F} = \phi_1 \mathscr{F}_2 + \phi_2 \mathscr{F}_1 - \left(\psi_{1L}^{\mathrm{T}} \epsilon \psi_{2L}\right) . \tag{26.3.29}$$

The presence of chiral superfields in a theory opens up an additional possibility for constructing supersymmetric actions. Inspection of the transformation rule (26.3.16) shows that a supersymmetry transformation changes the $\mathscr{F}$-term of a left-chiral superfield Φ by a derivative, so that the integral of the $\mathscr{F}$-term of any left-chiral superfield is supersymmetric. Thus we can form a supersymmetric action as

$$I = \int d^4x \left[f\right]_{\mathscr{F}} + \int d^4x \left[f\right]_{\mathscr{F}}^{*} + \frac{1}{2} \int d^4x \left[K\right]_D , \tag{26.3.30}$$

where f and K are any left-chiral superfield and general real superfield, respectively, formed from the elementary superfields.

On what can f and K depend? The function f will be left-chiral if it depends only on left-chiral elementary superfields Φ_n, but not their right-chiral complex conjugates. On the other hand, the superderivative of a chiral superfield is not chiral, so we cannot freely include superderivatives of the Φ_n in f. It is true that, acting on a superfield S that is not left-chiral (such as one involving complex conjugates of left-chiral superfields), a pair of right superderivatives gives a left-chiral superfield, because there are only two independent right superderivatives and they anticommute:

$$\mathscr{D}_{R\alpha}(\mathscr{D}_{R\beta}\mathscr{D}_{R\gamma}S) = 0 .$$

However, the $\mathscr{F}$-term of any function f that is constructed in this way makes a contribution to the action that is the same as that of a D term of some other composite superfield. Since the $\mathscr{D}$s anticommute, the most general left-chiral superfield formed by acting on a general superfield S with two $\mathscr{D}_R$s may be expressed in terms of $(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)S$. If one of the left-chiral superfields in the superpotential is of this form, then since each $\mathscr{D}_R$ annihilates all the other superfields in the superpotential, we can write the whole superpotential as $f = (\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)h$ for some other superfield h. Now,

$$\left(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R\right)\left(\theta_R^{\mathrm{T}}\epsilon\theta_R\right) = -4 ,$$

so, apart from spacetime derivatives that do not contribute to the action, $(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)h$ is the coefficient of $-(\theta_R^{\mathrm{T}}\epsilon\theta_R)/4$ in h. But, again apart from spacetime derivatives, $[f]_{\mathscr{F}}$ is the coefficient of $(\theta_L^{\mathrm{T}}\epsilon\theta_L)$ in f, so $[(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)h]_{\mathscr{F}}$ equals the coefficient of $-(\theta_L^{\mathrm{T}}\epsilon\theta_L)(\theta_R^{\mathrm{T}}\epsilon\theta_R)/4 = -(\bar{\theta}\gamma_5\theta)^2/4$ in h, and therefore

$$\int d^4x \left[(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)h\right]_{\mathscr{F}} = 2 \int d^4x \left[h\right]_D . \tag{26.3.31}$$

Thus we do not need to include terms in f that depend on left-chiral superfields of the form $\mathscr{D}_{R\beta}\mathscr{D}_{R\gamma}S$ — any such terms will be included

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

in the list of all possible D-terms. When f is expressed as a function only of elementary left-chiral superfields and not their superderivatives or spacetime derivatives, it is known as the *superpotential.*

In contrast, the function K is in general a real scalar function of both left-chiral superfields Φ_n and their right-chiral complex conjugates Φ_n^*, as well as their superderivatives and spacetime derivatives, known as the *Kahler potential.* (Any right-chiral superfield is the complex conjugate of a left-chiral superfield, so there is no loss of generality in supposing K to depend only on left-chiral superfields and their complex conjugates.) However, not all Ks obtained in this way will yield distinct actions. For one thing, chiral superfields have no D-terms, so two Ks that differ by a chiral superfield make the same contribution to the action.

It is also possible to change the form of K without changing the action by a partial integration in superspace. The D-term of the superderivative $\mathscr{D}_\alpha S$ of an arbitrary superfield makes no contribution to the action because

$$\int d^4x\, [\mathscr{D}_\alpha S]_D = 0\,. \tag{26.3.32}$$

To see this, recall that

$$\mathscr{D}_\alpha S = \sum_\beta \left(\gamma_5 \epsilon\right)_{\alpha\beta} \frac{\partial S}{\partial \theta_\beta} - (\gamma^\mu \theta)_\alpha \frac{\partial S}{\partial x^\mu}\,.$$

Since S is a polynomial at most of fourth order in θ, the first term in $\mathscr{D}_\alpha S$ is a polynomial in θ at most of third order and therefore cannot have a non-zero D-term that is not a derivative, while the second term is also a spacetime derivative, so that its D-term is also a spacetime derivative, and therefore neither the first nor the second terms in $\mathscr{D}_\alpha S$ can contribute to the integral in Eq. (26.3.32). Also, the superderivative acts distributively, so it follows from Eq. (26.3.32) that we can integrate by parts in superspace: for any two bosonic superfields S_1 and S_2,

$$\int d^4x\, [S_1 \mathscr{D}_\alpha S_2]_D = -\int d^4x\, [S_2 \mathscr{D}_\alpha S_1]_D\,. \tag{26.3.33}$$

In Sections 26.4 and 26.8 we will consider in detail the case where f and K depend only on elementary superfields, but not their superderivatives or ordinary derivatives.

We saw in the previous section that in theories in which parity is conserved, the effect of the space inversion operator on a general scalar superfield is to subject its arguments to the transformations $x^\mu \to (\Lambda_P)^\mu{}_\nu x^\nu$ and $\theta \to -i\beta\theta$, and perhaps to multiply the superfield by a phase η. Under these transformations, the arguments $x_\pm^\mu$ in Eqs. (26.3.21) and (26.3.22) are changed by

$$x_\pm^\mu \to (\Lambda_P x)^\mu \pm \tfrac{1}{2}\left(\bar{\theta}\beta\gamma_5\gamma^\mu\beta\theta\right) = (\Lambda_P x_\mp)^\mu\,, \tag{26.3.34}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

while $\theta_L \to -i\beta\theta_R$ and $\theta_R \to -i\beta\theta_L$. Thus space inversion takes left-chiral superfields into right-chiral superfields, and vice-versa. The only right-handed scalar superfield whose component fields involve creation and annihilation operators for the same particles that are created and destroyed by a left-handed scalar superfield Φ is $\tilde{\Phi} \propto \Phi^*$, so $\mathsf{P}^{-1}\Phi\mathsf{P}$ must be proportional to Φ^*. By a suitable choice of phase of Φ, we can arrange that this transformation rule reads

$$\mathsf{P}^{-1}\Phi(x,\theta)\mathsf{P} = \Phi^*(\Lambda_P x, -i\beta\theta)\,. \tag{26.3.35}$$

In terms of component fields, this transformation is

$$\begin{aligned} \mathsf{P}^{-1}\phi(x)\mathsf{P} &= \phi^*(\Lambda_P x)\,, \\ \mathsf{P}^{-1}\psi_L(x)\mathsf{P} &= -i\epsilon\gamma_5\beta\psi_L^*(\Lambda_P x)\,, \\ \mathsf{P}^{-1}\mathscr{F}(x)\mathsf{P} &= \mathscr{F}^*(\Lambda_P x)\,. \end{aligned} \tag{26.3.36}$$

There is another type of possible symmetry, known as *R-symmetry*, that is important in some models of spontaneous supersymmetry breaking discussed in Section 26.5, and that will also be used in proving no-renormalization theorems in Section 27.6. As mentioned in Section 25.2, in theories of simple $N=1$ supersymmetry an R-symmetry is invariance under a $U(1)$ transformation under which the left-handed components of the supersymmetry generator (called Q_a in Section 25.2) carry a non-vanishing quantum number, say -1, in which case their adjoints, the right-handed components of the supersymmetry generator carry the opposite quantum number $+1$. Inspection of Eq. (26.2.2) shows that the θ superspace coordinate has a non-trivial transformation property under R-symmetry transformations: θ_L carries R-quantum number $+1$ and θ_R, which is proportional to θ_L^*, carries R quantum number -1. In addition, the whole superfield may be given an R quantum number. If we give a left-chiral superfield Φ the R quantum number R_Φ, then its scalar component ϕ has the same R quantum number, while the left-handed spinor component ψ_L has $R_\psi = R_\Phi - 1$, and the auxiliary field $\mathscr{F}$ has $R_\mathscr{F} = R_\Phi - 2$. In particular, in order for the superpotential term $\int d^4x\,[f]_\mathscr{F}$ to conserve R, the superpotential itself must have $R_f = +2$, so if f depends on a single left-chiral superfield Φ, then it must be proportional to Φ^{2/R_Φ}. To put this another way, if $f(\Phi)$ is a pure mass term proportional to Φ^2 then we must choose $R_\Phi = +1$, while if $f(\Phi)$ is a pure interaction term proportional to Φ^3, then we must choose $R_\Phi = 2/3$. On the other hand, inspection of Eq. (26.2.10) shows that the D-term of a superfield has the same R value as the superfield, so in order for the term $\int d^4x[g]_D$ in the action to conserve R it is only necessary that K have $R = 0$, which will be the case if each term in K contains equal numbers of Φ and Φ^* factors, whatever R value we give to Φ. Of course, there is no general reason why R-symmetry

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

should be respected by the action, or why it should not be spontaneously broken.

* * *

There are other ways of constraining superfields, to yield other types of supermultiplets of fields. Among the more common are the *linear* superfields. To learn the conditions defining this sort of superfield, we note that if S is a general superfield, then we can form a chiral superfield

$$S' \equiv \frac{1}{4}\left(\bar{\mathscr{D}}\,\mathscr{D}\right) S \,. \tag{26.3.37}$$

This is a chiral superfield because it can be written as a sum of $\frac{1}{4}(\bar{\mathscr{D}}_L \mathscr{D}_L) S$, which is right-chiral, and $\frac{1}{4}(\bar{\mathscr{D}}_R \mathscr{D}_R) S$, which is left-chiral. Its components are given in terms of the components of S by

$$C' = N \,, \tag{26.3.38}$$
$$\omega' = \lambda + \not{\partial}\omega \,, \tag{26.3.39}$$
$$M' = -\partial_\mu V^\mu \,, \tag{26.3.40}$$
$$N' = D + \Box C \,, \tag{26.3.41}$$
$$V'_\mu = -\partial_\mu M \,, \tag{26.3.42}$$
$$\lambda' = D' = 0 \,. \tag{26.3.43}$$

A multiplet S is said to be linear if the superfield S' defined in this way vanishes

$$\left(\bar{\mathscr{D}}\,\mathscr{D}\right) S = 0 \,, \tag{26.3.44}$$

or in terms of its components,

$$N = M = \partial_\mu V^\mu = 0 \,, \quad \lambda = -\not{\partial}\omega \,, \quad D = -\Box C \,. \tag{26.3.45}$$

This leaves four independent bosonic fields, C and the three components of V_μ subject to the condition $\partial_\mu V^\mu = 0$, and four independent fermionic fields, the components of the Majorana four-spinor ω. We will see in Section 26.6 that the current superfields, whose V_μ-terms are the conserved currents associated with symmetry transformations, are linear superfields.

26.4 Renormalizable Theories of Chiral Superfields

We will now work out the details of a general renormalizable theory of scalar chiral superfields. This will provide some insight into the implications of supersymmetry, and the theory we obtain will be part of the supersymmetric standard model to be discussed in Chapter 28.

As discussed in Section 12.2, the Lagrangian density in a renormalizable theory can contain only operators with dimensionality (counting powers

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

of energy or momentum, with $\hbar = c = 1$) four or less. Eq. (26.2.6) shows that $\mathscr{Q}_\alpha$ and hence $\partial/\partial\theta_\alpha$ has dimensionality 1/2, so $\mathscr{D}_\alpha$ has dimensionality $+1/2$, and θ_α has dimensionality $-1/2$. The $\mathscr{F}$- and D-terms of a superfield S are the coefficients of two or four factors of θ, respectively, so if the superfield has dimensionality $d(S)$ then its $\mathscr{F}$- and D-terms have dimensionalities $d(\mathscr{F}^S) = d(S) + 1$ and $d(D^S) = d(S) + 2$. Thus in a renormalizable theory the functions f and K in Eq. (26.3.30) consist of operators with dimensionality at most three and two, respectively.

The dimensionality of an elementary scalar superfield Φ_n is that of an elementary scalar field, or $+1$, so in order for each term in the function f to have dimensionality three or less, it can contain at most three factors of Φ_n and/or derivatives $\partial/\partial x^\mu$ and/or pairs of spinor superderivatives $\mathscr{D}_\alpha$. As discussed in the previous section, any left-chiral term in f that involves superderivatives could be replaced with a term in K, so superderivatives can be omitted in f. Eq. (26.2.30) shows that spacetime derivatives can be expressed in terms of superderivatives, so these too can be omitted. (In any case, Lorentz invariance would rule out terms with one spacetime derivative, and terms with two derivatives in a renormalizable theory could involve only one Φ_n factor, on which these derivatives would have to act, so such terms would not contribute to the action.) We conclude that $f(\Phi)$ is at most a cubic polynomial in the Φ_n, without spacetime derivatives or superderivatives.

The same dimensional analysis shows that in a renormalizable theory K is at most a quadratic function of Φ_n and Φ_n^*, without derivatives. But any term in $K(\Phi, \Phi^*)$ that involves only Φ_n or only Φ_n^* would be a chiral superfield, and chiral superfields by definition have no D-terms, so $[K(\Phi, \Phi^*)]_D$ receives contributions only from terms in $K(\Phi, \Phi^*)$ that involve *both* Φ_n and Φ_n^*. Thus $K(\Phi, \Phi^*)$ must be of the form

$$K(\Phi, \Phi^*) = \sum_{mn} g_{nm} \Phi_n^* \Phi_m \,, \tag{26.4.1}$$

with constant coefficients g_{nm} forming a Hermitian matrix.

We must now calculate the $\mathscr{F}$- and D-components of $f(\Phi)$ and $K(\Phi, \Phi^*)$, respectively. To calculate the D-component of $K(\Phi, \Phi^*)$, we note that the term in $\Phi_n^* \Phi_m$ of fourth order in θ is

$$\begin{aligned}
\left[\Phi_n^* \Phi_m\right]_{\theta^4} = & -\frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\left[\phi_n^* \Box \phi_m + \left(\Box\phi_m^*\right)\phi_n\right] \\
& + \left(\bar\theta\gamma_5\,\theta\right)\left[\left(\overline{\psi_n}\,\theta\right)\left(\bar\theta\gamma^\mu\partial_\mu\psi_m\right) + \left(\left(\partial_\mu\overline{\psi_n}\right)\gamma^\mu\theta\right)\left(\bar\theta\,\psi_m\right)\right] \\
& + \frac{1}{4}\mathscr{F}_n^*\mathscr{F}_m\left(\bar\theta(1-\gamma_5)\theta\right)\left(\bar\theta(1+\gamma_5)\theta\right) \\
& - \frac{1}{4}\partial^\mu\phi_n^*\,\partial^\nu\phi_m\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\left(\bar\theta\gamma_5\gamma_\nu\theta\right) .
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Using (26.A.18) and (26.A.19) allows us to convert the θ dependence of this expression to an over-all factor $(\bar{\theta}\gamma_5\theta)^2$:

$$\left[\Phi_n^*\Phi_m\right]_{\theta^4} = -\frac{1}{4}\left(\bar{\theta}\gamma_5\theta\right)^2\left[\frac{1}{2}\phi_n^*\Box\phi_m + \frac{1}{2}\left(\Box\phi_m^*\right)\phi_n - \left(\overline{\psi_n}\,\gamma^\mu\partial_\mu\psi_m\right)\right.$$
$$\left.+\left(\left(\partial_\mu\overline{\psi_n}\right)\gamma^\mu\psi_m\right) + 2\mathscr{F}_n^*\mathscr{F}_m - \partial^\mu\phi_n^*\,\partial_\mu\phi_m\right].$$

The D-term of a superfield is the coefficient of $-\frac{1}{4}\left(\bar{\theta}\gamma_5\theta\right)^2$ minus $\frac{1}{2}\Box$ acting on the θ-independent term, which for $\Phi_n^*\Phi_m$ is $\phi_n^*\phi_m$, so

$$\frac{1}{2}\left[K(\Phi,\Phi^*)\right]_D = \sum_{nm} g_{nm}\left[-\partial_\mu\phi_n^*\,\partial^\mu\phi_m + \mathscr{F}_n^*\mathscr{F}_m\right.$$
$$\left.-\frac{1}{2}\left(\overline{\psi_{nL}}\,\gamma^\mu\partial_\mu\psi_{mL}\right) + \frac{1}{2}\left(\partial_\mu(\overline{\psi_{nL}})\,\gamma^\mu\psi_{mL}\right)\right]. \quad (26.4.2)$$

If we write Φ_n as a linear combination $\sum_m N_{nm}\Phi'_m$ of new superfields Φ'_m, then $K(\Phi,\Phi^*)$ is given in terms of the new superfields by a formula which is the same as Eq. (26.4.1), except that g_{nm} is replaced with $g'_{nm} = (N^\dagger g N)_{nm}$. In order for the kinematic terms for the scalar and spinor fields to have a sign consistent with the quantum commutation and anticommutation relations, it is necessary that the Hermitian matrix g_{nm} be positive-definite and, as shown in Section 12.5, this means that we can choose N so that $g'_{nm} = \delta_{nm}$. Dropping primes, the term (26.4.2) is now

$$\frac{1}{2}\left[K(\Phi,\Phi^*)\right]_D = \sum_n\left[-\partial_\mu\phi_n^*\,\partial^\mu\phi_n + \mathscr{F}_n^*\mathscr{F}_n\right.$$
$$\left.-\frac{1}{2}\left(\overline{\psi_{nL}}\,\gamma^\mu\partial_\mu\psi_{nL}\right) + \frac{1}{2}\left(\partial_\mu(\overline{\psi_{nL}})\,\gamma^\mu\psi_{nL}\right)\right]. \quad (26.4.3)$$

We can still use a unitary transformation to redefine the superfields without changing the form of Eq. (26.4.3), a freedom we will need to exercise shortly.

The terms in Eq. (26.4.3) involving ϕ_n and ψ_{nL} are the correct kinematic Lagrangians for conventionally normalized complex scalar and Majorana spinor fields. We will rewrite the fermion terms in a more familiar form after we have had a chance to consider mass terms.

To calculate the $\mathscr{F}$-term of $f(\Phi)$, it is most convenient to use the representation (26.3.21) of the superfields, and pick out the term of second

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

order in θ_L:

$$\Big[f\Big(\Phi(x,\theta)\Big)\Big]_{\theta_L^2} = \sum_{nm}\Big(\theta_L^{\mathrm{T}}\epsilon\,\psi_{nL}(x)\Big)\Big(\theta_L^{\mathrm{T}}\epsilon\,\psi_{mL}(x)\Big)\frac{\partial^2 f\Big(\phi(x)\Big)}{\partial\phi_n(x)\,\partial\phi_m(x)} + \sum_n \mathscr{F}_n(x)\,\frac{\partial f\Big(\phi(x)\Big)}{\partial\phi_n(x)}\Big(\theta_L^{\mathrm{T}}\epsilon\theta_L\Big)\,.$$

(We have replaced x_+ here with x because the term $(\theta_R^{\mathrm{T}}\epsilon\gamma^\mu\theta_L)$ in Eq. (26.3.21) vanishes when multiplied by an expression with two θ_L factors.) The θ dependence of the first term on the right-hand side can be put into a standard form by using Eq. (26.A.11) to write*

$$\Big(\theta_L^{\mathrm{T}}\epsilon\,\psi_{nL}\Big)\Big(\theta_L^{\mathrm{T}}\epsilon\,\psi_{mL}\Big) = \left(\psi_{nL}^{\mathrm{T}}\epsilon\left(\frac{1+\gamma_5}{2}\right)\theta\right)\left(\theta^{\mathrm{T}}\epsilon\left(\frac{1+\gamma_5}{2}\right)\psi_{mL}\right) = -\frac{1}{2}\Big(\bar\psi_{nL}\,\psi_{mL}\Big)\Big(\theta_L^{\mathrm{T}}\epsilon\,\theta_L\Big)\,.$$

The $\mathscr{F}$-term of any left-chiral superfield is the coefficient of $(\theta_L^{\mathrm{T}}\epsilon\theta_L)$, so here

$$\Big[f(\Phi)\Big]_{\mathscr{F}} = -\frac{1}{2}\sum_{nm}\frac{\partial^2 f(\phi_n)}{\partial\phi_n\,\partial\phi_m}\Big(\bar\psi_{nL}\,\psi_{mL}\Big) + \sum_n \mathscr{F}_n\,\frac{\partial f(\phi)}{\partial\phi_n}\,. \tag{26.4.4}$$

The complete Lagrangian density is the sum of the terms (26.4.3), (26.4.4), and the complex conjugate of (26.4.4):

$$\begin{aligned}\mathscr{L} = \sum_n \Bigg[&-\partial_\mu\phi_n^*\,\partial^\mu\phi_n + \mathscr{F}_n^*\mathscr{F}_n \\ &-\frac{1}{2}\Big(\overline{\psi_{nL}}\,\gamma^\mu\partial_\mu\,\psi_{nL}\Big) + \frac{1}{2}\Big((\partial_\mu\overline{\psi_{nL}})\,\gamma^\mu\,\psi_{nL}\Big)\Bigg] \\ &-\frac{1}{2}\sum_{nm}\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\Big(\bar\psi_{nL}\,\psi_{mL}\Big) - \frac{1}{2}\sum_{nm}\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^*\Big(\bar\psi_{nL}\,\psi_{mL}\Big)^* \\ &+\sum_n \mathscr{F}_n\,\frac{\partial f(\phi)}{\partial\phi_n} + \sum_n \mathscr{F}_n^*\left(\frac{\partial f(\phi)}{\partial\phi_n}\right)^*\,.\end{aligned} \tag{26.4.5}$$

The auxiliary fields $\mathscr{F}_n$ enter quadratically in the action, with constant coefficients for the second-order terms, so they can be eliminated by setting $\mathscr{F}_n$ equal to the value at which the Lagrangian density (26.4.5) is stationary with respect to $\mathscr{F}_n$ and $\mathscr{F}_n^*$:

$$\mathscr{F}_n = -\left(\frac{\partial f(\phi)}{\partial\phi_n}\right)^*\,. \tag{26.4.6}$$

* Note that $\bar\psi_{nL}$ is the left-handed component of $\bar\psi_n$, not $\overline{\psi_{nL}}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Inserting this in Eq. (26.4.5) gives

$$\mathscr{L} = \sum_n \left[-\partial_\mu \phi_n^* \partial^\mu \phi_n - \frac{1}{2}\left(\overline{\psi_{nL}}\,\gamma^\mu \partial_\mu \psi_{nL}\right) + \frac{1}{2}\left((\partial_\mu \overline{\psi_{nL}})\,\gamma^\mu \psi_{nL}\right)\right]$$
$$-\frac{1}{2}\sum_{nm} \frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\left(\bar{\psi}_{nL}\,\psi_{mL}\right) - \frac{1}{2}\sum_{nm}\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^*\left(\bar{\psi}_{nL}\,\psi_{mL}\right)^*$$
$$-\sum_n \left(\frac{\partial f(\phi)}{\partial\phi_n}\right)^* \frac{\partial f(\phi)}{\partial\phi_n}\,. \tag{26.4.7}$$

Thus the scalar field potential is $V(\phi) = \sum_n |\partial f(\phi)/\partial\phi_n|^2$.

With the auxiliary fields eliminated in this way, the action is no longer invariant under the supersymmetry transformations (26.3.15), (26.3.17) of the remaining fields ψ_{nL} and ϕ_n:

$$\delta\psi_{nL} = \sqrt{2}\partial_\mu\phi_n\gamma^\mu\alpha_R - \sqrt{2}\left(\frac{\partial f(\phi)}{\partial\phi_n}\right)^*\alpha_L\,, \qquad \delta\phi_n = \sqrt{2}\left(\overline{\alpha_R}\psi_{nL}\right).$$

This is because the expression (26.4.6) does not obey the transformation rule $\delta\mathscr{F}_n = \sqrt{2}(\overline{\alpha_L}\,\partial\!\!\!/\psi_{nL})$ for $\mathscr{F}_n$ given by Eq. (26.3.16), but instead

$$\delta\left(-\frac{\partial f(\phi)}{\partial\phi_n}\right)^* = -\sum_m\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^*\delta\phi_m^* = -\sqrt{2}\sum_m\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^*\left(\overline{\alpha_L}\psi_{mR}\right).$$

For the same reason, after eliminating the auxiliary fields the commutators of the supersymmetry transformations of ϕ_n and ψ_{nL} are no longer given by the supersymmetry anticommutation relations, and in fact do not form a closed Lie superalgebra. But this is not inconsistent with the existence of quantum mechanical operators Q_α that satisfy the anticommutation relations of supersymmetry. These operators generate supersymmetry transformations, in the sense that the commutator of $-i(\bar{\alpha}Q)$ with any Heisenberg-picture quantum field ϕ_n or ψ_{nL} equals the change in that field under a supersymmetry transformation with infinitesimal parameter α. With $\mathscr{F}_n$ given by Eq. (26.4.6) the commutator of $-i(\bar{\alpha}Q)$ with $\mathscr{F}_n$ *is* given by $\delta\mathscr{F}_n = \sqrt{2}(\overline{\alpha_L}\,\partial\!\!\!/\psi_{nL})$, because in the Heisenberg picture the quantum field ψ_{nL} satisfies the field equation derived from the Lagrangian (26.4.7):

$$\partial\!\!\!/\psi_{nL} = -\sum_m\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^*\psi_{mR}\,.$$

Likewise, the supersymmetry transformations of the quantum fields ϕ_n and ψ_{nL} do form a closed Lie superalgebra when the field equations are taken into account. Such algebras are often called *on-shell*.

The zeroth-order expectation values ϕ_{n0} of the scalar fields ϕ_n must be at the maximum of the last term in Eq. (26.4.7). Since this term is always negative or zero, the maximum will be at spacetime-independent

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

field values ϕ_{n0} at which this term vanishes, so that

$$\left.\frac{\partial f(\phi)}{\partial \phi_n}\right|_{\phi=\phi_0} = 0\,, \tag{26.4.8}$$

provided of course that a solution of this equation exists. Eq. (26.4.8) not only maximizes the last term in Eq. (26.4.7) — it is also the condition for supersymmetry to be unbroken. Invariance of the vacuum under supersymmetry transformations requires that the vacuum expectation value of the change in any field under a supersymmetry transformation should vanish. The change in a bosonic field is a fermionic field, which of course has zero expectation value anyway, but Eq. (26.3.15) shows that the vacuum expectation value of $\delta\psi_{nL}$ is proportional to the vacuum expectation value of the auxiliary field $\mathscr{F}_n$, which therefore must vanish if supersymmetry is unbroken. According to Eq. (26.4.6), in zeroth-order perturbation theory this condition requires that Eq. (26.4.8) must be satisfied. We will see in Section 27.6 that if Eq. (26.4.8) is satisfied then supersymmetry is unbroken to all orders of perturbation theory.

For a single left-chiral scalar superfield Φ, the fundamental theorem of algebra tells us that the polynomial $\partial f(\phi)/\partial\phi$ always has at least one zero somewhere in the complex plane. This is not necessarily true when there is more than one superfield. If we *assume* that there is a solution ϕ_{n0} of Eq. (26.4.8), we can evaluate the physical degrees of freedom of the theory by setting

$$\phi_n = \phi_{n0} + \varphi_n\,, \tag{26.4.9}$$

and expanding in powers of φ_n. The masses of the particles of this theory can be calculated by inspecting the terms of second order in φ and ψ:

$$\begin{aligned}\mathscr{L}_0 = \sum_n \Big[&- \partial_\mu \varphi_n^* \, \partial^\mu \varphi_n - \frac{1}{2}\Big(\overline{\psi_{nL}}\,\gamma^\mu \partial_\mu \psi_{nL}\Big) + \frac{1}{2}\Big(\partial_\mu(\overline{\psi_{nL}})\,\gamma^\mu\,\psi_{nL}\Big)\Big] \\ &- \frac{1}{2}\sum_{nm} \mathscr{M}_{nm}\Big(\bar{\psi}_{nL}\,\psi_{mL}\Big) - \frac{1}{2}\sum_{nm} \mathscr{M}^*_{nm}\Big(\bar{\psi}_{nL}\,\psi_{mL}\Big)^* \\ &- \sum_{nm}\Big(\mathscr{M}^\dagger \mathscr{M}\Big)_{mn} \varphi_m^* \varphi_n\,,\end{aligned} \tag{26.4.10}$$

where $\mathscr{M}$ is the symmetric complex matrix

$$\mathscr{M}_{mn} \equiv \left(\frac{\partial^2 f(\phi)}{\partial\phi_n \partial\phi_m}\right)_{\phi=\phi_0}\,. \tag{26.4.11}$$

Now if we redefine the fields by a unitary transformation

$$\varphi_n = \sum_m \mathscr{U}_{nm}\varphi'_m\,, \qquad \psi_{nL} = \sum_m \mathscr{U}_{nm}\psi'_{mL}\,, \tag{26.4.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

then the free-field Lagrangian (26.4.10) will take the same form, but with $\mathscr{M}$ replaced with $\mathscr{M}'$, where

$$\mathscr{M}' = \mathscr{U}^{\mathrm{T}} \mathscr{M} \mathscr{U} \,. \tag{26.4.13}$$

According to a theorem of matrix algebra, for any complex symmetric matrix $\mathscr{M}$ it is always possible to find a unitary matrix $\mathscr{U}$ such that the matrix $\mathscr{M}'$ defined by Eq. (26.4.13) is diagonal, with real positive elements m_n on the diagonal. (For future use, we note that $\mathscr{M}'^{\dagger}\mathscr{M}' = \mathscr{U}^{\dagger}\mathscr{M}^{\dagger}\mathscr{M}\mathscr{U}$, so the quantities m_n^2 are just the eigenvalues of the positive Hermitian matrix $\mathscr{M}^{\dagger}\mathscr{M}$.) Redefining the fields in this way and dropping the primes, the quadratic part of the Lagrangian is now

$$\begin{aligned}\mathscr{L}_0 = \sum_n \Big[&- \partial_\mu \varphi_n^* \, \partial^\mu \varphi_n - \frac{1}{2}\left(\overline{\psi_{nL}}\, \gamma^\mu \partial_\mu \psi_{nL}\right) + \frac{1}{2}\left(\partial_\mu(\overline{\psi_{nL}})\, \gamma^\mu \psi_{nL}\right)\Big] \\ &- \frac{1}{2}\sum_n m_n \left(\bar{\psi}_{nL}\, \psi_{nL}\right) - \frac{1}{2}\sum_n m_n \left(\bar{\psi}_{nL}\, \psi_{nL}\right)^* \\ &- \sum_n m_n^2 \, \varphi_n^* \, \varphi_n \,. \end{aligned} \tag{26.4.14}$$

To put the fermion mass terms in a more familiar form, we introduce fields $\psi_n(x)$ defined as the *Majorana* fields whose left-handed components are $\psi_{nL}(x)$. Then, using the symmetry properties (26.A.7) of Majorana bilinears:

$$\begin{aligned} &-\frac{1}{2}\left(\overline{\psi_{nL}}\, \gamma^\mu \partial_\mu \psi_{nL}\right) + \frac{1}{2}\left(\partial_\mu(\overline{\psi_{nL}})\, \gamma^\mu \psi_{nL}\right) \\ &\quad = -\frac{1}{2}\left(\overline{\psi_n}\, \gamma^\mu \left(\frac{1+\gamma_5}{2}\right) \partial_\mu \psi_n\right) + \frac{1}{2}\left(\partial_\mu(\overline{\psi_n})\, \gamma^\mu \left(\frac{1+\gamma_5}{2}\right) \psi_n\right) \\ &\quad = -\frac{1}{2}\left(\overline{\psi_n}\, \gamma^\mu \left(\frac{1+\gamma_5}{2}\right) \partial_\mu \psi_n\right) - \frac{1}{2}\left(\overline{\psi_n}\, \gamma^\mu \left(\frac{1-\gamma_5}{2}\right) \partial_\mu \psi_n\right) \\ &\quad = -\frac{1}{2}\left(\overline{\psi_n}\, \gamma^\mu \partial_\mu \psi_n\right), \end{aligned}$$

while the reality properties (26.A.21) give

$$\left(\bar{\psi}_{nL}\, \psi_{nL}\right) + \left(\bar{\psi}_{nL}\, \psi_{nL}\right)^* = 2\,\mathrm{Re}\left(\overline{\psi_n}\left(\frac{1+\gamma_5}{2}\right)\psi_n\right) = \left(\overline{\psi_n}\, \psi_n\right).$$

The complete quadratic Lagrangian is then

$$\begin{aligned}\mathscr{L}_0 = \sum_n \Big[&- \partial_\mu \varphi_n^* \, \partial^\mu \varphi_n - \sum_n m_n^2 \, \varphi_n^* \, \varphi_n \\ &- \frac{1}{2}\left(\overline{\psi_n}\, \gamma^\mu \partial_\mu \psi_n\right) - \frac{m_n}{2}\left(\overline{\psi_n}\, \psi_n\right)\Big] \,. \end{aligned} \tag{26.4.15}$$

The factor 1/2 in the fermion terms is correct because these are Majorana

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

fermion fields, while there is no factor $1/2$ in the scalar terms because these are complex scalars. We see that the spinless and spin $1/2$ particles have equal masses m_n, as required by the unbroken supersymmetry of the theory.

The interaction part $\mathscr{L}'$ of the Lagrangian density is given by the terms in Eq. (26.4.7) of higher than second order in φ_n and ψ_n. Since the superpotential $f(\phi_0+\varphi)$ is being assumed to be a cubic polynomial and stationary at $\varphi_n = 0$, with φ defined so that the second-order terms are $\frac{1}{2}\sum_n m_n\varphi_n^2$, we may write the superpotential (apart from an inconsequential constant term) as

$$f(\phi_0+\varphi) = \frac{1}{2}\sum_n m_n\,\varphi_n^2 + \frac{1}{6}\sum_{nm\ell} f_{nm\ell}\,\varphi_n\,\varphi_m\,\varphi_\ell\ . \tag{26.4.16}$$

Using this in Eq. (26.4.7) gives the interaction as

$$\begin{aligned}\mathscr{L}' = &-\frac{1}{2}\sum_{nm\ell} f_{nm\ell}\,\varphi_n\left(\overline{\psi_m}\left(\frac{1+\gamma_5}{2}\right)\psi_\ell\right)\\ &-\frac{1}{2}\sum_{nm\ell} f^*_{nm\ell}\,\varphi^*_n\left(\overline{\psi_m}\left(\frac{1-\gamma_5}{2}\right)\psi_\ell\right)\\ &-\frac{1}{2}\sum_{nm\ell} m_n\,f_{nm\ell}\varphi^*_n\varphi_m\varphi_\ell - \frac{1}{2}\sum_{nm\ell} m_n\,f^*_{nm\ell}\varphi_n\varphi^*_m\varphi^*_\ell\\ &-\frac{1}{4}\sum_{nm\ell m'\ell'} f_{nm\ell}f^*_{nm'\ell'}\varphi_m\varphi_\ell\varphi^*_{m'}\varphi^*_{\ell'}\ .\end{aligned} \tag{26.4.17}$$

We see that a knowledge of the masses m_n and the 'Yukawa' couplings $f_{nm\ell}$ of the scalars and fermions is enough to determine all the cubic and quartic self-couplings of the spinless fields.

As an illustration, consider the case of a single left-chiral superfield. For comparison with earlier results, let us write the single coefficient f in Eq. (26.4.16) as

$$f \equiv 2\sqrt{2}\,e^{i\alpha}\,\lambda\ , \tag{26.4.18}$$

where λ is real and α is some real phase. We will also introduce a pair of real spinless fields $A(x)$ and $B(x)$ by writing the single complex scalar here as

$$\varphi \equiv e^{-i\alpha}\left(\frac{A+iB}{\sqrt{2}}\right)\ . \tag{26.4.19}$$

The total Lagrangian density is then given by the sum of Eqs. (26.4.15) and (26.4.17) as

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{2}\partial_\mu A\,\partial^\mu A - \tfrac{1}{2}\partial_\mu B\,\partial^\mu B - \tfrac{1}{2}m^2(A^2+B^2)\\ &-\tfrac{1}{2}\left(\bar\psi\gamma^\mu\partial_\mu\psi\right) - \tfrac{1}{2}m\left(\bar\psi\psi\right)\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$$-\lambda A\left(\bar{\psi}\psi\right) - i\lambda B\left(\bar{\psi}\gamma_5\psi\right)$$
$$-m\lambda A(A^2+B^2) - \tfrac{1}{2}\lambda^2(A^2+B^2)^2 . \qquad (26.4.20)$$

which is the same as the Lagrangian density (24.2.9) originally found by Wess and Zumino.[2] It is noteworthy that in this simple case the Lagrangian turns out to be invariant under a space inversion transformation

$$A(x) \to A(\Lambda_P x) , \qquad B(x) \to -B(\Lambda_P x) , \qquad \psi(x) \to i\beta\psi(\Lambda_P x) , \quad (26.4.21)$$

even though we did not assume parity conservation in deriving it. The appearance of parity conservation as an 'accidental' symmetry is a familiar feature of various renormalizable gauge theories (see Sections 12.5 and 18.7) but not of theories involving spinless fields, so this is a special consequence of supersymmetry in the renormalizable theory of a single scalar superfield.

26.5 Spontaneous Supersymmetry Breaking in the Tree Approximation

We saw in the previous section that supersymmetry is unbroken (at least in the tree approximation) in renormalizable theories of chiral superfields if Eq. (26.4.8) has a solution, that is, if there is a value ϕ_0 of the fields at which the superpotential $f(\phi)$ is stationary:

$$\left.\frac{\partial f(\phi)}{\partial \phi_n}\right|_{\phi=\phi_0} = 0 . \qquad (26.5.1)$$

There are as many independent variables here as there are equations to satisfy, so we generally expect there to be solutions of Eq. (26.5.1). In order for supersymmetry to be spontaneously broken in these theories, it is necessary to impose restrictions on the form of the superpotential.

To see how a choice of superpotential may allow supersymmetry to be spontaneously broken, we will consider a generalization of a class of models due to O'Raifeartaigh.[3] Suppose that the superpotential is a linear combination of a set Y_i of left-chiral superfields, with coefficients given by functions $f_i(X)$ of a second set of left-chiral superfields X_n:

$$f(X,Y) = \sum_i Y_i f_i(X) . \qquad (26.5.2)$$

The conditions for supersymmetry to be unbroken by the values x_n and y_i of the scalar components of these superfields are that

$$0 = \frac{\partial f(x,y)}{\partial y_i} = f_i(x) , \qquad (26.5.3)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$$0 = \frac{\partial f(x,y)}{\partial x_n} = \sum_i y_i \frac{\partial f_i(x)}{\partial x_n} \,. \tag{26.5.4}$$

Eq. (26.5.4) can always be solved by taking $y_i = 0$, with no effect on the problem of solving Eq. (26.5.3). On the other hand, if the number of X_n superfields is smaller than the number of Y_i superfields, then Eq. (26.5.3) imposes more conditions on the x_n than there are variables, so without fine-tuning a solution is impossible, and supersymmetry is broken.

It may appear that the initial assumption (26.5.2) itself represents a radical form of fine-tuning, but in fact this form can be imposed on the superpotential by assuming a suitable R-symmetry. As discussed in Section 26.3, in theories with $N = 1$ supersymmetry an R-symmetry is a $U(1)$ symmetry for which the θ superspace coordinate has a non-trivial transformation property. If we assume an R-symmetry for which θ_L carries quantum number $+1$, then the $\mathscr{F}$-term of any superpotential has a quantum number equal to that of the superpotential itself minus 2, so R invariance requires the superpotential itself to have $R = 2$. We can therefore impose the structure (26.5.2) by requiring R invariance, with the Y_i and X_n superfields given R quantum numbers $+2$ and 0, respectively.

The scalar fields in this sort of model have a potential

$$V(x,y) = \sum_i |f_i(x)|^2 + \sum_n \left| \sum_i y_i \frac{\partial f_i(x)}{\partial x_n} \right|^2 \,. \tag{26.5.5}$$

The potential is always minimized by choosing the x_n to minimize the first term; whatever values this gives the x_n, the second term can always be minimized by taking $y_i = 0$. Whether or not supersymmetry can be spontaneously broken, these models have the peculiar feature that there are always directions in the space of the fields in which the minimum of the potential is flat. Whatever values x_{n0} of the x_n minimize the first term in Eq. (26.5.5), the second term vanishes not only for $y_i = 0$, but for any vector y_i in a direction orthogonal to all the vectors $(v^n)_i = (\partial f_i(x)/\partial x_n)_{x=x_0}$. If there are N_X superfields X_n and N_Y superfields Y_i with $N_Y > N_X$, then the v^n cannot span the space of the ys, and there will be at least $N_Y - N_X$ of these flat directions. For any non-vanishing values $y_i = y_{0i}$ along any one of these flat directions, the R-symmetry of the Lagrangian density is spontaneously broken, and the Goldstone boson field ϕ associated with this global symmetry breakdown corresponds to a term $\propto y_{0i}$ in the y_i.

The simplest example of this class of models is provided by the case where there is just one X superfield and two Y superfields. Renormalizability requires the coefficient functions $f_i(X)$ to be quadratic functions of X, and by taking suitable linear combinations of the Y_i and shifting and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

rescaling X we can choose these functions so that

$$f_1(X) = X - a\,, \qquad f_2(X) = X^2\,, \tag{26.5.6}$$

with an arbitrary constant a. There clearly is no simultaneous solution of the two equations (26.5.1) unless the superpotential is fine-tuned so that $a = 0$. The potential (26.5.5) here is

$$V(x,y) = |x|^4 + |x-a|^2 + |y_1 + 2xy_2|^2\,. \tag{26.5.7}$$

The sum of the first two terms has a unique global minimum x_0. The flat direction here is the one for which $y_1 + 2x_0y_2 = 0$. For $a = 0$ we have $x_0 = 0$, and the minima of the potential are along the line with $y_1 = 0$ and y_2 arbitrary.

Whatever the reason for a spontaneous breakdown of supersymmetry, this phenomenon always entails the existence of a massless spin 1/2 particle, the *goldstino*, analogous to the Goldstone bosons associated with the spontaneous breakdown of ordinary global symmetries. (The one exception, discussed in Section 31.3, is that in supergravity theories, where supersymmetry is a local symmetry, the goldstino appears as the helicity $\pm 1/2$ states of a massive particle of spin 3/2, the gravitino.) In renormalizable theories of chiral superfields, the tree-approximation vacuum expectation values ϕ_{n0} of the scalar fields must be at a minimum of the potential $\sum_n |\partial f(\phi)/\partial\phi_n|^2$ in Eq. (26.4.7), so

$$\sum_m \mathscr{M}_{nm} \left(\left.\frac{\partial f(\phi)}{\partial \phi_m}\right|_{\phi=\phi_0} \right)^* = 0\,, \tag{26.5.8}$$

where

$$\mathscr{M}_{nm} \equiv \left.\frac{\partial^2 f(\phi)}{\partial\phi_n \partial\phi_m}\right|_{\phi=\phi_0}\,. \tag{26.5.9}$$

If Eq. (26.5.1) is not satisfied, then Eq. (26.5.8) tells us that the matrix $\mathscr{M}_{nm}$ has at least one eigenvector with eigenvalue zero, so according to Eq. (26.4.10) there must be at least one linear combination of the spin 1/2 particles described by ψ_n with zero mass. For instance, for the model defined by Eqs. (26.5.2) and (26.5.6), the matrix $\mathscr{M}$ has the non-vanishing components

$$\mathscr{M}_{xy_1} = \mathscr{M}_{y_1x} = 1\,, \qquad \mathscr{M}_{xy_2} = \mathscr{M}_{y_2x} = 2x_0\,, \tag{26.5.10}$$

so this matrix has eigenvalues $\pm 2x_0$ and 0, with the last eigenvalue corresponding to the goldstino mode. In Chapter 29 we will show without the use of perturbation theory that the spontaneous breakdown of supersymmetry requires the existence of goldstinos, and explore their general properties.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

26.6 Superspace Integrals, Field Equations, and the Current Superfield

The '$\mathscr{F}$-terms' and 'D-terms' from which we construct the Lagrangian density may be expressed as integrals over the superspace coordinates θ_α. The rules for integrals over fermionic parameters originally given by Berezin[4] are derived in Section 9.5. Briefly, because the square of any fermionic parameter vanishes, any function of a set of N fermionic parameters ξ_n may be expressed as

$$f(\xi) = \left(\prod_{n=1}^{N} \xi_n\right) c \; + \; \text{terms with fewer } \xi \text{ factors}\,, \tag{26.6.1}$$

and its integral over the ξs is defined simply by

$$\int d^N\xi\, f(\xi) \equiv c\,. \tag{26.6.2}$$

The coefficient c may itself depend on other unintegrated c-number variables that anticommute with the ξs over which we integrate, in which case it is important to standardize the definition of c by moving all ξs to the left of c before integrating over them, as we have done in Eq. (26.6.1). With this definition, integration over fermionic variables is a linear operation. It resembles the integral over a real variable in the sense that, since a shift $\xi_n \to \xi_n + a_n$ of the variable ξ_n by a constant a_n changes the product $\prod_n \xi_n$ only by terms involving fewer ξ factors, it does not affect the value of the integral

$$\int d^N\xi\, f(\xi + a) = \int d^N\xi\, f(\xi)\,. \tag{26.6.3}$$

Also, as a special case of Eq. (26.6.2), the integral over N fermionic parameters of a polynomial of order $< N$ vanishes. Integrals over fermionic and bosonic parameters are strikingly different in the way they respond to change of variables: for bosonic parameters x_n we have $d^N x' = \mathrm{Det}\,(\partial x'/\partial x)\, d^N x$, while for fermionic parameters

$$d^N\xi' = [\mathrm{Det}\,(\partial\xi'/\partial\xi)]^{-1} d^N\xi\,. \tag{26.6.4}$$

In particular, the dimensionality of $d\xi$ is *opposite* to the dimensionality of ξ.

According to Eq. (26.2.10), the D-term of a general superfield $S(x,\theta)$ (which may be elementary or composite) is equal, up to a derivative term, to the coefficient of $-(\bar\theta\gamma_5\theta)^2/4 = -(\theta^{\mathrm{T}}\epsilon\theta)^2/4$. Any one of the four θs can be θ_1, and each possibility gives an equal contribution, so we may assume that θ_1 is the left-most, and pick up a factor of 4. Then θ_2 must be the next-to-left-most. Any of the two remaining θs may be θ_3, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

each possibility gives an equal contribution, so we may assume that θ_3 is third from the left and pick up a factor of 2, and θ_4 must then be the right-most. That is,

$$-\tfrac{1}{4}(\bar{\theta}\gamma_5\theta)^2 = -\tfrac{1}{4} \times 4 \times 2 \times \theta_1\theta_2\theta_3\theta_4 ,$$

so the coefficient of this function of θ is $-1/2$ the integral over $d^4\theta$. Since this is the D-term up to a derivative, we have then

$$\int d^4x \, [S]_D = -\frac{1}{2} \int d^4x \int d^4\theta \, S(x,\theta) . \tag{26.6.5}$$

In the same way, using Eq. (26.3.11), we find that the spacetime integral of the $\mathscr{F}$-term of a general left-chiral superfield Φ (again, either elementary or composite) may be expressed as

$$\int d^4x \, [\Phi]_{\mathscr{F}} = \frac{1}{2} \int d^4x \int d^2\theta_L \, \Phi(x,\theta) . \tag{26.6.6}$$

Since we are now integrating over θs, it is convenient to introduce a delta function, defined as usual by the condition that for an arbitrary function $f(\theta)$

$$\int d^4\theta' \, \delta^4(\theta' - \theta) f(\theta') = f(\theta) . \tag{26.6.7}$$

According to Eq. (9.5.40), this condition is satisfied by

$$\begin{aligned} \delta^4(\theta' - \theta) &= (\theta'_1 - \theta_1)(\theta'_2 - \theta_2)(\theta'_3 - \theta_3)(\theta'_4 - \theta_4) \\ &= \frac{1}{4}\left[\left(\theta_L - \theta'_L\right)^{\mathrm{T}} \epsilon \left(\theta_L - \theta'_L\right)\right] \left[\left(\theta_R - \theta'_R\right)^{\mathrm{T}} \epsilon \left(\theta_R - \theta'_R\right)\right] . \end{aligned} \tag{26.6.8}$$

The representation of the action as an integral over superspace allows an easy derivation of the field equations in superfield form. Consider, for instance, the action for a set of left-chiral scalar superfields Φ_n (which includes as a special case the general renormalizable theory of left-chiral superfields Φ_n):

$$I = \frac{1}{2} \int d^4x \left[K(\Phi, \Phi^*)\right]_D + 2 \, \mathrm{Re} \int d^4x \, [f(\Phi)]_{\mathscr{F}} , \tag{26.6.9}$$

with K an arbitrary function of Φ_n and Φ_n^* without derivatives, and f an arbitrary function of Φ_n, also without derivatives. (The motivation for this form of the action and the expression of the action in terms of component fields is described in Section 26.8.) We cannot derive correct field equations by simply demanding that this is stationary with respect to arbitrary variations in Φ, because the Φ_n are constrained by the requirement that they satisfy the requirement $\mathscr{D}_R\Phi_n = 0$ for left-chiral superfields. To make sure that this condition is preserved by any variation, we can use a trick that will also turn out to be useful in deriving

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

superspace Feynman rules in Chapter 30. We write the Φ_n in terms of *potential superfields* $S_n(x,\theta)$ as

$$\Phi_n = \mathscr{D}_R^2 S_n \,, \tag{26.6.10}$$

from which it follows (using Eq. (26.A.21)) that

$$\Phi_n^* = -\mathscr{D}_L^2 S_n^* \,, \tag{26.6.11}$$

where $\mathscr{D}_R^2$ and $\mathscr{D}_L^2$ are abbreviations for $(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R) = -(\bar{\mathscr{D}}_R\mathscr{D}_R)$ and $(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L) = (\bar{\mathscr{D}}_L\mathscr{D}_L)$, respectively. To see that it is always possible to find an S_n (not necessarily local) that satisfies Eq. (26.6.10), note that for any left-chiral superfields Φ_n,

$$\mathscr{D}_R^2\mathscr{D}_L^2\Phi_n = -16\Box\Phi_n \,, \tag{26.6.12}$$

so that Eq. (26.6.10) is satisfied by the solution of

$$-16\Box S_n = \mathscr{D}_L^2\Phi_n \,. \tag{26.6.13}$$

The expression $\mathscr{D}_R^2 S$ is left-chiral for any S, so the action must be stationary with respect to arbitrary variations in the S_n. Using Eq. (26.6.5), the action may be expressed in terms of the S_n and S_n^* as

$$I = -\frac{1}{4}\int d^4x \int d^4\theta\, K(-\mathscr{D}_L^2 S^*, \mathscr{D}_R^2 S) + 2\,\mathrm{Re}\int d^4x \left[f(\mathscr{D}_R^2 S)\right]_{\mathscr{F}} . \tag{26.6.14}$$

The variation of the first term under infinitesimal changes δS_n in the S_n (but not the S_n^*) is easily calculated by an integration by parts in superspace:

$$\begin{aligned} &-\delta\frac{1}{4}\int d^4x \int d^4\theta\, K(-\mathscr{D}_L^2 S^*, \mathscr{D}_R^2 S) \\ &\qquad = -\sum_n \int d^4\theta\, \delta S_n \mathscr{D}_R^2 \frac{\delta K(-\mathscr{D}_L^2 S^*, \mathscr{D}_R^2 S)}{\delta \mathscr{D}_R^2 S_n} \,. \end{aligned}$$

Eqs. (26.3.31) and (26.6.5) allow us to express the variation of the integral in the superpotential term under infinitesimal changes δS_n in the S_n as

$$\begin{aligned} \delta\int d^4x \left[f(\mathscr{D}_R^2 S)\right]_{\mathscr{F}} &= \sum_n \int d^4x \left[\left.\frac{\partial f(\Phi)}{\partial \Phi_n}\right|_{\Phi=\mathscr{D}_R^2 S} \mathscr{D}_R^2 \delta S_n\right]_{\mathscr{F}} \\ &= \sum_n \int d^4x \left[\mathscr{D}_R^2\left(\left.\frac{\partial f(\Phi)}{\partial \Phi_n}\right|_{\Phi=\mathscr{D}_R^2 S} \delta S_n\right)\right]_{\mathscr{F}} \\ &= 2\sum_n \int d^4x \left[\left.\frac{\partial f(\Phi)}{\partial \Phi_n}\right|_{\Phi=\mathscr{D}_R^2 S} \delta S_n\right]_D \\ &= -\sum_n \int d^4x \int d^4\theta \left.\frac{\partial f(\Phi)}{\partial \Phi_n}\right|_{\Phi=\mathscr{D}_R^2 S} \delta S_n \,. \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

The condition for Eq. (26.6.14) to be stationary with respect to arbitrary variations in S_n is then

$$\mathscr{D}_R^2 \frac{\delta K(-\mathscr{D}_L^2 S^*, \mathscr{D}_R^2 S)}{\delta \mathscr{D}_R^2 S_n} = -4 \left. \frac{\partial f(\Phi)}{\partial \Phi_n} \right|_{\Phi = \mathscr{D}_R^2 S},$$

or in terms of the chiral superfields

$$\mathscr{D}_R^2 \frac{\delta K(\Phi, \Phi^*)}{\delta \Phi_n} = -4 \frac{\partial f(\Phi)}{\partial \Phi_n}. \tag{26.6.15}$$

The complex conjugate yields

$$\mathscr{D}_L^2 \frac{\delta K(\Phi, \Phi^*)}{\delta \Phi_n^*} = 4 \left(\frac{\partial f(\Phi)}{\partial \Phi_n} \right)^*. \tag{26.6.16}$$

It can readily be checked that the components of these equations yield the field equations for the components of Φ_n^* and Φ_n. For instance, recalling that $\mathscr{D}_R^2(\theta_R^{\mathrm{T}} \epsilon \theta_R) = -4$, the θ-independent part of $\mathscr{D}_R^2 \Phi_n^*$ is $4\mathscr{F}_n^*$, while the θ-independent part of $\partial f(\Phi)/\partial \Phi_n$ is $\partial f(\phi)/\partial \phi_n$, so the θ-independent part of Eq. (26.6.15) for $K = \sum_n \Phi_n^* \Phi_n$ yields the relation $\mathscr{F}_n^* = -\partial f(\phi)/\partial \phi_n$, in agreement with Eq. (26.4.6).

As an example of the use of this formalism, let us consider the superfields to which conserved currents belong. Suppose that the superpotential and Kahler potential in the action (26.6.9) are invariant under an infinitesimal global transformation

$$\delta \Phi_n = i\epsilon \sum_m \mathscr{T}_{nm} \Phi_m, \qquad \delta \Phi_n^* = -i\epsilon \sum_m \mathscr{T}_{mn} \Phi_m^*, \tag{26.6.17}$$

with ϵ a real infinitesimal constant and $\mathscr{T}_{nm}$ a Hermitian matrix, perhaps part of a Lie algebra of similar transformation matrices. Since the superpotential depends only on the Φ_n, it is automatically also invariant under the extended transformations

$$\delta \Phi_n = i\epsilon \Lambda \sum_m \mathscr{T}_{nm} \Phi_m, \qquad \delta \Phi_n^* = -i\epsilon \Lambda^* \sum_m \mathscr{T}_{mn} \Phi_m^*, \tag{26.6.18}$$

where $\Lambda(x, \theta)$ is a superfield that must be taken left-chiral in order that the $\delta \Phi_n$ be left-chiral, but is otherwise unrestricted. On the other hand, other terms such as the Kahler potential are not in general invariant under these transformations, because $\Lambda \neq \Lambda^*$. For general fields the change in the action must therefore be of the form

$$\delta I = i\epsilon \int d^4x \int d^4\theta \, [\Lambda - \Lambda^*] \mathscr{J}, \tag{26.6.19}$$

where $\mathscr{J}(x, \theta)$ is some real superfield, known as *the current superfield*. But if the field equations are satisfied then the action is stationary under *any* variation in the superfields, so the integral (26.6.19) must vanish for any

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

left-chiral superfield $\Lambda(x,\theta)$. Any such Λ may be put in the form $\Lambda = \mathscr{D}_R^2 S$, so this means that the current superfield must satisfy

$$\mathscr{D}_R^2 \mathscr{J} = \mathscr{D}_L^2 \mathscr{J} = 0\,. \tag{26.6.20}$$

That is, $\mathscr{J}$ is a *linear* superfield. As we saw in Section 26.3, this means that its components satisfy

$$N^{\mathscr{J}} = M^{\mathscr{J}} = \partial^\mu V_\mu^{\mathscr{J}} = 0\,, \quad \lambda^{\mathscr{J}} = -\not{\partial}\omega^{\mathscr{J}}\,, \quad D^{\mathscr{J}} = -\Box C^{\mathscr{J}}\,. \tag{26.6.21}$$

This allows us to identify the V-component $V_\mu^{\mathscr{J}}$ as the conserved current associated with this symmetry.

For the particular action (26.6.9), the current superfield takes the form

$$\mathscr{J} = \sum_{nm} \frac{\partial K(\Phi,\Phi^*)}{\partial \Phi_n} \mathscr{T}_{nm} \Phi_m = \sum_{nm} \frac{\partial K(\Phi,\Phi^*)}{\partial \Phi_n^*} \mathscr{T}_{mn} \Phi_m^*\,, \tag{26.6.22}$$

with the equality of these two expressions a consequence of symmetry under the transformation (26.6.17). Then, using the field equation (26.6.15),

$$\mathscr{D}_R^2 \mathscr{J} = \sum_{nm} \left[\mathscr{D}_R^2 \frac{\partial K(\Phi,\Phi^*)}{\partial \Phi_n} \right] \mathscr{T}_{nm} \Phi_m = -4 \sum_{nm} \frac{\partial f(\Phi)}{\partial \Phi_n} \mathscr{T}_{nm} \Phi_m\,, \tag{26.6.23}$$

which vanishes because of the assumed invariance of the superpotential under the transformation (26.6.17). In the same way, using the second of the two expressions for $\mathscr{J}$ and the field equation (26.6.16), we find that $\mathscr{D}_L^2 \mathscr{J} = 0$, thus verifying the conservation condition (26.6.20).

26.7 The Supercurrent

Like any other continuous global symmetry, supersymmetry leads to the existence of a conserved current.[5] The conservation and commutation properties of the supersymmetry current are operator equations that will remain valid even when supersymmetry is spontaneously broken, and that will therefore be useful to us in Chapter 29, when we consider theories of spontaneous supersymmetry breaking in a non-perturbative context. Also, the supersymmetry current is related to components of a superfield known as the *supercurrent*,[6] that will be of fundamental importance in our treatment of supergravity in Chapter 31.

As we saw in Section 7.3, the existence of an ordinary global symmetry of the Lagrangian density under an infinitesimal transformation $\chi^\ell \to \chi^\ell + \epsilon \mathscr{F}^\ell$ (with χ^ℓ a generic canonical or auxiliary boson or fermion field and $\mathscr{F}^\ell$ a function of the canonical and auxiliary fields) leads to the existence of a current

$$J^\mu(x) \propto \sum_\ell \frac{\partial \mathscr{L}(x)}{\partial(\partial\chi^\ell(x)/\partial x^\mu)} \mathscr{F}^\ell(x)\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

which is conserved for fields satisfying the field equations, and generates the symmetry in the sense that the canonical commutation relations give

$$\left[\int d^3x\, J^0(x)\,,\ \chi^\ell(y)\right] \propto \mathscr{F}^\ell(y)\,.$$

The supersymmetry current requires a somewhat more complicated treatment for two reasons. One is that supersymmetry is only a symmetry of the action, not of the Lagrangian density or the Lagrangian. Instead, the variation of the Lagrangian density under an infinitesimal supersymmetry transformation is a spacetime derivative, which we may write in the form

$$\delta\mathscr{L} = \sum_\ell \left(\bar{\alpha}\,\partial_\mu K^\mu\right)\,, \tag{26.7.1}$$

with K^μ a four-vector of Majorana spinors. In consequence, the supersymmetry current is not the usual Noether current. The Noether current is a four-vector N^μ of Majorana spinors defined by

$$\sum_\ell \frac{\partial_R\mathscr{L}}{\partial(\partial_\mu\chi^\ell)}\delta\chi^\ell \equiv -\left(\bar{\alpha}\,N^\mu\right)\,, \tag{26.7.2}$$

whose divergence is given by the Euler–Lagrange equations as

$$\begin{aligned}\left(\bar{\alpha}\,\partial_\mu N^\mu\right) &= -\sum_\ell \frac{\partial_R\mathscr{L}}{\partial\chi^\ell}\delta\chi^\ell - \sum_\ell \frac{\partial_R\mathscr{L}}{\partial(\partial_\mu\chi^\ell)}\partial_\mu\delta\chi^\ell \\ &= -\delta\mathscr{L}\,.\end{aligned} \tag{26.7.3}$$

(Here ∂_R denotes the right partial derivative, defined by moving differentiated fermionic variables to the right before differentiating.) Instead, we must define the supersymmetry current as

$$S^\mu \equiv N^\mu + K^\mu\,, \tag{26.7.4}$$

which Eqs. (26.7.1) and (26.7.3) tell us *is* conserved:

$$\partial_\mu S^\mu = 0\,. \tag{26.7.5}$$

The second complication is that the change $\delta\chi^\ell$ in a canonical field χ^ℓ under a supersymmetry transformation is not just a function of canonical fields, but also involves their canonical conjugates. For instance, Eq. (26.3.15) shows that the change in the ψ-component of a chiral scalar superfield involves the time derivative of the ϕ-component. As a result the commutator of the Noether charge $\int d^3x\, N^0$ with a general canonical field does not give the supersymmetry transformation of that field. Fortunately, this complication is cancelled by the first complication: when the commutators of $\int d^3x\, K^0$ as well as $\int d^3x\, N^0$ with the fields are taken into account, the operator $\int d^3x\, S^0$ does generate supersymmetry

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

transformations* in the sense that

$$\left[\int d^3x \left(\bar{\alpha} S^0\right), \chi^\ell\right] = i\,\delta\chi^\ell\,, \tag{26.7.6}$$

consistently with Eqs. (26.2.1) and (26.2.8).

For example, we may derive an explicit formula for the supersymmetry current in the general renormalizable theory of left-chiral superfields Φ_n, which can be used to check that it does generate supersymmetry transformations, in the sense of Eq. (26.7.6). The Lagrangian density (26.4.7) for

* This is a general result for currents constructed in this way. For instance, consider a Lagrangian L (not Lagrangian density) that depends on a set of canonical variables q^n and their time derivatives $\dot{q}^n$, with no constraints of any class. In quantum field theory the label n includes spatial coordinates as well as discrete spin and species labels, and $L = \int d^3x\,\mathscr{L}$. Our assumption here that the Lagrangian density is invariant up to spacetime derivatives under some infinitesimal transformation δ means that δL is the time derivative of some functional F. That is,

$$\sum_n \frac{\partial L}{\partial q^n}\delta q^n + \sum_n \frac{\partial L}{\partial \dot{q}^n}\delta\dot{q}^n = \frac{d}{dt}F\,.$$

Using the canonical equations of motion, this can be written as a conservation law $\dot{Q} = 0$, where the conserved charge is

$$Q = -\sum_n \frac{\partial L}{\partial \dot{q}^n}\delta q^n + F\,.$$

In our case here, $Q = \int d^3x\,[N^0 + K^0]$. We assume the usual unconstrained commutation relations

$$\left[\frac{\partial L}{\partial \dot{q}^n}, q^m\right] = -i\,\delta_n^m\,, \qquad \left[q^n, q^m\right] = 0\,,$$

and find the commutator

$$\left[Q, q^m\right] = i\,\delta q^m - \sum_{nl} \frac{\partial L}{\partial \dot{q}^l}\frac{\partial \delta q^l}{\partial \dot{q}^n}\left[\dot{q}^n, q^m\right] + \sum_n \frac{\partial F}{\partial \dot{q}^n}\left[\dot{q}^n, q^m\right].$$

To evaluate the second and third terms, we note that the second time derivatives $\ddot{q}^n$ appear linearly in the invariance condition, so their coefficients must match: even without using the equations of motion, we have

$$\sum_l \frac{\partial L}{\partial \dot{q}^l}\frac{\partial \delta q^l}{\partial \dot{q}^n} = \frac{\partial F}{\partial \dot{q}^n}\,.$$

The second and third terms in the commutator therefore cancel, leaving us with the desired result

$$\left[Q, q^m\right] = i\,\delta q^m\,.$$

Taking the time derivative gives also

$$\left[Q, \dot{q}^m\right] = i\,\delta\dot{q}^m\,.$$

This result has been extended to theories with constraints.[7]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

this theory may be put in the form

$$\mathscr{L} = \sum_n \left[-\partial_\mu \phi_n^* \partial^\mu \phi_n - \frac{1}{2}\left(\overline{\psi_{Ln}}\,\gamma^\mu \partial_\mu \psi_{Ln}\right) - \frac{1}{2}\left(\overline{\psi_{Rn}}\,\gamma^\mu \partial_\mu \psi_{Rn}\right)\right] + \text{non-derivative terms}\,. \tag{26.7.7}$$

Using the transformation rules (26.3.15), (26.3.17), (26.3.18), and (26.3.20) (with $\tilde{\phi} = \phi^*$), the Noether current defined by Eq. (26.7.2) is

$$N^\mu = \frac{1}{\sqrt{2}} \sum_n \left[2\,(\partial^\mu \phi_n^*)\,\psi_{nL} + 2\,(\partial^\mu \phi_n)\,\psi_{nR} + (\not{\partial}\phi_n)\,\gamma^\mu \psi_{nR} + (\not{\partial}\phi_n^*)\,\gamma^\mu \psi_{nL} - \mathscr{F}_n\,\gamma^\mu \psi_{nR} - \mathscr{F}_n^*\,\gamma^\mu \psi_{nL}\right]. \tag{26.7.8}$$

We can calculate the change in the Lagrangian density either directly, or, more easily, by noting that the changes of D-terms and $\mathscr{F}$-terms under supersymmetry transformations are given respectively by Eqs. (26.2.17) and (26.3.16). Either way, we find that the current K^μ in Eq. (26.7.1) is

$$K^\mu = \frac{1}{\sqrt{2}} \sum_n \gamma^\mu \left[-(\not{\partial}\phi_n)\,\psi_{nR} - (\not{\partial}\phi_n^*)\,\psi_{nL} + \mathscr{F}_n^*\,\psi_{nL} + \mathscr{F}_n\,\psi_{nR} + 2\left(\frac{\partial f(\phi)}{\partial \phi_n}\right)\psi_{nL} + 2\left(\frac{\partial f(\phi)}{\partial \phi_n}\right)^* \psi_{nR}\right]. \tag{26.7.9}$$

Adding (26.7.8) and (26.7.9) gives the supersymmetry current for this class of theories

$$S^\mu = \sqrt{2} \sum_n \left[(\not{\partial}\phi_n)\gamma^\mu \psi_{nR} + (\not{\partial}\phi_n^*)\gamma^\mu \psi_{nL} + \left(\frac{\partial f}{\partial \phi_n}\right)\gamma^\mu \psi_{nL} + \left(\frac{\partial f}{\partial \phi_n}\right)^* \gamma^\mu \psi_{nR}\right]. \tag{26.7.10}$$

It is straightforward then to use the canonical commutation and anti-commutation relations to verify that $\int d^3x\, S^0$ satisfies the commutation relations (26.7.6).

There is another definition of symmetry currents, in terms of the response of the matter action to a local symmetry transformation, which is particularly useful when the associated symmetries are 'gauged,' as supersymmetry will be when we turn to supergravity theories in Chapter 31. In the absence of supergravity fields, the action is not invariant under local supersymmetry transformations. If we make such a transformation with a spacetime-dependent parameter $\alpha(x)$, the action will change by an amount that, in order to vanish when $\alpha(x)$ is constant, must (even when the field equations are not satisfied) be of the form

$$\delta I = -\int d^4x \left((\partial_\mu \bar{\alpha}(x))\, S^\mu(x) \right), \tag{26.7.11}$$

with $S^\mu(x)$ a four-vector of Majorana spinor operator coefficients. This

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

does not define $S^\mu(x)$ uniquely, because when we generalize global supersymmetry transformations to local transformations, we might in general give the change $\delta\chi$ of a field χ under a local supersymmetry transformation an arbitrary dependence on the derivatives of $\alpha(x)$. There is, however, one way of defining local supersymmetry transformations that guarantees that the coefficient $S^\mu(x)$ in Eq. (26.7.11) is the same as the current defined by Eq. (26.7.4), which as we have seen generates the symmetry transformations in the sense of Eq. (26.7.6). It is to specify that *derivatives of* $\alpha(x)$ *do not appear in the supersymmetry transformation of the canonical or auxiliary fields* χ^ℓ. For instance, the local versions of the transformation rules (26.3.15)–(26.3.17) for the components of a left-chiral superfield are

$$\delta\psi_L(x) = \sqrt{2}\partial_\mu\phi(x)\,\gamma^\mu\,\alpha_R(x)\,\phi(x) + \sqrt{2}\mathcal{F}(x)\,\alpha_L(x)\,, \quad (26.7.12)$$

$$\delta\mathcal{F}(x) = \sqrt{2}\left(\overline{\alpha_L}(x)\,\partial\!\!\!/\psi_L(x)\right)\,, \quad (26.7.13)$$

$$\delta\phi(x) = \sqrt{2}\left(\overline{\alpha_R}(x)\psi_L(x)\right)\,. \quad (26.7.14)$$

Eq. (26.3.21) shows that the superfield may be expressed in terms of its component fields at x_+^μ without derivatives, so the transformation rule for the superfield may be expressed as

$$\delta\Phi(x,\theta) = \left(\bar{\alpha}(x_+)\,\mathcal{Q}\right)\Phi(x,\theta)\,, \quad (26.7.15)$$

where $\mathcal{Q}$ is the operator (26.2.2).

With local supersymmetry transformations defined in this way, the change of the action that they induce consists of two terms. First, although the variation in canonical fields under a supersymmetry transformation does not involve derivatives of $\alpha(x)$, the variation in derivatives of canonical fields does. This produces a change in the Lagrangian density which is the same as Eq. (26.7.2), except that $\bar{\alpha}$ is replaced with $\partial_\mu\bar{\alpha}$:

$$\delta_1 I = -\int d^4x\left([\partial_\mu\bar{\alpha}(x)]\,N^\mu(x)\right).$$

The second term in the change of the action arises from the fact that the Lagrangian density is not invariant even under the part of the supersymmetry transformation that does not involve derivatives of $\alpha(x)$. According to Eq. (26.7.1), this produces a change in the action

$$\delta_2 I = \int d^4x\left(\bar{\alpha}(x)\,\partial_\mu K^\mu(x)\right) = -\int d^4x\left((\partial_\mu\bar{\alpha}(x))\,K^\mu(x)\right)\,.$$

Adding $\delta_1 I$ and $\delta_2 I$ gives a total change in the action of the form (26.7.11), with $S^\mu(x)$ given by Eq. (26.7.4), as was to be shown.

Even with this specification of the transformation property of the component fields, the supersymmetry current $S^\mu(x)$ is not uniquely specified

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

by Eq. (26.7.11), because we could always introduce a modified current

$$S^{\mu}_{\text{new}} = S^{\mu} + \partial_{\nu} A^{\mu\nu} , \tag{26.7.16}$$

where $A^{\mu\nu} = -A^{\nu\mu}$ is an arbitrary antisymmetric tensor of Majorana spinors. The term $\partial_{\nu} A^{\mu\nu}$ is conserved whether or not the field equations are satisfied, and its time component is a space derivative, so $\int d^3x \, S^0_{\text{new}} = \int d^3x \, S^0$, leaving Eq. (26.7.6) unchanged.

There is in fact a particular choice of $A^{\mu\nu}$ with the convenient feature that $\gamma_{\mu} S^{\mu}_{\text{new}}$ turns out to be a measure of the violation of scale invariance by the theory. By using the Dirac equations derived from the Lagrangian density (26.4.7):

$$\not{\partial} \psi_{mL} = -\sum_n \left(\frac{\partial^2 f(\phi)}{\partial \phi_m \partial \phi_n} \right)^* \psi_{nR} , \qquad \not{\partial} \psi_{mR} = -\sum_n \left(\frac{\partial^2 f(\phi)}{\partial \phi_m \partial \phi_n} \right) \psi_{nL} , \tag{26.7.17}$$

it is straightforward to calculate that

$$\begin{aligned} \gamma_{\mu} S^{\mu} = -2\sqrt{2} \sum_n \Bigg\{ & \not{\partial} \left(\phi_n \psi_{nR} + \phi_n^* \psi_{nL} \right) \\ & + \left(\sum_m \phi_m \frac{\partial^2 f(\phi)}{\partial \phi_n \partial \phi_m} - 2 \frac{\partial f(\phi)}{\partial \phi_n} \right) \psi_{nL} \\ & + \left(\sum_m \phi_m \frac{\partial^2 f(\phi)}{\partial \phi_n \partial \phi_m} - 2 \frac{\partial f(\phi)}{\partial \phi_n} \right)^* \psi_{nR} \Bigg\} . \end{aligned}$$

We can eliminate the first term by introducing a modified supersymmetry current of the general type of Eq. (26.7.16):

$$S^{\mu}_{\text{new}} = S^{\mu} + \frac{\sqrt{2}}{3} [\gamma^{\mu}, \gamma^{\nu}] \sum_n \partial_{\nu} \left(\phi_n \psi_{nR} + \phi_n^* \psi_{nL} \right) , \tag{26.7.18}$$

for which

$$\begin{aligned} \gamma_{\mu} S^{\mu}_{\text{new}} = -2\sqrt{2} \sum_n \Bigg\{ & \left(\sum_m \phi_m \frac{\partial^2 f(\phi)}{\partial \phi_n \partial \phi_m} - 2 \frac{\partial f(\phi)}{\partial \phi_n} \right) \psi_{nL} \\ & + \left(\sum_m \phi_m \frac{\partial^2 f(\phi)}{\partial \phi_n \partial \phi_m} - 2 \frac{\partial f(\phi)}{\partial \phi_n} \right)^* \psi_{nR} \Bigg\} . \end{aligned} \tag{26.7.19}$$

The right-hand side vanishes for a scale-invariant Lagrangian density, with $f(\Phi)$ a homogeneous polynomial of third order in the Φ_n.

We now turn to the supersymmetry transformation property of the supersymmetry current. It is straightforward to check that the current given by Eqs. (26.7.18) and (26.7.10) is related to the ω-component ω_{μ}^{Θ} of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

a real non-chiral superfield Θ_μ by**

$$S^\mu_{\text{new}} = -2\omega^{\Theta\,\mu} + 2\gamma^\mu\,\gamma^\nu\,\omega^\Theta_\nu\,, \tag{26.7.20}$$

where

$$\Theta_\mu = \frac{i}{12}\sum_n \left[4\Phi^*_n\partial_\mu\Phi_n - 4\Phi_n\partial_\mu\Phi^*_n + \left((\bar{\mathscr{D}}\Phi^*_n)\gamma_\mu(\mathscr{D}\Phi_n)\right)\right]. \tag{26.7.21}$$

The superfield Θ^μ is known as the *supercurrent*.

The supercurrent obeys a conservation law that incorporates the conservation of the supersymmetry current (26.7.20) and much else besides. To derive it, we may use the anticommutation relation (26.2.30) to write†

$$[\mathscr{D}_R\,,\,(\bar{\mathscr{D}}_L\mathscr{D}_L)] = -4\,\not{\partial}\mathscr{D}_L\,.$$

Together with the chirality conditions $\mathscr{D}_R\Phi_n = \mathscr{D}_L\Phi^*_n = 0$, this gives

$$\gamma^\mu\mathscr{D}_L\sum_n\left[\Phi^*_n\,\partial_\mu\Phi_n - \Phi_n\,\partial_\mu\Phi^*_n\right] = -\tfrac{1}{4}\sum_n\Phi^*_n\mathscr{D}_R\left(\bar{\mathscr{D}}_L\mathscr{D}_L\right)\Phi_n - \sum_n(\not{\partial}\Phi^*_n)\,\mathscr{D}_L\Phi_n$$

and

$$\gamma^\mu\mathscr{D}_L\sum_n\left((\bar{\mathscr{D}}\Phi^*_n)\,\gamma_\mu(\mathscr{D}\Phi_n)\right) = 4\sum_n(\not{\partial}\Phi^*_n)\,\mathscr{D}\Phi_n + 2\sum_n\mathscr{D}\Phi^*_n\left(\bar{\mathscr{D}}_L\mathscr{D}_L\right)\Phi_n\,,$$

so that the superfield (26.7.21) satisfies

$$\gamma_\mu\,\mathscr{D}_L\Theta^\mu = \tfrac{1}{6}\,i\sum_n(\mathscr{D}_R\Phi^*_n)\,\left(\bar{\mathscr{D}}_L\mathscr{D}_L\right)\Phi_n - \tfrac{1}{12}\,i\sum_n\Phi^*_n\mathscr{D}_R\left(\bar{\mathscr{D}}_L\mathscr{D}_L\right)\Phi_n\,. \tag{26.7.22}$$

We saw in Section 26.6 that the field equations for the Lagrangian density (26.4.7) may be expressed in the form

$$\left(\bar{\mathscr{D}}_L\mathscr{D}_L\right)\Phi_n = -4\left(\frac{\partial f(\Phi)}{\partial\Phi_n}\right)^*\,. \tag{26.7.23}$$

** Here we introduce a notation that will be used extensively in Chapter 31; following Eq. (26.2.10), the components C^S, ω^S, M^S, N^S, V^S_ν, λ^S, and D^S of an arbitrary superfield $S(x,\theta)$ are defined by the expansion

$$\begin{aligned}S(x,\theta) = {}& C^S(x) - i\left(\bar\theta\,\gamma_5\,\omega^S(x)\right) - \frac{i}{2}\left(\bar\theta\,\gamma_5\,\theta\right)M^S(x) - \frac{1}{2}\left(\bar\theta\,\theta\right)N^S(x)\\ &+\frac{i}{2}\left(\bar\theta\,\gamma_5\,\gamma^\nu\,\theta\right)V^S_\nu(x) - i\left(\bar\theta\,\gamma_5\,\theta\right)\left(\bar\theta\left[\lambda^S(x) + \frac{1}{2}\,\not{\partial}\omega^S(x)\right]\right)\\ &-\frac{1}{4}\left(\bar\theta\,\gamma_5\,\theta\right)^2\left[D^S(x) + \frac{1}{2}\Box C^S(x)\right].\end{aligned}$$

† It should be noted that $\bar{\mathscr{D}}_L$ and $\bar{\mathscr{D}}_R$ are the left- and right-handed components of the covariant adjoint $\bar{\mathscr{D}}$, rather than the covariant adjoints $\overline{\mathscr{D}_L}$ and $\overline{\mathscr{D}_R}$ of $\mathscr{D}_L$ and $\mathscr{D}_R$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Using this in Eq. (26.7.22) gives finally

$$\gamma^\mu \mathscr{D}_L \Theta_\mu = -\frac{2}{3} i \sum_n (\mathscr{D}_R \Phi_n^*) \left(\frac{\partial f}{\partial \Phi_n}\right)^* + \frac{1}{3} i \sum_n \Phi_n^* \mathscr{D}_R \left(\frac{\partial f}{\partial \Phi_n}\right)^*$$
$$= \frac{1}{3} i \mathscr{D}_R \left[\sum_n \Phi_n \frac{\partial f(\Phi)}{\partial \Phi_n} - 3 f(\Phi)\right]^* . \tag{26.7.24}$$

The Hermitian adjoint of Eq. (26.7.24) is

$$\gamma^\mu \mathscr{D}_R \Theta_\mu = -\frac{1}{3} i \mathscr{D}_L \left[\sum_n \Phi_n \frac{\partial f(\Phi)}{\partial \Phi_n} - 3 f(\Phi)\right] . \tag{26.7.25}$$

The sum of this and Eq. (26.7.24) then gives the conservation law

$$\gamma^\mu \mathscr{D} \Theta_\mu = \mathscr{D} X , \tag{26.7.26}$$

where X is a real chiral superfield given (up to an additive constant) in this class of theories by

$$X = \frac{2}{3} \mathrm{Im} \left[\sum_n \Phi_n \frac{\partial f(\Phi)}{\partial \Phi_n} - 3 f(\Phi)\right] . \tag{26.7.27}$$

Although it has been derived here only for renormalizable theories of chiral superfields, we may expect that the conservation law (26.7.26) holds more generally, though of course with X not necessarily given by Eq. (26.7.27), because of the other conservation laws that it incorporates. (A generalized formula for X will be given in Section 31.4.) To derive these relations, we must use Eq. (26.2.10) to express Θ_μ in terms of components C_μ^Θ, ω_μ^Θ, etc. and use Eq. (26.3.9) to express the chiral superfield X in terms of components A^X, ψ^X, etc. With the aid of Eqs. (26.A.9), (26.A.16), (26.A.17), and the Dirac matrix identities

$$[\gamma^\rho , \gamma^\sigma] = -\tfrac{1}{2} i \epsilon^{\rho\sigma\mu\nu} \gamma_5 [\gamma_\mu , \gamma_\nu] , \tag{26.7.28}$$

$$\gamma^\mu \gamma^\rho \gamma^\nu = \eta^{\mu\rho} \gamma^\nu - \eta^{\mu\nu} \gamma^\rho + \eta^{\nu\rho} \gamma^\mu + i \gamma_5 \epsilon^{\mu\nu\rho\sigma} \gamma_\sigma , \tag{26.7.29}$$

we can then expand both sides of Eq. (26.7.26) in the terms

$$1 , \quad \theta , \quad \gamma_5 \theta , \quad \gamma^\nu \theta , \quad \gamma_5 \gamma^\nu \theta , \quad \gamma_5 [\gamma^\mu , \gamma^\nu] \theta ,$$
$$\left(\bar\theta \theta\right) , \quad \left(\bar\theta \gamma_5 \theta\right) , \quad \left(\bar\theta \gamma_5 \gamma^\nu \theta\right) ,$$
$$\theta \left(\bar\theta \gamma_5 \theta\right) , \quad \gamma_5 \theta \left(\bar\theta \gamma_5 \theta\right) , \quad \gamma^\nu \theta \left(\bar\theta \gamma_5 \theta\right) ,$$
$$\gamma^\nu \gamma_5 \theta \left(\bar\theta \gamma_5 \theta\right) , \quad [\gamma^\rho , \gamma^\sigma] \theta \left(\bar\theta \gamma_5 \theta\right) , \quad \left(\bar\theta \gamma_5 \theta\right)^2 .$$

Matching the coefficients of 1, θ, $\gamma_5\theta$, $\gamma^\nu\theta$, $\gamma_5\gamma^\nu\theta$, $\gamma_5[\gamma^\mu, \gamma^\nu]\theta$, respectively,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

yields the results[††]

$$\psi^X = -i\gamma_5\gamma^\mu\omega_\mu^\Theta \,, \tag{26.7.30}$$

$$F^X = \partial^\mu C_\mu^\Theta \,, \tag{26.7.31}$$

$$G^X = (V^\Theta)^\mu{}_\mu \,, \tag{26.7.32}$$

$$\partial_\mu A^X = -N_\mu^\Theta \,, \tag{26.7.33}$$

$$\partial_\mu B^X = M_\mu^\Theta \,, \tag{26.7.34}$$

$$0 = V_{\mu\nu}^\Theta - V_{\nu\mu}^\Theta + \epsilon_{\mu\nu\rho\sigma}\partial^\sigma C^{\Theta\rho} \,. \tag{26.7.35}$$

Matching the coefficients of either $(\bar\theta\theta)$ or $(\bar\theta\gamma_5\theta)$ yields the same result:

$$0 = \gamma^\mu \lambda_\mu^\Theta \,, \tag{26.7.36}$$

and matching the coefficients of $(\bar\theta\gamma_5\gamma^\nu\theta)$ yields the result

$$-i\gamma_5[\gamma^\nu, \not\partial]\psi^X = 2\gamma^\mu\gamma^\nu\lambda_\mu^\Theta + \gamma^\mu[\gamma^\nu, \not\partial]\omega_\mu^\Theta \,. \tag{26.7.37}$$

From Eqs. (26.7.30), (26.7.36), and (26.7.37) we obtain the conservation of the supersymmetry current (26.7.20):

$$0 = \partial_\mu S^\mu_{\text{new}} = -2\partial^\mu\omega_\mu^\Theta + 2\not\partial\gamma^\mu\omega_\mu^\Theta \,, \tag{26.7.38}$$

and a relation among λ_μ^Θ and ω_μ^Θ:

$$\lambda_\nu^\Theta = -\not\partial\omega_\nu^\Theta + \partial_\nu\gamma^\mu\omega_\mu^\Theta \,. \tag{26.7.39}$$

Matching the coefficients of $\theta(\bar\theta\gamma_5\theta)$ and $\gamma_5\theta(\bar\theta\gamma_5\theta)$ gives relations that can be obtained by taking the divergence of Eqs. (26.7.34) and (26.7.33), respectively. Matching the coefficients of $\gamma^\rho\theta(\bar\theta\gamma_5\theta)$ gives

$$\partial_\rho G^X = \partial^\mu V_{\mu\rho}^\Theta + \partial^\mu V_{\rho\mu}^\Theta - \partial_\rho V^{\Theta\lambda}{}_\lambda \,, \tag{26.7.40}$$

which, combined with Eq. (26.7.32), yields the conservation law

$$\partial_\mu T^{\mu\nu} = 0 \,, \tag{26.7.41}$$

where $T^{\mu\nu}$ is the symmetric tensor

$$T_{\mu\nu} \equiv -\tfrac{1}{2}V_{\mu\nu}^\Theta - \tfrac{1}{2}V_{\nu\mu}^\Theta + \eta_{\mu\nu}V^{\Theta\lambda}{}_\lambda \,. \tag{26.7.42}$$

Matching the coefficients of $\gamma^\rho\gamma_5\theta(\bar\theta\gamma_5\theta)$ gives

$$\partial_\mu F^X = 2D_\mu^\Theta + \Box C_\mu^\Theta + \epsilon_{\rho\nu\sigma\mu}\partial^\nu V^{\Theta\rho\sigma} \,, \tag{26.7.43}$$

which, with Eqs. (26.7.31) and (27.6.35), yields a relation between D_μ^Θ and C_μ^Θ:

$$D_\mu^\Theta = -\Box C_\mu^\Theta + \partial_\mu\partial^\nu C_\nu^\Theta \,. \tag{26.7.44}$$

[††] Note that $V_{\mu\nu}^\Theta$ is the V_ν-component of Θ_μ, not the V_μ-component of Θ_ν.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Matching the coefficients of $[\gamma^\rho,\gamma^\sigma]\theta(\bar\theta\gamma_5\theta)$ and $(\bar\theta\gamma_5\theta)^2$ gives results that already follow from Eq. (26.7.34) and from Eqs. (26.7.38) and (26.7.39), respectively.

The conserved symmetric tensor $T^{\mu\nu}$ may be identified as the energy-momentum tensor of the system. To check this, we use Eqs. (26.1.18) and (26.2.12) to write the change in $\omega_\mu^\Theta(x)$ under a supersymmetry transformation with infinitesimal parameter α as

$$\begin{aligned}\delta\omega_\mu^\Theta &= -i\left[(\bar Q\alpha),\,\omega_\mu^\Theta\right] = +i\left[\omega_\mu^\Theta,\,(\bar Q\alpha)\right]\\ &= \left(-i\gamma_5\,\partial\!\!\!/ C_\mu^\Theta - M_\mu^\Theta + i\gamma_5 N_\mu^\Theta + \gamma^\nu V_{\mu\nu}^\Theta\right)\alpha\,.\end{aligned}$$

Eqs. (26.7.33)–(26.7.35) allow us to put this in the form

$$i\left\{\omega_\mu^\Theta,\,\bar Q\right\} = \tfrac{1}{2}\gamma^\nu(V_{\mu\nu}^\Theta + V_{\nu\mu}^\Theta) - \partial_\mu(B^X + \gamma_5 A^X) - i\gamma_5\,\partial\!\!\!/ C_\mu^\Theta + \tfrac{1}{2}\epsilon_{\mu\nu\kappa\sigma}\gamma^\nu\partial^\kappa C^{\Theta\,\sigma}\,.$$

In terms of the currents (26.7.20) and (26.7.42), this reads

$$\begin{aligned}i\{S^\mu_{\rm new},\bar Q\} &= 2\gamma_\nu T^{\mu\nu} + 2(\partial^\mu - \gamma^\mu\,\partial\!\!\!/)(B^X + \gamma_5 A^X) - \epsilon^{\mu\nu\kappa\sigma}\gamma_\nu\partial_\kappa C_\sigma^\Theta\\ &\quad + 2i\gamma_5\Big(\partial\!\!\!/ C^{\Theta\,\mu} - \gamma^\mu\gamma^\lambda\,\partial\!\!\!/ C_\lambda^\Theta - \tfrac{1}{2}\gamma^\mu[\partial\!\!\!/,\gamma^\sigma]C_\sigma^\Theta\Big)\,. \qquad (26.7.45)\end{aligned}$$

For $\mu = 0$ all the terms on the right except the first are space derivatives, and therefore vanish when we integrate over space, leaving us with

$$i\left\{\int d^3x\; S^0_{\rm new}\,,\,\bar Q\right\} = 2\gamma_\nu\int d^3x\; T^{0\nu}\,. \qquad (26.7.46)$$

We have defined the supersymmetry current $S^\mu_{\rm new}$ to give $\int d^3x\; S^0_{\rm new} = Q$, so the fundamental anticommutation relation (25.2.36) tells us that

$$\int d^3x\; T^{0\nu} = P^\nu\,, \qquad (26.7.47)$$

which, with the conservation condition (26.7.41), allows us to identify $T^{\mu\nu}$ as the energy-momentum tensor.

It is important to note *which* energy-momentum tensor we have constructed in this way. Either directly from Eq. (26.7.21), or by considering the supersymmetry transformation of the current (26.7.18), we can calculate that the energy-momentum tensor $T^{\mu\nu}$ for renormalizable theories of chiral superfields is

$$\begin{aligned}T^{\mu\nu} &= \sum_n\left[\partial^\mu\phi_n^*\partial^\nu\phi_n + \partial^\nu\phi_n^*\partial^\mu\phi_n\right] - \eta^{\mu\nu}\sum_n\left[\partial^\lambda\phi_n^*\partial_\lambda\phi_n + \left|\frac{\partial f(\phi)}{\partial\phi_n}\right|^2\right]\\ &\quad + \frac{1}{3}\left(\eta^{\mu\nu}\Box - \partial^\mu\partial^\nu\right)\sum_n|\phi_n|^2 + \cdots\,, \qquad (26.7.48)\end{aligned}$$

where the dots denote terms involving fermions, which do not concern us here. We see that the last term, which is related by supersymmetry

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

to the correction term in Eq. (26.7.18), has the effect that the energy-momentum tensor is traceless for the massless free-field theory with no superpotential, in which case $\Box\phi_n = 0$. A simple calculation shows that $T^{\mu\nu}$ is also traceless more generally for scale-invariant theories, with $f(\phi)$ a homogeneous polynomial of third order in the ϕ_n.

Supersymmetry also imposes an interesting relation among violations of scale invariance and R conservation. Eqs. (26.7.30)–(26.7.32) show that $\gamma_\mu S^\mu_{\text{new}} = 6\gamma_\mu \omega^{\Theta\,\mu}$, $\partial^\mu C^\Theta_\mu$, and $T^\lambda{}_\lambda = 2V^{\Theta\,\mu}{}_\mu$ (which measures the violation of scale invariance) are proportional to components of a chiral superfield X, so if any one of these vanishes as an operator equation (that is, not just for some particular field configuration), then they all do. In this case, we can show that $C^{\Theta\,\rho}$ is proportional to the current of an R quantum number. To see this, note that Eq. (26.2.11) gives

$$\delta C^\Theta_\sigma = i\left[C^\Theta_\sigma\,,\,(\bar\alpha\, Q)\right] = i\left(\bar\alpha\,\gamma_5\,\omega^\Theta_\sigma\right),$$

so that in general

$$\left[C^\Theta_\sigma\,,\,Q\right] = \gamma_5\omega^\Theta_\sigma\,. \tag{26.7.49}$$

We have seen that if C^Θ_σ is conserved then $\gamma_\mu S^\mu = 0$, so that Eq. (26.7.20) gives $S_\sigma = -2\omega^\Theta_\sigma$. Setting $\sigma = 0$ and integrating over $\mathbf{x}$ in Eq. (26.7.49) then gives

$$\left[\int d^3x\, C^{0\,\Theta}\,,\,Q\right] = -\tfrac{1}{2}\gamma_5\, Q\,. \tag{26.7.50}$$

We can therefore introduce a current

$$\mathscr{R}^\mu \equiv 2\, C^{\mu\,\Theta}\,, \tag{26.7.51}$$

which *if* conserved is the current of a quantum number $\mathscr{R} \equiv \int d^3x\, \mathscr{R}^0$ for which Q_L and Q_R destroy the values $+1$ and -1, respectively. Since the commutator of Q_L with a scalar superfield Φ involves a term $\partial\Phi/\partial\theta_L$, this means that θ_L carries the $\mathscr{R}$ value $+1$, in accord with the usual definition. A theory in which the superfield X vanishes, or, equivalently, in which $T^\mu{}_\mu$, $\gamma_\mu S^\mu$, and $\partial_\mu \mathscr{R}^\mu$ all vanish, is invariant under an enlarged set of supersymmetry transformations, generated by the superconformal algebra described at the end of Section 25.2.

In scale-invariant theories the value of the $\mathscr{R}$ quantum number carried by various superfields is fixed by the structure of the Lagrangian. For instance, in a scale-invariant theory of chiral scalar superfields, the superpotential must be a homogeneous polynomial of third order in the superfields. The $\mathscr{F}$-term of the superpotential is proportional to the coefficient of θ_L^2, which has $\mathscr{R}$ quantum number $+2$, so the $\mathscr{R}$ quantum number of the $\mathscr{F}$-term of the superpotential is the $\mathscr{R}$ quantum number of the superpotential itself minus two. $\mathscr{R}$ invariance then requires that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

we give the scalar superfields an $\mathscr{R}$ quantum number $+2/3$, so that the superpotential will have $\mathscr{R}$ quantum number $+2$, and its $\mathscr{F}$-term will have $\mathscr{R}$ quantum number zero. That is, the scalar components ϕ_n have $\mathscr{R} = 2/3$ and the spinor components ψ_{nL} (proportional to the coefficient of θ_L in the superfield) have $\mathscr{R} = -1/3$. This can be verified by calculating the current $\mathscr{R}^\mu$ from the C-term of the supercurrent (26.7.21) for this class of theories:

$$\mathscr{R}_\mu = \tfrac{2}{3} i [\phi^* \partial_\mu \phi - \phi \partial_\mu \phi^*] - \tfrac{1}{6} i \left(\bar{\psi} \gamma_\mu \gamma_5 \psi\right) . \tag{26.7.52}$$

(The second term contains an extra factor of $1/2$ because ψ is a Majorana spinor.)

Quantum corrections can introduce violations of $\mathscr{R}$ invariance (through Adler–Bell–Jackiw anomalies) and of scale invariance (through the renormalization group running of the coupling constants) but supersymmetry continues to impose a relation between these symmetry violations despite these corrections.[7a] We will see an example of this in Section 29.3.

* * *

The conservation condition (26.7.26) does not uniquely determine either the supercurrent Θ^μ or the associated chiral superfield X. In particular, we may add to Θ^μ a change

$$\Delta\Theta^\mu = \partial^\mu Y , \tag{26.7.53}$$

with Y an arbitrary chiral superfield. The left-hand side of Eq. (26.7.26) is then changed by

$$\gamma_\mu \mathscr{D} \Delta\Theta^\mu = \bar{\not{\partial}} \mathscr{D} Y .$$

For a left-chiral superfield Y_L the chirality condition $\mathscr{D}_R Y_L = 0$ and the anticommutation relations (26.2.30) give

$$\begin{aligned} \bar{\not{\partial}} \mathscr{D}_\alpha Y &= -\tfrac{1}{2} \left[\left\{ \mathscr{D}_L, \bar{\mathscr{D}}_R \right\} \mathscr{D}_R \right]_\alpha Y_L \\ &= -\tfrac{1}{2} \left[\mathscr{D}_{L\alpha} (\bar{\mathscr{D}}_R \mathscr{D}_R) Y_L + \sum_\beta \bar{\mathscr{D}}_{L\beta} \mathscr{D}_{R\alpha} \mathscr{D}_{L\beta} Y_L \right] . \end{aligned}$$

The matrix $\epsilon\gamma_5$ in $\bar{\mathscr{D}}_{L\beta}$ in the second term on the right of the above expression may be moved to the final operator $\mathscr{D}_{L\beta}$, so that the chirality condition and anticommutation relations give

$$\sum_\beta \bar{\mathscr{D}}_{L\beta} \mathscr{D}_{R\alpha} \mathscr{D}_{L\beta} Y_L = -\sum_\beta \mathscr{D}_{L\beta} \mathscr{D}_{R\alpha} \bar{\mathscr{D}}_{L\beta} Y_L = 2(\bar{\not{\partial}} \mathscr{D})_\alpha Y_L ,$$

and hence

$$\bar{\not{\partial}} \mathscr{D} Y_L = -\tfrac{1}{2} \left[\mathscr{D} \left(\bar{\mathscr{D}} \mathscr{D} \right) Y_L + 2 \, \bar{\not{\partial}} \mathscr{D} Y_L \right] = -\tfrac{1}{4} \mathscr{D} \left(\bar{\mathscr{D}} \mathscr{D} \right) Y_L .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

The same result can be derived in the same way for any right-chiral superfield, and hence it holds also for an arbitrary sum Y of a left-chiral and a right-chiral superfield

$$\gamma_\mu \mathscr{D} \Delta\Theta^\mu = \not{\partial} \mathscr{D} Y = -\tfrac{1}{4} \mathscr{D} \left(\bar{\mathscr{D}} \mathscr{D} \right) Y \ . \tag{26.7.54}$$

This is of the same form as the conservation condition (26.7.26), with the associated chiral superfield X changed by the chiral superfield

$$\Delta X = -\tfrac{1}{4} (\bar{\mathscr{D}} \mathscr{D}) Y \ . \tag{26.7.55}$$

It is easy to check that the addition of $\Delta\Theta^\mu$ to Θ^μ changes $T^{\mu 0}$ and S^0_{new} only by space derivatives, and therefore does not change the energy-momentum four-vector P^μ or the supercharge Q.

We saw in Section 26.6 that any chiral superfield X may be expressed in the form $X = (\bar{\mathscr{D}}\mathscr{D})S$, and hence can be removed by adding a term of the form (26.7.55) to Θ^μ with $Y = 4S$. But in general S and the new Θ^μ constructed in this way will not be local. This situation is already familiar from our experience with triangle anomalies, discussed in Chapter 22 — we saw there that, although it is always possible to construct terms that if added to the Lagrangian density would cancel these anomalies, in general these terms would not be local, and hence must be excluded from the Lagrangian density. There are chiral superfields that can be expressed as $(\bar{\mathscr{D}}\mathscr{D})S$ with S local, and that therefore, if present in the associated chiral superfield X, could be eliminated by adding local terms of the form (26.7.53) to Θ^μ. They include, for instance, a term of the form $\text{Re}\,(k\partial f(\Phi)/\partial\Phi)$, with k an arbitrary complex constant, because the field equations (26.6.15) and (26.6.16) show that $(\bar{\mathscr{D}}\mathscr{D})\text{Re}\,(k^*\Phi) = 4\text{Re}\,(k\partial f(\Phi)/\partial\Phi)$. But in general the changes in X that can be made in this way are quite limited.

26.8 General Kahler Potentials*

There are several circumstances in which it is necessary to consider non-renormalizable Lagrangian densities of the general form (26.3.30)

$$\mathscr{L} = 2\,\text{Re}\left[f(\Phi) \right]_{\mathscr{F}} + \tfrac{1}{2} \left[K(\Phi, \Phi^*) \right]_D \ , \tag{26.8.1}$$

where the superpotential f is an arbitrary function of left-chiral scalar superfields Φ_n but not their derivatives, and the Kahler potential K is an arbitrary function of the Φ_n and Φ_n^* but not their derivatives.

* This section lies somewhat out of the book's main line of development, and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

This situation arises in effective field theories whose symmetries rule out any renormalizable interactions, or in which the renormalizable interactions all happen to be small. It is often then possible to calculate scattering amplitudes at low energy from tree graphs, using a Lagrangian with the smallest value for some combination of the numbers of derivatives, fermion fields, and any small renormalizable couplings. In Section 19.5 we examined such an effective field theory with no renormalizable couplings, involving nucleons and soft pions. The dynamically broken gauge theories discussed in Section 21.4 provide examples of effective field theories of this sort with small renormalizable couplings. This circumstance also arises in supersymmetric theories whose symmetries do not allow a superpotential, or where the superpotential is for some reason small. We will encounter an example of this sort when we consider the extended $N = 2$ supersymmetric theory of Abelian gauge superfields and gauge-neutral chiral scalar superfields in Section 29.5. We will show there that low-energy scattering amplitudes in this theory are generated by tree graphs using a Lagrangian density of the form (26.8.1), with $f = 0$ and K a function only of the Φ_n and Φ_n^* but no derivatives, plus $\mathscr{F}$-terms quadratic in gauge superfields. The inclusion of Kahler potentials having arbitrary dependence on Φ_n and Φ_n^* but not on their derivatives is particularly important in effective field theories in which some scalar fields are of the same order of magnitude as the fundamental energy scale of the underlying theory, although all other field values and all energies are much smaller. This will be of interest, for instance, in connection with theories of gravity-mediated supersymmetry breaking, discussed in Section 31.6.

Let us consider how to express the Lagrangian density (26.8.1) in terms of component fields. We did not use the assumption that $f(\Phi)$ is a *cubic* polynomial in deriving Eq. (26.4.4), so this continues to give the $\mathscr{F}$-terms in the Lagrangian contributed by an arbitrary superpotential. To derive the D-terms, we note that the term in the Kahler potential of fourth order in θ is

$$
\begin{aligned}
K(\Phi,\Phi^*)_{\theta^4} = &-\frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2 \sum_n \left[\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\,\Box\phi_n + \frac{\partial K(\phi,\phi^*)}{\partial\phi_n^*}\,\Box\phi_n^*\right] \\
&+\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\partial\phi_m^*}\left(\bar\theta\gamma_5\theta\right)\Big[\left(\bar\theta\psi_{mR}\right)\left(\bar\theta\ \not\partial\psi_{nL}\right) \\
&\qquad -\left(\bar\theta\psi_{nL}\right)\left(\bar\theta\ \not\partial\psi_{mR}\right)\Big] \\
&+2\,\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\left(\theta_L^{\mathrm{T}}\epsilon\psi_{nL}\right)\left(\theta_L^{\mathrm{T}}\epsilon\psi_{mL}\right)\left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)^*\mathscr{F}_l^*
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$$
\begin{aligned}
&+2\,\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\left(\bar\theta\psi_{mL}\right)\left(\bar\theta\psi_{lR}\right)\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\phi_n\\
&+\sum_{nmlk}\frac{\partial^4 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_k^*\,\partial\phi_\ell^*}\left(\bar\theta\psi_{nL}\right)\left(\bar\theta\psi_{mL}\right)\left(\bar\theta\psi_{lR}\right)\left(\bar\theta\psi_{kR}\right)\\
&-\frac{1}{4}\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m^*}\mathscr{F}_n\mathscr{F}_m^*\left(\bar\theta(1+\gamma_5)\theta\right)\left(\bar\theta(1-\gamma_5)\theta\right)\\
&+\frac{1}{4}\left(\bar\theta\gamma_5\gamma^\mu\theta\right)\left(\bar\theta\gamma_5\gamma^\nu\theta\right)\sum_{mn}\left[-\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\partial\phi_m^*}\partial_\mu\phi_n\,\partial_\nu\phi_m^*\right.\\
&\left.\quad+\frac{1}{2}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\phi_m}\partial_\mu\phi_n\,\partial_\nu\phi_m+\frac{1}{2}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n^*\partial\phi_m^*}\partial_\mu\phi_n^*\,\partial_\nu\phi_m^*\right].
\end{aligned}
\tag{26.8.2}
$$

We can again use Eqs. (26.A.18) and (26.A.19) as well as Eq. (26.A.9) to put the θ dependence of this expression in the form of an over-all factor $(\bar\theta\gamma_5\theta)^2$, and find

$$
\begin{aligned}
K(\Phi,\Phi^*)_{\theta^4}=\frac{1}{4}\left(\bar\theta\gamma_5\theta\right)^2\Bigg\{&-\frac{1}{2}\sum_n\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\Box\phi_n-\frac{1}{2}\sum_n\frac{\partial K(\phi,\phi^*)}{\partial\phi_n^*}\Box\phi_n^*\\
&+\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m^*}\left[\left(\overline{\psi_m}\,\partial\!\!\!/\psi_{nL}\right)+\left(\overline{\psi_n}\,\partial\!\!\!/\psi_{mR}\right)-2\mathscr{F}_n\mathscr{F}_m^*\right]\\
&+2\,\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\left(\overline{\psi_n}\psi_{mL}\right)\mathscr{F}_\ell^*\\
&-2\,\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\left(\overline{\psi_m}\gamma^\mu\psi_{\ell R}\right)\partial_\mu\phi_n\\
&-\frac{1}{2}\sum_{nmlk}\frac{\partial^4 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*\,\partial\phi_k^*}\left(\overline{\psi_n}\psi_{mL}\right)\left(\overline{\psi_k}\psi_{lR}\right)\\
&+\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m^*}\partial_\mu\phi_n\,\partial^\mu\phi_m^*-\frac{1}{2}\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m}\partial_\mu\phi_n\,\partial^\mu\phi_m\\
&-\frac{1}{2}\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n^*\,\partial\phi_m^*}\partial_\mu\phi_n^*\,\partial^\mu\phi_m^*\Bigg\}.
\end{aligned}
\tag{26.8.3}
$$

To make the reality properties of the fermion kinematic terms transparent, we can use Eq. (26.A.21) to write

$$
\left(\overline{\psi_n}\,\partial\!\!\!/\psi_{mR}\right)=\left(\overline{\psi_n}\,\partial\!\!\!/\psi_{mL}\right)^*.
$$

The D-term of $K(\Phi,\Phi^*)$ is the coefficient of $-(\bar\theta\gamma_5\theta)^2/4$ minus half the d'Alembertian of the θ-independent term in $K(\Phi,\Phi^*)$, which is just

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

$K(\phi,\phi^*)$, so

$$\frac{1}{2}\Big[K(\Phi,\Phi^*)\Big]_D = \mathrm{Re}\sum_{nm}\mathscr{G}_{nm}\Bigg[-\frac{1}{2}\Big(\overline{\psi_m}\,\not{\partial}(1+\gamma_5)\psi_n\Big) + \mathscr{F}_n\mathscr{F}^*_m - \partial_\mu\phi_n\,\partial^\mu\phi^*_m\Bigg]$$
$$-\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l}\Big(\overline{\psi_n}\psi_{mL}\Big)\mathscr{F}^*_\ell$$
$$+\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l}\Big(\overline{\psi_m}\gamma^\mu\psi_{\ell R}\Big)\partial_\mu\phi_n$$
$$+\frac{1}{4}\sum_{nmlk}\frac{\partial^4 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l\,\partial\phi^*_k}\Big(\overline{\psi_n}\psi_{mL}\Big)\Big(\overline{\psi_k}\psi_{lR}\Big)\,, \qquad (26.8.4)$$

where $\mathscr{G}(\phi,\phi^*)$ is the *Kahler metric*

$$\mathscr{G}_{nm}(\phi,\phi^*) \equiv \frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi^*_m}\,. \qquad (26.8.5)$$

Note that the constant matrix g_{nm} in Eq. (26.4.2) is replaced here with the Kahler metric $\mathscr{G}_{nm}(\phi,\phi^*)$. Because the Kahler metric is field-dependent, we cannot in general make it equal to the unit matrix by a field redefinition, so the total Lagrangian must be left in the form

$$\mathscr{L} = \mathrm{Re}\sum_{nm}\mathscr{G}_{nm}\Bigg[-\frac{1}{2}\Big(\overline{\psi_m}\,\not{\partial}(1+\gamma_5)\psi_n\Big) + \mathscr{F}_n\mathscr{F}^*_m - \partial_\mu\phi_n\,\partial^\mu\phi^*_m\Bigg]$$
$$-\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l}\Big(\overline{\psi_n}\psi_{mL}\Big)\mathscr{F}^*_\ell$$
$$+\mathrm{Re}\sum_{nml}\frac{\partial^3 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l}\Big(\overline{\psi_m}\gamma^\mu\psi_{\ell R}\Big)\partial_\mu\phi_n$$
$$+\frac{1}{4}\sum_{nmlk}\frac{\partial^4 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi^*_l\,\partial\phi^*_k}\Big(\overline{\psi_n}\psi_{mL}\Big)\Big(\overline{\psi_k}\psi_{lR}\Big)$$
$$-\mathrm{Re}\sum_{nm}\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\Big(\overline{\psi_n}\,\psi_{mL}\Big) + 2\,\mathrm{Re}\sum_n\mathscr{F}_n\frac{\partial f(\phi)}{\partial\phi_n}\,. \qquad (26.8.6)$$

The bilinear $(\overline{\psi_m}\;\not{\partial}\gamma_5\psi_n)$ is a total derivative, and could be discarded if $\mathscr{G}_{nm}$ were a constant, but must be kept for general Kahler potentials. This result will be extended to include gauge superfields at the end of Section 27.4.

* * *

As discussed in Section 19.6, the spontaneous breakdown of a global

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

symmetry group G to a subgroup H entails the existence of a set of real massless Goldstone bosons with scalar fields π_k, for which the term in the Lagrangian with the minimum number of derivatives takes the form

$$\mathscr{L}_{G/H} = -\sum_{k\ell} G_{k\ell}(\pi)\partial_\mu \pi_k \partial^\mu \pi_\ell \,, \tag{26.8.7}$$

where $G_{k\ell}(\pi)$ is the metric of the coset space G/H. (Theories with Lagrangian densities of this general form are known as *non-linear σ-models.*) By writing the complex fields ϕ_n in terms of their real and imaginary parts, the term $-\sum_{nm} \mathscr{G}_{nm}(\phi,\phi^*)\partial_\mu\phi_n\partial^\mu\phi_m^*$ in the Lagrangian density (26.8.6) can be put in the form (26.8.7), but the converse is not generally true: the condition that a set of real coordinates like the Goldstone boson fields π_k can be interpreted as the real and imaginary parts of a set of complex coordinates like the fields ϕ_n with the metric in these coordinates locally given by Eq. (26.8.5) defines what is called a *Kahler manifold.*** But it should not be thought that in the common cases where G/H is *not* a Kahler manifold, it is impossible for G to be spontaneously broken to H while keeping supersymmetry unbroken. What happens in these cases is the appearance of extra massless bosons, which, together with the Goldstone bosons, do form a Kahler manifold.

This is because the superpotential $f(\phi)$ depends on ϕ but not ϕ^*, so if the whole Lagrangian is invariant under a global symmetry group G then the superpotential is automatically invariant under a group $G_{\mathbb{C}}$, the complexification of G: if G consists of transformations $\exp(i\sum_A \theta_A t_A)$ with generators t_A and arbitrary real parameters θ_A, then $G_{\mathbb{C}}$ consists of transformations $\exp(i\sum_A z_A t_A)$ with the same generators and arbitrary complex parameters z_A. (For instance, if G is $U(n)$ then $G_{\mathbb{C}}$ is $GL(n,\mathbb{C})$, the group of all complex non-singular matrices, while if G is $SU(n)$ then $G_{\mathbb{C}}$ is $SL(n,\mathbb{C})$, the group of all complex non-singular matrices with unit determinant.) Likewise, if some stationary point $\phi^{(0)}$ of $f(\phi)$ is left invariant by some subgroup H of G, then it will also be left invariant by a

** The significance of Kahler manifolds in this context was pointed out in an early paper by Zumino.[8] Note that it is not necessary for the metric to be expressible in the form (26.8.5) over the whole manifold with a *single* Kahler potential $K(\phi,\phi^*)$; it is only necessary that the manifold can be covered in finite overlapping patches in which this holds, with different Kahler potentials in each patch. The simplest example of a Kahler manifold is the flat complex plane, with Kahler potential $|z|^2$. As an example of a coset space G/H that is a Kahler manifold, Zumino gave the case where $G = GL(p,\mathbb{C}) \times GL(p+q,\mathbb{C})$ and $H = GL(p,\mathbb{C})$, with p and q arbitrary positive-definite integers, and $GL(N,\mathbb{C})$ the group of complex non-singular $N \times N$ matrices. The coset space G/H here has complex coordinates ϕ_n that may be taken as the components of a complex $p \times (p+q)$ matrix A, which under G and H undergoes the transformations $A \to BAC$ and $A \to BA$, respectively, where B and C are square non-singular complex matrices of dimensionality p and $p+q$, respectively. The Kahler potential in this case is simply $K \propto \ln \mathrm{Det}\, A A^\dagger$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

subgroup $H_{\mathbb{C}}$ of $G_{\mathbb{C}}$, the complexification of H. Whether or not G/H is a Kahler manifold, the complexified coset space $G_{\mathbb{C}}/H_{\mathbb{C}}$ is always a Kahler manifold. This follows because $G_{\mathbb{C}}/H_{\mathbb{C}}$ is a complex submanifold of the flat complex space of the ϕ_n, which is a Kahler manifold, and it is a theorem that any complex submanifold of a Kahler manifold is a Kahler manifold.[9] If parameterized by the values of $\phi_n(z) = [\exp(i\sum_A z_A t_A)\phi^{(0)}]_n$, the Kahler manifold $G_{\mathbb{C}}/H_{\mathbb{C}}$ has the metric obtained by embedding it in the flat complex space of the ϕ_n, usually taken with line element $\sum_n d\phi_n\, d\phi_n^*$.

It is true that $G_{\mathbb{C}}$ is not a symmetry of the whole Lagrangian, but the Goldstone bosons associated with the breakdown of $G_{\mathbb{C}}$ to $H_{\mathbb{C}}$ are nevertheless exactly massless. This is guaranteed by the non-renormalization theorem of Section 27.6, or, more simply, by the result of Section 25.4, that massless spin zero particles must come in *pairs* that are related by supersymmetry transformations and that therefore have the same transformation under any global symmetry group G that commutes with supersymmetry.

Appendix Majorana Spinors

This appendix summarizes some algebraic properties of Majorana spinors that are needed in dealing with superfields.

Consider a four-component fermionic Majorana spinor s that like Q or θ may be expressed in the form

$$s = \begin{pmatrix} e\varsigma^* \\ \varsigma \end{pmatrix}, \tag{26.A.1}$$

where ς is some two-component spinor and e is the 2×2 matrix

$$e \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i\sigma_2 .$$

Such a spinor is related to its complex conjugate by

$$s^* = \begin{pmatrix} 0 & e \\ -e & 0 \end{pmatrix} s = -\beta\gamma_5 \epsilon\, s , \tag{26.A.2}$$

where ϵ is the 4×4 matrix

$$\epsilon \equiv \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix} \tag{26.A.3}$$

and as usual γ_5 and β are 4×4 matrices

$$\gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

with 1 and 0 understood here as 2×2 submatrices. Taking the transpose of Eq. (26.A.2) and multiplying on the right by β gives the equivalent formula

$$\bar{s} \equiv s^\dagger \beta = s^{\mathrm{T}} \epsilon \gamma_5 \,. \tag{26.A.4}$$

The anticommutation of the spinor components limits the variety of covariants that can be formed from a Majorana spinor. To see this, it will be convenient first to consider the symmetry properties of bilinear covariants, which will be of some interest in themselves. For a pair of Majorana spinors s_1 and s_2 and any 4×4 numerical matrix M, Eq. (26.A.4) gives

$$\begin{aligned}\overline{s_1}\, M\, s_2 &= \sum_{\alpha\beta} s_{1\alpha}\, s_{2\beta}\, (\epsilon\, \gamma_5\, M)_{\alpha\beta} = -\sum_{\alpha\beta} s_{2\alpha}\, s_{1\beta}\, (\epsilon\, \gamma_5\, M)_{\beta\alpha} \\ &= +\sum_{\alpha\beta} s_{2\alpha}\, s_{1\beta}\, (M^{\mathrm{T}} \epsilon\, \gamma_5)_{\alpha\beta} = \overline{s_2}\, (\epsilon\gamma_5)^{-1} M^{\mathrm{T}} \epsilon\gamma_5\, s_1 \,,\end{aligned}$$

with the minus sign following the second equal sign arising from the fermionic nature of these spinors. In Section 5.4 we found that the 16 covariant matrices formed from Dirac matrices satisfy

$$M^{\mathrm{T}} = \begin{cases} +\mathscr{C}\, M\, \mathscr{C}^{-1} & M = 1,\ \gamma_5\gamma_\mu,\ \gamma_5 \\ -\mathscr{C}\, M\, \mathscr{C}^{-1} & M = \gamma_\mu,\ [\gamma_\mu, \gamma_\nu] \end{cases} , \tag{26.A.5}$$

where $\mathscr{C}$ is the matrix

$$\mathscr{C} = \gamma_2 \beta = -\epsilon\gamma_5 = \begin{pmatrix} -e & 0 \\ 0 & e \end{pmatrix} . \tag{26.A.6}$$

It follows that

$$(\overline{s_1}\, M\, s_2) = \begin{cases} +(\overline{s_2}\, M\, s_1) & M = 1,\ \gamma_5\gamma_\mu,\ \gamma_5 \\ -(\overline{s_2}\, M\, s_1) & M = \gamma_\mu,\ [\gamma_\mu, \gamma_\nu] \end{cases} . \tag{26.A.7}$$

In particular, setting $s_1 = s_2 = s$, we find that

$$\bar{s}\, \gamma_\mu\, s = \bar{s}\, [\gamma_\mu\,,\, \gamma_\nu]\, s = 0 \,, \tag{26.A.8}$$

so the only bilinear covariants formed from a single Majorana spinor s are $\bar{s}\, s$, $\bar{s}\, \gamma_5 \gamma_\mu\, s$, and $\bar{s}\, \gamma_5\, s$.

In considering the form of the most general superfields, we need to have expressions for the products of two or more Majorana spinors. For two spinors, we recall that any 4×4 matrix may be expanded as a sum of the 16 covariant matrices 1, γ_μ, $[\gamma_\mu\,,\, \gamma_\nu]$, $\gamma_5\gamma_\mu$, γ_5. Lorentz invariance tells us that for the matrix $s_\alpha\, \bar{s}_\beta$ this expansion must take the form

$$\begin{aligned} s\, \bar{s} = {} & k_S\, (\bar{s}\, s) + k_V\, \gamma_\mu\, (\bar{s}\, \gamma^\mu\, s) + k_T\, [\gamma_\mu\,,\, \gamma_\nu]\, (\bar{s}\, [\gamma^\mu\,,\, \gamma^\nu]\, s) \\ & + k_A\, \gamma_5 \gamma_\mu\, (\bar{s}\, \gamma_5 \gamma^\mu\, s) + k_P\, \gamma_5\, (\bar{s}\, \gamma_5\, s) \,, \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

where the ks are constants to be determined. Eq. (26.A.8) shows that we may take $k_V = k_T = 0$. The remaining coefficients may be calculated by multiplying on the right with 1, $\gamma_5\gamma^\mu$, and γ_5 and taking the trace, which gives $k_S = -1/4$, $k_A = +1/4$, and $k_P = -1/4$. In this way we find that

$$s\,\bar{s} = -\tfrac{1}{4}(\bar{s}\,s) + \tfrac{1}{4}\gamma_5\gamma_\mu(\bar{s}\,\gamma_5\gamma^\mu\,s) - \tfrac{1}{4}\gamma_5(\bar{s}\,\gamma_5\,s)\,. \tag{26.A.9}$$

By multiplying on the right with $-\epsilon\gamma_5$ and using Eq. (26.A.4), we may put this in the form

$$s_\alpha\,s_\beta = \tfrac{1}{4}(\epsilon\gamma_5)_{\alpha\beta}(\bar{s}\,s) + \tfrac{1}{4}(\gamma_\mu\epsilon)_{\alpha\beta}(\bar{s}\,\gamma_5\gamma^\mu\,s) + \tfrac{1}{4}\epsilon_{\alpha\beta}(\bar{s}\,\gamma_5\,s)\,, \tag{26.A.10}$$

or, equivalently,

$$s_\alpha\,s_\beta = \tfrac{1}{4}(\epsilon\gamma_5)_{\alpha\beta}(s^{\mathrm{T}}\epsilon\,\gamma_5\,s) + \tfrac{1}{4}(\gamma_\mu\epsilon)_{\alpha\beta}(s^{\mathrm{T}}\epsilon\,\gamma^\mu\,s) + \tfrac{1}{4}\epsilon_{\alpha\beta}(s^{\mathrm{T}}\epsilon\,s)\,. \tag{26.A.11}$$

Now consider the product $s_\alpha s_\beta s_\gamma$ of three components of a Majorana spinor s. We can divide s into left- and right-handed parts

$$s = s_L + s_R\,, \qquad s_L = \tfrac{1}{2}(1+\gamma_5)s\,, \qquad s_R = \tfrac{1}{2}(1-\gamma_5)s\,. \tag{26.A.12}$$

Each of s_L and s_R has only two independent components, so since the square of any fermionic c-number vanishes, we have $s_{L\alpha}s_{L\beta}s_{L\gamma} = 0$ and $s_{R\alpha}s_{R\beta}s_{R\gamma} = 0$ for all α, β, and γ, and therefore

$$s_\alpha s_\beta s_\gamma = s_{L\alpha}s_{L\beta}s_{R\gamma} + s_{L\alpha}s_{R\beta}s_{L\gamma} + s_{R\alpha}s_{L\beta}s_{L\gamma} + L \leftrightarrow R\,,$$

with '$L \leftrightarrow R$' denoting the sum of previous terms with labels L and R interchanged. To evaluate this expression, we multiply Eq. (26.A.11) by suitable factors of $(1+\gamma_5)/2$, and find

$$s_{L\alpha}\,s_{L\beta} = \tfrac{1}{4}[\epsilon(1+\gamma_5)]_{\alpha\beta}(s_L^{\mathrm{T}}\epsilon s_L)\,.$$

If we now multiply this with $s_{R\gamma}$, since $(s_R^{\mathrm{T}}\epsilon s_R)\,s_{R\gamma} = 0$ we can drop the label L on the spinors in the bilinear $(s_L^{\mathrm{T}}\epsilon s_L)$:

$$s_{L\alpha}\,s_{L\beta}\,s_{R\gamma} = \tfrac{1}{4}[\epsilon(1+\gamma_5)]_{\alpha\beta}(s^{\mathrm{T}}\epsilon\,s)\,s_{R\gamma}\,.$$

The same arguments also yield

$$s_{R\alpha}\,s_{R\beta}\,s_{L\gamma} = \tfrac{1}{4}[\epsilon(1-\gamma_5)]_{\alpha\beta}(s^{\mathrm{T}}\epsilon\,s)\,s_{L\gamma}\,.$$

Adding the sum of these two expressions to the same quantity with γ replaced with α or β yields finally

$$\begin{aligned} s_\alpha s_\beta s_\gamma = {} & \tfrac{1}{4}\left(s^{\mathrm{T}}\epsilon s\right)\Big[\epsilon_{\alpha\beta}\,s_\gamma - (\epsilon\gamma_5)_{\alpha\beta}(\gamma_5 s)_\gamma - \epsilon_{\alpha\gamma}\,s_\beta \\ & + (\epsilon\gamma_5)_{\alpha\gamma}(\gamma_5 s)_\beta + \epsilon_{\beta\gamma}\,s_\alpha - (\epsilon\gamma_5)_{\beta\gamma}(\gamma_5 s)_\alpha\Big]\,. \end{aligned} \tag{26.A.13}$$

To calculate the products of four Majorana spinor components, we note that $(s^{\mathrm{T}}\epsilon s)$ contains only terms with two s_Ls or two s_Rs, so

$$(s^{\mathrm{T}}\epsilon s)s_\gamma s_\delta = (s^{\mathrm{T}}\epsilon s)[s_{R\gamma}s_{R\delta} + s_{L\gamma}s_{L\delta}]\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Using Eq. (26.A.11) to evaluate the sum in square brackets, and noting that

$$(s^{\mathrm{T}}\epsilon s)\,(s^{\mathrm{T}}\epsilon\gamma_5 s) = (s_L^{\mathrm{T}}\epsilon s_L)(s_R^{\mathrm{T}}\epsilon s_R) - (s_R^{\mathrm{T}}\epsilon s_R)(s_L^{\mathrm{T}}\epsilon s_L) = 0\,,$$

we find that

$$(s^{\mathrm{T}}\epsilon s)\,s_\gamma s_\delta = \tfrac{1}{4}\,\epsilon_{\gamma\delta}(s^{\mathrm{T}}\epsilon s)^2\,. \tag{26.A.14}$$

Multiplying Eq. (26.A.13) with s_δ therefore yields the result

$$\begin{aligned} s_\alpha s_\beta s_\gamma s_\delta = {} & \tfrac{1}{16}\left(s^{\mathrm{T}}\epsilon s\right)^2\Big[\epsilon_{\alpha\beta}\,\epsilon_{\gamma\delta} - (\epsilon\gamma_5)_{\alpha\beta}\,(\epsilon\gamma_5)_{\gamma\delta} - \epsilon_{\alpha\gamma}\,\epsilon_{\beta\delta} \\ & +(\epsilon\gamma_5)_{\alpha\gamma}\,(\epsilon\gamma_5)_{\beta\delta} + \epsilon_{\beta\gamma}\,\epsilon_{\alpha\delta} - (\epsilon\gamma_5)_{\beta\gamma}\,(\epsilon\gamma_5)_{\alpha\delta}\Big]\,. \end{aligned} \tag{26.A.15}$$

Any product of five components of s vanishes, so this completes the list of formulas for products of components of a Majorana spinor.

We can use these formulas to derive some additional relations that will be useful in dealing with superfields. By contracting Eq. (26.A.13) with $(\epsilon\gamma_5)_{\beta\gamma}$ and $(\epsilon\gamma_\mu)_{\beta\gamma}$, we find

$$s_\alpha\left(\bar{s}\,s\right) = -(\gamma_5\,s)_\alpha\left(\bar{s}\,\gamma_5\,s\right) \tag{26.A.16}$$

and

$$s_\alpha\left(\bar{s}\,\gamma_5\gamma_\mu\,s\right) = -(\gamma_\mu\,s)_\alpha\left(\bar{s}\,\gamma_5\,s\right)\,. \tag{26.A.17}$$

From Eqs. (26.A.16) and (26.A.17) we may derive 'Fierz' identities

$$\left(\bar{s}\,s\right)^2 = -\left(\bar{s}\,\gamma_5\,s\right)^2\,, \qquad \left(\bar{s}\,\gamma_5\gamma_\mu\,s\right)\left(\bar{s}\,\gamma_5\gamma_\nu\,s\right) = -\eta_{\mu\nu}\left(\bar{s}\,\gamma_5\,s\right)^2\,. \tag{26.A.18}$$

Also, Eq. (26.A.14) may be put in a covariant form

$$(\bar{s}\,\gamma_5\,s)\,s\,\bar{s} = -\tfrac{1}{4}\gamma_5\,(\bar{s}\,\gamma_5\,s)^2\,. \tag{26.A.19}$$

It will also be useful to record the reality properties of bilinear products of Majorana spinors. For any pair of Majorana spinors s_1, s_2 satisfying the phase convention (26.A.1), Eqs. (26.A.2) and (26.A.4) give

$$(\overline{s_1}\,M\,s_2)^* = -(s_1^\dagger\epsilon\gamma_5\,M^*\,s_2^*) = (\overline{s_1}\,\beta\,\epsilon\,\gamma_5\,M^*\,\beta\,\epsilon\,\gamma_5\,s_2)\,.$$

(The minus sign in the middle expression arises from reversing the interchange of s_1 and s_2 that occurs when we take the complex conjugate.) But Eqs. (5.4.40) and (26.A.6) give $\beta\epsilon\gamma_5\gamma_\mu^*\beta\epsilon\gamma_5 = \gamma_\mu$, so

$$\beta\,\epsilon\,\gamma_5\,M^*\,\beta\,\epsilon\,\gamma_5 = \begin{cases} +M & M = 1\,,\ \gamma_\mu\,,\ [\gamma_\mu\,,\ \gamma_\mu] \\ -M & M = \gamma_\mu\gamma_5\,,\ \gamma_5 \end{cases}\,, \tag{26.A.20}$$

and therefore

$$(\overline{s_1}\,M\,s_2)^* = \begin{cases} +(\overline{s_1}\,M\,s_2) & M = 1\,,\ \gamma_\mu\,,\ [\gamma_\mu\,,\ \gamma_\mu] \\ -(\overline{s_1}\,M\,s_2) & M = \gamma_\mu\gamma_5\,,\ \gamma_5 \end{cases}\,. \tag{26.A.21}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

Finally, we mention that any spinor u may be written in terms of a pair of Majorana spinors $s_\pm$ as

$$u = s_+ + i s_- \,, \tag{26.A.22}$$

where

$$s_+ \equiv \frac{1}{2}\left(u - \beta\epsilon\gamma_5 u^*\right) \,, \qquad s_- \equiv \frac{1}{2i}\left(u + \beta\epsilon\gamma_5 u^*\right) \,. \tag{26.A.23}$$

To check that $s_\pm$ are Majorana spinors satisfying Eq. (26.A.2), it is only necessary to recall that $\beta\epsilon\gamma_5$ is real, and that $(\beta\epsilon\gamma_5)^2 = 1$.

Problems

1. Using the direct technique of Section 26.1 in the case of $N = 2$ supersymmetry, find the supersymmetry transformation rules of the massive field supermultiplet with just one Majorana spinor field and two complex scalars.

2. Calculate the component fields of a time-reversed superfield

   $$\mathsf{T}^{-1} S(x,\theta) \mathsf{T}$$

   in terms of the components of the superfield $S(x,\theta)$. What sort of superfield do we get by the time reversal of a left-chiral superfield? Of a linear superfield?

3. Consider the $N = 1$ supersymmetric theory of a single left-chiral superfield Φ. In superfield notation, list all the terms of dimensionality 5 involving Φ and/or Φ^* that might be added to the Lagrangian density.

4. Consider the theory of three left-chiral scalar superfields Φ_1, Φ_2, and Φ_3, with a conventional kinematic term, and with a superpotential

   $$f(\Phi_1,\Phi_2,\Phi_3) = \Phi_1\Phi_3^2 + \Phi_2\left(\Phi_3^2 + a\right) \,,$$

   where a is a non-zero real constant. Show that this is a theory with spontaneously broken supersymmetry. Find the minimum value of the potential. Express the field of the goldstino in terms of the fermionic components of Φ_1, Φ_2, and Φ_3.

5. Find all the components of the current superfield for the action (26.6.9), in terms of the components of the left-chiral superfields Φ_n and derivatives of the superpotential f and the Kahler potential K.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

6. Check that the supersymmetry current given by Eqs. (26.7.18) and (26.7.10) is related to the ω-component of the superfield (26.7.21) by Eq. (26.7.20).

References

1. A. Salam and J. Strathdee, *Nucl. Phys.* **B76**, 477 (1974). This article is reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987).
2. J. Wess and B. Zumino, *Nucl. Phys.* **B70**, 13 (1974). This article is reprinted in *Supersymmetry*, Ref. 1.
3. L. O'Raifeartaigh, *Nucl. Phys.* **B96**, 331 (1975). This article is reprinted in *Supersymmetry*, Ref. 1.
4. F. A. Berezin, *The Method of Second Quantization* (Academic Press, New York, 1966).
5. J. Iliopoulos and B. Zumino, *Nucl. Phys.* **B76**, 310 (1974); S. Ferrara and B. Zumino, *Nucl. Phys.* **B87**, 207 (1975). These articles are reprinted in *Supersymmetry*, Ref. 1.
6. This section follows the approach of S. Ferrara and B. Zumino, Ref. 5.
7. X. Gràcia and J. Pons, *J. Phys.* **A25**, 6357 (1992). I am grateful to J. Gomis for suggesting the use of the equation matching the coefficients of $\ddot{q}^n$.

7*a*. M. Grisaru, in *Recent Developments in Gravitation – Cargèse 1978*, M. Lévy and S. Deser, eds. (Plenum Press, New York, 1979): 577.

8. B. Zumino, *Phys. Lett.* **87B**, 203 (1979). This article is reprinted in *Supersymmetry*, Ref. 1.
9. P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1978): 109. I thank D. Freed for telling me of this application of the general theorem.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.005

27

Supersymmetric Gauge Theories

The successful theories of strong, weak, and electromagnetic interactions described in the first two volumes are all gauge theories. In order to see how simple supersymmetry may make contact with reality, we must therefore consider how to construct actions that satisfy both supersymmetry and gauge invariance.[1]

27.1 Gauge-Invariant Actions for Chiral Superfields

Consider a set of Abelian or non-Abelian gauge transformations that leave the supersymmetry generator Q invariant. (For simple supersymmetry there is just one Majorana spinor supersymmetry generator, which can only furnish a trivial representation of any semi-simple gauge group.) Each component field in a supermultiplet must transform in the same way under such gauge transformations. In particular, for a left-chiral superfield we have

$$\begin{aligned}
\phi_n(x) &\to \sum_m \left[\exp\left(i\sum_A t_A\Lambda^A(x)\right)\right]_{nm} \phi_m(x)\,, \\
\psi_{nL}(x) &\to \sum_m \left[\exp\left(i\sum_A t_A\Lambda^A(x)\right)\right]_{nm} \psi_{mL}(x)\,, \\
\mathscr{F}_n(x) &\to \sum_m \left[\exp\left(i\sum_A t_A\Lambda^A(x)\right)\right]_{nm} \mathscr{F}_m(x)\,,
\end{aligned} \tag{27.1.1}$$

where t_A are Hermitian matrices representing the generators of the gauge algebra, and $\Lambda^A(x)$ are real functions of x^μ that parameterize a finite gauge transformation. (We are using the same notation for gauge transformations as in Section 15.1, except that in order to avoid confusion with Dirac indices we label the gauge generators and gauge transformation parameters with the letters A, B, etc. instead of α, β, etc.)

The left-chiral superfield (26.3.11) involves derivatives of some of the component fields, so its transformation is more complicated than

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Eq. (27.1.1). However, Eq. (26.3.21) shows that the superfield does not involve derivatives if expressed in terms of θ_L and the variable x_+ defined by Eq. (26.3.23). It therefore has the transformation property

$$\Phi_n(x,\theta) \to \sum_m \left[\exp\left(i\sum_A t_A \Lambda^A(x_+)\right)\right]_{nm} \Phi_m(x,\theta)\,. \tag{27.1.2}$$

If a term in the action depends only on left-chiral superfields, and not their derivatives or complex conjugates, like the term $\int d^4x\,[f(\Phi)]_{\mathscr{F}}$ in Eq. (26.3.30), then it (and its complex conjugate) will be invariant under the local transformation (27.1.2) if it is invariant under global transformations with $\Lambda^A(x)$ independent of x^μ. The need for introducing gauge fields in renormalizable theories of chiral superfields arises only in the D-terms, which involve both Φ_n and Φ_n^*. Because the matrices t_A are Hermitian, the Hermitian adjoint of Eq. (27.1.2) is

$$\Phi_n^\dagger(x,\theta) \to \sum_m \Phi_m^\dagger(x,\theta)\left[\exp\left(-i\sum_A t_A \Lambda^A(x_+)^*\right)\right]_{mn}\,. \tag{27.1.3}$$

If it were not for the difference between $\Lambda^A(x_+)^* = \Lambda^A(x_-)$ and $\Lambda^A(x_+)$, this would just say that $\Phi^\dagger$ transforms according to the representation of the gauge group that is contragredient to the representation furnished by Φ, and any function of Φ and $\Phi^\dagger$ that is invariant under global gauge transformations would also be invariant under local gauge transformations. Because x_+ and x_- are different, we must introduce a gauge connection matrix $\Gamma_{nm}(x,\theta)$, with the transformation property

$$\Gamma(x,\theta) \to \exp\left(+i\sum_A t_A \Lambda^A(x_+)^*\right)\Gamma(x,\theta)\exp\left(-i\sum_A t_A\Lambda^A(x_+)\right)\,. \tag{27.1.4}$$

Then by multiplying $\Phi^\dagger$ on the right with Γ we obtain a superfield that transforms as

$$\left[\Phi^\dagger(x,\theta)\,\Gamma(x,\theta)\right]_n \to \sum_m \left[\Phi^\dagger(x,\theta)\,\Gamma(x,\theta)\right]_m \left[\exp\left(-i\sum_A t_A\Lambda^A(x_+)\right)\right]_{mn}\,, \tag{27.1.5}$$

so that any globally gauge-invariant function constructed from Φ and $\Phi^\dagger\Gamma$ (and not their derivatives or complex conjugates) will also be locally gauge-invariant. One obvious example is the gauge-invariant version $(\Phi^\dagger\Gamma\Phi)_D$ of the D-term in the Lagrangian constructed in Section 26.4.

Any $\Gamma(x,\theta)$ that transforms as in Eq. (27.1.4) will allow us to construct gauge-invariant Lagrangians of chiral superfields. The choice is not unique; if Γ transforms as in Eq. (27.1.4), and we multiply on the right with any left-chiral superfield Υ_L with the transformation rule

$$\Upsilon_L(x,\theta) \to \exp\left(i\sum_A t_A\Lambda^A(x_+)\right)\Upsilon_L(x,\theta)\exp\left(-i\sum_A t_A\Lambda^A(x_+)\right)\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

then we obtain a new gauge connection that also satisfies Eq. (27.1.4). One simplification is to take $\Gamma(x,\theta)$ Hermitian:

$$\Gamma^\dagger(x,\theta) = \Gamma(x,\theta)\,. \tag{27.1.6}$$

This is always possible if there is any $\Gamma(x,\theta)$ that satisfies Eq. (27.1.4), for then by taking the Hermitian adjoint of Eq. (27.1.4) we easily see that $\Gamma^\dagger(x,\theta)$ transforms in the same way as $\Gamma(x,\theta)$, so if $\Gamma(x,\theta)$ is not Hermitian then we can replace it with its Hermitian part $(\Gamma+\Gamma^\dagger)/2$ (or, if that vanishes, by its anti-Hermitian part $(\Gamma-\Gamma^\dagger)/2i$.) Another simplification of great physical importance is to express $\Gamma(x,\theta)$ in terms of fields whose gauge transformation properties do not depend on the specific representation t_A of the gauge algebra under which the chiral superfield $\Phi(x,\theta)$ transforms, so that these fields can be used to form a suitable matrix $\Gamma(x,\theta)$ for chiral superfields that transform according to any representation of the gauge group. For this purpose, it is useful to recall the Baker–Hausdorff formula, which states that, for arbitrary matrices a and b,

$$e^a\,e^b = \exp\left(a+b+\tfrac{1}{2}[a,b]+\tfrac{1}{12}[a,[a,b]]+\tfrac{1}{12}[b,[b,a]]+\cdots\right), \tag{27.1.7}$$

where '$\cdots$' denotes higher-order terms that can be written as multiple commutators of as and bs, like the second- and third-order terms that are shown explicitly. It follows from this that for any representation of a Lie algebra, we have

$$\exp\Big(\sum_A a^A t_A\Big)\,\exp\Big(\sum_A b^A t_A\Big) = \exp\Big(\sum_A f^A(a,b)\,t_A\Big)\,, \tag{27.1.8}$$

where

$$\begin{aligned} f^A(a,b) = {} & a^A + b^A + \tfrac{1}{2}i\sum_{BC} C^A{}_{BC}a^B b^C - \tfrac{1}{12}\sum_{BCDE} C^A{}_{BC}C^C{}_{DE}a^B a^D b^E \\ & - \tfrac{1}{12}\sum_{BCDE} C^A{}_{BC}C^C{}_{DE}b^B b^D a^E + \cdots\,, \end{aligned} \tag{27.1.9}$$

which depends on the Lie algebra through its structure constants $C^A{}_{BC}$, defined as usual by

$$[t_B,\,t_C] = i\sum_A C^A{}_{BC}\,t_A\,,$$

but does not depend on the particular representation furnished by the t_A. We will therefore take $\Gamma(x,\theta)$ in the form

$$\Gamma(x,\theta) = \exp\left(-2\sum_A t_A\,V^A(x,\theta)\right)\,, \tag{27.1.10}$$

where $V^A(x,\theta)$ are a set of real superfields (so that Γ is Hermitian), not depending on the representation of the gauge algebra furnished by the t_A.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

We can achieve an important further simplification by noting an additional symmetry of supersymmetric gauge theories. If some function of Φ and $\Phi^\dagger\Gamma$ is invariant under global gauge transformations, then it will automatically be invariant not only under the local gauge transformations (27.1.2)–(27.1.4), but also under the larger group of extended gauge transformations

$$\Phi_{nL}(x,\theta) \to \sum_m \left[\exp\left(i\sum_A t_A \Omega^A(x,\theta)\right)\right]_{nm} \Phi_{mL}(x,\theta) \tag{27.1.11}$$

and

$$\Gamma(x,\theta) \to \exp\left(-i\sum_A t_A\Omega^A(x,\theta)\right)\Gamma(x,\theta)\exp\left(+i\sum_A t_A\Omega^A(x,\theta)^*\right), \tag{27.1.12}$$

where $\Omega^A(x,\theta)$ is an arbitrary left-chiral superfield — that is, an arbitrary function of θ_L as well as x_+. Under this transformation,

$$V^A(x,\theta) \to V^A(x,\theta) + \frac{i}{2}\left[\Omega^A(x,\theta) - \Omega^A(x,\theta)^*\right] + \cdots, \tag{27.1.13}$$

where '$\cdots$' denotes terms arising from the commutators in Eq. (27.1.7), which are of first or higher order in gauge coupling constants. As a general left-chiral superfield, Ω may be written in the form (26.3.11)

$$\begin{aligned}\Omega^A(x,\theta) = {} & W^A(x) - \sqrt{2}\left(\bar\theta\left(\frac{1+\gamma_5}{2}\right)w^A(x)\right) + \mathscr{W}^A(x)\left(\bar\theta\left(\frac{1+\gamma_5}{2}\right)\theta\right) \\ & + \frac{1}{2}\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu W^A(x) - \frac{1}{\sqrt{2}}\left(\bar\theta\gamma_5\theta\right)\left(\bar\theta\,\not\partial\left(\frac{1+\gamma_5}{2}\right)w^A(x)\right) \\ & - \frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box W^A(x),\end{aligned} \tag{27.1.14}$$

in which $W^A(x)$ and $\mathscr{W}^A(x)$ are arbitrary complex functions of x^μ, and we introduced Majorana spinors $w^A(x)$ defined so that the left-handed spinor components of the superfields are $\frac{1}{2}(1+\gamma_5)w^A(x)$. Using the complex conjugation properties (26.A.21) of Majorana bilinears, the complex conjugate of Eq. (27.1.14) gives

$$\begin{aligned}\Omega^A(x,\theta)^* = {} & W^{A*}(x) - \sqrt{2}\left(\bar\theta\left(\frac{1-\gamma_5}{2}\right)w^A(x)\right) + \mathscr{W}^{A*}(x)\left(\bar\theta\left(\frac{1-\gamma_5}{2}\right)\theta\right) \\ & - \frac{1}{2}\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu W^{A*}(x) \frac{1}{\sqrt{2}}\left(\bar\theta\gamma_5\theta\right)\left(\bar\theta\,\not\partial\left(\frac{1-\gamma_5}{2}\right)w^A(x)\right) \\ & - \frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box W^{A*}(x).\end{aligned} \tag{27.1.15}$$

We write the real superfields $V^A(x,\theta)$ in terms of component fields as in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Eq. (26.2.10):

$$\begin{aligned} V^A(x,\theta) = {} & C^A(x) - i\bigl(\bar\theta\,\gamma_5\,\omega^A(x)\bigr) - \frac{i}{2}\bigl(\bar\theta\,\gamma_5\,\theta\bigr)M^A(x) - \frac{1}{2}\bigl(\bar\theta\,\theta\bigr)N^A(x) \\ & + \frac{i}{2}\bigl(\bar\theta\,\gamma_5\,\gamma^\mu\,\theta\bigr)V_\mu^A(x) - i\bigl(\bar\theta\,\gamma_5\,\theta\bigr)\left(\bar\theta\Bigl[\lambda^A(x) + \frac{1}{2}\,\partial\!\!\!/\omega^A(x)\Bigr]\right) \\ & - \frac{1}{4}\bigl(\bar\theta\,\gamma_5\,\theta\bigr)^2\left(D^A(x) + \frac{1}{2}\Box C^A(x)\right), \end{aligned} \tag{27.1.16}$$

where $C^A(x)$, $M^A(x)$, $N^A(x)$, and $V_\mu^A(x)$ are all real, and $\omega^A(x)$ and $\lambda^A(x)$ are Majorana spinors. Using Eqs. (27.1.14)–(27.1.16) in Eq. (27.1.13), we find that the component fields of the gauge superfield undergo the extended gauge transformation

$$\begin{aligned} & C^A(x) \to C^A(x) - \operatorname{Im} W^A(x) + \cdots, \\ & \omega^A(x) \to \omega^A(x) + \frac{1}{\sqrt{2}} w^A(x) + \cdots, \\ & V_\mu^A(x) \to V_\mu^A(x) + \partial_\mu \operatorname{Re} W^A(x) + \cdots, \\ & M^A(x) \to M^A(x) - \operatorname{Re} \mathscr{W}^A(x) + \cdots, \\ & N^A(x) \to N^A(x) + \operatorname{Im} \mathscr{W}^A(x) + \cdots, \\ & \lambda^A(x) \to \lambda^A(x) + \cdots, \\ & D^A(x) \to D^A(x) + \cdots, \end{aligned} \tag{27.1.17}$$

where again '$\cdots$' denotes terms that arise from the structure constants in Eq. (27.1.9) and that are therefore proportional to one or more factors of gauge coupling constants. We can use such an extended gauge transformation to put the gauge superfields into a convenient form, known as *Wess–Zumino gauge*,[1] in which

$$C^A(x) = \omega^A(x) = M^A(x) = N^A(x) = 0\,, \tag{27.1.18}$$

so that

$$\begin{aligned} V^A(x,\theta) = {} & \frac{i}{2}\bigl(\bar\theta\,\gamma_5\,\gamma^\mu\,\theta\bigr)V_\mu^A(x) - i\bigl(\bar\theta\,\gamma_5\,\theta\bigr)\bigl(\bar\theta\lambda^A(x)\bigr) \\ & - \frac{1}{4}\bigl(\bar\theta\,\gamma_5\,\theta\bigr)^2 D^A(x)\,. \end{aligned} \tag{27.1.19}$$

To accomplish this to zeroth order in the coupling constants, it is only necessary to set $\operatorname{Im} W^A(x) = C^A(x)$, $w^A(x) = -\sqrt{2}\omega^A(x)$, and $\mathscr{W}^A(x) = M^A(x) - iN^A(x)$. For Abelian gauge theories, in which the structure constants vanish, this ends our task. For non-Abelian gauge theories it is necessary to add terms to $\operatorname{Im} W^A(x)$, $w^A(x)$, and $\mathscr{W}^A(x)$ of first order in gauge coupling constants to cancel the terms arising from commutators of zeroth-order terms, then add terms to $\operatorname{Im} W^A(x)$, $w^A(x)$, and $\mathscr{W}^A(x)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

of second order in gauge coupling constants to cancel the terms arising from commutators of first-order terms with zeroth-order terms, and so on. It is not easy to calculate the series of terms in $\mathrm{Im}\, W^A(x)$, $w^A(x)$, and $\mathscr{W}^A(x)$ needed to satisfy the gauge conditions (27.1.18) to all orders in gauge couplings, but there is no need — the important thing is that it is possible.

Inspection of the transformation rules (26.2.11)–(26.2.14) shows that the Wess–Zumino gauge condition (27.1.18) is not invariant under supersymmetry transformations unless $V_\mu^A = \lambda^A = 0$, and the condition $\lambda^A = 0$ is not supersymmetric unless also $D^A = 0$, in which case the whole superfield vanishes. Once we adopt Wess–Zumino gauge, the action is no longer invariant under either general extended gauge transformations or under supersymmetry, but it is invariant under supersymmetry transformations, which take us out of Wess–Zumino gauge, followed by suitable extended gauge transformations that take us back to Wess–Zumino gauge. (We will go into this explicitly in Section 27.8.) As we shall now see, it is also invariant under the ordinary gauge transformations (27.1.2)–(27.1.4), which preserve Wess–Zumino gauge.

With the gauge superfield satisfying the Wess–Zumino gauge condition (27.1.18), it becomes relatively easy to calculate its behavior under ordinary infinitesimal gauge transformations. In this case, $\Omega^A(x_+)$ are left-chiral superfields of the form (26.3.11), but with no ψ_L- or $\mathscr{F}$-components, and with ϕ-components given by *real* infinitesimal functions $\Lambda^A(x)$:

$$\Omega^A(x_+) = \Lambda^A(x) + \frac{1}{2}\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\Lambda^A(x) - \frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box\Lambda^A(x)\,. \qquad (27.1.20)$$

To calculate the product of exponentials in the transformation rule (27.1.4), we use a version of the Baker–Hausdorff formula:

$$\exp(a)\exp(X)\exp(b) = \exp\left[X + L_X\cdot(b-a) + (L_X\coth L_X)\cdot(b+a) + \cdots\right], \qquad (27.1.21)$$

where a, b, and X are arbitrary matrices, L_X is the operator

$$L_X\cdot f = \tfrac{1}{2}[X, f]\,, \qquad (27.1.22)$$

and '$\cdots$' here denotes terms of second and higher order in a and/or b. In our case we have

$$b + a = 2\sum_A t_A\,\mathrm{Im}\,\Lambda^A(x_+) = -i\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\sum_A t_A\,\partial^\mu\Lambda^A(x)\,,$$

$$b - a = -2i\sum_A t_A\,\mathrm{Re}\,\Lambda^A(x_+) = -2i\sum_A t_A\left[\Lambda^A(x) - \frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box\Lambda^A(x)\right],$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$X = -2\sum_A t_A\, V^A(x,\theta) = -2\sum_A t_A \left[\frac{i}{2}\left(\bar\theta\,\gamma_5\,\gamma^\mu\,\theta\right)V_\mu^A(x) - i\left(\bar\theta\,\gamma_5\,\theta\right)\left(\bar\theta\lambda^A(x)\right) - \frac{1}{4}\left(\bar\theta\,\gamma_5\,\theta\right)^2 D^A(x)\right].$$

Now, every term in X involves at least one factor of θ_L and at least one factor of θ_R, while $a+b$ has just one factor of θ_L and one factor of θ_R, so we can drop any terms in $L_X \coth L_X$ of second or higher order in L_X. Since $L_X \coth L_X$ is an *even* function of L_X, this means that we can replace it with its term of zeroth order in L_X, which is just unity. Also, we may drop the term in $b-a$ proportional to $(\bar\theta\gamma_5\theta)^2$, since when acted on by L_X it would yield at least three factors of either θ_L or θ_R. Thus the argument of the exponential on the right-hand side of Eq. (27.1.21) may be replaced with

$$X + \tfrac{1}{2}[X, b-a] + b + a = -2\sum_A t_A \left[V^A(x,\theta) + \sum_{BC} C^A{}_{BC}\, V^B(x,\theta)\,\Lambda^C(x) + \tfrac{1}{2}i\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\Lambda^A(x)\right].$$

Thus for infinitesimal gauge transformations, the transformation rule (27.1.4) yields

$$V^A(x,\theta) \to V^A(x,\theta) + \sum_{BC} C^A{}_{BC}\, V^B(x,\theta)\,\Lambda^C(x) + \frac{1}{2}i\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\Lambda^A(x)\,. \tag{27.1.23}$$

It is important to note that under ordinary gauge transformations, a gauge superfield in Wess–Zumino gauge remains in Wess–Zumino gauge. In terms of the component fields in Eq. (27.1.19), Eq. (27.1.23) reads

$$V_\mu^A(x) \to \sum_{BC} C^A{}_{BC}\, V_\mu^B(x)\,\Lambda^C(x) + \partial_\mu\Lambda^A(x)\,, \tag{27.1.24}$$

$$\lambda^A(x) \to \sum_{BC} C^A{}_{BC}\, \lambda^B(x)\,\Lambda^C(x)\,, \tag{27.1.25}$$

$$D^A(x) \to \sum_{BC} C^A{}_{BC}\, D^B(x)\,\Lambda^C(x)\,. \tag{27.1.26}$$

We recognize Eq. (27.1.24) as the usual Yang–Mills gauge transformation rule (15.1.9) of a gauge field, while Eqs. (27.1.25) and (27.1.26) tell us that the fields $\lambda^A(x)$ and $D^A(x)$ transform as 'matter' fields that belong to the adjoint representation of the gauge group. The Majorana spinors λ^A are known as *gaugino* fields, while the real scalars D^A will turn out to be another set of auxiliary fields.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Next we must evaluate the matrix Γ that is needed in constructing gauge-invariant functions of chiral superfields. Because all terms with more than four factors of θ vanish, the expansion of the exponential is very simple in Wess–Zumino gauge:

$$\begin{aligned}\Gamma(x,\theta) &= \exp\left(-2\sum_A t_A\, V^A(x,\theta)\right)\\ &= 1 - i\left(\bar\theta\,\gamma_5\,\gamma^\mu\,\theta\right)\sum_A t_A\, V_\mu^A(x)\\ &\quad -\frac{1}{2}\left(\bar\theta\,\gamma_5\,\gamma^\mu\,\theta\right)\left(\bar\theta\,\gamma_5\,\gamma^\nu\,\theta\right)\sum_{AB} t_A\, t_B\, V_\mu^A(x)\, V_\nu^B(x)\\ &\quad +2i\left(\bar\theta\,\gamma_5\,\theta\right)\sum_A t_A\left(\bar\theta\,\lambda^A(x)\right) + \frac{1}{2}\left(\bar\theta\,\gamma_5\,\theta\right)^2\sum_A t_A\, D^A(x)\,.\end{aligned}$$

We can construct a gauge-invariant density by multiplying this on the right with a column vector of left-chiral superfields of the form (26.3.11):

$$\begin{aligned}\Phi_n(x,\theta) = {}& \phi_n(x) - \sqrt{2}\left(\bar\theta\psi_{nL}(x)\right) + \mathscr{F}_n(x)\left(\bar\theta\left(\frac{1+\gamma_5}{2}\right)\theta\right)\\ &+\frac{1}{2}\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\phi_n(x) - \frac{1}{\sqrt{2}}\left(\bar\theta\gamma_5\theta\right)\left(\bar\theta\,\partial\!\!\!/\psi_{nL}(x)\right)\\ &-\frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box\phi_n(x)\,,\end{aligned}$$

and multiplying on the left with the column

$$\begin{aligned}\Phi_n(x,\theta)^* = {}& \phi_n^*(x) - \sqrt{2}\left(\overline{\psi_{nL}(x)}\theta\right) + \mathscr{F}_n^*(x)\left(\bar\theta\left(\frac{1-\gamma_5}{2}\right)\theta\right)\\ &-\frac{1}{2}\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\partial^\mu\phi_n^*(x) - \frac{1}{\sqrt{2}}\left(\bar\theta\gamma_5\theta\right)\partial_\mu\left(\overline{\psi_{nL}(x)}\gamma^\mu\theta\right)\\ &-\frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\Box\phi_n^*(x)\,.\end{aligned}$$

The term in this product of fourth order in θ is

$$\begin{aligned}\left[\Phi^\dagger\,\Gamma\,\Phi\right]_{\theta^4} = {}& -\frac{1}{8}\left(\bar\theta\gamma_5\theta\right)^2\left\{\left[\phi^\dagger\Box\phi\right] + \left[\left(\Box\phi^\dagger\right)\phi\right]\right\}\\ &+\left(\bar\theta\gamma_5\,\theta\right)\left\{\left[\left(\overline{\psi_L}\,\theta\right)\left(\bar\theta\gamma^\mu\partial_\mu\psi_L\right)\right] + \left[\left((\partial_\mu\overline{\psi_L})\,\gamma^\mu\theta\right)\left(\bar\theta\,\psi_L\right)\right]\right\}\\ &+\frac{1}{4}\left(\bar\theta(1-\gamma_5)\theta\right)\left(\bar\theta(1+\gamma_5)\theta\right)\left[\mathscr{F}^\dagger\,\mathscr{F}\right]\\ &-\frac{1}{4}\left(\bar\theta\gamma_5\gamma^\mu\theta\right)\left(\bar\theta\gamma_5\gamma^\nu\theta\right)\left[\partial_\mu\phi^\dagger\,\partial_\nu\phi\right]\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$
\begin{aligned}
&-\frac{i}{2}\left(\bar\theta\gamma_5\gamma^\mu\theta\right)\left(\bar\theta\gamma_5\gamma^\nu\theta\right)\sum_A V_\mu^A\left\{\left[\phi^\dagger t_A\,\partial_\nu\phi\right]-\left[(\partial_\nu\phi^\dagger)\,t_A\,\phi\right]\right\}\\
&-\frac{1}{2}\left(\bar\theta\gamma_5\gamma^\mu\theta\right)\left(\bar\theta\gamma_5\gamma^\nu\theta\right)\sum_{AB} V_\mu^A V_\nu^B\left[\phi^\dagger t_A\,t_B\,\phi\right]\\
&-2i\left(\bar\theta\gamma_5\gamma_\mu\theta\right)\sum_A V_\mu^A\left[\left(\overline{\psi_L}\,\theta\right)t_A\left(\bar\theta\,\psi_L\right)\right]\\
&-2i\sqrt{2}\left(\bar\theta\,\gamma_5\,\theta\right)\sum_A\left[\left(\overline{\psi_L}\,\theta\right)t_A\left(\bar\theta\lambda^A\right)\phi\right]\\
&-2i\sqrt{2}\left(\bar\theta\,\gamma_5\,\theta\right)\sum_A\left[\phi^\dagger\left(\overline{\lambda^A}\,\theta\right)t_A\left(\bar\theta\,\psi_L\right)\right]\\
&+\frac{1}{2}\left(\bar\theta\gamma_5\,\theta\right)^2\sum_A D_A\left[\phi^\dagger t_A\,\phi\right],
\end{aligned}
$$

in which we use square brackets to indicate scalar products in the flavor indices n, m, and continue to use round brackets to indicate scalar products in Dirac indices. Just as in Section 26.4, we may use the identities (26.A.17)–(26.A.19) to put all θ dependence in the form of an over-all factor $(\bar\theta\gamma_5\theta)^2$:

$$
\begin{aligned}
\left[\Phi^\dagger\,\Gamma\,\Phi\right]_{\theta^4} = \left(\bar\theta\gamma_5\theta\right)^2\Bigg\{&-\frac{1}{8}\left[\phi^\dagger\Box\phi\right]-\frac{1}{8}\left[\left(\Box\phi^\dagger\right)\phi\right]\\
&+\frac{1}{4}\left[\left(\overline{\psi_L}\,\gamma^\mu\partial_\mu\psi_L\right)\right]-\frac{1}{4}\left[\left((\partial_\mu\overline{\psi_L})\,\gamma^\mu\,\psi_L\right)\right]\\
&-\frac{1}{2}\left[\mathscr{F}^\dagger\,\mathscr{F}\right]+\frac{1}{4}\left[\partial_\mu\phi^\dagger\,\partial^\mu\phi\right]\\
&+\frac{i}{2}\sum_A V_\mu^A\left[\phi^\dagger t_A\,\partial^\mu\phi\right]-\frac{i}{2}\sum_A V_\mu^A\left[(\partial^\mu\phi^\dagger)\,t_A\,\phi\right]\\
&+\frac{1}{2}\sum_{AB} V_\mu^A V^{B\mu}\left[\phi^\dagger t_A\,t_B\,\phi\right]-\frac{i}{2}\sum_A V_\mu^A\left[\left(\overline{\psi_A}\gamma^\mu t_A\,\psi_A\right)\right]\\
&-\frac{i}{\sqrt{2}}\sum_A\left[\left(\overline{\psi_L}\,t_A\,\lambda^A\right)\phi\right]+\frac{i}{\sqrt{2}}\sum_A\left[\phi^\dagger\left(\overline{\lambda^A}\,t_A\,\psi_L\right)\right]\\
&+\frac{1}{2}\sum_A D_A\left[\phi^\dagger t_A\,\phi\right]\Bigg\}.
\end{aligned}
$$

The D-term is the coefficient of $-\frac{1}{4}(\bar\theta\gamma_5\theta)^2$ minus $\frac{1}{2}\Box$ acting on the θ-independent term, which for $[\Phi^\dagger\Gamma\Phi]$ is $[\phi^\dagger\phi]$, so

$$
\begin{aligned}
\left[\Phi^\dagger\,\Gamma\,\Phi\right]_D = &-2[\partial_\mu\phi^\dagger\,\partial^\mu\phi]\\
&-\left[\left(\overline{\psi_L}\,\gamma^\mu\partial_\mu\psi_L\right)\right]+\left[\left((\partial_\mu\overline{\psi_L})\,\gamma^\mu\,\psi_L\right)\right]+2\left[\mathscr{F}^\dagger\,\mathscr{F}\right]
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$
\begin{aligned}
&-2i\sum_A V_\mu^A\left[\phi^\dagger t_A\,\partial^\mu\phi\right]+2i\sum_A V_\mu^A\left[(\partial^\mu\phi^\dagger)\,t_A\,\phi\right]\\
&-2\sum_{AB} V_\mu^A V^{B\mu}\left[\phi^\dagger t_A t_B\,\phi\right]+2i\sum_A V_\mu^A\left[\left(\overline{\psi_A}\gamma^\mu t_A\,\psi_A\right)\right]\\
&+2i\sqrt{2}\sum_A\left[\left(\overline{\psi_L}\,t_A\,\lambda^A\right)\phi\right]-2i\sqrt{2}\sum_A\left[\phi^\dagger\left(\overline{\lambda^A}\,t_A\,\psi_L\right)\right]\\
&-2\sum_A D_A\left[\phi^\dagger t_A\,\phi\right].
\end{aligned}
$$

To see that this *is* gauge-invariant, we note that it may be written as

$$
\begin{aligned}
\frac{1}{2}\left[\Phi^\dagger\,\Gamma\,\Phi\right]_D=&-\left[(D_\mu\phi)^\dagger D^\mu\phi\right]\\
&-\frac{1}{2}\left[\left(\overline{\psi_L}\,\gamma^\mu D_\mu\psi_L\right)\right]+\frac{1}{2}\left[\left(\overline{(D_\mu\psi_L)}\,\gamma^\mu\,\psi_L\right)\right]+\left[\mathscr{F}^\dagger\,\mathscr{F}\right]\\
&+i\sqrt{2}\sum_A\left[\left(\overline{\psi_L}\,t_A\,\lambda^A\right)\phi\right]-i\sqrt{2}\sum_A\left[\phi^\dagger\left(\overline{\lambda^A}\,t_A\,\psi_L\right)\right]\\
&-\sum_A D_A\left[\phi^\dagger t_A\,\phi\right],
\end{aligned}
\tag{27.1.27}
$$

where D_μ is the gauge-invariant derivative (15.1.10):

$$
D_\mu\psi_L\equiv\partial_\mu\psi_L-i\sum_A t_A\,V_\mu^A\,\psi_L\,,\qquad D_\mu\phi\equiv\partial_\mu\phi-i\sum_A t_A\,V_\mu^A\,\phi\,. \tag{27.1.28}
$$

Eq. (27.1.27) is thus a suitable gauge-invariant kinematic Lagrangian for the scalar and spinor components of a left-chiral superfield, now supplemented with Yukawa couplings of gaugino fields to the scalar and spinor components of chiral superfields as well as terms involving auxiliary fields $\mathscr{F}_n$ and D_A.

27.2 Gauge-Invariant Action for Abelian Gauge Superfields

We must now consider how to construct a gauge-invariant supersymmetric action for the gauge superfields $V^A(x,\theta)$ that contain the gauge fields $V_\mu^A(x)$. In order to motivate this construction we will first consider the case of a single Abelian gauge field (dropping the superscript A), and then return to the general case in the next section.

In an Abelian gauge theory like quantum electrodynamics, the gauge-invariant field constructed from $V_\mu(x)$ is the familiar field-strength tensor

$$
f_{\mu\nu}(x)=\partial_\mu V_\nu(x)-\partial_\nu V_\mu(x)\,. \tag{27.2.1}
$$

The supersymmetry transformation rule for $f_{\mu\nu}(x)$ is then given by the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

transformation rule (26.2.15) for $V_\mu(x)$ as

$$\delta f_{\mu\nu} = \left(\bar{\alpha}(\partial_\mu \gamma_\nu - \partial_\nu \gamma_\mu)\lambda\right) . \tag{27.2.2}$$

Eq. (26.2.16) gives the transformation rule for $\lambda(x)$ as

$$\delta\lambda = \left(-\tfrac{1}{4} f_{\mu\nu}[\gamma^\mu, \gamma^\nu] + i\gamma_5 D\right)\alpha , \tag{27.2.3}$$

while Eq. (26.2.17) gives the transformation rule for $D(x)$:

$$\delta D = i\left(\bar{\alpha}\gamma_5 \not{\partial}\lambda\right) . \tag{27.2.4}$$

None of this depends on whether or not the superfield $V^A(x,\theta)$ is taken to be in Wess–Zumino gauge. We see that the fields $f_{\mu\nu}(x)$, $\lambda(x)$, and $D(x)$ form a complete supersymmetry multiplet.

It is not hard to construct a suitable kinematic Lagrangian density for the fields of this supermultiplet. The only Lorentz-invariant, parity conserving, and gauge-invariant functions of these fields with dimensionality four are $f_{\mu\nu}f^{\mu\nu}$, $\bar{\lambda}\not{\partial}\lambda$, and D^2. We can make V^μ a conventionally normalized vector field by taking the coefficient of $f_{\mu\nu}f^{\mu\nu}$ to be $-\frac{1}{4}$, so we may tentatively take the kinematic Lagrangian density as

$$\mathscr{L}_{\text{gauge}} = -\tfrac{1}{4} f_{\mu\nu}f^{\mu\nu} - c_\lambda\left(\bar{\lambda}\not{\partial}\lambda\right) - c_D D^2 ,$$

with coefficients c_λ and c_D to be determined from the condition that $\int \mathscr{L}_{\text{gauge}} d^4x$ is supersymmetric. Using Eqs. (27.2.2)–(27.2.4), an infinitesimal supersymmetry transformation changes the operators in the Lagrangian by

$$\begin{aligned}
\delta\left(f_{\mu\nu}f^{\mu\nu}\right) &= 2f^{\mu\nu}\left(\bar{\alpha}(\gamma_\nu\partial_\mu - \gamma_\mu\partial_\nu)\lambda\right) , \\
\delta\left(\bar{\lambda}\not{\partial}\lambda\right) &= 2\left(\bar{\alpha}\left[+\tfrac{1}{4} f_{\mu\nu}[\gamma^\mu, \gamma^\nu] + i\gamma_5 D\right]\not{\partial}\lambda\right) , \\
\delta D^2 &= 2iD\left(\bar{\alpha}\gamma_5 \not{\partial}\lambda\right) ,
\end{aligned}$$

in which we drop derivative terms that do not contribute to the variation of the action. In order to see how these terms cancel, it is necessary to use an identity for gamma matrices*

$$[\gamma^\mu, \gamma^\nu]\gamma^\rho = -2\eta^{\mu\rho}\gamma^\nu + 2\eta^{\nu\rho}\gamma^\mu - 2i\epsilon^{\mu\nu\rho\sigma}\gamma_\sigma\gamma_5 . \tag{27.2.5}$$

The term $-i\epsilon^{\mu\nu\rho\sigma} f_{\mu\nu}(\bar{\alpha}\gamma_\sigma\gamma_5\partial_\rho\lambda)$ does not contribute to $\int d^4x\, \delta\mathscr{L}$, because integrating by parts yields a contribution proportional to $\epsilon^{\mu\nu\rho\sigma}\partial_\rho f_{\mu\nu}$, which

* To derive this, use the fact that any 4×4 matrix may be expressed as a linear combination of the 16 independent covariant matrices described in Section 5.4, which in our case is limited by Lorentz invariance and space inversion invariance to the terms shown here. The coefficients of these terms can be calculated by giving $\mu\nu\rho$ the values 121 and 123.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

vanishes for an $f_{\mu\nu}$ of the form (27.2.1). This identity then allows us to rewrite the variation of the λ-term as

$$\delta\left(\bar{\lambda}\,\partial\!\!\!/\lambda\right) = -f^{\mu\nu}\left(\bar{\alpha}(\gamma_\nu\partial_\mu - \gamma_\mu\partial_\nu)\lambda\right) + 2iD\left(\bar{\alpha}\,\gamma_5\,\partial\!\!\!/\lambda\right).$$

The cancellation of terms proportional to $f^{\mu\nu}\lambda$ requires that $c_\lambda = 1/2$, while the cancellation of terms proportional to $D\lambda$ requires that $c_D = -c_\lambda$, so the supersymmetric Lagrangian density takes the form

$$\mathscr{L}_{\text{gauge}} = -\tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} - \tfrac{1}{2}\left(\bar{\lambda}\,\partial\!\!\!/\lambda\right) + \tfrac{1}{2}D^2\,. \tag{27.2.6}$$

This shows that with V^μ canonically normalized, the field λ related to V^μ by the transformation rules (27.2.2) and (27.2.3) is also canonically normalized.

In addition, for Abelian gauge theories there is a superrenormalizable term, known as a *Fayet–Iliopoulos term*:[2]

$$\mathscr{L}_{\text{FI}} = \xi\,D\,, \tag{27.2.7}$$

with ξ an arbitrary constant. Its variation under a supersymmetry transformation is shown by Eq. (27.2.4) to be a derivative, so that it yields another supersymmetric term in the action. As we will see in Section 27.5, the presence of such a term can provide a mechanism for the spontaneous breakdown of supersymmetry.

Both for its own interest and as a tool in constructing supersymmetric interactions involving the fields $f_{\mu\nu}$, λ, and D, it is interesting to ask what sort of superfield has these as component fields. Somewhat surprisingly, it turns out to be a *spinor* superfield $W_\alpha(x)$, with component fields (in the notation of Eq. (26.2.10)) given by

$$\begin{aligned}
&C_{(\alpha)}(x) = \lambda_\alpha(x)\,,\\
&\omega_{(\alpha)\beta}(x) = \tfrac{1}{2}i\left(\gamma^\mu\gamma^\nu\epsilon\right)_{\alpha\beta}f_{\mu\nu}(x) + (\gamma_5\epsilon)_{\alpha\beta}D(x)\,,\\
&V_{(\alpha)\mu}(x) = -i\partial_\mu\left(\gamma_5\lambda(x)\right)_\alpha\,,\\
&M_{(\alpha)}(x) = -i\left(\partial\!\!\!/\gamma_5\lambda(x)\right)_\alpha\,, \qquad N_{(\alpha)}(x) = -\left(\partial\!\!\!/\lambda(x)\right)_\alpha\,,\\
&\lambda_{(\alpha)\beta}(x) = D_{(\alpha)}(x) = 0\,.
\end{aligned} \tag{27.2.8}$$

(The subscript α on these component fields is put inside parentheses to emphasize that it labels the whole superfield.) It is straightforward to use Eqs. (27.2.2)–(27.2.4) to check directly that the superfield components given by Eq. (27.2.8) do transform as in Eqs. (26.2.11)–(26.2.17).

Inserting the component fields (27.2.8) in Eq. (26.2.10) and using

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Eq. (26.A.5), we find that the superfield W_α takes the form

$$\begin{aligned} W_\alpha(x,\theta) = \Big[& \lambda(x) + \tfrac{1}{2}\gamma^\mu\gamma^\nu\theta\, f_{\mu\nu}(x) - i\gamma_5\theta\, D(x) - \tfrac{1}{2}\Big(\theta^{\mathrm{T}}\epsilon\theta\Big)\, \partial\!\!\!/\gamma_5\lambda(x) \\ & + \tfrac{1}{2}\Big(\theta^{\mathrm{T}}\epsilon\gamma_5\theta\Big)\, \partial\!\!\!/\lambda(x) + \tfrac{1}{2}\Big(\theta^{\mathrm{T}}\epsilon\gamma^\mu\theta\Big)\gamma_5\partial_\mu\lambda(x) \\ & - \tfrac{1}{4}\Big(\theta^{\mathrm{T}}\epsilon\theta\Big)\gamma_5\gamma^\mu\gamma^\nu\gamma^\sigma\theta\, \partial_\sigma f_{\mu\nu}(x) \\ & + \tfrac{1}{2}i\Big(\theta^{\mathrm{T}}\epsilon\theta\Big)\gamma^\sigma\theta\, \partial_\sigma D(x) - \tfrac{1}{8}\Big(\theta^{\mathrm{T}}\epsilon\theta\Big)^2\Box\lambda(x)\Big]_\alpha . \end{aligned} \tag{27.2.9}$$

As we showed in Section 26.3, a superfield like this with zero λ- and D-components is *chiral* — that is, it is the sum of left-chiral and right-chiral superfields

$$W(x,\theta) = W_L(x,\theta) + W_R(x,\theta)\,. \tag{27.2.10}$$

Here the left- and right-chiral superfields are simply the projections of W on the subspaces with $\gamma_5 = +1$ and $\gamma_5 = -1$, respectively:

$$\begin{aligned} W_L(x,\theta) &= \tfrac{1}{2}(1+\gamma_5)W(x,\theta) \\ &= \lambda_L(x_+) + \tfrac{1}{2}\gamma^\mu\gamma^\nu\theta_L\, f_{\mu\nu}(x_+) + \Big(\theta_L^{\mathrm{T}}\epsilon\theta_L\Big)\, \partial\!\!\!/\lambda_R(x_+) - i\theta_L D(x_+)\,, \end{aligned} \tag{27.2.11}$$

$$\begin{aligned} W_R(x,\theta) &= \tfrac{1}{2}(1-\gamma_5)W(x,\theta) \\ &= \lambda_R(x_-) + \tfrac{1}{2}\gamma^\mu\gamma^\nu\theta_R\, f_{\mu\nu}(x_-) - \Big(\theta_R^{\mathrm{T}}\epsilon\theta_R\Big)\, \partial\!\!\!/\lambda_L(x_-) - i\theta_R D(x_-)\,, \end{aligned} \tag{27.2.12}$$

with $x_\pm^\mu$ given by Eq. (26.3.23).

As we saw in Section 26.3, we can construct suitable Lagrangian densities from the $\mathscr{F}$-term of any scalar function of a left-chiral superfield, plus its Hermitian adjoint. The simplest scalar function of the left-chiral superfield (27.2.11) is $\sum_{\alpha\beta}\epsilon_{\alpha\beta}W_{L\alpha}W_{L\beta}$. To calculate the $\mathscr{F}$-term, we note that, when expressed as a function of θ_L and x_+, the term in $\sum_{\alpha\beta}\epsilon_{\alpha\beta}W_{L\alpha}W_{L\beta}$ of second order in θ_L is

$$\begin{aligned} -\Big[\sum_{\alpha\beta}\epsilon_{\alpha\beta}W_{L\alpha}W_{L\beta}\Big]_{\theta_L^2} = \Big(\theta_L^{\mathrm{T}}\epsilon\theta_L\Big)\Big[-2\Big(\lambda_L^{\mathrm{T}}(x)\,\epsilon\, \partial\!\!\!/\lambda_R(x)\Big) + D^2(x)\Big] \\ + \frac{1}{16}\Big(\overline{\theta_L}[\gamma^\mu,\gamma^\nu]\,[\gamma^\rho,\gamma^\sigma]\theta_L\Big) f_{\mu\nu}(x) f_{\rho\sigma}(x)\,. \end{aligned}$$

(The argument of the fields here can be taken as x^μ rather than x_+^μ because the difference would yield terms that contain at least three factors of θ_L, which therefore vanish.) Lorentz invariance together with the fact that $(\bar{s}[\gamma_\mu,\gamma_\nu]s)$ and $(\bar{s}[\gamma_\mu,\gamma_\nu]\gamma_5 s)$ vanish for any Majorana fermion s tell us

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

that the bilinear $(\overline{\theta_L}[\gamma^\mu, \gamma^\nu][\gamma^\rho, \gamma^\sigma]\theta_L)$ must be proportional to a linear combination of $(\overline{\theta_L}\theta_L)(\eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho})$ and $(\overline{\theta_L}\theta_L)\epsilon^{\mu\nu\rho\sigma}$. We can find the coefficients by giving $\mu\nu\rho\sigma$ the values 1212 or 1230, and in this way we find that

$$\left(\overline{\theta_L}[\gamma^\mu, \gamma^\nu][\gamma^\rho, \gamma^\sigma]\theta_L\right) = 4\left(\overline{\theta_L}\theta_L\right)\left[-\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho} + i\epsilon^{\mu\nu\rho\sigma}\right].$$

The $\mathscr{F}$-term is the coefficient of $(\overline{\theta_L}\theta_L)$, so

$$-\left[\sum_{\alpha\beta}\epsilon_{\alpha\beta}W_{L\alpha}W_{L\beta}\right]_{\mathscr{F}} = -2\left(\overline{\lambda_R}\,\not\partial\lambda_R\right) - \frac{1}{2}f_{\mu\nu}f^{\mu\nu} + \frac{i}{4}\epsilon^{\mu\nu\rho\sigma}f_{\mu\nu}f_{\rho\sigma} + D^2\,. \tag{27.2.13}$$

Eq. (26.A.21) shows that $(\bar\lambda\,\not\partial\lambda)$ is real, while $(\bar\lambda\,\not\partial\gamma_5\lambda)$ is imaginary, so the real part of Eq. (27.2.13) yields the Lagrangian (27.2.6) for the gauge and gaugino fields

$$-\frac{1}{2}\,\mathrm{Re}\left[\sum_{\alpha\beta}\epsilon_{\alpha\beta}W_{L\alpha}W_{L\beta}\right]_{\mathscr{F}} = -\frac{1}{2}\left(\bar\lambda\,\not\partial\lambda\right) - \frac{1}{4}f_{\mu\nu}f^{\mu\nu} + \frac{1}{2}D^2\,. \tag{27.2.14}$$

The physical significance of the imaginary part will be discussed in a more general context in the next section.

There is another way to derive the form of the spinor superfield, which will turn out to provide a more convenient way of deriving the components of the gauge superfield in non-Abelian gauge theories. A tedious but straightforward calculation shows that the gauge-invariant superfield (27.2.9) may be expressed in terms of the gauge superfield (27.1.16) as

$$W_\alpha(x,\theta) = \frac{i}{4}\left(\mathscr{D}^{\mathrm{T}}\epsilon\mathscr{D}\right)\mathscr{D}_\alpha V(x,\theta)\,, \tag{27.2.15}$$

where $\mathscr{D}_\alpha$ is the superderivative introduced in Eq. (26.2.26):

$$\mathscr{D}_\alpha \equiv \sum_\beta(\gamma_5\epsilon)_{\alpha\beta}\frac{\partial}{\partial\theta_\beta} - (\gamma^\mu\theta)_\alpha\frac{\partial}{\partial x^\mu} = -\frac{\partial}{\partial\bar\theta_\alpha} - (\gamma^\mu\theta)_\alpha\frac{\partial}{\partial x^\mu}\,.$$

This result might have been obtained (aside from the normalization factor) by noting that the function (27.2.15) has the desired property of being a gauge-invariant chiral spinor superfield. First, note that Eq. (27.2.15) *is* a superfield, because it is formed by acting on the superfield V with superderivatives. Also, it follows from the anticommutation of the $\mathscr{D}$s that the product of any three or more $\mathscr{D}_L$s or three or more $\mathscr{D}_R$s vanishes, so that

$$\left(\mathscr{D}^{\mathrm{T}}\epsilon\mathscr{D}\right)\mathscr{D} = \left(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L\right)\mathscr{D}_R + \left(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R\right)\mathscr{D}_L\,. \tag{27.2.16}$$

Because $\mathscr{D}_L(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L) = \mathscr{D}_R(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R) = 0$, the superfield (27.2.15) is chiral,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

with

$$W_{L\alpha}(x,\theta) = \frac{i}{4}\left(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R\right)\mathscr{D}_{L\alpha}V(x,\theta), \quad W_{R\alpha}(x,\theta) = \frac{i}{4}\left(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L\right)\mathscr{D}_{R\alpha}V(x,\theta). \tag{27.2.17}$$

Finally, we can show that (27.2.15) is invariant under the generalized gauge transformation (27.1.13), which for a single Abelian gauge superfield is simply

$$V(x,\theta) \to V(x,\theta) + \frac{i}{2}\left[\Omega(x,\theta) - \Omega^*(x,\theta)\right], \tag{27.2.18}$$

where $\Omega(x,\theta)$ is an arbitrary left-chiral superfield. Since $\mathscr{D}_L\Omega^* = 0$, the change in $W_{L\alpha}$ is proportional to $(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R)\mathscr{D}_{L\alpha}\Omega$. But $\mathscr{D}_R\Omega = 0$ and

$$\left[(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R),\, \mathscr{D}_{L\alpha}\right] = -2\left[(1+\gamma_5)\,\not{\partial}\mathscr{D}_R\right]_\alpha,$$

so the change in $W_{L\alpha}$ vanishes. A similar argument shows that $W_{R\alpha}$ is also gauge-invariant. (The work of checking Eq. (27.2.15) is greatly reduced by using this gauge-invariance property to put $V(x,\theta)$ in Wess–Zumino gauge.)

The chiral superfields (27.2.11) and (27.2.12) are evidently not of the most general form for left- and right-chiral superfields. To put the constraints satisfied by these superfields in a manifestly supersymmetric form, we note by using the anticommutation relation (26.2.30) that

$$\begin{aligned}\epsilon_{\alpha\beta}\mathscr{D}_{L\alpha}\left(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R\right)\mathscr{D}_{L\beta} &= -2\mathscr{D}_{R\alpha}\mathscr{D}_{L\beta}\left(\epsilon(1+\gamma_5)\,\not{\partial}\right)_{\beta\alpha} + \left(\mathscr{D}_R^{\mathrm{T}}\epsilon\mathscr{D}_R\right)\left(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L\right)\\ &= \epsilon_{\alpha\beta}\mathscr{D}_{R\alpha}\left(\mathscr{D}_L^{\mathrm{T}}\epsilon\mathscr{D}_L\right)\mathscr{D}_{R\beta}\,.\end{aligned} \tag{27.2.19}$$

From Eq. (27.2.17) it follows then that W_L and W_R are related by the constraint

$$\epsilon_{\alpha\beta}\mathscr{D}_{L\alpha}W_{L\beta} = \epsilon_{\alpha\beta}\mathscr{D}_{R\alpha}W_{R\beta}\,. \tag{27.2.20}$$

It is straightforward to show that the most general chiral spinor superfields satisfying Eq. (27.2.20) are of the form (27.2.11) and (27.2.12), with $f_{\mu\nu}$ constrained to satisfy the 'Bianchi' identities $\epsilon^{\mu\nu\rho\sigma}\partial_\rho f_{\mu\nu} = 0$.

27.3 Gauge-Invariant Action for General Gauge Superfields

Our experience in the previous section with supersymmetric Abelian gauge theories suggests immediately that in a general non-Abelian gauge theory the kinematic Lagrangian for the fields $V_\mu^A(x)$, $\lambda^A(x)$, and $D^A(x)$ should appear as part of the gauge-invariant generalization of Eq. (27.2.6):

$$\mathscr{L}_{\text{gauge}} = -\tfrac{1}{4}\sum_A f_{A\mu\nu}f_A^{\mu\nu} - \tfrac{1}{2}\sum_A\left(\overline{\lambda_A}(\not{p}\lambda)_A\right) + \tfrac{1}{2}\sum_A D_A D_A\,. \tag{27.3.1}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

We are now using a basis for the Lie algebra with totally antisymmetric structure constants, and we are consequently not preserving the distinction between upper and lower group indices, writing all indices A, B, etc. as subscripts. Also, $f_{A\mu\nu}$ is the gauge-covariant field-strength tensor

$$f_{A\mu\nu} = \partial_\mu V_{A\nu} - \partial_\nu V_{A\mu} + \sum_{BC} C_{ABC} V_{B\mu} V_{C\nu} , \tag{27.3.2}$$

and $D_\mu\lambda$ is the gauge-covariant derivative of the gaugino field, which in the adjoint representation is

$$(D_\mu\lambda)_A = \partial_\mu\lambda_A + \sum_{BC} C_{ABC} V_{B\mu}\lambda_C . \tag{27.3.3}$$

The question is: does Eq. (27.3.1) yield a supersymmetric action?

Since the Lagrangian density (27.3.1) is manifestly gauge-invariant, we can test whether the action is supersymmetric in any convenient gauge. To find out whether $\delta\mathscr{L}_{\text{gauge}}$ is a derivative at some point X^μ, it is convenient to adopt a specific version of Wess–Zumino gauge in which $V_A^\mu(X) = 0$. Then at X the changes in the component fields are given by Eqs. (26.2.15)–(26.2.17) at $x = X$ as

$$\delta V_{A\mu} = \left(\bar{\alpha}\gamma_\mu\lambda_A\right) , \tag{27.3.4}$$

$$\delta\lambda_A = \left(\frac{1}{4} f_{A\mu\nu}\,[\gamma^\nu , \gamma^\mu] + i\gamma_5 D_A\right)\alpha , \tag{27.3.5}$$

$$\delta D_A = i\left(\bar{\alpha}\gamma_5 \not{\partial}\lambda_A\right) . \tag{27.3.6}$$

(We must set x^μ in these expressions equal to X^μ *after* calculating the change under a supersymmetry transformation, not before.) Also, the non-linear terms in $f_A^{\mu\nu}$ are quadratic in the Vs and therefore at $x = X$ have zero variation, so at $x = X$

$$\delta f_{A\mu\nu} = \left(\bar{\alpha}(\gamma_\nu\partial_\mu - \gamma_\mu\partial_\nu)\lambda_A\right) . \tag{27.3.7}$$

With one exception, the terms in Eq. (27.3.1) and their transformations under supersymmetry transformations are thus just a number of copies (labelled by A) of the Abelian theory discussed in the previous section, and therefore give a supersymmetric action. The one exception that might disturb the supersymmetry of the action arises from the second term in the gauge-covariant derivative (27.3.3) of the gaugino field:

$$\mathscr{L}_{\lambda\lambda V} = -\tfrac{1}{2} \sum_{ABC} C_{ABC}\left(\overline{\lambda_A}\not{V}_B\lambda_C\right) , \tag{27.3.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

whose variation at $x = X$ is

$$\delta \mathscr{L}_{\lambda\lambda V} = -\tfrac{1}{2} \sum_{ABC} C_{ABC} \left(\overline{\lambda_A} (\delta \not{V}_B) \lambda_C \right) = -\tfrac{1}{2} \sum_{ABC} C_{ABC} \left(\overline{\lambda_A} \gamma_\mu \lambda_C \right) \left(\bar{\alpha} \gamma^\mu \lambda_B \right) . \tag{27.3.9}$$

We can write the product of bilinears on the right-hand side as the sum of two terms

$$\left(\overline{\lambda_A} \gamma_\mu \lambda_C \right) \left(\bar{\alpha} \gamma^\mu \lambda_B \right) = X_{ABC} + Y_{ABC} ,$$

with

$$X_{ABC} \equiv \tfrac{1}{4} \sum_{\pm} \left(\overline{\lambda_A} (1 \pm \gamma_5) \gamma_\mu \lambda_C \right) \left(\bar{\alpha} \gamma^\mu (1 \pm \gamma_5) \lambda_B \right) ,$$
$$Y_{ABC} \equiv \tfrac{1}{4} \sum_{\pm} \left(\overline{\lambda_A} (1 \pm \gamma_5) \gamma_\mu \lambda_C \right) \left(\bar{\alpha} \gamma^\mu (1 \mp \gamma_5) \lambda_B \right) .$$

By using standard Fierz identities and the anticommutativity of the spinor fields, we have

$$\left(\overline{\lambda_A} (1 \pm \gamma_5) \gamma_\mu \lambda_B \right) \left(\bar{\alpha} (1 \pm \gamma_5) \gamma^\mu \lambda_C \right) = \left(\overline{\lambda_A} (1 \pm \gamma_5) \gamma_\mu \lambda_C \right) \left(\bar{\alpha} (1 \pm \gamma_5) \gamma^\mu \lambda_B \right) ,$$
$$\left(\overline{\lambda_A} (1 \pm \gamma_5) \gamma_\mu \lambda_B \right) \left(\bar{\alpha} (1 \mp \gamma_5) \gamma^\mu \lambda_C \right) = \left(\overline{\lambda_C} (1 \pm \gamma_5) \gamma_\mu \lambda_B \right) \left(\bar{\alpha} (1 \mp \gamma_5) \gamma^\mu \lambda_A \right) .$$

(To derive the first of these relations, we note that $[(1 \pm \gamma_5)\gamma_\mu]_{\alpha\gamma} [(1 \pm \gamma_5)\gamma^\mu]_{\delta\beta}$ may be thought of as the $\alpha\beta$ matrix element of a matrix depending on δ and γ, and may therefore be expanded in a linear combination of $1_{\alpha\beta}$, $\gamma^\mu_{\alpha\beta}$, $[\gamma^\mu, \gamma^\kappa]_{\alpha\beta}$, $(\gamma_5 \gamma^\mu)_{\alpha\beta}$, and $(\gamma_5)_{\alpha\beta}$. Because of the factors $(1 \pm \gamma_5)$, the only term in the expansion is proportional to $[(1 \pm \gamma_5)\gamma^\mu]_{\alpha\beta}$. Lorentz invariance and the presence of the other $1 \pm \gamma_5$ factor tell us that this expansion takes the form

$$[(1 \pm \gamma_5)\gamma_\mu]_{\alpha\gamma} [(1 \pm \gamma_5)\gamma^\mu]_{\delta\beta} = k [(1 \pm \gamma_5)\gamma_\mu]_{\alpha\beta} [(1 \pm \gamma_5)\gamma^\mu]_{\delta\gamma} .$$

To determine the proportionality constant k, we may contract both sides with $(\gamma_\nu)_{\gamma\alpha}$ and find $k = -1$. The minus sign is cancelled by a minus sign arising from the anticommutation of λ_C and $\bar{\alpha}$. The other Fierz identity is proved in the same way, except that we also need to use the symmetry property (26.A.7) for Majorana bilinears.) Hence X_{ABC} is symmetric under interchange of B and C, while Y_{ABC} is symmetric under interchange of A and B. Since C_{ABC} is totally antisymmetric, both X_{ABC} and Y_{ABC} make a vanishing contribution to the sum in Eq. (27.3.9), leaving us $\delta \mathscr{L}_{\lambda\lambda V} = 0$, so that Eq. (27.3.1) gives a supersymmetric action, as was to be shown.

We can understand *why* Eq. (27.3.1) gives a supersymmetric action by identifying the superfield that has $f_{A\mu\nu}$, λ_A, and D_A as component fields. Recall that under a generalized gauge transformation, the vector superfield

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$V_A(x,\theta)$ has the transformation property (27.1.12):

$$\exp\left(-2\sum_A t_A\, V_A(x,\theta)\right) \to \exp\left(-i\sum_A t_A \Omega_A(x,\theta)\right) \times \exp\left(-2\sum_A t_A\, V_A(x,\theta)\right) \exp\left(+i\sum_A t_A \Omega_A^*(x,\theta)\right) , \quad (27.3.10)$$

where $\Omega_A(x,\theta)$ is a general left-chiral superfield. This is not a gauge-covariant transformation rule, because $\Omega_A^* \neq \Omega_A$. To eliminate the factor involving Ω_A^*, we note that Ω_A^* is a right-chiral superfield, so that $\mathscr{D}_{L\alpha}\Omega_A^* = 0$, and therefore

$$\exp\left(-2\sum_A t_A\, V_A(x,\theta)\right) \mathscr{D}_{L\alpha} \exp\left(+2\sum_A t_A\, V_A(x,\theta)\right) \to \exp\left(-i\sum_A t_A \Omega_A(x,\theta)\right) \exp\left(-2\sum_A t_A\, V_A(x,\theta)\right) \times \mathscr{D}_{L\alpha}\left[\exp\left(+2\sum_A t_A\, V_A(x,\theta)\right) \exp\left(+i\sum_A t_A \Omega_A(x,\theta)\right)\right] . \quad (27.3.11)$$

This is still not gauge-covariant, because the left-superderivative $\mathscr{D}_{L\alpha}$ acts on $\exp(+i\sum_A t_A \Omega_A(x,\theta))$ as well as on $\exp(+2\sum_A t_A\, V_A(x,\theta))$. This is eliminated if we follow the lead of the Abelian theory discussed in the previous section, and define a spinor superfield

$$2\sum_A t_A W_{AL\alpha}(x,\theta) \equiv \sum_{\beta\gamma} \epsilon_{\beta\gamma} \mathscr{D}_{R\beta}\mathscr{D}_{R\gamma}\left[\exp\left(-2\sum_A t_A\, V_A(x,\theta)\right) \times \mathscr{D}_{L\alpha} \exp\left(+2\sum_A t_A\, V_A(x,\theta)\right)\right] . \quad (27.3.12)$$

Because the product of any three $\mathscr{D}_R$s vanishes, $W_{AL\alpha}$ is left-chiral

$$\mathscr{D}_{R\beta} W_{AL\alpha}(x,\theta) = 0 , \quad (27.3.13)$$

and because $\mathscr{D}_{R\beta}\mathscr{D}_{R\gamma}\mathscr{D}_{L\alpha}\Omega_A \propto \mathscr{D}_{R\delta}\Omega_A = 0$, $W_{AL\alpha}$ is gauge-covariant in the sense that, for a generalized gauge transformation,

$$\sum_A t_A\, W_{AL\alpha}(x,\theta) \to \exp\left(-i\sum_A t_A \Omega_A(x,\theta)\right) \sum_A t_A\, W_{AL\alpha}(x,\theta) \times \exp\left(+i\sum_A t_A \Omega_A(x,\theta)\right) . \quad (27.3.14)$$

To calculate the spinor superfield at a point $x^\mu = X^\mu$, we can again adopt a version of Wess–Zumino gauge in which $V_A(X) = 0$, and after a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

straightforward calculation find that in this gauge

$$W_{AL}(X,\theta) = \lambda_{AL}(X_+) + \tfrac{1}{2}\gamma^\mu\gamma^\nu\theta_L\left(\partial_\mu V_{A\nu}(X_+) - \partial_\nu V_{A\mu}(X_+)\right) + \left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\not{\partial}\lambda_{RA}(X_+) - i\theta_L D_A(X_+)\,.$$

Since W_{AL} is gauge-covariant, at a general point in a general gauge it must have the value

$$W_{AL}(x,\theta) = \lambda_{AL}(x_+) + \tfrac{1}{2}\gamma^\mu\gamma^\nu\theta_L f_{A\mu\nu}(x_+) + \left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\not{D}\lambda_{RA}(x_+) - i\theta_L D_A(x_+)\,. \tag{27.3.15}$$

From this, we can construct a Lorentz- and gauge-invariant $\mathscr{F}$-term bilinear in W

$$-\Big[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{AL\alpha}W_{AL\beta}\Big]_{\mathscr{F}} = \sum_A\left[-\left(\overline{\lambda_A}\not{D}(1-\gamma_5)\lambda_A\right) - \frac{1}{2}f_{A\mu\nu}f_A^{\mu\nu} + \frac{i}{4}\epsilon_{\mu\nu\rho\sigma}f_A^{\mu\nu}f_A^{\rho\sigma} + D_A^2\right]. \tag{27.3.16}$$

Just as in the previous section, the gauge-invariant Lagrangian (27.3.1) is obtained from the real part of this F-term:

$$-\frac{1}{2}\,\mathrm{Re}\Big[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{AL\alpha}W_{AL\beta}\Big]_{\mathscr{F}} = \mathscr{L}_{\text{gauge}}\,. \tag{27.3.17}$$

What about the imaginary part? This is given by

$$-\,\mathrm{Im}\Big[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{AL\alpha}W_{AL\beta}\Big]_{\mathscr{F}} = -i\sum_A\left(\overline{\lambda_A}\not{D}\gamma_5\lambda_A\right) + \frac{1}{4}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu}f_A^{\rho\sigma}\,. \tag{27.3.18}$$

Eq. (26.A.7) and the antisymmetry of the structure constants show that $(\overline{\lambda_A}\not{D}\gamma_5\lambda_A) = \frac{1}{2}\partial_\mu(\overline{\lambda_A}\gamma^\mu\gamma_5\lambda_A)$, so the first term is a total derivative, while Eq. (23.5.4) tells us that the second term is a total derivative also. In Abelian gauge theories, this means that a term like (27.3.18) would have no effect. But as discussed in Sections 23.5 and 23.6, in non-Abelian gauge theories the existence of instanton solutions allows the density (27.3.18) to have a non-vanishing integral over spacetime. We therefore must allow for the possibility of a new term in the Lagrangian density

$$\mathscr{L}_\theta = -\frac{g^2\theta}{16\pi^2}\,\mathrm{Im}\Big[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{AL\alpha}W_{AL\beta}\Big]_{\mathscr{F}}\,, \tag{27.3.19}$$

where θ is a new real parameter, and g is a gauge coupling, which can be conveniently defined for a simple gauge group so that if t_A, t_B, and t_C are in the 'standard' $SU(2)$ subalgebra of the gauge algebra used in calculating instanton effects, we have $C_{ABC} = g\,\epsilon_{ABC}$. With this definition

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

of the gauge coupling, Eq. (23.5.20) gives for simple gauge groups

$$\int d^4x\, \epsilon_{\mu\nu\rho\sigma} \sum_A f_A^{\mu\nu} f_A^{\rho\sigma} = 64\pi^2 \nu / g^2 \,, \tag{27.3.20}$$

where $\nu = 0, \pm 1, \pm 2, \ldots$ is an integer, the winding number, which characterizes the topological class of the gauge field configuration. Thus for instantons of winding number ν, the Lagrangian density $\mathscr{L}_\theta$ contributes a phase to path integrals, given by

$$\left[\exp\left(i \int d^4x\, \mathscr{L}_\theta\right)\right]_\nu = \exp(i\nu\theta) \,, \tag{27.3.21}$$

so the effects of $\mathscr{L}_\theta$ are periodic in θ, with period 2π.

It is often convenient to absorb a factor g into the gauge field, so that the structure constants do not depend on g, and instead the Lagrangian density for the gauge field is multiplied by an over-all factor $1/g^2$. In this notation, the complete Lagrangian density for the gauge field may be written in terms of the rescaled gauge fields and structure constants as

$$\mathscr{L}_{\text{gauge}} + \mathscr{L}_\theta = -\text{Re}\left[\frac{\tau}{8\pi i} \sum_{A\alpha\beta} \epsilon_{\alpha\beta} W_{AL\alpha} W_{AL\beta}\right]_{\mathscr{F}} \,, \tag{27.3.22}$$

where τ is the complex coupling parameter

$$\tau \equiv \frac{4\pi i}{g^2} + \frac{\theta}{2\pi} \,. \tag{27.3.23}$$

According to Eq. (23.5.19), the contribution of instantons of winding number ν to path integrals is suppressed by a factor $\exp(-8\pi^2|\nu|/g^2)$, which together with the factor (27.3.21) yields an over-all factor

$$\exp\left[i\nu\theta - \frac{8\pi^2|\nu|}{g^2}\right] = \begin{cases} \exp(2\pi i \nu \tau) & \nu \geq 0 \\ \exp(2\pi i \nu \tau^*) & \nu \leq 0 \end{cases} \,. \tag{27.3.24}$$

27.4 Renormalizable Gauge Theories with Chiral Superfields

We will now put together the pieces assembled in the previous three sections, to construct the most general renormalizable action for chiral superfields interacting with general gauge fields. Adding the terms (27.1.27), (27.2.7), and (27.3.1) and the superpotential terms in Eq. (26.4.5) gives the Lagrangian density

$$\mathscr{L} = \frac{1}{2}\left[\Phi^\dagger \exp\left(-2\sum_A t_A V_A\right)\Phi\right]_D - \frac{1}{2}\text{Re}\sum_A \left(W_{AL}^{\text{T}} \epsilon W_{AL}\right)_{\mathscr{F}}$$
$$- \frac{g^2\theta}{16\pi^2}\sum_A \text{Im}\left(W_{AL}^{\text{T}} \epsilon W_{AL}\right)_{\mathscr{F}} - \sum_A \xi_A [V_A]_D + 2\,\text{Re}\,[f]_{\mathscr{F}}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$
\begin{aligned}
&= -\sum_n (D_\mu \phi)_n^* (D^\mu \phi)_n - \frac{1}{2}\sum_n \left(\overline{\psi_n} \gamma^\mu (D_\mu \psi)_n \right) + \sum_n \mathscr{F}_n^* \mathscr{F}_n \\
&\quad - \mathrm{Re} \sum_{nm} \frac{\partial^2 f(\phi)}{\partial \phi_n \partial \phi_m} \left(\psi_{nL}^{\mathrm{T}} \epsilon \psi_{mL} \right) + 2\mathrm{Re} \sum_n \frac{\partial f(\phi)}{\partial \phi_n} \mathscr{F}_n \\
&\quad - 2\sqrt{2}\,\mathrm{Im} \sum_{Anm} (t_A)_{nm} \left(\overline{\psi_{nL}} \lambda_A \right) \phi_m + 2\sqrt{2}\,\mathrm{Im} \sum_{Anm} (t_A)_{mn} \left(\overline{\psi_{nR}} \lambda_A \right) \phi_m^* \\
&\quad - \sum_{Anm} \phi_n^* (t_A)_{nm} \phi_m D_A - \sum_A \xi_A D_A + \frac{1}{2} \sum_A D_A D_A \\
&\quad - \frac{1}{4} \sum_A f_{A\mu\nu} f_A^{\mu\nu} - \frac{1}{2} \sum_A \left(\overline{\lambda_A} (\not{D} \lambda)_A \right) + \frac{g^2 \theta}{64\pi^2} \epsilon_{\mu\nu\rho\sigma} \sum_A f_A^{\mu\nu} f_A^{\rho\sigma} \,. \qquad (27.4.1)
\end{aligned}
$$

Here $f(\phi)$ is the superpotential, a gauge-invariant complex function of ϕ_n (but not of ϕ_n^*) which the condition of renormalizability requires to be a cubic polynomial; the ξ_A are constants which gauge invariance requires to vanish except where t_A is a $U(1)$ generator; the gauge-covariant derivatives are

$$
D_\mu \psi_L \equiv \partial_\mu \psi_L - i \sum_A t_A V_{A\mu} \psi_L \,, \qquad (27.4.2)
$$

$$
D_\mu \phi \equiv \partial_\mu \phi - i \sum_A t_A V_{A\mu} \phi \,, \qquad (27.4.3)
$$

$$
(D_\mu \lambda)_A = \partial_\mu \lambda_A + \sum_{BC} C_{ABC} V_{B\mu} \lambda_C \,, \qquad (27.4.4)
$$

and $f_{A\mu\nu}$ is the gauge-covariant gauge field-strength tensor

$$
f_{A\mu\nu} = \partial_\mu V_{A\nu} - \partial_\nu V_{A\mu} + \sum_{BC} C_{ABC} V_{B\mu} V_{C\nu} \,. \qquad (27.4.5)
$$

The auxiliary fields enter quadratically, with field-independent constants as the coefficients of the second-order term, so they can be eliminated by setting them equal to the values at which the Lagrangian density is stationary:

$$
\mathscr{F}_n = -\left(\partial f(\phi) / \partial \phi_n \right)^* \,, \qquad (27.4.6)
$$

$$
D_A = \xi_A + \sum_{nm} \phi_n^* (t_A)_{nm} \phi_m \,. \qquad (27.4.7)
$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Using these back in Eq. (27.4.1), the Lagrangian density becomes

$$\begin{aligned}\mathscr{L} = &-\sum_n (D_\mu\phi)_n^* (D^\mu\phi)_n \\ &-\frac{1}{2}\sum_n \left(\overline{\psi_{nL}}\,\gamma^\mu (D_\mu\psi_L)_n\right) + \frac{1}{2}\sum_n \left(\overline{(D_\mu\psi_L)_n}\,\gamma^\mu\,\psi_{nL}\right) \\ &-\frac{1}{2}\sum_{nm}\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\left(\psi_{nL}^{\mathrm{T}}\epsilon\,\psi_{mL}\right) - \frac{1}{2}\sum_{nm}\left(\frac{\partial^2 f(\phi)}{\partial\phi_n\,\partial\phi_m}\right)^* \left(\psi_{nL}^{\mathrm{T}}\epsilon\,\psi_{mL}\right)^* \\ &-\sum_n \left|\frac{\partial f(\phi)}{\partial\phi_n}\right|^2 \\ &+i\sqrt{2}\sum_{Anm}\left(\overline{\psi_{nL}}\,(t_A)_{nm}\,\lambda_A\right)\phi_m - i\sqrt{2}\sum_{Anm}\phi_n^*\left(\overline{\lambda_A}\,(t_A)_{nm}\,\psi_{mL}\right) \\ &-\frac{1}{2}\sum_A\left(\xi_A + \sum_{nm}\phi_n^*\,(t_A)_{nm}\,\phi_m\right)^2 - \frac{1}{4}\sum_A f_{A\mu\nu}f_A^{\mu\nu} \\ &-\frac{1}{2}\sum_A\left(\overline{\lambda_A}\,(\not{D}\lambda)_A\right) + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu}f_A^{\rho\sigma}\,. \end{aligned} \tag{27.4.8}$$

Lorentz invariance requires the fields ψ_{nL}, λ_A, and $f_{A\mu\nu}$ to have vanishing vacuum expectation values, while the tree-value vacuum expectation values of the ϕ_n are at the minimum of the potential

$$V(\phi) = \sum_n \left|\frac{\partial f(\phi)}{\partial\phi_n}\right|^2 + \frac{1}{2}\sum_A\left(\xi_A + \sum_{nm}\phi_n^*\,(t_A)_{nm}\,\phi_m\right)^2 . \tag{27.4.9}$$

This potential is positive, so *if* there is a set of field values at which $V(\phi)$ vanishes, then this is automatically also a minimum of the potential. For $V(\phi)$ to vanish at some field value $\phi_n = \phi_{n0}$, it is necessary and sufficient that

$$\mathscr{F}_{n0} = -\left[\frac{\partial f(\phi)}{\partial\phi_n}\right]^*_{\phi=\phi_0} = 0 \tag{27.4.10}$$

and

$$D_{A0} = \xi_A + \sum_{nm}\phi_{n0}^*\,(t_A)_{nm}\,\phi_{m0} = 0\,. \tag{27.4.11}$$

This in turn is the necessary and sufficient condition for supersymmetry not to be spontaneously broken, since Eq. (26.3.15) gives $\langle\delta\psi_{nL}\rangle_{\mathrm{VAC}} = \sqrt{2}\langle\mathscr{F}_n\rangle_{\mathrm{VAC}}\,\alpha_L$, and Eq. (26.2.16) gives $\langle\delta\lambda_A\rangle_{\mathrm{VAC}} = i\langle D_A\rangle_{\mathrm{VAC}}\,\gamma_5\alpha$.

It is worth stressing here that spontaneous symmetry breaking is more difficult for supersymmetry than for other symmetries. For most symmetries of the action there will be field configurations at which the symmetry is unbroken and the potential is stationary, but the symmetry will

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

nevertheless be spontaneously broken if none of these configurations are minima of the potential. In contrast, any supersymmetric field configuration gives the potential a value of zero, which is necessarily lower than the value of the potential for any non-supersymmetric configuration, so the existence of *any* supersymmetric field configuration insures that supersymmetry is unbroken. As we will see in Section 27.6, this conclusion goes beyond the tree approximation used in this section; it is unaffected by corrections of any finite order in perturbation theory.

It may appear that Eqs. (27.4.10) and (27.4.11) impose too many conditions on the scalar fields to expect a solution, without some fine-tuning of the superpotential. However, for a gauge group of dimensionality D, the superpotential $f(\phi)$ is subject to the D constraints

$$\sum_m \frac{\partial f(\phi)}{\partial \phi_m} \left(t_A \phi\right)_m = 0 \,, \tag{27.4.12}$$

for all A and all ϕ. Hence if ϕ has N independent components, then the number of *independent* conditions (27.4.10) is $N - D$, while the number of conditions (27.4.11) is D, so there are just N conditions altogether. With the number of conditions equal to the number of free variables, it is likely to find solutions for generic superpotentials. In fact, it is more usual to find solutions than not. For instance, for chiral scalar superfields in a non-trivial representation of a semi-simple gauge group we have $\xi_A = 0$, while $f(\phi)$ can have no terms linear in the ϕ_n, so both Eqs. (27.4.10) and (27.4.11) are satisfied for $\phi_{n0} = 0$. There may be other solutions of Eqs. (27.4.10) and (27.4.11) which break the gauge symmetries, but in such a theory supersymmetry cannot be broken, at least not in the tree approximation, and, as we shall see in Section 27.6, not in any order of perturbation theory.

More generally, it is easy to see that, even if the gauge group has $U(1)$ factors and even if the superpotential involves gauge-invariant superfields, if there exists a set of scalar field values ϕ_{n0} that satisfy Eq. (27.4.10), then there is another that satisfies both Eq. (27.4.10) *and* Eq. (27.4.11), provided only that the Fayet–Iliopoulos constants ξ_A all vanish. To show this, we note that since the superpotential $f(\phi)$ does not involve ϕ^*, it is invariant not only under ordinary gauge transformations $\phi \to \exp(i \sum_A \Lambda_A t_A)\phi$ with Λ_A arbitrary real numbers, but also under transformations with Λ_A arbitrary complex numbers. Under all these transformations the $\mathscr{F}$-terms in Eq. (27.4.10) transform linearly, so if ϕ_0 satisfies Eq. (27.4.10), then so does $\phi^\Lambda \equiv \exp(i \sum_A \Lambda_A t_A)\phi_0$. On the other hand, the scalar product $[\phi^\dagger \phi]$ is not invariant under transformations with Λ_A complex, but $[\phi^{\Lambda\dagger} \phi^\Lambda]$ does remain real and positive for complex Λ_A, so it is bounded below, and therefore has a minimum. For $\xi_A = 0$, the condition that $[\phi^{\Lambda\dagger} \phi^\Lambda]$ be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

stationary at this minimum is just that ϕ^Λ should satisfy Eq. (27.4.11). We see then that in the absence of Fayet–Ilioupoulos D-terms, the question of whether supersymmetry is unbroken in gauge theories is entirely a matter of whether the superpotential allows solutions of Eq. (27.4.10). The same result applies even in non-renormalizable theories.[3]

Now let us assume that there is a set of values ϕ_{n0} at which $V(\phi_0) = 0$, so that supersymmetry is unbroken. The spin 0 degrees of freedom are described by a shifted field

$$\varphi_n = \phi_n - \phi_{n0} \,. \tag{27.4.13}$$

There is then a cross-term between φ and the gauge fields, arising from the first term in Eq. (27.4.1):

$$2\sum_{nA} \mathrm{Im}\left(\partial_\mu \varphi_n \,(t_A\phi_0)^*_n\right) V^\mu_A \,.$$

As shown in Section 21.1, it is always possible to eliminate this term by adopting a 'unitarity gauge,' in which ϕ_n satisfies a constraint that makes this term vanish:

$$\sum_n \mathrm{Im}\left(\phi_n \,(t_A\phi_0)^*_n\right) = 0 \,. \tag{27.4.14}$$

This will have the effect of eliminating the Goldstone bosons associated with broken gauge symmetries.

We shall now work out the masses of the particles of spin 0, spin 1/2, and spin 1 that arise in this theory if supersymmetry is unbroken, taking into account the possibility that the gauge symmetries may be spontaneously broken.

Spin 0

Because $\partial f(\phi)/\partial\phi_n$ and $\xi_A + \sum_{nm} \phi^*_n (t_A)_{nm}\, \phi_m$ must both vanish at $\phi_n = \phi_{n0}$, the terms in $V(\phi)$ of second order in $\varphi_n \equiv \phi_n - \phi_{n0}$ and/or φ^*_n are of the form

$$\begin{aligned} V_{\mathrm{quad}}(\phi) = &\sum_{nm} (\mathscr{M}^*\mathscr{M})_{nm} \varphi^*_n \varphi_m + \sum_{Anm} \left(t_A\phi_0\right)_n \left(t_A\phi_0\right)^*_m \varphi^*_n \varphi_m \\ &+ \frac{1}{2}\sum_{Anm} \left(t_A\phi_0\right)^*_n \left(t_A\phi_0\right)^*_m \varphi_n \varphi_m + \frac{1}{2}\sum_{Anm} \left(t_A\phi_0\right)_n \left(t_A\phi_0\right)_m \varphi^*_n \varphi^*_m \,, \end{aligned} \tag{27.4.15}$$

where $\mathscr{M}$ is the complex symmetric matrix (26.4.11):

$$\mathscr{M}_{nm} \equiv \left(\frac{\partial^2 f(\phi)}{\partial\phi_n \partial\phi_m}\right)_{\phi=\phi_0} \,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

This can be written as

$$V_{\text{quad}} = \frac{1}{2}\begin{bmatrix} \varphi \\ \varphi^* \end{bmatrix}^\dagger M_0^2 \begin{bmatrix} \varphi \\ \varphi^* \end{bmatrix}, \tag{27.4.16}$$

where M_0^2 is the block matrix

$$M_0^2 = \begin{bmatrix} \mathscr{M}^*\mathscr{M} + \sum_A (t_A\phi_0)(t_A\phi_0)^\dagger & \sum_A (t_A\phi_0)(t_A\phi_0)^{\mathrm{T}} \\ \sum_A (t_A\phi_0)^*(t_A\phi_0)^\dagger & \mathscr{M}\mathscr{M}^* + \sum_A (t_A\phi_0)^*(t_A\phi_0)^{\mathrm{T}} \end{bmatrix}. \tag{27.4.17}$$

Now we must look for the eigenvalues of this mass-squared matrix. Differentiating Eq. (27.4.12) with respect to ϕ_n gives

$$\sum_m \frac{\partial^2 f(\phi)}{\partial\phi_n\partial\phi_m}\left(t_A\phi\right)_m + \sum_m \frac{\partial f(\phi)}{\partial\phi_m}\left(t_A\right)_{mn} = 0\,. \tag{27.4.18}$$

But as we have seen, $\partial f(\phi)/\partial\phi_m$ vanishes at $\phi = \phi_0$, so by setting ϕ at this value in Eq. (27.4.18), we find

$$\sum_m \mathscr{M}_{nm}(t_A\phi_0)_m = 0\,. \tag{27.4.19}$$

It follows that

$$M_0^2\begin{bmatrix} t_B\phi_0 \\ \pm(t_B\phi_0)^* \end{bmatrix} = \sum_A \left(\phi_0^\dagger[t_A t_B \pm t_B t_A]\phi_0\right)\begin{bmatrix} t_B\phi_0 \\ \pm(t_B\phi_0)^* \end{bmatrix}.$$

But the vanishing of D_A at $\phi = \phi_0$ and the global gauge invariance of ξ_A tell us that

$$\left(\phi_0^\dagger[t_A, t_B]\phi_0\right) = i\sum_C C_{ABC}\left(\phi_0^\dagger t_C\phi_0\right) = -i\left(\phi_0^\dagger\phi_0\right)\sum_C C_{ABC}\xi_C = 0\,. \tag{27.4.20}$$

The matrix (27.4.17) therefore has a pair of eigenvectors for each gauge symmetry

$$u = \begin{bmatrix} \sum_B c_B\, t_B\phi_0 \\ \sum_B c_B\,(t_B\phi_0)^* \end{bmatrix}, \qquad v = \begin{bmatrix} \sum_B c_B\, t_B\phi_0 \\ -\sum_B c_B\,(t_B\phi_0)^* \end{bmatrix}, \tag{27.4.21}$$

for which

$$M_0^2 u = \mu^2 u\,, \qquad M_0^2 v = 0\,, \tag{27.4.22}$$

where μ^2 and c_A are any real solutions of the eigenvalue problem*

$$\sum_B \left(\phi_0^\dagger\{t_A, t_B\}\phi_0\right)c_B = \mu^2 c_A\,, \tag{27.4.23}$$

* The presence of a factor 1/2 in Eq. (21.1.17) which does not appear in Eq. (27.4.23) is due to a difference in the way that the scalar fields are normalized.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

with the exception that if the eigenvalue μ^2 vanishes then $\sum_B c_B t_B \phi_0 = 0$, so that the eigenvectors u and v are absent. The massless particles associated with the v eigenvectors are Goldstone bosons, which are eliminated from the physical spectrum by the unitarity gauge condition (27.4.14). In addition to these mass eigenstates, there is another set that are orthogonal to all the us and vs and that therefore take the form

$$w_\pm = \begin{bmatrix} \zeta \\ \pm\zeta^* \end{bmatrix} , \tag{27.4.24}$$

where

$$\sum_n (t_A \phi_0)^*_n \zeta_n = 0 . \tag{27.4.25}$$

Eq. (27.4.19) shows that the space of ζs satisfying Eq. (27.4.25) is invariant under multiplication with the Hermitian matrix $\mathscr{M}^\dagger \mathscr{M}$, so it is spanned by the eigenvectors of this matrix, satisfying

$$\mathscr{M}^\dagger \mathscr{M} \zeta = m^2 \zeta \tag{27.4.26}$$

with m^2 a set of real positive (or zero) eigenvalues. Eq. (27.4.26) and its complex conjugate together with Eq. (27.4.25) show that $w_\pm$ are eigenvectors of M_0^2 with eigenvalues m^2:

$$M_0^2 w_\pm = m^2 w_\pm . \tag{27.4.27}$$

We thus have *two* self-charge-conjugate spinless bosons of each mass m satisfying Eq. (27.4.27), and one self-charge-conjugate spinless boson of each non-zero mass μ satisfying Eq. (27.4.23).

Spin 1/2

The fermion masses arise from the non-derivative terms in Eq. (27.4.8) that are of second order in the fermion fields ψ_n and λ_A:

$$\mathscr{L}_{1/2} = -\frac{1}{2} \sum_{nm} \mathscr{M}_{nm} \left(\psi^{\mathrm{T}}_{nL} \epsilon \psi_{mL} \right) - i\sqrt{2} \sum_{Am} (t_A \phi_0)^*_m \left(\lambda^{\mathrm{T}}_{LA} \epsilon \psi_{mL} \right) + \text{H.c.} \tag{27.4.28}$$

We saw in Section 26.4 that if the fermion mass term in the Lagrangian for a column χ of Majorana spinor fields is put in the form

$$\mathscr{L}_{1/2} = -\frac{1}{2} \left(\chi^{\mathrm{T}}_L \epsilon M \chi_L \right) + \text{H.c.} , \tag{27.4.29}$$

then the fermion squared masses are the eigenvalues of the Hermitian matrix $M^\dagger M$. Eq. (27.4.28) gives the elements of the matrix M here as

$$M_{nm} = \mathscr{M}_{nm} , \quad M_{nA} = M_{An} = i\sqrt{2} (t_A \phi_0)^*_n , \quad M_{AB} = 0 , \tag{27.4.30}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

for which, using Eqs. (27.4.19) and (27.4.20),

$$\begin{aligned}(M^\dagger M)_{nm} &= (\mathscr{M}^\dagger \mathscr{M})_{nm} + 2\sum_A (t_A\phi_0)_n (t_A\phi_0)^*_m \,,\\ (M^\dagger M)_{nA} &= (M^\dagger M)_{An} = 0 \,,\\ (M^\dagger M)_{AB} &= 2(\phi_0^\dagger t_B t_A \phi_0) = (\phi_0^\dagger \{t_B \,, t_A\}\phi_0) \,.\end{aligned} \tag{27.4.31}$$

The eigenvectors of the matrix (27.4.30) are of three types. First are those of the form

$$z = \begin{bmatrix} \zeta \\ 0 \end{bmatrix} \,, \tag{27.4.32}$$

with eigenvalues m^2, where ζ_n and m^2 are any eigenvectors and corresponding eigenvalues of $\mathscr{M}^\dagger \mathscr{M}$. Next are those of the form

$$g = \begin{bmatrix} 0 \\ c \end{bmatrix} \,, \tag{27.4.33}$$

with eigenvalues μ^2, where c_B and μ^2 are any eigenvectors and eigenvalues of the matrix $(\phi_0^\dagger \{t_B \,, t_A\}\phi_0)$. Finally, there are those of the form

$$h = \begin{bmatrix} \sum_B c_B t_B \phi_0 \\ 0 \end{bmatrix} \,, \tag{27.4.34}$$

with eigenvalues μ^2, where c_B and μ^2 are again any eigenvectors and eigenvalues of the matrix $(\phi_0^\dagger \{t_B \,, t_A\}\phi_0)$. The only exception is that the eigenvectors c of this matrix with eigenvalue zero have $\sum_A c_A t_A \phi_0 = 0$, corresponding to unbroken symmetries, so that in this case the vector (27.4.34) vanishes and we have only the eigenvector (27.4.33). Thus there is one Majorana fermion of each mass m satisfying Eq. (27.4.26), two Majorana fermions of each non-zero mass μ satisfying Eq. (27.4.22), and one Majorana fermion of zero mass for each unbroken gauge symmetry.

Spin 1

The mass terms in the Lagrangian for the gauge fields arise from the part of the first term in Eq. (27.4.1) that is of second order in the gauge field V_A^μ:

$$\mathscr{L}_V = -\sum_{nAB} (t_A\phi_0)^*_n (t_B\phi_0)_n V_{A\mu} V_B^\mu \,. \tag{27.4.35}$$

Since the fields $V_{A\mu}$ are real, their mass-squared matrix is the matrix in Eq. (27.4.23):

$$(\mu^2)_{AB} = \left(\phi_0^\dagger \{t_B \,, t_A\}\phi_0\right) \,. \tag{27.4.36}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

There is one spin 1 particle of mass μ for each eigenvalue μ^2 of the matrix (27.4.36).

Putting this all together, we see that for each eigenvalue m^2 of the matrix $\mathscr{M}^*\mathscr{M}$ there are two self-charge-conjugate spinless particles and one Majorana fermion of mass m; for each non-zero eigenvalue of the matrix μ^2_{AB} there are one self-charge-conjugate spinless boson, two Majorana fermions, and one self-charge-conjugate spin 1 boson of mass μ; and for each zero eigenvalue of this matrix there is one Majorana fermion and one self-charge-conjugate spin 1 boson of zero mass. It is not surprising that the particle multiplets for each zero or non-zero mass are just the same as we found by direct use of the supersymmetry algebra in Sections 25.4 and 25.5. What *is* a bit surprising is that the masses of the gauge and chiral particles are unaffected by each other. The masses m that are given by eigenvalues of $(\mathscr{M}^*\mathscr{M})_{nm}$ and the particles with these masses are just what they would be in a theory of chiral superfields without gauge superfields, and the masses μ that are given by eigenvalues of μ^2_{AB} and the particles with these masses are just what they would be in a theory of gauge superfields with no chiral superfields.

For future use in Section 27.9 we will now apply the method described in Section 26.7 to construct the supersymmetry current for the supersymmetric gauge Lagrangian (27.4.1). In the gauge used earlier, an infinitesimal supersymmetry transformation changes V_A, λ_a, and D_A by the amounts (27.3.4)–(27.3.6). Adding the Noether supersymmetry current given by Eq. (26.7.2) for these fields to the Noether current for ϕ_n, ψ_n, and $\mathscr{F}_n$ already given in Eq. (26.7.8), with derivatives replaced by gauge-invariant derivatives, gives the total Noether supersymmetry current:

$$\begin{aligned} N^\mu = &\sum_A f_A^{\mu\nu}\gamma_\nu\lambda_A - \frac{1}{8}\sum_A f_{A\rho\sigma}[\gamma^\rho,\gamma^\sigma]\gamma^\mu\lambda_A - \frac{1}{2}i\sum_A D_A\gamma_5\gamma^\mu\lambda_A \\ &+\frac{1}{\sqrt{2}}\sum_n \Big[2(D^\mu\phi)^*_n\psi_{nL} + 2(D^\mu\phi)_n\psi_{nR} + (\not{D}\phi)_n\gamma^\mu\psi_{nR} \\ &+(\not{D}\phi)^*_n\gamma^\mu\psi_{nL} - \mathscr{F}_n\gamma^\mu\psi_{nR} - \mathscr{F}^*_n\gamma^\mu\psi_{nL}\Big]\,. \end{aligned} \tag{27.4.37}$$

This is not the supersymmetry current, because the Lagrangian density is not invariant under supersymmetry; instead, its change is the derivative

$$\delta\mathscr{L} = \partial_\mu\left(\bar{\alpha}K^\mu\right)\,, \tag{27.4.38}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

where**

$$K^\mu = \frac{1}{2} i \sum_A \epsilon^{\rho\sigma\mu\nu} f_{A\rho\sigma} \gamma_\nu \gamma_5 \lambda_A + \frac{1}{8} \sum_A [\gamma^\rho, \gamma^\sigma] \gamma^\mu \lambda_A f_{A\rho\sigma} + \frac{1}{2} i \sum_A D_A \gamma_5 \gamma^\mu \lambda_A$$
$$- i \sum_{Anm} (t_A)_{nm} \gamma_5 \gamma^\mu \lambda_A \phi_n^* \phi_m$$
$$+ \frac{1}{\sqrt{2}} \sum_n \gamma^\mu \Big[- (\not{D}\phi)_n \psi_{nR} - (\not{D}\phi)_n^* \psi_{nL} + \mathscr{F}_n^* \psi_{nL} + \mathscr{F}_n \psi_{nR}$$
$$+ 2 \left(\frac{\partial f(\phi)}{\partial \phi_n} \right) \psi_{nL} + 2 \left(\frac{\partial f(\phi)}{\partial \phi_n} \right)^* \psi_{nR} \Big] . \qquad (27.4.39)$$

The first two terms are derived using the identity (27.2.5). Using the same identity again with Eq. (26.7.4) gives the total supersymmetry current:

$$S^\mu = N^\mu + K^\mu$$
$$= -\frac{1}{4} \sum_A f_{A\rho\sigma} [\gamma^\rho, \gamma^\sigma] \gamma^\mu \lambda_A - i \sum_{Anm} (t_A)_{nm} \gamma_5 \gamma^\mu \lambda_A \phi_n^* \phi_m$$
$$+ \frac{1}{\sqrt{2}} \sum_n \Big[(\not{D}\phi)_n \gamma^\mu \psi_{nR} + (\not{D}\phi^*)_n \gamma^\mu \psi_{nL}$$
$$+ 2 \left(\frac{\partial f(\phi)}{\partial \phi_n} \right) \gamma^\mu \psi_{nL} + 2 \left(\frac{\partial f(\phi)}{\partial \phi_n} \right)^* \gamma^\mu \psi_{nR} \Big] . \qquad (27.4.40)$$

* * *

In Section 26.8 we considered a class of supersymmetric theories with a superpotential $f(\Phi)$ that has an arbitrary dependence on a set of left-chiral scalar superfields Φ_n but not their derivatives, and with a Kahler potential $K(\Phi, \Phi^*)$ that has an arbitrary dependence on the Φ_n and Φ_n^*, but not their derivatives. We can extend the same considerations to gauge theories, again with the dependence of the Lagrangian on the chiral superfields limited only by supersymmetry, but without introducing new superderivatives or spacetime derivatives. The renormalizable Lagrangian density is then replaced with

$$\mathscr{L} = \frac{1}{2} \Big[K \Big(\Phi, \Phi^\dagger \exp(-2 \sum_A t_A V_A) \Big) \Big]_D + 2 \,\mathrm{Re} \Big[f(\Phi) \Big]_{\mathscr{F}}$$
$$- \frac{1}{2} \mathrm{Re} \sum_{AB} \Big[h_{AB}(\Phi) \Big(W_{AL}^{\mathrm{T}} \epsilon W_{BL} \Big) \Big]_{\mathscr{F}} , \qquad (27.4.41)$$

** The easiest way to calculate the change in the term $[\Phi^\dagger \exp(-2\sum_A t_A V_A)\Phi]_D$ is to calculate the λ-component of $\Phi^\dagger \exp(-2\sum_A t_A V_A)\Phi$ and use Eq. (26.2.17). In this way of doing the calculation, the important term on the second line of the right-hand side of Eq. (27.4.39) arises from the λ-component of $\exp(-2\sum_A t_A V_A)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

where $h_{AB}(\Phi)$ is a new function of the Φ_n, but not of the Φ_n^* or derivatives.

The chiral gauge and scalar superfields are given by the expansions (26.3.21) and (27.3.15):

$$\begin{aligned} W_{AL}(x,\theta) &= \lambda_{AL}(x_+) + \frac{1}{2}\gamma^\mu\gamma^\nu\theta_L f_{A\mu\nu}(x_+) + \left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\not{\partial}\lambda_{AR}(x_+) \\ &\quad -i\theta_L D_A(x_+)\,, \\ \Phi_n(x,\theta) &= \phi_n(x_+) - \sqrt{2}\left(\theta_L^{\mathrm{T}}\epsilon\psi_{nL}(x_+)\right) + \mathscr{F}_n(x_+)\left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\,, \end{aligned}$$

where x_+^μ is the shifted coordinate (26.3.23). The terms of second order in θ_L (and independent of θ_R) in $\sum_{AB} h_{AB}(\Phi)(W_{AL}^{\mathrm{T}}\epsilon W_{BL})$ are then

$$\begin{aligned} &-\left[\sum_{AB} h_{AB}(\Phi)\left(W_{AL}^{\mathrm{T}}\epsilon W_{BL}\right)\right]_{\theta_L^2} = \\ &\left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\sum_{AB}\left(\lambda_{AL}^{\mathrm{T}}\epsilon\lambda_{BL}\right)\left[\frac{1}{2}\sum_{nm}\left(\psi_{nL}^{\mathrm{T}}\epsilon\psi_{mL}\right)\frac{\partial^2 h_{AB}(\phi)}{\partial\phi_n\partial\phi_m} - \sum_n \mathscr{F}_n\frac{\partial h_{AB}(\phi)}{\partial\phi_n}\right] \\ &+\left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\sum_{AB} h_{AB}(\phi)\left[-\left(\overline{\lambda_A}\not{\partial}(1-\gamma_5)\lambda_B\right) - \frac{1}{2}f_{A\mu\nu}f_B^{\mu\nu}\right. \\ &\qquad\left. +\frac{i}{4}\epsilon_{\mu\nu\rho\sigma}f_A^{\mu\nu}f_B^{\rho\sigma} + D_A D_B\right] \\ &+\sqrt{2}\sum_{ABn}\frac{\partial h_{AB}(\phi)}{\partial\phi_n}\left(\theta_L^{\mathrm{T}}\epsilon\psi_{nL}\right)\left[-\left(\lambda_{BL}^{\mathrm{T}}\epsilon\gamma^\mu\gamma^\nu\theta_L\right)f_{A\mu\nu} + 2i\left(\lambda_{BL}^{\mathrm{T}}\epsilon\theta_L\right)\right]\,, \end{aligned}$$

with all fields now understood to be evaluated at x^μ rather than x_+^μ. (The first and second terms on the right-hand side are taken from Eqs. (26.4.4) and (27.3.16), respectively.) Also, by writing $\theta_{L\alpha}\theta_{L\beta}$ as $\frac{1}{2}\epsilon_{\alpha\beta}(\theta_L^{\mathrm{T}}\epsilon\theta_L)$, the third term on the right-hand side can also be expressed as proportional to $(\theta_L^{\mathrm{T}}\epsilon\theta_L)$:

$$\begin{aligned} &\left(\theta_L^{\mathrm{T}}\epsilon\psi_{nL}\right)\left[\left(\overline{\psi_B}\gamma^\mu\gamma^\nu\theta_L\right)f_{A\mu\nu} - 2i\left(\overline{\psi_B}\theta_L\right)\right] = \\ &\qquad\frac{1}{2}\left(\theta_L^{\mathrm{T}}\epsilon\theta_L\right)\left[\left(\overline{\psi_B}\gamma^\mu\gamma^\nu\psi_{nL}\right) - 2i\left(\overline{\psi_B}\psi_{nL}\right)D_A\right]\,. \end{aligned}$$

The $\mathscr{F}$-term is the coefficient of $(\theta_L^{\mathrm{T}}\epsilon\theta_L)$, so

$$\begin{aligned} &-\left[\sum_{AB} h_{AB}(\Phi)\left(W_{AL}^{\mathrm{T}}\epsilon W_{BL}\right)\right]_{\mathscr{F}} = \\ &\qquad\sum_{AB}\left(\lambda_{AL}^{\mathrm{T}}\epsilon\lambda_{BL}\right)\left[\frac{1}{2}\sum_{nm}\left(\psi_{nL}^{\mathrm{T}}\epsilon\psi_{mL}\right)\frac{\partial^2 h_{AB}(\phi)}{\partial\phi_n\partial\phi_m} - \sum_n \mathscr{F}_n\frac{\partial h_{AB}(\phi)}{\partial\phi_n}\right] \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$+\sum_{AB} h_{AB}(\phi)\Big[-\Big(\overline{\lambda_A}\,\not{D}(1-\gamma_5)\lambda_B\Big)-\frac{1}{2}f_{A\mu\nu}f_B^{\mu\nu}+\frac{i}{4}\epsilon_{\mu\nu\rho\sigma}f_A^{\mu\nu}f_B^{\rho\sigma}$$
$$+D_AD_B\Big]$$
$$+\frac{\sqrt{2}}{2}\sum_{ABn}\frac{\partial h_{AB}(\phi)}{\partial\phi_n}\Big[-\Big(\overline{\psi_B}\gamma^\mu\gamma^\nu\psi_{nL}\Big)f_{A\mu\nu}+2i\Big(\overline{\psi_B}\psi_{nL}\Big)D_A\Big]\ .$$

The other terms in Eq. (27.4.41) are just given by the gauge-invariant version of the Lagrangian density (26.8.6). Putting this together gives the Lagrangian density

$$\mathscr{L}=\mathrm{Re}\sum_{nm}\mathscr{G}_{nm}(\phi,\phi^*)\Big[-\frac{1}{2}\Big(\overline{\psi_m}\,\not{D}(1+\gamma_5)\psi_n\Big)+\mathscr{F}_n\mathscr{F}_m^*-D_\mu\phi_n\,D^\mu\phi_m^*\Big]$$
$$-2\mathrm{Re}\sum_i\frac{\partial K(\phi,\phi^*)}{\partial\phi_i^*}D_A(\phi^*t_A)_i$$
$$+i\sqrt{2}\frac{\partial^2K(\phi,\phi^*)}{\partial\phi_i\partial\phi_j^*}\big[(t_A\phi)_i\overline{\psi_j}\lambda_{AR}-(\phi^*t_A)_j\overline{\psi_i}\lambda_{AL}\big]$$
$$-\mathrm{Re}\sum_{nml}\frac{\partial^3K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\Big(\overline{\psi_n}\psi_{mL}\Big)\mathscr{F}_\ell^*$$
$$+\mathrm{Re}\sum_{nml}\frac{\partial^3K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*}\Big(\overline{\psi_m}\gamma^\mu\psi_{\ell R}\Big)D_\mu\phi_n$$
$$+\frac{1}{4}\sum_{nmlk}\frac{\partial^4K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m\,\partial\phi_l^*\,\partial\phi_k^*}\Big(\overline{\psi_n}\psi_{mL}\Big)\Big(\overline{\psi_k}\psi_{lR}\Big)$$
$$-\mathrm{Re}\sum_{nm}\frac{\partial^2f(\phi)}{\partial\phi_n\,\partial\phi_m}\Big(\overline{\psi_n}\,\psi_{mL}\Big)+2\,\mathrm{Re}\sum_n\mathscr{F}_n\frac{\partial f(\phi)}{\partial\phi_n}$$
$$+\frac{1}{4}\mathrm{Re}\sum_{ABnm}\Big(\overline{\lambda_A}\lambda_{BL}\Big)\Big(\overline{\psi_n}\psi_{mL}\Big)\frac{\partial^2h_{AB}(\phi)}{\partial\phi_n\partial\phi_m}-\frac{1}{2}\,\mathrm{Re}\sum_{ABn}\Big(\overline{\lambda_A}\lambda_{BL}\Big)\mathscr{F}_n\frac{\partial h_{AB}(\phi)}{\partial\phi_n}$$
$$+\mathrm{Re}\sum_{AB}h_{AB}(\phi)\Big[-\Big(\overline{\lambda_A}\,\not{D}\lambda_{BR}\Big)-\frac{1}{4}f_{A\mu\nu}f_B^{\mu\nu}+\frac{1}{8}i\,\epsilon_{\mu\nu\rho\sigma}f_A^{\mu\nu}f_B^{\rho\sigma}+\frac{1}{2}D_AD_B\Big]$$
$$+\frac{\sqrt{2}}{4}\mathrm{Re}\sum_{ABn}\frac{\partial h_{AB}(\phi)}{\partial\phi_n}\Big[-\Big(\overline{\lambda_B}\gamma^\mu\gamma^\nu\psi_{nL}\Big)f_{A\mu\nu}+2i\Big(\overline{\lambda_B}\psi_{nL}\Big)D_A\Big]\ . \tag{27.4.42}$$

One interesting feature of this result is the appearance of a gaugino mass in theories with ϕ_n-dependent functions $h_{AB}(\phi)$, when supersymmetry is broken by a non-vanishing value of $\mathscr{F}_n$. This mechanism is used to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

generate gaugino masses in some theories of gravitationally mediated supersymmetry breaking, discussed in Section 31.7.

27.5 Supersymmetry Breaking in the Tree Approximation Resumed

We saw in the previous section that if the Fayet–Iliopoulos constants ξ_A all vanish and if there exists a solution of the equations $\partial f(\phi)/\partial\phi_n = 0$, then there is also some solution of these equations where the D-components of the gauge superfields all vanish, so that supersymmetry is unbroken. It follows that there are only two (non-exclusive) ways that supersymmetry can be spontaneously broken in the tree approximation in renormalizable theories of gauge and chiral superfields: the superpotential $f(\phi)$ may be arranged so that there are no solutions of all the equations $\partial f(\phi)/\partial\phi_n = 0$, or for gauge groups with $U(1)$ factors there may be Fayet–Iliopoulos terms in the action.

We have already seen in Section 26.5 how it can happen that there might not be any value of ϕ for which $\partial f(\phi)/\partial\phi_n = 0$. No change in that discussion is needed when the chiral superfields interact with gauge fields, so let's turn to the other possibility: spontaneous supersymmetry breaking produced by Fayet–Iliopoulos terms. Since this can only arise for gauge groups with $U(1)$ factors, the simplest case is a theory with a single $U(1)$ gauge group. As discussed in Section 22.4, to avoid $U(1)$-$U(1)$-$U(1)$ and $U(1)$-graviton-graviton anomalies it is necessary that the sum of the $U(1)$ quantum numbers of all left-chiral superfields and the sum of their cubes should vanish. We will consider the simplest possibility: two left-chiral superfields $\Phi_\pm$, with $U(1)$ quantum numbers $\pm e$. (This is a supersymmetric version of quantum electrodynamics, with the spinor components ψ_{-L} and ψ_{+L} of the two superfields providing the left-handed parts of the electron field and of its charge conjugate.) The most general $U(1)$-invariant superpotential in a renormalizable theory is just $f(\Phi) = m\Phi_+\Phi_-$. The scalar potential (27.4.9) for the scalar components $\phi_\pm$ of these superfields is then

$$V(\phi_+,\phi_-) = m^2|\phi_+|^2 + m^2|\phi_-|^2 + \left(\xi + e^2|\phi_+|^2 - e^2|\phi_-|^2\right)^2 . \quad (27.5.1)$$

Unless the Fayet–Iliopoulos constant ξ vanishes, it is evidently not possible to find a supersymmetric vacuum with $V = 0$. For $\xi > m^2/2e^2$ or $\xi < -m^2/2e^2$ the potential (27.5.1) has a minimum with either $\phi_+ = 0$ and $|\phi_-|^2 = (2e^2\xi - m^2)/2e^4$ or $\phi_- = 0$ and $|\phi_+|^2 = (-2e^2\xi - m^2)/2e^4$, so the $U(1)$ gauge symmetry is broken along with supersymmetry. For $|\xi| < m^2/2e^2$ the minimum of the potential is at $\phi_+ = \phi_- = 0$, so

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

here the gauge symmetry is unbroken. There is in general no necessary connection between the possible breakdown of supersymmetry and of gauge symmetries.

Whether supersymmetry is spontaneously broken by the Fayet–Iliopoulos mechanism discussed here or by the O'Raifeartaigh mechanism of Section 26.5 or by some combination of the two, supersymmetry leaves a remnant in the pattern of tree-approximation masses. Inspection of the Lagrangian (27.4.8) for a general renormalizable supersymmetric theory of gauge and chiral superfields shows that the spontaneous breakdown of supersymmetry in this theory produces the following corrections to the masses calculated in Section 27.4.

Spin 0 Masses

If the $\mathscr{F}$-terms $\mathscr{F}_n = -(\partial f(\phi)/\partial\phi_n)^*$ do not vanish at the minimum ϕ_0 of the potential, then the terms in the potential of second order in $\varphi_n \equiv \phi_n - \phi_{n0}$ have additional terms beyond those listed in Eq. (27.4.15):

$$\begin{aligned} V_{\text{quad}}(\varphi) = &\sum_{nm} (\mathscr{M}^* \mathscr{M})_{nm} \varphi_n^* \varphi_m + \sum_{Anm} \left(t_A \phi_0\right)_n \left(t_A \phi_0\right)_m^* \varphi_n^* \varphi_m \\ &+ \frac{1}{2} \sum_{Anm} \left(t_A \phi_0\right)_n^* \left(t_A \phi_0\right)_m^* \varphi_n \varphi_m + \frac{1}{2} \sum_{Anm} \left(t_A \phi_0\right)_n \left(t_A \phi_0\right)_m \varphi_n^* \varphi_m^* , \\ &+ \frac{1}{2} \sum_{nm} \mathscr{N}_{nm} \varphi_n \varphi_m + \frac{1}{2} \sum_{nm} \mathscr{N}_{nm}^* \varphi_n^* \varphi_m^* \\ &+ \sum_{Anm} D_{A0} (t_A)_{nm} \varphi_n^* \varphi_m , \end{aligned} \tag{27.5.2}$$

where $\mathscr{M}$ is again the complex symmetric matrix (26.4.11):

$$\mathscr{M}_{nm} \equiv \left(\frac{\partial^2 f(\phi)}{\partial\phi_n \partial\phi_m} \right)_{\phi=\phi_0} ,$$

$\mathscr{N}_{nm}$ is a new ingredient

$$\mathscr{N}_{nm} \equiv -\sum_\ell \mathscr{F}_{\ell 0} \left(\frac{\partial^3 f(\phi)}{\partial\phi_n \partial\phi_m \partial\phi_\ell} \right)_{\phi=\phi_0} , \tag{27.5.3}$$

and $\mathscr{F}_0$ and D_{A0} are again the $\mathscr{F}$-terms and D-terms of the chiral scalar and gauge superfields at the minimum of the potential:

$$\mathscr{F}_{n0} = -\left[\frac{\partial f(\phi)}{\partial\phi_n} \right]_{\phi=\phi_0}^* , \qquad D_{A0} = \xi_A + \sum_{nm} \phi_{n0}^* (t_A)_{nm} \phi_{m0} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

If we write the quadratic part (27.5.2) of the potential in the form (27.4.16):

$$V_{\text{quad}} = \frac{1}{2}\begin{bmatrix} \varphi \\ \varphi^* \end{bmatrix}^\dagger M_0^2 \begin{bmatrix} \varphi \\ \varphi^* \end{bmatrix},$$

then, instead of Eq. (27.4.17), we now have the scalar mass matrix

$$M_0^2 = \begin{bmatrix} \mathscr{M}^*\mathscr{M} + \mathscr{A} + \sum_A D_{A0}\, t_A & \mathscr{B} + \mathscr{N}^* \\ \mathscr{B}^* + \mathscr{N} & \mathscr{M}\mathscr{M}^* + \mathscr{A}^* + \sum_A D_{A0}\, t_A^{\mathrm{T}} \end{bmatrix}, \tag{27.5.4}$$

where

$$\mathscr{A} \equiv \sum_A (t_A\phi_0)(t_A\phi_0)^\dagger\,, \qquad \mathscr{B} \equiv \sum_A (t_A\phi_0)(t_A\phi_0)^{\mathrm{T}}\,.$$

Spin 1/2 Masses

The fermion mass matrix M here is given again by Eq. (27.4.30):

$$M_{nm} = \mathscr{M}_{nm}\,, \qquad M_{nA} = M_{An} = i\sqrt{2}(t_A\phi_0)_n^*\,, \qquad M_{AB} = 0\,.$$

However, now instead of Eq. (27.4.19), the gauge invariance condition (27.4.18) yields

$$\sum_m \mathscr{M}_{nm}(t_A\phi_0)_m = \sum_m \mathscr{F}_{m0}(t_A)_{mn}\,. \tag{27.5.5}$$

Thus the Hermitian positive matrix whose eigenvalues are the squared masses of the fermions is given by

$$\begin{aligned} (M^\dagger M)_{nm} &= (\mathscr{M}^\dagger\mathscr{M})_{nm} + 2\sum_A (t_A\phi_0)_n (t_A\phi_0)_m^*\,, \\ (M^\dagger M)_{AB} &= 2(\phi_0^\dagger t_B t_A \phi_0)\,, \\ (M^\dagger M)_{An} &= (M^\dagger M)_{nA}^* = i\sqrt{2}\sum_m \mathscr{F}_{m0}(t_A)_{mn}\,. \end{aligned} \tag{27.5.6}$$

Spin 1 Masses

The squared masses of the vector bosons are again given by the eigenvalues of the matrix (27.4.36):

$$(\mu^2)_{AB} = \left(\phi_0^\dagger, \{t_B\,, t_A\}\phi_0\right). \tag{27.5.7}$$

With the exception of the D-terms in Eq. (27.5.4), all of the changes in the mass-squared matrices are in their off-diagonal components. Therefore

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Eqs. (27.5.4), (27.5.6), and (27.5.7) yield particularly simple results for the traces of these matrices: for spin 0

$$\mathrm{Tr}\, M_0^2 = 2\mathrm{Tr}\,(\mathscr{M}^*\mathscr{M}) + \mathrm{Tr}\,\mu^2 + 2\sum_A D_{A0}\mathrm{Tr}\, t_A\,, \tag{27.5.8}$$

and for spin 1/2

$$\mathrm{Tr}\,(M^\dagger M) = \mathrm{Tr}\,(\mathscr{M}^*\mathscr{M}) + 2\mathrm{Tr}\,\mu^2\,. \tag{27.5.9}$$

Since the trace is the sum of the eigenvalues, we obtain from this a *mass sum rule*:

$$\sum_{\text{spin }0}\text{mass}^2 - 2\sum_{\text{spin }1/2}\text{mass}^2 + 3\sum_{\text{spin }1}\text{mass}^2 = -2\sum_A D_{A0}\mathrm{Tr}\, t_A\,. \tag{27.5.10}$$

The trace of t_A automatically vanishes unless t_A is a $U(1)$ generator, and, as mentioned in Section 22.4, the trace must also vanish (when taken over all left-handed fermions) for $U(1)$ gauge generators to avoid gravitational contributions to an anomaly that would violate conservation of the $U(1)$ current. Thus (27.5.10) leads to the simpler result[4]

$$\sum_{\text{spin }0}\text{mass}^2 - 2\sum_{\text{spin }1/2}\text{mass}^2 + 3\sum_{\text{spin }1}\text{mass}^2 = 0\,. \tag{27.5.11}$$

Of course, the unbroken conservation of charge, color, and baryon and lepton numbers prevents the mass matrices from having elements linking particles with different values of these quantum numbers, so all these results hold separately for each set of conserved quantum numbers.

The sum rule (27.5.11) is often quoted as providing evidence against models in which supersymmetry is spontaneously broken in the tree approximation within the minimum supersymmetric extension of the standard model. We will discuss this along with other arguments in Section 28.3.

As already observed in Section 26.5 (and discussed in greater generality in Sections 29.1 and 29.2), the spontaneous breakdown of supersymmetry necessarily entails the existence of a massless fermion, the goldstino. For renormalizable gauge theories in the tree approximation, the goldstino field g appears as a term in the spinor components ψ_n and λ_A of the chiral and gauge superfields, with coefficients given by

$$\psi_{nL} = i\sqrt{2}\,\mathscr{F}_{n0}\,g_L + \cdots\,, \qquad \lambda_{AL} = D_{A0}\,g_L + \cdots\,, \tag{27.5.12}$$

where the dots denote terms involving spinor fields of definite non-zero mass. To check this we have to confirm that $(i\sqrt{2}\mathscr{F}_{n0}\,,\, D_{A0})$ is an eigenvector of the fermion mass-squared matrix $M^\dagger M$ with eigenvalue zero. For this purpose we will need to use the condition that the potential (27.4.9)

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

is stationary at $\phi = \phi_0$:

$$0 = \left.\frac{\partial V}{\partial \phi_n}\right|_{\phi=\phi_0} = -\sum_m \mathscr{M}_{nm}\mathscr{F}_{m0} + \sum_A D_{A0}(\phi_0^\dagger t_A)_n \,. \tag{27.5.13}$$

We also need the gauge invariance condition (27.4.12), which at $\phi = \phi_0$ reads

$$\sum_n \mathscr{F}_{n0}\,(t_A\phi_0)_n = 0\,. \tag{27.5.14}$$

Combining Eqs. (27.5.13) and (27.5.14) with Eqs. (27.5.5) and (27.5.6) then yields

$$i\sqrt{2}\sum_m (M^\dagger M)_{nm}\mathscr{F}_{m0} = i\sqrt{2}\sum_A D_A(t_A\mathscr{F}_0^*)_n = -\sum_A (M^\dagger M)_{nA}D_{A0} \tag{27.5.15}$$

and

$$i\sqrt{2}\sum_m (M^\dagger M)_{Am}\mathscr{F}_{m0} = -2\sum_{nm}\mathscr{F}_{n0}\,(t_A)_{nm}\,\mathscr{F}_{m0} = -\sum_B (M^\dagger M)_{AB}\,D_{B0}\,. \tag{27.5.16}$$

That is,

$$M^\dagger M\begin{pmatrix} i\sqrt{2}\mathscr{F}_0 \\ D_0 \end{pmatrix} = 0\,, \tag{27.5.17}$$

as was to be shown.

27.6 Perturbative Non-Renormalization Theorems

From the beginning, several of the ultraviolet divergences in ordinary renormalizable quantum field theories were found to be absent in the supersymmetric versions of these theories. With the development in 1975 of supergraph techniques, in which all particles in each supermultiplet are considered together, it became possible to show that some radiative corrections are not only finite, but are absent altogether in perturbation theory.[5] Supergraphs will be described in detail in Chapter 30, but as it happens they are not needed to prove the most important non-renormalization theorems. This section will give a version of a method developed by Seiberg[6] in 1993, which showed how the non-renormalization theorems may be easily obtained from simple considerations of symmetry and analyticity.

Consider a general renormalizable supersymmetric gauge theory with a number of left-chiral superfields Φ_n and/or gauge superfields V_A. As mentioned in Section 27.3, if we remove the factor g in the t_A and C_{ABC} and include it instead in the gauge superfields, then the Lagrangian density

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

will be of the form

$$\mathscr{L} = \left[\Phi^\dagger e^{-V}\Phi\right]_D + 2\,\mathrm{Re}\left[f(\Phi)\right]_{\mathscr{F}} + \frac{1}{2g^2}\mathrm{Re}\left[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{A\alpha L}\,W_{A\beta L}\right]_{\mathscr{F}}, \tag{27.6.1}$$

where the superpotential $f(\Phi)$ is a gauge-invariant cubic polynomial in the left-chiral superfields. (We are ignoring a possible θ-term, which has no effect in perturbation theory.)

Suppose we impose an ultraviolet cut-off λ on the momenta circulating in loop graphs. As discussed in Section 12.4, we can find a *local* 'Wilsonian' effective Lagrangian density $\mathscr{L}_\lambda$ that, with this cut-off, gives precisely the same results as the original Lagrangian density for S-matrix elements of processes at momenta below λ. The effective Lagrangian density has masses and coupling parameters that now depend on λ, and usually there will be an infinite number of coupling terms in the effective Lagrangian density, all possible terms allowed by the symmetries of the theory. But things are much simpler in supersymmetric theories. The non-renormalization theorems tell us that, as long as the cut-off preserves supersymmetry and gauge invariance, to all orders in perturbation theory the effective Lagrangian will have the structure

$$\mathscr{L}_\lambda = \left[\mathscr{A}_\lambda(\Phi,\Phi^\dagger,V,\mathscr{D}\cdots)\right]_D + 2\,\mathrm{Re}\left[f(\Phi)\right]_{\mathscr{F}} + \frac{1}{2g_\lambda^2}\mathrm{Re}\left[\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{A\alpha L}\,W_{A\beta L}\right]_{\mathscr{F}}, \tag{27.6.2}$$

where $\mathscr{A}_\lambda$ is a general Lorentz- and gauge-invariant function; '$\mathscr{D}\cdots$' denotes terms involving superderivatives or spacetime derivatives of the preceding arguments; and g_λ is the *one-loop* effective gauge coupling, given by the same formula as the one-loop renormalized gauge coupling constant

$$g_\lambda^{-2} = \text{constant} - 2b\ln\lambda\,, \tag{27.6.3}$$

where b is the coefficient of g^3 in the Gell-Mann–Low function $\beta(g)$, discussed in Chapter 18. This is for a simple gauge group, with a single gauge coupling, but the extension to a direct product of simple and $U(1)$ gauge groups is trivial. Note in particular that the effective superpotential is not only finite in the limit $\lambda \to \infty$, but at least in perturbation theory it contains no terms beyond those in the original superpotential, and there is no change in the coefficients of the terms it does contain.

To prove this theorem, we shall interpret this theory as a special case of one with two additional external gauge-invariant left-chiral superfields

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

X and Y, with Lagrangian density

$$\mathscr{L}^{\sharp} = \frac{1}{2}\left[\Phi^{\dagger} e^{-V} \Phi\right]_D + 2\mathrm{Re}\left[Y f(\Phi)\right]_{\mathscr{F}} + \frac{1}{2}\mathrm{Re}\left[X \sum_{A\alpha\beta} \epsilon_{\alpha\beta} W_{A\alpha L} W_{A\beta L}\right]_{\mathscr{F}} . \tag{27.6.4}$$

This Lagrangian density becomes equal to the original one when the scalar components x and y of X and Y are given the values $x = 1/g^2$ and $y = 1$, and the spinor and auxiliary components of X and Y are set equal to zero. Since supersymmetry and gauge invariance are assumed to be preserved in the cut-off procedure, the effective Lagrangian density in the presence of these external superfields must be the sum of the D-term of a general superfield and the real part of the $\mathscr{F}$-term of a left-chiral superfield:

$$\mathscr{L}^{\sharp}_{\lambda} = \left[\mathscr{A}_{\lambda}(\Phi, \Phi^{\dagger}, V, X, X^{\dagger}, Y, Y^{\dagger}, \mathscr{D}\cdots)\right]_D + 2\,\mathrm{Re}\left[\mathscr{B}_{\lambda}(\Phi,\, W_L, X, Y)\right]_{\mathscr{F}} , \tag{27.6.5}$$

with $\mathscr{A}_{\lambda}$ and $\mathscr{B}_{\lambda}$ both gauge-invariant functions of the displayed arguments. We do not include any superderivatives or spacetime derivatives in the $\mathscr{F}$-term because, as in Section 26.3, terms involving derivatives of any of the left-chiral superfields or their adjoints may be rewritten as contributions to $[\mathscr{A}_{\lambda}]_D$. (It is true that Eq. (27.3.12) gives W_L itself in terms of two $\mathscr{D}_R$s acting on a superfield $\exp(-2V)\mathscr{D}_L \exp(2V)$, but this superfield is not gauge-invariant, and we are requiring that $\mathscr{A}_{\lambda}$ be gauge-invariant.)

The dependence of $\mathscr{B}_{\lambda}$ on X and Y is severely limited by two additional symmetries of the action obtained from the Lagrangian density (27.6.4). (Both of these symmetries are broken by non-perturbative effects, which will be considered in Chapter 29.) The first symmetry is a perturbative $U(1)$ R-symmetry, of the sort discussed in Section 26.3, for which θ_L and θ_R are given the R values $+1$ and -1, the superfields Φ, V, and X are R-neutral, and Y has the R value $+2$. (Recall that $f_{\mathscr{F}}$ is the coefficient of θ_L^2 in f, so in order for $f_{\mathscr{F}}$ to have the R value 0, f must have the R value 2.) Because W_L is given by two $\mathscr{D}_R$s and one $\mathscr{D}_L$ acting on R-neutral superfields, it has the R value $+1$. Now, R invariance requires $\mathscr{B}_{\lambda}$, like the superpotential, to have the R value $+2$. It cannot depend on any superfields with negative R values, such as adjoints of left-chiral superfields, because it is *holomorphic*, so the terms in $\mathscr{B}_{\lambda}$ can only be of first order in Y or of second order in W_L, with coefficients depending only on the R-neutral superfields Φ and/or X:

$$\mathscr{B}_{\lambda}(\Phi, W_L, X, Y) = Y f_{\lambda}(\Phi, X) + \sum_{\alpha\beta AB} \epsilon_{\alpha\beta} W_{A\alpha L} W_{B\beta L} h_{\lambda AB}(\Phi, X) . \tag{27.6.6}$$

(Lorentz invariance requires the spinor indices on the W_Ls to be contracted with $\epsilon_{\alpha\beta}$.) The other symmetry is translation of X by an imaginary numerical constant, $X \to X + i\xi$, with ξ real. This changes the Lagrangian

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

density (27.6.4) by an amount proportional to Im $\sum_{A\alpha\beta} W_{A\alpha L} W_{A\beta L}$, which as we saw in Section 27.3 is a spacetime derivative, and therefore can have no effect in perturbation theory. This translation symmetry prevents X from appearing anywhere in the effective Lagrangian density (27.6.5) except where it appears in the original Lagrangian density (27.6.4). We therefore conclude that f_λ is independent of X, while $h_{\lambda AB}$ consists of a Φ-independent term proportional to $X\delta_{AB}$, plus a term that is independent of X. That is,

$$\mathscr{B}_\lambda(\Phi, W_L, X, Y) = Y\, f_\lambda(\Phi) + \sum_{\alpha\beta AB} \epsilon_{\alpha\beta} W_{A\alpha L} W_{B\beta L} \left[c_\lambda \delta_{AB} X + \ell_{\lambda AB}(\Phi) \right], \tag{27.6.7}$$

where c_λ is a real cut-off-dependent constant.

The point of introducing the external auxiliary superfields X and Y is that, by giving them suitable values, we can make use of weak-coupling approximations to determine the coefficients in Eq. (27.6.7). If we set the spinor and auxiliary components of X and Y equal to zero, and take their scalar components x and y to approach infinity and zero, respectively, then the gauge coupling constant vanishes as $1/\sqrt{x}$, and all Yukawa and scalar couplings derived from the superpotential vanish as y. In this limit, the only graph that contributes to the term in (27.6.7) proportional to Y has a single vertex arising from the term $2\,\mathrm{Re}\,[Y f(\Phi)]_{\mathscr{F}}$ in Eq. (27.6.4), so

$$f_\lambda(\Phi) = f(\Phi)\,. \tag{27.6.8}$$

Also, with $Y = 0$ there is a conservation law which requires every term in $\mathscr{L}_\lambda^\sharp$ to have equal numbers of Φs and $\Phi^\dagger$s, so since $\Phi^\dagger$ cannot occur in $\ell_{\lambda AB}$, neither can Φ. Gauge invariance then requires the constant $\ell_{\lambda AB}$ to be proportional to δ_{AB} for a simple group:

$$\ell_{\lambda AB} = \delta_{AB} L_\lambda\,. \tag{27.6.9}$$

Now, since gauge propagators go as $1/x$ while pure gauge interactions go as x and scalar propagators and interactions are x-independent, with $y = 0$ the number of powers of x in a diagram with V_W pure gauge boson vertices, I_W internal gauge boson lines, and any number of scalar-gauge boson vertices and scalar propagators is

$$N_x = V_W - I_W\,. \tag{27.6.10}$$

The number of loops is given by

$$L = I_W + I_\Phi - V_W - V_\Phi + 1\,, \tag{27.6.11}$$

where I_Φ is the number of internal Φ lines and V_Φ is the number of Φ–V interaction vertices. All the Φ–V vertices have two Φ lines attached, so with no external Φ lines I_Φ and V_Φ are equal, and therefore cancel in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Eq. (27.6.11), so that Eq. (27.6.10) may be written

$$N_x = 1 - L\,. \tag{27.6.12}$$

Thus the coefficient c_λ of X in Eq. (27.6.7) is correctly given by the tree approximation and is therefore what it was in the original Lagrangian, simply $c_\lambda = 1$, while the coefficient L_λ of the X-independent term is given by one-loop diagrams only. Putting this all together, we have

$$\mathscr{L}^\sharp_\lambda = \left[\mathscr{A}_\lambda(\Phi, \Phi^\dagger, V, X, X^\dagger, Y, Y^\dagger, \mathscr{D}\cdots)\right]_D + 2\,\mathrm{Re}\left[Y f(\Phi)\right]_{\mathscr{F}} + \frac{1}{2}\mathrm{Re}\left[\left(X + L_\lambda\right)\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{A\alpha L}W_{A\beta L}\right]_{\mathscr{F}}, \tag{27.6.13}$$

where L_λ is the one-loop contribution. Setting $Y = 1$ and $X = 1/g^2$ then gives Eq. (27.6.2), with $g_\lambda^{-2} = g^{-2} + L_\lambda$. As shown in Section 18.3, the leading order contribution to $\lambda dg_\lambda/d\lambda$ is the same function of g_λ whatever renormalization scheme is used to define this coupling, so to one-loop order we must have

$$\lambda\, dg_\lambda/d\lambda = b\, g_\lambda^3\,, \tag{27.6.14}$$

where b is the same coefficient of g^3 as in the renormalization group equation of Gell-Mann and Low. The solution is Eq. (27.6.3), completing the proof.

In theories with a $U(1)$ gauge superfield V_1, the Lagrangian may contain a Fayet–Iliopoulos term (27.2.7):

$$\mathscr{L}_{\mathrm{FI}} = \xi\left[V_1\right]_D\,. \tag{27.6.15}$$

It is easy to see that the coefficient ξ of such a term is not renormalized.[7] If the corresponding coefficient ξ_λ in the Wilsonian Lagrangian density did depend on the gauge couplings or the couplings in the superpotential, then when we replace the original Lagrangian (27.6.1) with a Lagrangian (27.6.4) involving the external superfields X and Y, this term in the Wilsonian Lagrangian would be required by supersymmetry to take the form

$$\mathscr{L}^\sharp_{\mathrm{FI}\,\lambda} = \left[\xi_\lambda(X, Y, X^*, Y^*)\, V_1\right]_D\,, \tag{27.6.16}$$

with ξ_λ a function with a non-trivial dependence on X and/or Y and/or their adjoints. But such a term would not be gauge-invariant, because, according to Eq. (27.2.18), a gauge transformation shifts V_1 by a chiral superfield $i(\Omega - \Omega^*)/2$, and although the D-term of a chiral superfield vanishes, the product of $i(\Omega - \Omega^*)/2$ and ξ_λ is not chiral for general gauge transformations if ξ_λ has any dependence on other superfields. There actually are diagrams that make contributions to ξ_λ that are independent

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

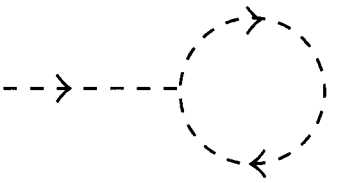

Figure 27.1. A one-loop diagram that could be quadratically divergent in theories with supersymmetry broken by trilinear couplings among scalar fields and their adjoints. The lines all represent complex scalar fields.

of all coupling constants. For the Lagrangian (27.6.1) there are no factors of the gauge coupling g at vertices at which the gauge superfield interacts with chiral matter, but instead a factor g^{-2} for each gauge propagator, so a graph with no internal gauge lines and no self-couplings of the chiral superfield will have no dependence on coupling constants. The only such graphs that contribute to ξ_λ are those in which a single external gauge line is attached to a chiral loop. (See Figure 27.1.) The contribution of all such graphs is proportional to the sum of the gauge couplings of all chiral superfields — that is, to the trace of the $U(1)$ generator. But as discussed in Section 22.4, this trace must vanish (if the $U(1)$ symmetry is unbroken) in order to avoid gravitational anomalies that violate the conservation of the $U(1)$ current.

The most important application of these theorems is a corollary, which tells us that if there is no Fayet–Iliopoulos term and if the superpotential $f(\Phi)$ allows solutions of the equations $\partial f(\phi)/\partial\phi_n = 0$, then supersymmetry is not broken in any finite order of perturbation theory.

To test this we must examine Lorentz-invariant field configurations, in which the Φ_n have only constant scalar components ϕ_n and constant auxiliary components $\mathscr{F}_n$, while (in Wess–Zumino gauge) the coefficients V_A of the gauge generators t_A in the matrix gauge superfield V have only auxiliary components D_A. Supersymmetry is unbroken if there are values of ϕ_n for which $\mathscr{L}_\lambda$ has no terms of first order in $\mathscr{F}_n$ or D_A, in which case there is sure to be an equilibrium solution with $\mathscr{F}_n = D_A = 0$. (In Section 29.2 we will see that this is the sufficient as well as the necessary condition for supersymmetry to be unbroken.) In the absence of Fayet–Iliopoulos terms, this will be the case if for all A

$$\sum_{nm} \frac{\partial K_\lambda(\phi,\phi^*)}{\partial\phi_n^*}(t_A)_{mn}\phi_m^* = 0 \tag{27.6.17}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

and for all n

$$\frac{\partial f(\phi)}{\partial \phi_n} = 0\,, \tag{27.6.18}$$

where the effective Kahler potential $K_\lambda(\phi, \phi^*)$ is

$$K_\lambda(\phi, \phi^*) = \mathscr{A}_\lambda(\phi, \phi^*, 0, 0 \cdots)\,, \tag{27.6.19}$$

with $\mathscr{A}_\lambda(\phi, \phi^*, 0, 0 \cdots)$ obtained from $\mathscr{A}_\lambda$ by setting the gauge superfield and all superderivatives equal to zero. (With superderivatives required to vanish by Lorentz invariance, the only dependence of $\mathscr{A}_\lambda$ on V is a factor $\exp(-V)$ following every factor $\Phi^\dagger$.) We now use a trick that we have already employed in Section 27.4. If there is any solution $\phi^{(0)}$ of Eq. (27.6.18), then the gauge symmetry tells us that there is a continuum of such solutions, with ϕ_n replaced with

$$\phi_n(z) = \left[\exp(i \sum_A t_A z_A)\right]_{nm} \phi_m^{(0)}\,, \tag{27.6.20}$$

where (since f depends only on ϕ, not ϕ^*) the z_A are an arbitrary set of *complex* parameters. If $K_\lambda(\phi, \phi^*)$ has a stationary point anywhere on the surface $\phi = \phi(z)$, then at that point

$$0 = \sum_{nmA} \frac{\partial K_\lambda(\phi, \phi^*)}{\partial \phi_n} (t_A)_{nm} \phi_m \, \delta z_A - \sum_{nmA} \frac{\partial K_\lambda(\phi, \phi^*)}{\partial \phi_n^*} (t_A)_{mn} \phi_m^* \, \delta z_A^*\,. \tag{27.6.21}$$

Since this must be satisfied for all infinitesimal *complex* δz_A, the coefficients of both δz_A and δz_A^* must both vanish, and therefore Eq. (27.6.17) as well as Eq. (27.6.18) is satisfied at this point. Thus the existence of a stationary point of $K_\lambda(\phi, \phi^*)$ on the surface $\phi = \phi(z)$ would imply that supersymmetry is unbroken to all orders of perturbation theory. The zeroth-order Kahler potential $(\phi^\dagger \phi)$ is bounded below and goes to infinity as $\phi \to \infty$, so it certainly has a minimum on the surface $\phi = \phi(z)$, where of course it is stationary. If there were no flat directions in which K_λ is constant at this minimum then any sufficiently small perturbation to the Kahler potential might shift the minimum, but would not destroy it. At the minimum of the Kahler potential on the surface $\phi = \phi(z)$ there are flat directions: ordinary global gauge transformations $\delta\phi = i \sum_A \delta z_A t_A \phi$ with z_A real. But these are also flat directions for the perturbation $K_\lambda(\phi, \phi^*) - (\phi^\dagger \phi)$, so there is still a local minimum of K_λ on the surface $\phi = \phi(z)$ for any perturbation in at least a finite range, and thus to all orders in whatever couplings appear in $K_\lambda(\phi, \phi^*)$. As we have seen, this is a set of scalar field values at which $\mathscr{F}_n = 0$ and $D_A = 0$ for all n and A, which means that supersymmetry is unbroken.

* * *

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

These results may be extended to non-renormalizable theories.[3] In such theories the first term $[\Phi^\dagger e^{-V}\Phi]_D$ in Eq. (27.6.1) is replaced with the D-term of an arbitrary real gauge-invariant scalar function of $\Phi^\dagger$, Φ, V, and their superderivatives and spacetime derivatives, while the second and third terms in Eq. (27.6.1) are replaced with the $\mathcal{F}$-term of an arbitrary globally gauge-invariant scalar function $f(\Phi, W)$ of Φ_n and W_α. It has been shown that, to all orders of perturbation theory, the function $f_\lambda(\Phi, W)$ appearing in the $\mathcal{F}$-term of the Wilsonian Lagrangian is the same as $f(\Phi, W)$, except for the one-loop renormalization of the term quadratic in W.

27.7 Soft Supersymmetry Breaking*

We will see in the next chapter that, even if supersymmetry is an exact symmetry of the action, the spontaneous breakdown of supersymmetry at very high energy can produce superrenormalizable terms that violate supersymmetry conservation in the effective action that describes physics at lower energies. These superrenormalizable terms may explain the lack of supersymmetry observed in phenomena at accessible energies. In this section we will consider the radiative corrections that can be produced by such supersymmetry-breaking superrenormalizable terms, in part to see whether this provides a criterion for including or rejecting such terms in supersymmetric versions of the standard model.

The sign of supersymmetry breaking is the appearance of expectation values of D-terms of general superfields or $\mathcal{F}$-terms of chiral superfields. Any operator $\epsilon\mathcal{O}$ in the Lagrangian density that breaks supersymmetry can be written in a supersymmetric form, as a D-term

$$\epsilon\,\mathcal{O} = \left[Z\,S\right]_D\,, \tag{27.7.1}$$

where S is a non-chiral superfield that has $\mathcal{O}$ as its C-term, and Z is a non-chiral external superfield whose only non-vanishing component is $[Z]_D = \epsilon$. Some but not all operators $\epsilon\mathcal{O}$ that break supersymmetry can also be written as $\mathcal{F}$-terms,

$$\epsilon\,\mathcal{O} = \left[\Omega\,O\right]_{\mathcal{F}}\,, \tag{27.7.2}$$

or their adjoints, where O is the left-chiral superfield whose $\mathcal{F}$-term is $\mathcal{O}$ and Ω is an external left-chiral superfield whose only non-vanishing component is $[\Omega]_{\mathcal{F}} = \epsilon$. We can count the order in ϵ in which a given

* This section lies somewhat out of the book's main line of development and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

correction to the effective Lagrangian will occur by counting the powers of Z or Ω needed to construct this correction in a supersymmetric way. We will find interesting limitations on the radiative corrections that can be produced by those interactions that can be written in the form (27.7.2) as well as (27.7.1).

According to the results of the previous section, there are no radiative corrections to $\mathscr{F}$-terms, so all supersymmetry-breaking radiative corrections to the Wilsonian Lagrangian density must take the form of D-terms. This theorem does not prevent any given operator from appearing in the Wilsonian Lagrangian density, because even if an operator $\epsilon\Delta\mathscr{L}$ cannot be expressed in the form $[Z\Lambda]_D$, where Λ is the general superfield whose C-term is $\Delta\mathscr{L}$, yet $\epsilon^2\Delta\mathscr{L}$ *can* be expressed in the form

$$\epsilon^2\Delta\mathscr{L} = 2\left[\Omega^*\Omega\Lambda\right]_D . \tag{27.7.3}$$

But not all operators can be produced by radiative corrections of *first order* in Ω or Ω^*. In particular, a function only of the ϕ-term of a left-chiral superfield Φ, but not of ϕ^*, cannot be written as the D-term of a superfield linear in Ω. (Note that $[\Omega h(\Phi)]_D$ is a derivative, while $[\Omega^* h(\Phi)]_D = 2[\Phi]_{\mathscr{F}}\partial h(\phi)/\partial\phi$ is not a function of ϕ alone.) We conclude then that *the supersymmetry-breaking terms in the Wilsonian Lagrangian that depend on ϕ alone cannot be produced by radiative corrections that are of first order in supersymmetry-breaking interactions of the form (27.7.2).*

This result is significant because the most divergent radiative corrections are those that are of lowest order in superrenormalizable couplings. To be more specific, the coefficient of an interaction of dimensionality $\mathscr{D}$ has dimensionality (in powers of energy) $4 - \mathscr{D}$, so dimensional analysis indicates that the contribution of a set of interactions of dimensionality d_1, d_2, etc. to the coefficient of an interaction of dimensionality d can contain the ultraviolet cut-off to at most a power

$$p = 4 - d - (4 - d_1) - (4 - d_2) - \cdots \tag{27.7.4}$$

and is therefore finite if $p < 0$. (This argument ignores possible ultraviolet divergences in subintegrations; for a thorough treatment of this topic, see Reference 8.) Superrenormalizable interactions are 'soft,' in the sense that they reduce the degree of divergence of the graphs in which they appear. In particular, in a renormalizable theory where all interactions have $d_i \leq 4$, and the strictly renormalizable interactions with $d_i = 4$ are supersymmetric, the contribution of one or more superrenormalizable interactions to the coefficient of an interaction with $d = 4$ will always have $p < 0$, so even if they are not supersymmetric the superrenormalizable interactions will not produce supersymmetry-violating ultraviolet-divergent corrections to the coefficients of the supersymmetric $d = 4$ interactions.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

On the other hand, in such a theory there may be divergent radiative corrections to the superrenormalizable interactions themselves.[9] The most worrisome are quadratic (or higher) divergences, which if cut off at some high energy scale M_X may require a fine-tuning of the bare coupling constants to preserve supersymmetry as a good approximate symmetry at energies below M_X. According to Eq. (27.7.4), in renormalizable theories where all interactions with $d_i = 4$ are supersymmetric, radiative corrections can produce quadratic or more highly divergent ($p \geq 2$) supersymmetry-violating operators of dimensionality d only if they involve insertion of a superrenormalizable supersymmetry-violating interaction of dimensionality $d_1 \geq 2 + d$. This allows either $d = 0$ and $d_1 \geq 2$, which arises only when we calculate the cosmological constant, or $d = 1$ and $d_1 = 3$, which arises only when we calculate the 'tadpole' graphs in which a scalar field line disappears into the vacuum. The cosmological constant raises fine-tuning problems for all known theories[10] and will not be considered further here. The tadpole graphs represent operators linear in ϕ or ϕ^* and, as we have seen, can not be produced to first order in supersymmetry-violating interactions that can be put in the form (27.7.2). Such superrenormalizable interactions are therefore 'soft' in the sense that they do not induce quadratic or higher divergences. Along with the superrenormalizable interactions with $d \leq 2$, including arbitrary quadratic polynomials in ϕ and ϕ^*, the supersymmetry-breaking interactions that are soft in this sense include terms of third order in the ϕ, which can be expressed as $\phi^3 = [\Omega\Phi^3]_{\mathscr{F}}$, and likewise terms of third order in the ϕ^*, and also $d = 3$ gaugino mass terms, which can be written as $[\Omega\, \epsilon_{\alpha\beta} W_\alpha W_\beta]_{\mathscr{F}}$, but not terms like $\phi^2\phi^*$ or $\phi\phi^{2*}$, which in general *can* produce quadratically divergent tadpole graphs.[9]

Nevertheless, tadpoles can only arise for scalar fields that are neutral with respect to all exact symmetries. In theories without such neutral scalars, like the supersymmetric standard model discussed in the next chapter, *all* superrenormalizable interactions may be considered to be soft.

27.8 Another Approach: Gauge-Invariant Supersymmetry Transformations*

It is somewhat disturbing that the supersymmetry transformation rules discussed so far involve ordinary spacetime derivatives, not gauge-invariant

* This section lies somewhat out of the book's main line of development and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

derivatives. For instance, in a $U(1)$ gauge theory, the transformation of the component fields of a chiral scalar superfield is given by Eqs. (26.3.15)–(26.3.17) as

$$\begin{aligned}\delta\psi_L &= \sqrt{2}\partial_\mu\phi\,\gamma^\mu\,\alpha_R\,\phi + \sqrt{2}\mathcal{F}\,\alpha_L\,,\\ \delta\mathcal{F} &= \sqrt{2}\left(\overline{\alpha_L}\,\not{\partial}\psi_L\right)\,,\\ \delta\phi &= \sqrt{2}\left(\overline{\alpha_R}\psi_L\right)\,.\end{aligned} \tag{27.8.1}$$

One might have thought that in the transformation of a chiral superfield that carries $U(1)$ charge q, the ordinary spacetime derivatives in Eq. (27.8.1) should be replaced with gauge-covariant derivatives, given in terms of the $U(1)$ gauge field V_μ by

$$D_\mu = \partial_\mu - iqV_\mu\,. \tag{27.8.2}$$

With such gauge-invariant supersymmetry transformations of the chiral superfields, one would still attempt to formulate supersymmetry transformations of the gauge supermultiplet that involve only the physical and auxiliary fields V_μ, λ, and D:

$$\begin{aligned}\tilde{\delta}V_\mu &= \left(\bar{\alpha}\gamma_\mu\lambda\right)\,,\\ \tilde{\delta}\lambda &= iD\gamma_5\alpha + \frac{1}{2}\left[\partial_\mu\not{V}\,,\,\gamma^\mu\right]\alpha\,,\\ \tilde{\delta}D &= i\left(\bar{\alpha}\gamma_5\,\not{\partial}\lambda\right)\,,\end{aligned} \tag{27.8.3}$$

in which ordinary spacetime derivatives appear because the gauge superfield carries no $U(1)$ charge.

This doesn't work. The algebra of these transformations does not close: the commutator of two of the modified supersymmetry transformations is not a linear combination of bosonic symmetry transformations, such as spacetime translations and gauge transformations. It follows that it is not possible to construct a Lagrangian for the chiral and gauge superfields that would be invariant under these modified supersymmetry transformations, because if there were such a Lagrangian then it would have to be also invariant under the commutators of these transformations, so that these commutators would have to be bosonic symmetries of the Lagrangian.

In 1973 de Wit and Freedman[11] showed that the supersymmetry algebra could be made to close by modifying the supersymmetry transformation properties of the chiral superfields not only by changing ordinary derivatives to gauge-invariant derivatives, but by also adding an extra term in the transformation of the $\mathcal{F}$-component, so that for $U(1)$ gauge theories

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

the modified supersymmetry transformation rules read

$$\begin{aligned}\tilde{\delta}\psi_L &= \sqrt{2}D_\mu\phi\,\gamma^\mu\,\alpha_R\,\phi + \sqrt{2}\mathscr{F}\,\alpha_L\,,\\ \tilde{\delta}\mathscr{F} &= \sqrt{2}\left(\overline{\alpha_L}\,\not{D}\psi_L\right) - 2\,i\,q\,\phi\left(\overline{\alpha_L}\lambda_R\right),\\ \tilde{\delta}\phi &= \sqrt{2}\left(\overline{\alpha_R}\psi_L\right).\end{aligned} \tag{27.8.4}$$

With this change, they were also able to construct a Lagrangian that is invariant under the transformations (27.8.3)–(27.8.4) and that turned out to be just the one we have found in Sections 27.1 and 27.2.

There is nothing wrong with continuing to use the conventional transformation rules (27.8.1), so we do not need the de Wit–Freedman formalism to deal with supersymmetric gauge theories. Nevertheless this formalism is of some interest, because in supergravity theories the analog of the conventional formalism is very cumbersome. As described in Chapter 31, the formalism that has been chiefly used to derive physically interesting results in supergravity theories follows an approach like that of de Wit and Freedman, with supersymmetry transformation rules involving covariant derivatives in place of ordinary derivatives, rather than an approach based on conventional supersymmetry transformations like those of Eq. (27.8.1). It is therefore of some interest to understand the relation between the de Wit–Freedman formalism and the conventional approach in the relatively simple context of $U(1)$ gauge theory, and in particular to explain the origin of the extra term in the transformation rule for $\mathscr{F}$.

In writing supersymmetry transformations (27.8.3) that did not involve the components C, M, N, or ω of the gauge superfield V, de Wit and Freedman were implicitly adopting the Wess–Zumino gauge discussed in Section 27.1. But the choice of Wess–Zumino gauge is not invariant under either the conventional supersymmetry transformations (26.2.11)–(26.2.17) or the extended gauge transformations (27.1.17), so once we adopt this gauge both symmetries are lost. We can, however, define a *combined* transformation, acting on fields in Wess–Zumino gauge, which consists of a conventional supersymmetry transformation followed by an extended gauge transformation that takes us back to Wess–Zumino gauge. *This is the de Wit–Freedman transformation* $\tilde{\delta}$.**

To construct the de Wit–Freedman transformations in this way, note that for a gauge superfield that satisfies the Wess–Zumino gauge conditions

** This was not shown explicitly by de Wit and Freedman. But, in fact, although the point of their paper was to emphasize that the details of the transformations (27.8.3) and (27.8.4) could be inferred from the requirement of a closed supersymmetry algebra (also for non-Abelian gauge theories), they remarked that they had actually found these transformations by identifying the fermionic transformations that survive in Wess–Zumino gauge.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$C = M = N = \omega = 0$, the transformation rules (26.2.11)–(26.2.14) give

$$\delta C = 0\,, \quad \delta\omega = \not{\nabla}\alpha\,, \quad \delta M = -\left(\bar{\alpha}\lambda\right), \quad \delta N = i\left(\bar{\alpha}\gamma_5\lambda\right). \tag{27.8.5}$$

According to Eqs. (27.1.17), we can get back to Wess–Zumino gauge by performing an infinitesimal extended gauge transformation (27.1.13):

$$V \to V + \frac{i}{2}\left[\Omega - \Omega^*\right], \tag{27.8.6}$$

where Ω is a left-chiral superfield with components

$$\phi^\Omega = 0\,, \quad \psi_L^\Omega = -\sqrt{2}\not{\nabla}\alpha_R\,, \quad \mathscr{F}^\Omega = -\left(\bar{\alpha}(1-\gamma_5)\lambda\right). \tag{27.8.7}$$

According to Eq. (27.1.11), this extended gauge transformation induces on a chiral superfield of charge q the transformation

$$\delta'\Phi = iq\Omega\Phi\,. \tag{27.8.8}$$

Using the multiplication rules (26.3.27)–(26.3.29), the transformation of the components of Φ is

$$\begin{aligned}
&\delta'\psi_L = -i\sqrt{2}q\phi\not{\nabla}\alpha_R\,,\\
&\delta'\mathscr{F} = -2iq\phi\left(\overline{\alpha_L}\lambda_R\right) - i\sqrt{2}q\left(\overline{\alpha_L}\not{\nabla}\psi_L\right),\\
&\delta'\phi = 0\,.
\end{aligned} \tag{27.8.9}$$

Adding this to Eq. (27.8.1) and comparing with Eq. (27.8.4) shows that the de Wit–Freedman transformation is indeed the combination of a conventional supersymmetry transformation and the corresponding extended gauge transformation (27.8.8):

$$\tilde{\delta}\Phi = \delta\Phi + \delta'\Phi\,. \tag{27.8.10}$$

27.9 Gauge Theories with Extended Supersymmetry*

Theories with unbroken extended supersymmetry are not considered to be good candidates for realistic extensions of the standard model, because of the non-chirality of particle multiplets discussed in Section 25.4. Nevertheless gauge theories with extended supersymmetry are worth some consideration here because they have provided paradigms for the use of powerful mathematical methods to solve dynamical problems.

There are a number of special formalisms that have been proposed to construct Lagrangians with $N = 2$ extended supersymmetry,[12] but

* This section lies somewhat out of the book's main line of development and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

fortunately we can get by with the tools already at hand. Any theory with $N = 2$ supersymmetry also has $N = 1$ supersymmetry, so its Lagrangian must be a special case of the Lagrangians already considered in this chapter. To construct a Lagrangian with $N = 2$ supersymmetry for some set of the $N = 2$ supermultiplets of particles constructed in Sections 25.4 and 25.5, we need only write down the most general Lagrangian with $N = 1$ supersymmetry whose $N = 1$ supermultiplets contain physical fields for the particles in the $N = 2$ supermultiplets, and then impose a discrete R-symmetry on the Lagrangian: a symmetry that acts differently on different components of $N = 2$ supermultiplets. The Lagrangian density will then be invariant under a second supersymmetry, whose supermultiplets are given by acting on the supermultiplets of ordinary $N = 1$ supersymmetry with the R-symmetry.

It will be convenient to choose the discrete R-transformation so that

$$Q_1 \to Q_2 , \qquad Q_2 \to -Q_1 . \tag{27.9.1}$$

If the central charge were zero then the supersymmetry algebra would be invariant under an $SU(2)$ R-symmetry group, which has the transformation (27.9.1) as one finite element $\exp(i\pi\tau_2/2)$, but symmetry under the discrete symmetry is sufficient for our purposes, so we do not need to assume a zero central charge. In fact, it will turn out that the Lagrangians we construct by this method will have an $SU(2)$ R-symmetry, not just symmetry under the discrete transformation (27.9.1).

Let us first consider the renormalizable theory of the gauge bosons of a general gauge group, together with the superpartners required by $N = 2$ extended supersymmetry. We saw in Section 25.4 that in $N = 2$ global supersymmetry theories a massless gauge boson can only belong to a multiplet also containing a pair of massless fermions of each helicity $\pm 1/2$ that transform as a doublet under the $SU(2)$ R-symmetry and a pair of $SU(2)$-singlet spinless bosons. Since $N = 2$ supersymmetry includes $N = 1$ supersymmetry, the renormalizable Lagrangian for this theory must be a special case of the general renormalizable Lagrangian density (27.4.1). One feature of this special case is that since the gauge boson belongs to the adjoint representation of the gauge group, so must also the fermions and scalar field. To furnish $N = 2$ supermultiplets of fields with the correct particle content, we must have one $N = 1$ chiral superfield Φ_A with component fields ϕ_A, ψ_A, $\mathscr{F}_A$ (with ψ_A Majorana and ϕ_A and $\mathscr{F}_A$ both complex) for each $N = 1$ gauge multiplet V_A^μ, λ_A, D_A. We impose a discrete R-symmetry under the transformation

$$\psi_A \to \lambda_A , \qquad \lambda_A \to -\psi_A , \tag{27.9.2}$$

(with all other fields invariant) because this is the effect of the transformation (27.9.1). Since a non-trivial superpotential would give ψ_A interactions

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

or mass terms that are absent for λ_A, the superpotential must vanish. The Lagrangian (27.4.1) therefore takes the special form

$$\begin{aligned}\mathscr{L} = &-\sum_A (D_\mu\phi)^*_A (D^\mu\phi)_A - \frac{1}{2}\sum_A \left(\overline{\psi_A}(\not{D}\psi)_A\right) + \sum_A \mathscr{F}^*_A \mathscr{F}_A \\ &-2\sqrt{2}\mathrm{Re}\sum_{ABC} C_{ABC}\left(\lambda^{\mathrm{T}}_{AL}\psi_{CL}\right)\phi^*_B \\ &+i\sum_{ABC} C_{ABC}\phi^*_B\phi_C D_A - \sum_A \xi_A D_A + \frac{1}{2}\sum_A D_A D_A \\ &-\frac{1}{4}\sum_A f_{A\mu\nu}f^{\mu\nu}_A - \frac{1}{2}\sum_A \left(\overline{\lambda_A}(\not{D}\lambda)_A\right) + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f^{\mu\nu}_A f^{\rho\sigma}_A ,\end{aligned} \tag{27.9.3}$$

where

$$(D_\mu\psi)_A = \partial_\mu\psi_A + \sum_{BC} C_{ABC}V_{B\mu}\psi_C , \tag{27.9.4}$$

$$(D_\mu\lambda)_A = \partial_\mu\lambda_A + \sum_{BC} C_{ABC}V_{B\mu}\lambda_C , \tag{27.9.5}$$

$$(D_\mu\phi)_A = \partial_\mu\phi_A + \sum_{BC} C_{ABC}V_{B\mu}\phi_C , \tag{27.9.6}$$

and

$$f_{A\mu\nu} = \partial_\mu V_{A\nu} - \partial_\nu V_{A\mu} + \sum_{BC} C_{ABC}V_{B\mu}V_{C\nu} . \tag{27.9.7}$$

(Recall that in the adjoint representation $(t_A)_{BC} = -iC_{ABC}$, where C_{ABC} is the real structure constant, defined as usual in this book to include factors of gauge couplings, and taken in a basis in which it is totally antisymmetric.) The Lagrangian density (27.9.3) has $N = 1$ supersymmetry with multiplets ϕ_A, ψ_A, $\mathscr{F}_A$ and V^μ_A, λ_A, D_A, because it is a special case of Eq. (27.4.1), and it has an $SU(2)$ symmetry relating ψ_A and λ_A, including invariance under the finite $SU(2)$ transformation (27.9.2), so it also has a *second* independent $N = 1$ supersymmetry with multiplets ϕ_A, λ_A, $\mathscr{F}_A$ and V^μ_A, $-\psi_A$, D_A. It therefore satisfies the conditions imposed by $N = 2$ supersymmetry.

We can eliminate the auxiliary fields by setting them equal to the values at which the Lagrangian density (27.9.3) is stationary:

$$\mathscr{F}_A = 0 , \qquad D_A = -i\sum_{BC} C_{ABC}\phi^*_B\phi_C . \tag{27.9.8}$$

(We are now assuming that the Fayet–Iliopoulos constants ξ_A all vanish.) Inserting these values back in Eq. (27.9.3) gives an equivalent Lagrangian

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

density

$$\begin{aligned}\mathscr{L} = &-\sum_A (D_\mu\phi)^*_A (D^\mu\phi)_A - \frac{1}{2}\sum_A \left(\overline{\psi_A}\,(\not{D}\psi)_A\right)\\ &+\sqrt{2}\sum_{ABC} C_{ABC}\left(\overline{\psi_B}\left(\frac{1-\gamma_5}{2}\right)\lambda_A\right)\phi_C\\ &-\sqrt{2}\sum_{ABC} C_{ABC}\left(\overline{\lambda_A}\left(\frac{1+\gamma_5}{2}\right)\psi_C\right)\phi^*_B - V(\phi,\phi^*)\\ &-\frac{1}{4}\sum_A f_{A\mu\nu}f^{\mu\nu}_A - \frac{1}{2}\sum_A\left(\overline{\lambda_A}\,(\not{D}\lambda)_A\right) + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f^{\mu\nu}_A f^{\rho\sigma}_A\,,\end{aligned} \tag{27.9.9}$$

where the potential is

$$V(\phi,\phi^*) = -\frac{1}{2}\sum_A\left[\sum_{BC} C_{ABC}\,\phi^*_B\phi_C\right]^2 = 2\sum_A\left[\sum_{BC} C_{ABC}\,\mathrm{Re}\,\phi_B\,\mathrm{Im}\,\phi_C\right]^2. \tag{27.9.10}$$

This potential has a minimum value of zero, which is reached not only for $\phi_A = 0$ but also for any set of ϕs for which $\sum_{BC} C_{ABC}\,\phi^*_B\phi_C = 0$ for all A or, in other words, for which

$$[t\cdot \mathrm{Re}\,\phi\,,\ t\cdot \mathrm{Im}\,\phi] = 0\,, \qquad \text{where}\quad t\cdot v \equiv \sum_B t_B v_B\,. \tag{27.9.11}$$

That is, the minimum of the potential is reached for those scalar fields for which all generators $t\cdot \mathrm{Re}\,\phi$ and $t\cdot \mathrm{Im}\,\phi$ belong to a Cartan subalgebra of the full gauge algebra, all of whose generators commute with one another. Though all such values of ϕ give zero potential, and hence unbroken $N = 2$ supersymmetry, they are not physically equivalent, as shown for instance by the different masses they give the gauge bosons associated with the broken gauge symmetries.

One remarkable feature of extended supersymmetry is that the central charges of the supersymmetry algebra in any state can be calculated in terms of 'charges' in that state to which bosonic fields are coupled.[13] The easiest way to do this calculation is to use the transformation properties under ordinary $N = 1$ supersymmetry of the extended supersymmetry currents $S^\mu_r(x)$ with $r = 2, 3, \ldots, N$ to calculate the anticommutators $\{Q_{1\alpha}, S^\mu_{r\beta}(x)\}$. We can then calculate the central charges from the anti-commutators

$$\{Q_{1\alpha}, Q_{r\beta}\} = \int d^3x\,\{Q_{1\alpha}, S^0_{r\beta}(x)\}\,. \tag{27.9.12}$$

It turns out that the integrand on the right-hand side is a derivative with

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

respect to space coordinates, but its integral does not vanish if the states have fields that do not vanish rapidly as $\mathbf{x} \to \infty$.

To see how this works in detail, let's consider the case of $N = 2$ supersymmetry with an $SU(2)$ gauge symmetry and a single $N = 2$ gauge supermultiplet, with no additional matter supermultiplets. The Lagrangian here is given by Eq. (27.9.3), with A, B, and C running over the values 1, 2, 3, and

$$C_{ABC} = e\,\epsilon_{ABC}\,, \qquad \xi_A = 0\,. \tag{27.9.13}$$

(The coupling constant here is denoted e because this is the charge with which the massless gauge field of the unbroken $U(1)$ gauge symmetry interacts.) The usual $N = 1$ supersymmetry current (distinguished now with a subscript 1) is given by Eq. (27.4.40) as

$$\begin{aligned} S_1^\mu = &-\frac{1}{4}\sum_A f_{A\rho\sigma}[\gamma^\rho,\gamma^\sigma]\gamma^\mu\lambda_A - e\sum_{ABC}\epsilon_{ABC}\,\gamma_5\gamma^\mu\lambda_A\,\phi_B^*\,\phi_C \\ &+\frac{1}{\sqrt{2}}\sum_A\left[(\not{D}\phi)_A\,\gamma^\mu\psi_{AR} + (\not{D}\phi^*)_A\,\gamma^\mu\psi_{AL}\right]. \end{aligned} \tag{27.9.14}$$

We can calculate the second supersymmetry current by subjecting S_1^μ to the finite $SU(2)$ R-symmetry used above, which simply amounts to making the replacements $\psi_A \to \lambda_A$, $\lambda_A \to -\psi_A$. This gives

$$\begin{aligned} S_2^\mu = &\frac{1}{4}\sum_A f_{A\rho\sigma}[\gamma^\rho,\gamma^\sigma]\gamma^\mu\psi_A + e\sum_{ABC}\epsilon_{ABC}\,\gamma_5\gamma^\mu\psi_A\,\phi_B^*\,\phi_C \\ &+\frac{1}{\sqrt{2}}\sum_A\left[(\not{D}\phi)_A\,\gamma^\mu\lambda_{AR} + (\not{D}\phi^*)_A\,\gamma^\mu\lambda_{AL}\right]. \end{aligned} \tag{27.9.15}$$

It will be enough for our purposes (and somewhat easier) to calculate only the change in the right-handed part of this current under an $N = 1$ supersymmetry transformation. After setting the auxiliary fields equal to their equilibrium values

$$\mathscr{F}_A = 0\,, \qquad D_A = -ie\sum_{BC}\epsilon_{ABC}\phi_B^*\phi_C\,,$$

we find

$$\begin{aligned} \delta S_{2R}^\mu = &\frac{\sqrt{2}}{4}\sum_A f_{A\rho\sigma}[\gamma^\rho\,,\,\gamma^\sigma]\gamma^\mu(\not{D}\phi)_A\alpha_R \\ &-\sqrt{2}\,e\sum_{ABC}\epsilon_{ABC}\gamma^\mu(\not{D}\phi)_A\alpha_R\phi_B^*\phi_C - \frac{\sqrt{2}}{4}f_{A\rho\sigma}(\not{D}\phi)_A\gamma^\mu[\gamma^\rho\,,\,\gamma^\sigma]\alpha_R \\ &-\sqrt{2}\,e\sum_{ABC}\epsilon_{ABC}\phi_B^*\phi_C(\not{D}\phi)_A\gamma^\mu\alpha_R + \cdots\,, \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

where the dots denote terms bilinear in fermion fields, which do not concern us here because we are interested in effects of long range boson fields. We can combine terms by using the Dirac anticommutation relations and the identity

$$[\gamma^\rho, \gamma^\sigma][\gamma^\mu, \gamma^\nu] + [\gamma^\mu, \gamma^\nu][\gamma^\rho, \gamma^\sigma] = -8\eta^{\mu\rho}\eta^{\nu\sigma} + 8\eta^{\sigma\mu}\eta^{\rho\nu} + 8i\epsilon^{\mu\nu\rho\sigma}\gamma_5$$

and find

$$\delta S^\mu_{2R} = -2\sqrt{2}\sum_A f^{\mu\nu}_A (D_\nu\phi)_A \alpha_R - i\sqrt{2}\sum_A \epsilon^{\mu\nu\rho\sigma} f_{A\rho\sigma}(D_\nu\phi)_A\alpha_R$$
$$-2\sqrt{2}e\sum_{ABC}\epsilon_{ABC}\phi^*_B\phi_C(D^\mu\phi)_A\alpha_R + \cdots .$$

In order to write this as a derivative, we need to use the Yang–Mills field equations given by Eqs. (15.3.6), (15.3.7), and (15.3.9):

$$D_\nu f^{\mu\nu}_A = J^\mu_A = e\sum_{BC}\epsilon_{ABC}\left((D^\mu\phi)^*_B\phi_C - \phi^*_B(D^\mu\phi)_C\right),$$
$$\epsilon_{\mu\nu\rho\sigma}(D^\nu f^{\rho\sigma})_A = 0 .$$

These allow us to write δS^μ_{2R} as a total derivative

$$\delta S^\mu_{2R} = D_\nu X^{\mu\nu}\alpha_R , \tag{27.9.16}$$

where

$$X^{\mu\nu} = -2\sqrt{2}\sum_A f^{\mu\nu}_A\phi_A - i\sqrt{2}\sum_A\epsilon^{\mu\nu\rho\sigma}f_{A\rho\sigma}\phi_A + \cdots , \tag{27.9.17}$$

with the dots again indicating irrelevant terms involving fermion fields. Eq. (26.1.18) allows us to write Eq. (27.9.16) as an anticommutation relation

$$\left\{Q_{R\alpha}, S^\mu_{R\beta}\right\} = i\left[\epsilon\left(\frac{1-\gamma_5}{2}\right)\right]_{\alpha\beta} D_\nu X^{\mu\nu} . \tag{27.9.18}$$

Since $X^{\mu\nu}$ is a gauge-invariant quantity, its gauge-covariant derivative is the same as its ordinary derivative. Also, $X^{\mu\nu}$ is antisymmetric, so $D_\nu X^{0\nu} = \partial_i X^{0i}$. From Eqs. (27.9.12) and (27.9.18), we have at last

$$\left\{Q_{R\alpha}, Q_{R\beta}\right\} = i\left[\epsilon\left(\frac{1-\gamma_5}{2}\right)\right]_{\alpha\beta}\int dS_i\, X^{0i} , \tag{27.9.19}$$

with the integral taken over a large closed surface enclosing the system in question, with surface area differential $d\mathbf{S}$ taken normal to the surface. Comparing this with Eq. (25.2.38) gives the central charge

$$Z_{12} = -i\int dS_i\, X^{0i} . \tag{27.9.20}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

If we choose a gauge in which ϕ_A (almost everywhere) has only the constant non-vanishing component $\phi_3 \equiv v$, then

$$\sum_A f_A^{0i}\phi_A = -vE^i\,, \qquad \frac{1}{2}\sum_A \epsilon^{0i\rho\sigma} f_{A\rho\sigma}\phi_A = vB^i\,, \tag{27.9.21}$$

where **E** and **B** are the electric and magnetic fields associated with the unbroken $U(1)$ subgroup of the $SU(2)$ gauge group. Therefore the central charge (27.9.20) here is

$$Z_{12} = 2\sqrt{2}\,v\left[iq - \mathscr{M}\right], \tag{27.9.22}$$

where q and $\mathscr{M}$ are the electric charge and magnetic monopole moment, defined by

$$q = \int dS_i\, E^i\,, \qquad \mathscr{M} = \int dS_i\, B^i\,. \tag{27.9.23}$$

As discussed in Section 23.3, this theory, with $SU(2)$ gauge symmetry spontaneously broken by the expectation value of an $SU(2)$ triplet of scalars, is one in which magnetic monopoles actually do occur.

The application of the results of Section 27.4 to the Lagrangian density (27.9.3) shows that, after the spontaneous breaking of the $SU(2)$ gauge symmetry, this theory will contain elementary particles of charge $\pm e$, zero magnetic monopole moment, and tree-approximation mass $M = \sqrt{2}\,|ev|$. Specifically, for each sign of the charge there is one such particle of spin 1, two of them with spin 1/2, and one with spin 0. It is a striking consequence of the results obtained here that *the mass value* $\sqrt{2}\,|ev|$ *is exact, being unaffected by either radiative corrections or non-perturbative effects, provided that the quantity* v *is defined by Eq. (27.9.22) for the central charge.*[13]

To see this, note that the massive one-particle states for each sign of the charge are 'short' $N = 2$ supermultiplets, which, as shown at the end of Section 25.5, have masses that saturate the lower bound (25.5.24):

$$M = |Z_{12}|/2\,. \tag{27.9.24}$$

Even if we did not trust the tree approximation to give us the precise value of the particle masses, we would not expect corrections to this approximation to turn short multiplets into the full multiplets, with many more states, that could have larger masses, so we can be confident that Eq. (27.9.24) is exactly valid. For particles with electric charge $q = \pm e$ and zero magnetic monopole moment, Eq. (27.9.20) gives $Z_{12} = \pm 2\sqrt{2}\,iv\,e$, so Eq. (27.9.24) tells us that their masses are

$$M = \sqrt{2}|ev|\,. \tag{27.9.25}$$

This is the result found in the tree approximation, but now we see that it is exact.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

The semi-classical calculations described in Section 23.3 show that electrically neutral magnetic monopoles in this theory have monopole strengths**

$$\mathscr{M} = \frac{4\pi v}{e}\,, \tag{27.9.26}$$

with v the winding number, a positive or negative integer. The formula (27.9.22) for the central charge together with the inequality (25.5.24) thus give a lower bound on the monopole masses

$$M \geq \frac{4\pi\sqrt{2}\,|v\,v|}{|e|}\,. \tag{27.9.27}$$

Interestingly, this is the same as the Bogomol'nyi lower bound[14] on monopole energies derived in Section 23.3.† In fact, the monopole solution for $v = 1$ that was described in Section 23.3 saturates this bound. More generally, the 'dyons' of this theory,[15] particles with both charge and magnetic moment, have masses given by[16]

$$M = 2\,|v|\,\sqrt{q^2 + \mathscr{M}^2}\,, \tag{27.9.28}$$

which again is the minimum value allowed by Eqs. (25.5.24) and (27.9.20). Indeed, *all* of the known particles in this theory have masses given in the semi-classical limit by Eq. (27.9.28).[17]

Returning now to general $N = 2$ gauge theories, we may also include additional 'matter' fields in the Lagrangian. For simplicity, we will restrict ourselves to 'short' massive hypermultiplets (with central charge $\mathscr{Z}$ saturating the inequality (25.5.24)), each consisting of a single fermion of spin 1/2 and an $SU(2)$ doublet of spin 0 particles, together with distinct antiparticles. This is the same spin content as is given under $N = 1$ supersymmetry by pairs of left-chiral scalar superfields Φ'_n and Φ''_n, together with their right-chiral adjoints, with the complex scalar field components ϕ'_n and ϕ''_n and their adjoints forming pairs of $SU(2)$ doublets, and the spinor fields all $SU(2)$ singlets. (We are using primes and double primes to distinguish these superfields and their components from Φ_A and its components.) If some of these hypermultiplets Φ'_n and Φ''_n are non-neutral

** Note that the magnetic moment $\mathscr{M}$ defined by Eq. (27.9.23) is related to the magnetic moment g defined in Section 23.3 by $\mathscr{M} = 4\pi g$.

† The canonically normalized field with a non-vanishing vacuum expectation value is $\sqrt{2}\,\mathrm{Re}\,\phi_3$ (for real v), so the quantity $\langle\phi\rangle$ appearing in the Bogomol'nyi inequality (23.3.19) is $\sqrt{2}v$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

under the gauge group, then a superpotential is allowed, of the form:††

$$f(\Phi,\Phi',\Phi'')=\frac{1}{2}\sum_{Anm}(s_A)_{nm}\Phi'_n\Phi''_m\Phi_A+\frac{1}{2}\sum_{nm}\mu_{nm}\Phi'_n\Phi''_m\,. \tag{27.9.29}$$

To the Lagrangian density (27.9.3) we then must add a Lagrangian density for these hypermultiplets, given by the first eight terms on the right-hand side of Eq. (27.4.1), and find a total Lagrangian density

$$\begin{aligned}
\mathscr{L}=&-\sum_n(D_\mu\phi')^*_n(D^\mu\phi')_n-\sum_n(D_\mu\phi'')^*_n(D^\mu\phi'')_n-\sum_A(D_\mu\phi)^*_A(D^\mu\phi)_A\\
&-\frac{1}{2}\sum_n\left(\overline{\psi'_n}(\not{D}\psi')_n\right)-\frac{1}{2}\sum_n\left(\overline{\psi''_n}(\not{D}\psi'')_n\right)\\
&-\frac{1}{2}\sum_A\left(\overline{\psi_A}(\not{D}\psi)_A\right)-\frac{1}{2}\sum_A\left(\overline{\lambda_A}(\not{D}\lambda)_A\right)\\
&+\sum_n\mathscr{F}'^*_n\mathscr{F}'_n+\sum_n\mathscr{F}''^*_n\mathscr{F}''_n+\sum_A\mathscr{F}^*_A\mathscr{F}_A\\
&-\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi_A\left(\psi'^{\mathrm{T}}_{nL}\epsilon\psi''_{mL}\right)-2\sqrt{2}\,\mathrm{Re}\sum_{ABC}C_{ABC}\left(\lambda^{\mathrm{T}}_{AL}\epsilon\psi_{CL}\right)\phi^*_B\\
&-\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi'_n\left(\psi''^{\mathrm{T}}_{mL}\epsilon\psi_{AL}\right)-\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi''_m\left(\psi'^{\mathrm{T}}_{nL}\epsilon\psi_{AL}\right)\\
&+2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{mn}\left(\psi'^{\mathrm{T}}_{nL}\epsilon\lambda_{AL}\right)\phi'^*_m+2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t''_A)_{mn}\left(\psi''^{\mathrm{T}}_{nL}\epsilon\lambda_{AL}\right)\phi''^*_m\\
&+\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi_A\phi'_n\mathscr{F}''_m+\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi_A\phi''_m\mathscr{F}'_n\\
&+\mathrm{Re}\sum_{Anm}(s_A)_{nm}\phi'_n\phi''_m\mathscr{F}_A\\
&+\mathrm{Re}\sum_{Anm}\mu_{nm}\phi'_n\mathscr{F}''_m+\mathrm{Re}\sum_{Anm}\mu_{nm}\phi''_m\mathscr{F}'_n-\mathrm{Re}\sum_{nm}\mu_{nm}\left(\psi'^{\mathrm{T}}_{nL}\epsilon\psi''_{mL}\right)
\end{aligned}$$

†† There still cannot be any terms in the superpotential that are of second or higher order in the Φ_A for the same reason as before: such terms would lead to scalar couplings or masses for the ψ_A, with no corresponding couplings or masses for their $SU(2)$ partners λ_A. Also, there cannot be any term that is trilinear in the Φ'_n and/or Φ''_n, because then the term in Eq. (27.4.1) involving the product of fermion bilinears with second derivatives of the superpotential would lead to a coupling of the $SU(2)$-singlet fermions to $SU(2)$ doublet fields ϕ'_n or ϕ''_n. Thus the only trilinear interaction terms in the superpotential must involve one factor of Φ_A and two factors of Φ'_n and/or Φ''_n. There cannot be any trilinear terms involving a Φ_A and two Φ'_ns or two Φ''_ns, because that would give the $SU(2)$ singlet auxiliary fields $\mathscr{F}_A$ an interaction with $SU(2)$-triplet products $\phi'_n\phi'_m$ or $\phi''_n\phi''_m$, and there cannot be any bilinear terms involving two Φ'_ns or two Φ''_ns, because that would yield $SU(2)$-triplet mass terms $(\psi'^{\mathrm{T}}_n\epsilon\psi'_m)$ or $(\psi''^{\mathrm{T}}_n\epsilon\psi''_m)$. The only remaining allowed bilinear or trilinear terms are of the form (27.9.29).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$-\sum_{Anm}(t'_A)_{nm}\phi'^{*}_n\phi'_m D_A - \sum_{Anm}(t''_A)_{nm}\phi''^{*}_n\phi''_m D_A + i\sum_{ABC} C_{ABC}\,\phi^*_B\phi_C D_A$$
$$-\sum_A \xi_A D_A + \frac{1}{2}\sum_A D_A D_A$$
$$-\frac{1}{4}\sum_A f_{A\mu\nu}f_A^{\mu\nu} + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu}f_A^{\rho\sigma}\,, \qquad (27.9.30)$$

where the $(t'_A)_{nm}$ and $(t''_A)_{nm}$ are the matrices (including coupling constant factors) representing the gauge group on the left-chiral scalar superfields Φ'_n and Φ''_n, respectively. The Yukawa couplings between fermions and scalars have a discrete R-symmetry under the transformation

$$\lambda_{AL}\to -\psi_{AL}\,,\quad \psi_{AL}\to +\lambda_{AL}\,,\quad \phi''_n\to -\phi'^{*}_n\,,\quad \phi'_n\to \phi''^{*}_n\,, \qquad (27.9.31)$$

provided that

$$s_A = -2\sqrt{2}\,i\,t'^{\,\mathrm{T}}_A = +2\sqrt{2}\,i\,t''_A\,. \qquad (27.9.32)$$

(Note in particular that Eq. (27.9.32) requires the representations of the gauge group furnished by Φ'_n and Φ''_n to be complex conjugates.) This is also a symmetry of all the other terms in the Lagrangian density (27.9.30), except for those involving the auxiliary fields.

It is not possible to extend the symmetry under the transformation (27.9.31) to the auxiliary fields, but the symmetry appears after the auxiliary fields are eliminated.‡ After setting D_A, $\mathscr{F}'_n$, and $\mathscr{F}''_n$ equal to values at which the Lagrangian density is stationary, and combining D- and $\mathscr{F}$-terms, the Lagrangian density (with s_A and t''_A given by Eq. (27.9.32), and ξ_A taken to vanish) takes the form

$$\mathscr{L} = -\sum_n (D_\mu\phi')^*_n(D^\mu\phi')_n - \sum_n (D_\mu\phi'')^*_n(D^\mu\phi'')_n - \sum_A (D_\mu\phi)^*_A(D^\mu\phi)_A$$
$$-\frac{1}{2}\sum_n\left(\overline{\psi'_n}(\not{D}\psi')_n\right) - \frac{1}{2}\sum_n\left(\overline{\psi''_n}(\not{D}\psi'')_n\right)$$
$$-\frac{1}{2}\sum_A\left(\overline{\psi_A}\,(\not{D}\psi)_A\right) - \frac{1}{2}\sum_A\left(\overline{\lambda_A}\,(\not{D}\lambda)_A\right)$$
$$-2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{mn}\phi_A\left(\psi'^{\mathrm{T}}_{nL}\epsilon\psi''_{mL}\right) - 2\sqrt{2}\,\mathrm{Re}\sum_{ABC}C_{ABC}\left(\lambda^{\mathrm{T}}_{AL}\epsilon\psi_{CL}\right)\phi^*_B$$

‡ After elimination of the auxiliary fields, the resulting action is invariant under the original $N = 2$ supersymmetry transformation only 'on-shell' — that is, only up to terms that vanish when the fields satisfy the interacting field equations. This doesn't hurt, because there still are two conserved supersymmetry currents whose integrated time-components satisfy the $N = 2$ supersymmetry anticommutation relations when the fields of which they are composed are required to satisfy the field equations of the Heisenberg picture. 'Off-shell' formulations of $N = 2$ supersymmetry exist, but are subject to various complications.[18]

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

$$
\begin{aligned}
&-2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{mn}\phi'_n\left(\psi''^{\mathrm{T}}_{mL}\epsilon\psi_{AL}\right)-2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{mn}\phi''_m\left(\psi'^{\mathrm{T}}_{nL}\epsilon\psi_{AL}\right)\\
&+2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{mn}\left(\psi'^{\mathrm{T}}_{nL}\epsilon\lambda_{AL}\right)\phi'^{*}_m-2\sqrt{2}\,\mathrm{Im}\sum_{Anm}(t'_A)_{nm}\left(\psi''^{\mathrm{T}}_{nL}\epsilon\lambda_{AL}\right)\phi''^{*}_m\\
&-\frac{1}{4}\sum_A f_{A\mu\nu}f_A^{\mu\nu}+\frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu}f_A^{\rho\sigma}\\
&-\sum_{ABnm}\{t'_A,t'_B\}_{mn}\phi_A\phi_B^{*}\left(\phi'_n\phi'^{*}_m+\phi''^{*}_n\phi''_m\right)\\
&-\frac{1}{2}\sum_A\left[\sum_{nm}(t'_A)_{nm}\left(\phi'^{*}_n\phi'_m-\phi''_n\phi''^{*}_m\right)\right]^2\\
&+\frac{1}{2}\sum_{ABCDE}C_{ABC}C_{ADE}\phi_B^{*}\phi_C\phi_D^{*}\phi_E-2\sum_A\left|\sum_{nm}(t'_A)_{nm}\phi'_n\phi''_m\right|^2\\
&-4\,\mathrm{Re}\sum_{nm}(t'_A\mu)_{nm}\phi'^{*}_n\phi'_m-4\,\mathrm{Re}\sum_{nm}(\mu\,t'_A)_{nm}\phi''_n\phi''^{*}_m\\
&-2\sum_{nm}(\mu^\dagger\mu)_{nm}\phi'^{*}_n\phi'_m-2\sum_{nm}(\mu\,\mu^\dagger)_{nm}\phi''_n\phi''^{*}_m\,.
\end{aligned}
\tag{27.9.33}
$$

The last five lines on the right-hand side come from the terms in Eq. (27.9.30) involving auxiliary fields, and now these too are invariant under the discrete transformation (27.9.31), provided that

$$[t'_A,\mu]=[\mu^\dagger,\mu]=0\,. \tag{27.9.34}$$

We can now go a step further and consider the case of $N=4$ extended global supersymmetry. (As remarked in Section 25.4, $N=3$ supersymmetry is the same as $N=4$ supersymmetry.) The only massless multiplets of $N=4$ supersymmetry that do not contain gravitons or gravitinos consist of a single particle of helicity 1, an $SU(4)$ quartet of particles of helicity $1/2$, and an $SU(4)$ sextet of particles of helicity 0, together with their CPT-conjugates of opposite helicity. There is one such supermultiplet for each generator t_A of the gauge group. These particles can be grouped into supermultiplets of $N=2$ supersymmetry: for each t_A there is one gauge supermultiplet consisting of one particle of helicity 1, two particles of helicity $\pm 1/2$, and one particle of helicity 0, together with their CPT-conjugates of opposite helicity, plus two hypermultiplets, each consisting of one particle of each helicity $\pm 1/2$ and two particles of helicity 0. The $N=2$ gauge superfield consists of an $N=1$ gauge superfield V_A and a left-chiral scalar superfield Φ_A and its complex conjugate, while the two $N=2$ hypermultiplets consist of two additional left-chiral scalar superfields Φ'_A and Φ''_A and their complex conjugates.

Since $N=4$ supersymmetry includes $N=2$ supersymmetry, the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

Lagrangian density after elimination of the auxiliary fields of $N = 1$ supersymmetry[18] must be a special case of Eq. (27.9.33), but with the labels n, m, etc. running over the indices A, B, C, etc. of the adjoint representation. Also, the coefficient μ_{nm} in the superpotential (27.9.29) must vanish here, because otherwise Eq. (27.9.33) would contain terms quadratic in the fermion fields ψ'_A and ψ''_A, with no counterpart for their $N = 4$ superpartners λ_A and ψ_A. Also setting $(t'_A)_{BC}$ equal to the generators $-iC_{ABC}$ in the adjoint representation, we find that the Lagrangian density must take the form

$$\begin{aligned}
\mathscr{L} = &-\sum_A (D_\mu \phi')^*_A (D^\mu \phi')_A - \sum_A (D_\mu \phi'')^*_A (D^\mu \phi'')_A - \sum_A (D_\mu \phi)^*_A (D^\mu \phi)_A \\
&-\frac{1}{2}\sum_A \left(\overline{\psi'_A}(\not{D}\psi')_A\right) - \frac{1}{2}\sum_A \left(\overline{\psi''_A}(\not{D}\psi'')_A\right) \\
&-\frac{1}{2}\sum_A \left(\overline{\psi_A}(\not{D}\psi)_A\right) - \frac{1}{2}\sum_A \left(\overline{\lambda_A}(\not{D}\lambda)_A\right) \\
&-2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\phi_A\left(\psi'^{\mathrm{T}}_{BL}\epsilon\psi''_{CL}\right) - 2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\left(\lambda^{\mathrm{T}}_{AL}\epsilon\psi_{CL}\right)\phi^*_B \\
&-2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\phi'_B\left(\psi''^{\mathrm{T}}_{CL}\epsilon\psi_{AL}\right) - 2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\phi''_C\left(\psi'^{\mathrm{T}}_{BL}\epsilon\psi_{AL}\right) \\
&+2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\left(\psi'^{\mathrm{T}}_{BL}\epsilon\lambda_{AL}\right)\phi'^*_C + 2\sqrt{2}\,\mathrm{Re}\sum_{ABC} C_{ABC}\left(\psi''^{\mathrm{T}}_{BL}\epsilon\lambda_{AL}\right)\phi''^*_C \\
&-\frac{1}{4}\sum_A f_{A\mu\nu}f^{\mu\nu}_A + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f^{\mu\nu}_A f^{\rho\sigma}_A - V\,,
\end{aligned} \tag{27.9.35}$$

where the potential is

$$\begin{aligned}
V = &\sum_{ABCDE} C_{ADE}C_{BCE}\left(\phi_A\phi^*_B + \phi_B\phi^*_A\right)\left(\phi'_C\phi'_D{}^* + \phi''_C{}^*\phi''_D\right) \\
&+\frac{1}{2}\sum_A\left|\sum_{BC} C_{ABC}\left(\phi'_B{}^*\phi'_C - \phi''_B\phi''_C{}^*\right)\right|^2 \\
&-\frac{1}{2}\sum_{ABCDE} C_{ABC}C_{ADE}\phi^*_B\phi_C\phi^*_D\phi_E + 2\sum_A\left|\sum_{BC} C_{ABC}\phi'_B\phi''_C\right|^2\,.
\end{aligned} \tag{27.9.36}$$

With no further constraints needed, this Lagrangian has an $SU(4)$ R-symmetry, which implies that it is invariant under $N = 4$ supersymmetry. To see this, we need to use the Jacobi identity to write the cross-term in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

the second line of the right-hand side of Eq. (27.9.36) in the form

$$\sum_{ABCDE} C_{ABC}C_{ADE}\phi'_B{}^*\phi'_C\phi''_D{}^*\phi''_E = -\sum_{ABCDE} C_{ABC}C_{ADE}\phi'_B{}^*\phi'_D\phi''_E{}^*\phi''_C - \sum_{ABCDE} C_{ABC}C_{ADE}\phi'_B{}^*\phi'_E\phi''_C{}^*\phi''_D ,$$

which allows us to write the potential (27.9.36) in a form symmetric among the scalars and their adjoints

$$\begin{aligned} V = &\sum_A \left|\sum_{BC} C_{ABC}\phi^*_B\phi'_C\right|^2 + \sum_A \left|\sum_{BC} C_{ABC}\phi^*_B\phi''_C{}^*\right|^2 + \sum_A \left|\sum_{BC} C_{ABC}\phi_B\phi'_C\right|^2 \\ &+ \sum_A \left|\sum_{BC} C_{ABC}\phi_B\phi''_C{}^*\right|^2 + \sum_A \left|\sum_{BC} C_{ABC}\phi'_B{}^*\phi''_C\right|^2 \\ &+ \sum_A \left|\sum_{BC} C_{ABC}\phi'_B\phi''_C\right|^2 + \frac{1}{2}\sum_A \left|\sum_{BC} C_{ABC}\phi'_B\phi'_C{}^*\right|^2 \\ &+ \frac{1}{2}\sum_A \left|\sum_{BC} C_{ABC}\phi''_B\phi''_C{}^*\right|^2 + \frac{1}{2}\sum_A \left|\sum_{BC} C_{ABC}\phi_B\phi_C{}^*\right|^2 . \end{aligned} \tag{27.9.37}$$

Now to make the $SU(4)$ symmetry apparent, we introduce an $SU(4)$ notation for the fields. We assemble the left-handed fermion fields into an $SU(4)$ vector:

$$\psi_{1AL} \equiv \psi_{AL} , \quad \psi_{2AL} \equiv \lambda_{AL} , \quad \psi_{3AL} \equiv \psi'_{AL} , \quad \psi_{4AL} \equiv \psi''_{AL} . \tag{27.9.38}$$

In order for the fermion kinematic terms in the Lagrangian density to be $SU(4)$-invariant, we then must assemble the right-handed fermion fields into a contragredient vector:

$$\psi^1_{AR} \equiv \psi_{AR} , \quad \psi^2_{AR} \equiv \lambda_{AR} , \quad \psi^3_{AR} \equiv \psi'_{AR} , \quad \psi^4_{AR} \equiv \psi''_{AR} . \tag{27.9.39}$$

The Majorana condition on the fermion fields then takes the $SU(4)$-invariant form

$$(\psi_{iAL})^* = -\beta\epsilon\psi^i_{AR} , \tag{27.9.40}$$

with indices i, j, etc. running over the values 1, 2, 3, 4. In order for the Yukawa couplings between the fermion and scalar fields to be $SU(4)$-invariant, we must give the scalars the transformation properties of an antisymmetric $SU(4)$ tensor

$$\begin{aligned} &\phi^{12}_A \equiv \phi^*_A , \quad \phi^{13}_A \equiv \phi''_A , \quad \phi^{14}_A \equiv -\phi'_A , \\ &\phi^{23}_A \equiv -\phi'^*_A , \quad \phi^{24}_A \equiv -\phi''^*_A , \quad \phi^{34}_A \equiv \phi_A , \end{aligned} \tag{27.9.41}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

which also obeys an $SU(4)$-invariant reality condition

$$\left(\phi_A^{ij}\right)^* = \frac{1}{2}\sum_{kl}\epsilon_{ijkl}\,\phi_A^{kl}\,. \tag{27.9.42}$$

The whole Lagrangian density (27.9.35) can then be written in the manifestly $SU(4)$-invariant form

$$\begin{aligned}\mathscr{L} = &-\frac{1}{2}\sum_{Aij}(D_\mu\phi^{ij})_A(D^\mu\phi^{ij})_A^* \\ &-\frac{1}{2}\sum_{Ai}\left(\psi_{iAL}^{\mathrm{T}}\epsilon(\not{D}\psi_R^i)_A\right)+\frac{1}{2}\sum_{Ai}\left(\psi_{AR}^{i\mathrm{T}}\epsilon(\not{D}\psi_{iL})_A\right) \\ &-\sqrt{2}\,\mathrm{Re}\sum_{ABCij}C_{ABC}\phi_A^{ij}\left(\psi_{iBL}^{\mathrm{T}}\epsilon\psi_{jCL}\right)-V \\ &-\frac{1}{4}\sum_A f_{A\mu\nu}f_A^{\mu\nu}+\frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu}f_A^{\rho\sigma}\,,\end{aligned} \tag{27.9.43}$$

where the potential is

$$V=\frac{1}{8}\sum_{Aijkl}\left|\sum_{BC}C_{ABC}\phi_B^{ij}\phi_C^{kl}\right|^2\,. \tag{27.9.44}$$

The potential has a minimum value zero, so that supersymmetry is not broken in this theory. The minimum is reached when the generators $\sum_A t_A\phi_A^{ij}$ all commute with one another.

For zero theta angle, the gauge theories with a simple gauge group and either $N = 2$ or $N = 4$ supersymmetry have just a single coupling constant, the gauge coupling constant g. Since these theories have $N = 1$ supersymmetry, they share the property discussed in Section 27.6, that the only infinity in higher orders of perturbation theory is in a one-loop correction to this coupling.‡‡ The function $\beta(g)$ in the renormalization group equation $\mu dg/d\mu = \beta(g)$ is then given to all orders of perturbation theory by the one-loop formula (18.7.2), with a suitable correction for the presence of scalar fields:

$$\beta(g) = -\frac{g^3}{4\pi^2}\left(\frac{11}{12}C_1-\frac{1}{6}C_2^f-\frac{1}{12}C_2^s\right)\,, \tag{27.9.45}$$

‡‡ The trilinear term in the superpotential (27.9.29) is proportional to the gauge coupling, and is therefore renormalized, despite the no-renormalization theorem of Section 27.6. This is because here we are renormalizing the left-chiral scalar superfields Φ_A, Φ'_n, and Φ''_n as well as the gauge superfield V_A so as to keep them canonically normalized. The bilinear term in Eq. (27.9.29) is renormalized for the same reason.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

where

$$\sum_{AB} C_{ABC}C_{ABD} = g^2 C_1 \delta_{CD} ,$$
$$\left[\mathrm{Tr}\,(t_C t_D)\right]_{\text{Majorana fermions}} = g^2 C_2^f \delta_{CD} , \tag{27.9.46}$$
$$\left[\mathrm{Tr}\,(t_C t_D)\right]_{\text{complex scalars}} = g^2 C_2^s \delta_{CD} .$$

In general theories with $N = 2$ supersymmetry we have two Majorana fermions λ_A and ψ_A in the adjoint representation and H pairs of Majorana fermions ψ'_n and ψ''_n whose left- and right-handed parts are in representations with generators either t'_A or $-{t'_A}^{\mathrm{T}}$, so

$$C_2^f = 2C_1 + 2HC'_2 , \tag{27.9.47}$$

where C'_2 is defined by

$$\mathrm{Tr}\, t'_C t'_D = g^2 C'_2 \delta_{CD} . \tag{27.9.48}$$

Also, we have one complex scalar ϕ_A in the adjoint representation and H pairs of complex scalars ϕ'_n and ϕ''_n in representations with generators t'_A or $-{t'_A}^{\mathrm{T}}$, so

$$C_2^s = C_1 + 2HC'_2 . \tag{27.9.49}$$

The beta function (27.9.45) is therefore

$$\beta(g) = -\frac{g^2}{8\pi^2}\left(C_1 - HC'_2\right) . \tag{27.9.50}$$

The case of $N = 4$ supersymmetry is just the special case with $H = 1$ pairs of $N = 2$ hypermultiplets in the adjoint representation, with $C'_2 = C_1$, so in this case the beta function vanishes. This is therefore a finite theory, with no renormalizations at all.[19]

Gauge theories with $N = 4$ supersymmetry have another remarkable property, known as *duality*. This was first conjectured by Montonen and Olive[17] for purely bosonic theories in which a simple gauge group is spontaneously broken to a $U(1)$ electromagnetic gauge group. They noticed that semi-classical calculations (of the sort described in Section 23.3) give the mass of particles with charge $q = ne$ and magnetic monopole moment $\mathscr{M} = 4\pi m/e$ (with n and m integers of any sign) as

$$M = \sqrt{2}\left|v\left(ne + \frac{4\pi i m}{e}\right)\right| , \tag{27.9.51}$$

which is invariant under the transformations

$$m \to n , \qquad n \to -m , \qquad e \leftrightarrow 4\pi/e . \tag{27.9.52}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

On this basis, they suggested that the theory with a weak gauge coupling e is fully equivalent to one with a strong gauge coupling $4\pi/e$. Neither the purely bosonic theory nor the simplest versions of the $N = 1$ and $N = 2$ extended supersymmetry theories really have this property;[20] for one thing, the massive charged elementary vector bosons of the broken gauge symmetries have spin 1, while all of the monopoles and dyons have spins 1/2 or 0. (We will see in Section 29.5 that $N = 2$ theories do have a duality property of a more subtle sort.) But for $N = 4$ supersymmetry the monopole states form multiplets with one particle of spin 1, four particles of spin 1/2, and two particles of spin 0, just like the elementary particles.[20] Evidence has accumulated[21] that $N = 4$ supersymmetric gauge theories are indeed invariant under the interchange of electric and magnetic quantum numbers and of e with $4\pi/e$. The equivalence of theories with large and small coupling constants has become an increasingly important theme in string theory, but this is beyond the scope of this book.

Problems

1. To second order in gauge coupling constants, calculate the components of the superfields Ω^A that are needed to put the gauge superfield V^A in Wess–Zumino gauge by the transformation (27.1.12).

2. Show that the most general chiral spinor superfield W_α satisfying the condition (27.2.20) has components $f_{\mu\nu}$ satisfying the homogeneous Maxwell equations $\epsilon^{\mu\nu\rho\sigma}\partial_\rho f_{\mu\nu} = 0$. What conditions on the other components of W_α are imposed by Eq. (27.2.20)?

3. Consider a general renormalizable $N = 1$ supersymmetric gauge theory with an $SU(2)$ gauge group and a single chiral superfield belonging to the 3-vector representation of $SU(2)$. What is the most general superpotential for this theory? Construct the Lagrangian density of the whole theory explicitly. Eliminate the auxiliary fields. Show that supersymmetry is not broken in this theory. What are the masses of the particles of this theory?

4. Express the gaugino and chiral fermion fields in the supersymmetric version of quantum electrodynamics described in Section 27.5 in terms of the goldstino field and other spinor fields of definite mass.

5. Consider the renormalizable $N = 2$ supersymmetric theory with an $SU(3)$ gauge symmetry and no hypermultiplets. What are the values of the scalar fields for which the potential vanishes? What are the massless gauge fields for non-zero values of these scalars? Calculate

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

the central charge in terms of the quantities to which these massless gauge fields are coupled.

References

1. Supersymmetry was applied first to Abelian gauge theories without using the superfield formalism by J. Wess and B. Zumino, *Nucl. Phys.* **B78**, 1 (1974). It was then extended to non-Abelian gauge theories by S. Ferrara and B. Zumino, *Nucl. Phys.* **B79**, 413 (1974); A. Salam and J. Strathdee, *Phys. Lett.* **51B**, 353 (1974). These references are reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987).
2. P. Fayet and J. Iliopoulos, *Phys. Lett.* **51B**, 461 (1974). This reference is reprinted in *Supersymmetry*, Ref. 1.
3. S. Weinberg, *Phys. Rev. Lett.* **80**, 3702 (1998).
4. S. Ferrara, L. Girardello, and F. Palumbo, *Phys. Rev.* **D20**, 403 (1979). This article is reprinted in *Supersymmetry*, Ref. 1. Special cases of this sum rule had been given by P. Fayet, *Phys. Lett.* **84B**, 416 (1979).
5. M. T. Grisaru, W. Siegel, and M. Roček, *Nucl. Phys.* **B159**, 429 (1979).
6. N. Seiberg, *Phys. Lett.* **B318**, 469 (1993).
7. This was first proved in the supergraph formalism by W. Fischler, H. P. Nilles, J. Polchinski, S. Raby, and L. Susskind, *Phys. Rev. Lett.* **47**, 757 (1981). The proof presented here was given by M. Dine, in *Fields, Strings, and Duality: TASI 96*, C. Efthimiou and B. Greene, eds. (World Scientific, Singapore, 1997); S. Weinberg, Ref. 3.
8. A detailed proof of the statement that superrenormalizable terms that violate some global symmetry do not introduce infinite symmetry breaking radiative corrections to the coefficients of renormalizable interactions is given by K. Symanzik, in *Cargèse Lectures in Physics*, Vol. 5, D. Bessis, ed. (Gordon and Breach, New York, 1972). This is discussed briefly in the footnote on p. 507 of Volume I.
9. L. Girardello and M. T. Grisaru, *Nucl. Phys.* **B194**, 65 (1982), reprinted in *Supersymmetry*, Ref. 1; K. Harada and N. Sakai, *Prob. Theor. Phys.* **67**, 67 (1982).
10. For a review, see S. Weinberg, *Rev. Mod. Phys.* **61**, 1–23 (1989).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

11. B. de Wit and D. Z. Freedman, *Phys. Rev.* **D12**, 2286 (1975). This article is reprinted in *Supersymmetry*, Ref. 1.

12. The first examples of gauge theories with $N = 2$ extended supersymmetry were given by P. Fayet, *Nucl. Phys.* **B113**, 135 (1976); reprinted in *Supersymmetry*, Ref. 1. The approach presented here is similar to that of Fayet. A superfield formalism was subsequently given by R. Grimm, M. Sohnius, and J. Scherk, *Nucl. Phys.* **B113**, 77 (1977). Gauge theories with $N = 2$ and $N = 4$ supersymmetry in four spacetime dimensions were constructed by dimensional reduction of higher-dimensional theories with simple supersymmetry by L. Brink, J. H. Schwarz, and J. Scherk, *Nucl. Phys.* **B113**, 77 (1977); M. F. Sohnius, K. S. Stelle, and P. C. West, *Nucl. Phys.* **B113**, 127 (1980). For other approaches, see M. F. Sohnius, *Nucl. Phys.* **B138**, 109 (1979); A. Halperin, E. A. Ivanov, and V. I. Ogievetsky, *Prima JETP* **33**, 176 (1981); P. Breitenlohner and M. F. Sohnius, *Nucl. Phys.* **B178**, 151 (1981); P. Howe, K. S. Stelle, and P. K. Townsend, *Nucl. Phys.* **B214**. 519 (1983).

13. E. Witten and D. Olive, *Phys. Lett.* **78B**, 97 (1978). Also see H. Osborn, *Phys. Lett.* **83B**, 321 (1979).

14. E. B. Bogomol'nyi, *Sov. J. Nucl. Phys.* **24**, 449 (1976).

15. D. Zwanziger, *Phys. Rev.* **176**, 1480, 1489 (1968); J. Schwinger, *Phys. Rev.* **144**, 1087 (1966); **173**, 1536 (1968); B. Julia and A. Zee, *Phys. Rev.* **D11**, 2227 (1974); F. A. Bais and J. R. Primack, *Phys. Rev.* **D13**, 819 (1975). (The last article was incorrectly attributed to Julia and Zee in Chapter 23 of the first printings of Volume II.)

16. M. K. Prasad and C. M. Sommerfield, *Phys. Rev. Lett.* **35**. 760 (1975); E. B. Bogomol'nyi, Ref. 14; S. Coleman, S. Parke, A. Neveu, and C. M. Sommerfield, *Phys. Rev.* **D15**, 544 (1977).

17. This was noted by C. Montonen and D. Olive, *Phys. Lett.* **72B**, 117 (1977). The absence of one-loop corrections to the masses was shown by A. D'Adda, R. Horsley, and P. Di Vecchia, *Phys. Lett.* **76B**, 298 (1978).

18. Obstacles to formulating theories with auxiliary fields for $N = 4$ supersymmetry have been analyzed by W. Siegel and M. Roček, *Phys. Lett.* **105B**, 275 (1981).

19. The finiteness of the $N = 4$ theory was shown by M. F. Sohnius and P. C. West, *Nucl. Phys.* **B100**, 245 (1981); P. S. Howe, K. S. Stelle, and P. Townsend, *Nucl. Phys.* **B214**, 519 (1983); S. Mandelstam, *Nucl. Phys.* **B213**, 149 (1983); L. Brink, O. Lindgren, and B. E. W. Nilsson, *Nucl. Phys.* **B212**, 401 (1983); *Phys. Lett.* **123B**, 328 (1983). The proof

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

of finiteness was extended to non-perturbative effects by N. Seiberg, *Phys. Lett.* **B206**, 75 (1988). Also see S. Kovacs, hep-th/9902047, to be published. There is a wide class of ultraviolet-finite N = 2 theories; see P. S. Howe, K. S. Stelle, and P. C. West, *Phys. Lett.* **B124**, 55 (1983).

20. H. Osborn, Ref. 13.

21. A. Sen, *Phys. Lett.* **B329**, 217 (1994); C. Vafa and E. Witten, *Nucl. Phys.* **B 431**, 3 (1994); L. Girardello, A. Giveon, M. Porrati, and A. Zaffaroni, *Phys. Lett.* **B334**, 331 (1994).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.006

28

Supersymmetric Versions of the Standard Model

Physical phenomena at energies accessible in today's accelerator laboratories are accurately described by the standard model, the renormalizable theory of quarks, leptons, and gauge bosons, governed by the gauge group $SU(3) \times SU(2) \times U(1)$, described in Sections 18.7 and 21.3. The standard model is today usually[1] understood as a low-energy approximation to some as-yet-unknown fundamental theory in which gravitation appears unified with the strong and electroweak forces at an energy somewhere in the range of 10^{16} to 10^{18} GeV. This raises the *hierarchy problem*: what accounts for the enormous ratio of this fundamental energy scale and the energy scale ≈ 300 GeV that characterizes the standard model?

The strongest theoretical motivation for supersymmetry is that it offers a hope of solving the hierarchy problem. Quarks, leptons, and gauge bosons are required by the $SU(3) \times SU(2) \times U(1)$ gauge symmetry to appear with zero masses in the Lagrangian of the standard model, so that the physical masses of these particles are proportional to the electroweak breaking scale, which in turn is proportional to the mass of the scalar fields responsible for the electroweak symmetry breakdown. The crux of the hierarchy problem[1a] is that the scalar fields, unlike the fermion and gauge boson fields, are not protected from acquiring large bare masses by any symmetry of the standard model, so it is difficult to see why their masses, and hence all other masses, are not in the neighborhood of 10^{16} to 10^{18} GeV.

It has been hoped that this problem could be solved by embedding the standard model in a supersymmetric theory. If the scalar fields appear in supermultiplets along with fermions in a chiral representation of some gauge group, then supersymmetry would require vanishing bare masses for the scalars as well as the fermions. All the masses of the standard model would then be tied to the energy scale at which supersymmetry is broken. The hope of a solution of the hierarchy problem along these lines has been the single strongest motivation for trying to incorporate supersymmetry in a realistic theory.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

Unfortunately, none of the new particles required by supersymmetric theories have been detected, and no entirely satisfactory supersymmetric version of the standard model has emerged so far. This chapter will describe the attempts that have been made in this direction.

28.1 Superfields, Anomalies, and Conservation Laws

In this section we shall attempt to decide at least tentatively what ingredients should appear in a supersymmetric version of the standard model.

None of the quark and lepton fields of the standard model belong to the adjoint representation of the $SU(3) \times SU(2) \times U(1)$ gauge group, so they cannot be the superpartners of known gauge bosons and must therefore be included in chiral scalar superfields. We will define U_i, D_i, $\bar{U}_i$, $\bar{D}_i$, N_i, E_i, and $\bar{E}_i$ as the left-chiral superfields whose ψ_L components are the left-handed fields of the quarks of charge $2e/3$ and $-e/3$, of the antiquarks of charge $-2e/3$ and $+e/3$, of the leptons of charge 0 and $-e$, and of the antileptons of charge $+e$, respectively, with i a generation label running over values 1, 2, and 3. (For instance, the spinor components of U_1, U_2, and U_3 are the left-handed fields of the u, c, and t quarks, respectively.) Of these superfields, U_i and D_i form $SU(2)$ doublets, N_i and E_i also form $SU(2)$ doublets, and the others are $SU(2)$ singlets. The quark superfields form $SU(3)$ triplets and the antiquark superfields form $SU(3)$ antitriplets, with color indices suppressed, and the leptons and antilepton superfields are $SU(3)$ singlets. As mentioned earlier, the particles described by the scalar components of these superfields are known as squarks, antisquarks, sleptons, and antisleptons. There are also the gauginos, the spin 1/2 superpartners of the gauge bosons of $SU(3)$, $SU(2)$, and $U(1)$, respectively known as the gluino, wino, and bino.*

We must also add some mechanism that produces a spontaneous breakdown of $SU(2) \times U(1)$ and gives mass to all the quarks and leptons as well as to the $W^{\pm}$ and Z^0. The simplest possibility is to suppose the existence

* As discussed in Section 28.3, the energy scale that characterizes supersymmetry breaking is expected to be considerably higher than the $\approx$ 300 GeV that characterizes the breaking of $SU(2) \times U(1)$, so there is a substantial range of energies in which supersymmetry but not $SU(2) \times U(1)$ may be considered to be broken. In this range, the gauginos have masses that are governed by $SU(2) \times U(1)$ symmetry, so the neutral electroweak gauginos of definite mass are superpartners of the $SU(2)$ triplet W^0 and the $SU(2)$ singlet B, known as the neutral wino and the bino, rather than superpartners of the Z^0 and the photon. When $SU(2) \times U(1)$ breaking is taken into account there is a small mixing of the neutral wino and bino.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

of just two more $SU(2)$ doublets of left-chiral superfields:

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix}, \qquad H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix}, \tag{28.1.1}$$

which appear in the Lagrangian density in linear combinations of the $SU(3) \times SU(2) \times U(1)$-invariant $\mathscr{F}$-terms:

$$\left[\left(D_i H_1^0 - U_i H_1^-\right)\bar{D}_j\right]_{\mathscr{F}}, \qquad \left[\left(E_i H_1^0 - N_i H_1^-\right)\bar{E}_j\right]_{\mathscr{F}}, \tag{28.1.2}$$

and

$$\left[\left(D_i H_2^+ - U_i H_2^0\right)\bar{U}_j\right]_{\mathscr{F}}, \tag{28.1.3}$$

with obvious contractions of color indices. According to Eq. (26.4.11), a non-vanishing expectation value of the scalar component of H_1^0 gives mass to the charged leptons and charge $-e/3$ quarks, while a non-vanishing expectation value of the scalar component of H_2^0 gives mass to the charge $+2e/3$ quarks. These expectation values of course also give mass to the $W^\pm$ and Z^0 vector bosons and, since the H_1 and H_2 are $SU(2)$ doublets, we automatically get the same successful results for these masses as found in Section 21.3. Note that supersymmetry does not allow the complex conjugates of the H_1 and H_2 left-chiral superfields to appear in the superpotential, so a vacuum expectation value of the scalar component of H_1^0 cannot give masses to the charge $+2e/3$ quarks, and a vacuum expectation value of the scalar component of H_2^0 cannot give mass to the charge $-e/3$ quarks or the charged leptons, which is why both H_1 and H_2 are needed to give masses to all the quarks and leptons.

Of course, there might be more than one of the H_1 and/or H_2 doublets. Their numbers are partly constrained by the condition of anomaly cancellation. We saw in Section 22.4 that the gauge symmetries of the non-supersymmetric standard model are anomaly-free, as they must be for quantum mechanical consistency, but now there are additional spinor fields in the Lagrangian. The gaugino fields don't create any problems, because their left-handed components belong to the adjoint representation of the gauge group, which is real for all gauge groups. The only problem can arise from the higgsinos — the spin 1/2 components of the superfields (H_1^0, H_1^-) and (H_2^+, H_2^0). The spinor components of each (H_1^0, H_1^-) doublet of superfields produces an $SU(2)$-$SU(2)$-$U(1)$ anomaly proportional to $\sum t_3^2 y = (\frac{1}{2}g)^2(\frac{1}{2}g') + (-\frac{1}{2}g)^2(\frac{1}{2}g') = \frac{1}{2}g^2 g'$, while the spinor components of each (H_2^+, H_2^0) doublet of superfields produces an $SU(2)$-$SU(2)$-$U(1)$ anomaly proportional to $\sum t_3^2 y = (\frac{1}{2}g)^2(-\frac{1}{2}g') + (-\frac{1}{2}g)^2(-\frac{1}{2}g') = -\frac{1}{2}g^2 g'$. *The cancellation of anomalies thus requires an equal number of* (H_1^0, H_1^-) *and* (H_2^+, H_2^0) *doublets.* In this case, all anomalies cancel, including the $U(1)^3$ and $U(1)$-graviton-graviton anomalies. The next section will give an argument that there is in fact just one of each type of doublet.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

In a theory constructed along these lines, we have to give up one of the attractive features of the non-supersymmetric standard model, that is, that it *automatically* excludes any renormalizable interactions that violate the conservation of baryon or lepton number. There are several renormalizable supersymmetric $SU(3) \times SU(2) \times U(1)$-invariant $\mathscr{F}$-terms that could be included in the Lagrangian density that would violate baryon and/or lepton number conservation without violating the $SU(3) \times SU(2) \times U(1)$ gauge symmetries:

$$\left[\left(D_i N_j - U_i E_j\right)\bar{D}_k\right]_{\mathscr{F}}, \qquad \left[\left(E_i N_j - N_i E_j\right)\bar{E}_k\right]_{\mathscr{F}}, \tag{28.1.4}$$

and also

$$\left[\bar{D}_i \bar{D}_j \bar{U}_k\right]_{\mathscr{F}}, \tag{28.1.5}$$

with the three suppressed color indices in Eq. (28.1.5) understood to be contracted with an antisymmetric ϵ-symbol to give a color singlet. With all of these interactions present, there would be no clever way to assign baryon and lepton numbers to the squarks and sleptons that would avoid an unsuppressed violation of baryon and lepton number conservation. For instance, the exchange of the scalar boson of the $\bar{D}$ superfield between vertices for interactions (28.1.4) and (28.1.5) would lead to the process $u_L d_R u_R \to \overline{e_R}$, observed for instance as $p \to \pi^0 + e^+$, at a catastrophic rate that is only suppressed by factors of coupling constants. To avoid this, it is necessary to make an independent assumption that would rule out some or all of the interactions (28.1.4)–(28.1.5).

Note that it is not necessary to rule out *all* of the interactions (28.1.4) and (28.1.5). For instance, suppose that we assume only that baryon number is conserved, with conventional baryon number assignments: the U_i and D_i left-chiral superfields are assigned baryon number $+1/3$; the $\bar{U}_i$ and $\bar{D}_i$ are assigned baryon number $-1/3$; and the L_i, $\bar{E}_i$, H_1, and H_2 all are assigned baryon number 0. This would allow the interactions (28.1.4) while forbidding the interactions (28.1.5). Despite appearances, the interactions (28.1.4) alone do not violate the conservation of lepton number, provided the scalar components of the superfields are assigned appropriate lepton numbers. This can be done by assigning lepton number 0 to the N_i and E_i superfields, lepton number -1 to the U_i, D_i, $\bar{U}_i$, and $\bar{D}_i$ superfields, lepton number -2 to the $\bar{E}_i$ superfields, lepton number 0 to the H_1 and H_2 superfields, and lepton numbers -1 and $+1$ to θ_L and θ_R, respectively. (Recall that such symmetries, under which θ transforms non-trivially, are known as *R-symmetries*.) Then all quarks and leptons have conventional lepton numbers: the fermion components ν_{iL} and e_{iL}, which are the coefficients of θ_L in N_i and E_i, have the lepton numbers $0 + 1 = +1$; the fermion components $\overline{e_{iR}}$ of the $\bar{E}_i$ superfields have lepton

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

numbers $-2+1=-1$, and the quarks and antiquarks have lepton numbers $-1+1=0$. The higgsinos (the fermion components of H_1 and H_2) have lepton numbers $0+1=+1$. On the other hand, the scalar components of the superfields have the same lepton numbers as the superfields themselves, which are unconventional. Further, the $\mathscr{F}$-term of a left-chiral superfield is the coefficient of θ_L^2, so the interactions in Eq. (28.1.4) have lepton numbers $-1+0-1+2=0$ and $0+0-2+2=0$; the H_1 interactions (28.1.2) have lepton number $-1+0-1+2=0$ and $0+0-2+2=0$, respectively; and the H_2 interaction (28.1.3) has lepton number $-1+0-1+2=0$; so none of these interactions violate the conservation of lepton number. Also, the scalar components of H_1 and H_2 have lepton number 0, so their vacuum expectation values also do not violate the conservation of lepton number. With this assignment of lepton numbers, lepton number conservation rules out any renormalizable interactions that would violate baryon number conservation: the interaction (28.1.5) has lepton number $-1-1-1+2=-1$, and so is forbidden.

The interactions (28.1.4) would allow an alternative mechanism for breaking $SU(2)\times U(1)$ and giving mass to the charged leptons and charge $-e/3$ quarks: the scalar components of the neutrino superfields N_i might have non-vanishing vacuum expectation values. (With the lepton number assignments of the previous paragraph this expectation value would not violate lepton number conservation, because these scalar components have the lepton number of the N_i superfields, which is zero.) But we cannot rely on this mechanism to let us do without the H_1 superfields altogether, because we still need the H_2 interactions (28.1.3) to give mass to the charge $+2e/3$ quarks and, as we have seen, the cancellation of anomalies requires equal numbers of H_1 and H_2 superfields.

It is usually assumed instead that some symmetries forbid both interactions (28.1.4) and (28.1.5). Obviously, these symmetries could be baryon and lepton number conservation, with conventional assignments of these numbers: U_i, D_i having baryon number $B=1/3$ and lepton number $L=0$, $\bar{U}_i$ and $\bar{D}_i$ having baryon number $B=-1/3$ and lepton number 0, N_i and E_i having lepton number $L=+1$ and baryon number 0, $\bar{E}_i$ having lepton number -1 and baryon number 0, and H_1^0, H_1^- and H_2^+, H_2^0 *and* θ_L and θ_R all having baryon and lepton numbers 0. The same results apply if we only assume the conservation of certain linear combinations of baryon and lepton number, such as the anomaly-free combination $B-L$ discussed in Section 22.4.

There are widespread doubts about whether it is possible to have exact continuous global symmetries, because in string theory the existence of any exact continuous symmetry would imply the existence of a massless spin 1 particle coupled to the symmetry current, so that the symmetry would have to be local, not global.[1b] But the interactions (28.1.4) and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

(28.1.5) may also be banned by assuming a *discrete* global symmetry, known as the conservation of R *parity*.[2] The R parity is defined to be $+1$ for quarks, leptons, gauge bosons, and Higgs scalars, and -1 for their superpartners. This R parity is equal to

$$\Pi_R = (-1)^F (-1)^{3(B-L)} , \tag{28.1.6}$$

where $(-1)^F$ is the fermion parity, which is $+1$ for all bosons and -1 for all fermions. Fermion parity is the same sign as is produced by a 2π rotation, and is therefore always conserved, so if $B - L$ is conserved then so is R parity.** It is possible that R parity may be conserved even if $B - L$ is not, but in fact the interactions (28.1.4) and (28.1.5) are forbidden by R parity conservation, so as far as renormalizable interactions are concerned R parity conservation implies the conservation of both baryon and lepton number. This is *not* true of the non-renormalizable supersymmetric interactions that are presumably produced by physical processes at very high energies. The baryon- and lepton-non-conserving processes produced by such interactions are discussed in Section 28.7.

All of the new 'sparticles' (squarks, sleptons, gauginos, and higgsinos) that are required by supersymmetry theories have negative R parity, so if R parity is exact and unbroken then the lightest of the new particles required by supersymmetry must be absolutely stable. All of the other new particles will then undergo a chain of decays, ultimately yielding ordinary particles and the lightest new particle. Much of the phenomenology of various supersymmetry models is governed by the choice of which of the new particles is the lightest.

With supersymmetry and either R parity or $B - L$ conserved, the most general renormalizable Lagrangian for the superfields discussed above consists of the usual gauge-invariant kinematic part for the chiral superfields, given by a sum of terms of the form $(\Phi^* \exp(-V)\Phi)_D$ for each of the quark, lepton, and Higgs chiral superfields, plus the usual gauge-invariant kinematic term for the gauge superfields, given by a sum of terms of the form $\epsilon_{\alpha\beta}(W_\alpha W_\beta)_{\mathscr{F}}$ for each of the $SU(3)$, $SU(2)$, and $U(1)$ field strength superfields, plus the supersymmetric Yukawa couplings, given by a linear combination of the interactions (28.1.2), (28.1.3), and a new $\mathscr{F}$-term

** The value of $(-1)^{3(B-L)}$ is -1 for quark and lepton superfields and $+1$ for all other superfields, so the conservation of R parity is equivalent to invariance under a transformation in which all quark and lepton superfields change sign with no change in other superfields. This invariance principle was introduced in Reference 3 in order to rule out the interactions (28.1.4) and (28.1.5).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

coupling H_1 and H_2:

$$\mathscr{L}_Y = \sum_{ij} h^D_{ij}\left[\left(D_i H^0_1 - U_i H^-_1\right)\bar{D}_j\right]_{\mathscr{F}} + \sum_{ij} h^E_{ij}\left[\left(E_i H^0_1 - N_i H^-_1\right)\bar{E}_j\right]_{\mathscr{F}} + \sum_{ij} h^U_{ij}\left[\left(D_i H^+_2 - U_i H^0_2\right)\bar{U}_j\right]_{\mathscr{F}} + \mu\left[H^+_2 H^-_1 - H^0_2 H^0_1\right]_{\mathscr{F}} + \text{H.c.} \tag{28.1.7}$$

As we will see in Section 28.3, more terms will have to be added to the Lagrangian in order to account for supersymmetry breaking.

The coefficient μ in Eq. (28.1.7) has the dimensions of mass, and is the only dimensional parameter that enters in the supersymmetric version of the standard model Lagrangian. It is somewhat disappointing to find that this term is still allowed, because it revives the hierarchy problem: why is μ not of order 10^{16} to 10^{18} GeV? The μ-term in Eq. (28.1.7) can be avoided if we assume that lepton number is conserved, with the unconventional lepton number assignments discussed above that would allow the interactions (28.1.4) but not the interactions (28.1.5). In this case, the μ-term carries lepton number $+2$ and is therefore also forbidden. This term can also be forbidden if we assume a $U(1)$ 'Peccei–Quinn symmetry,'[4] for which the superfields H_1 and H_2 carry equal quantum numbers, say $+1$, while θ_L and θ_R are neutral. The interactions (28.1.2) and (28.1.3) which give mass to the quarks and leptons are then allowed if, for instance, we give Peccei–Quinn quantum numbers -1 to the left-chiral superfields of antisquarks and antisleptons while the left-chiral superfields of squarks and sleptons are taken to be neutral. This choice then also forbids the dangerous interactions (28.1.4) and (28.1.5). Unfortunately, as we will see in Section 28.4, the μ-term in Eq. (28.1.7) seems to be needed for phenomenological reasons. The theories of gravity-mediated supersymmetry breaking discussed in Section 31.7 provide a natural mechanism for producing a μ-term of an acceptable magnitude.

We can obtain a crude upper bound on the masses of the new particles by assuming that supersymmetry does solve the hierarchy problem discussed at the beginning of this chapter. In accord with the theorem of Section 27.6, if supersymmetry were unbroken the contribution to the mass of the scalar component of H_1 or H_2 from one-loop diagrams with an intermediate quark, lepton, W or Z loop would be cancelled by the corresponding one-loop diagram with an intermediate squark, slepton, wino, or bino. Therefore with supersymmetry broken the contribution δm^2_H of such diagrams to the squared masses of the H_1 and H_2 scalars is a sum of terms of order $(\mathscr{G}^2_s/8\pi^2)\Delta m^2_s$, where $\mathscr{G}_s$ is the Yukawa or gauge coupling of the Higgs scalar to supermultiplet s, and Δm^2_s is the mass-squared splitting within the supermultiplet. To avoid having to fine-tune

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

these corrections, we need δm_H^2 not much larger than the coefficient of order $(300\ \text{GeV})^2$ of the term in the standard model Lagrangian density that gives the observed $SU(2) \times U(1)$ breaking in the tree approximation, so we will assume that $\delta m_H^2 < (1\ \text{TeV})^2$. For instance, the top quark and squark have couplings to H_2 of order unity, so we expect that the splittings Δm^2 should be less than about $8\pi^2\ \text{TeV}^2$, and hence the masses of the top squarks should be less than about 10 TeV. We will see in Section 28.4 that the rates of flavor-changing processes can be brought within experimental upper bounds by taking the masses of the squarks to be nearly equal, in which case this can be taken as a rough upper limit on the masses of all the squarks. (However it is possible that the rates of these processes may be suppressed instead by very large masses of the first two generations of squarks, while the mass of the top squark is below the 10 TeV naturalness bound.[4a]) The limits set by this sort of argument on the masses of other particles with $R = -1$ are somewhat weaker, but at least in the popular class of models discussed in Section 28.6, none of the masses of these particles are expected to be much greater than the squark masses, so 10 TeV can be taken as an upper bound on all of them. On the other hand, the fact that none of these particles have been observed only indicates that their masses are probably greater than about 100 GeV, so there is an ample mass range in which they may yet be found.

* * *

If R parity conservation or some other conservation law makes the lightest of the new particles predicted by supersymmetry stable, then some of these particles may be left from the early universe. The number density of these relics can be estimated using techniques that were originally applied to the cosmic density of massive neutrinos.[4b] To give one example of this sort of calculation, we shall show that for a broad range of plausible masses, the new stable particle of supersymmetry theories cannot be a charged and uncolored particle, like a charged slepton, wino, or higgsino.[4c]

Once the cosmic temperature T (in energy units, with the Boltzmann constant set equal to unity) drops below the mass m of any stable charged untrapped particle, their number nR^3 in a volume R^3 that expands with the universe is decreased by annihilation at a rate per particle equal to $\overline{v\sigma}n$, where $\overline{v\sigma}$ is the mean value of the product of the relative velocity and the annihilation cross-section. That is,

$$\frac{d(nR^3)}{dt} = -\overline{v\sigma}n^2R^3 ,$$

so that

$$\frac{1}{nR^3} = \left(\frac{1}{nR^3}\right)_0 + \int_{t_0}^{t} \frac{\overline{v\sigma}}{R^3}\, dt , \tag{28.1.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

where 0 labels the epoch at which $T \simeq m$. The annihilation process is exothermic, so $\overline{v\sigma}$ approaches a constant for $v \ll 1$. Also, in a radiation-dominated phase of the cosmic expansion $R \propto t^{1/2}$, so the integral converges, and gives

$$\left(\frac{1}{nR^3}\right)_{t\to\infty} = \left(\frac{1}{nR^3}\right)_0 + \overline{v\sigma}\int_{t_0}^{\infty} \frac{dt}{R_0^3\,(t/t_0)^{3/2}}$$
$$= \left(\frac{1}{nR^3}\right)_0 + \frac{2\,\overline{v\sigma}\,t_0}{R_0^3}\,. \tag{28.1.9}$$

The density n_B of baryon number (baryons minus antibaryons) goes as R^{-3}, so this can be rewritten as a formula for the present ratio of new particles to baryons:

$$(n/n_B)_\infty = \left[(n_B/n)_0 + 2\,\overline{v\sigma}\,n_{B0}\,t_0\right]^{-1}\,. \tag{28.1.10}$$

We expect that the ratio $(n/n_B)_0$ at the time that T drops to a value $\approx m$ is roughly of order unity, and since in any realistic theory the present ratio $(n/n_B)_\infty$ must be much less than unity, we can neglect the first term in the denominator on the right-hand side of Eq. (28.1.10), and write instead

$$(n/n_B)_\infty \simeq \frac{1}{\overline{v\sigma}\,n_{B0}\,t_0}\,. \tag{28.1.11}$$

The precise value of $\overline{v\sigma}$ depends on the particle spin and its interactions; keeping track only of factors of 2π, the particle mass m, and the electric charge, we can estimate it generally to be of order

$$\overline{v\sigma} \approx \frac{e^4\mathcal{N}}{2\pi m^2} \approx 10^{-3}\frac{\mathcal{N}}{m^2}\,, \tag{28.1.12}$$

where $\mathcal{N}$ is the number of charged particle spin states with mass less than m, into which this particle may annihilate. Also, the age of the universe at a temperature $T_0 \simeq m$ is $t_0 \approx m^4/m_{PL}$, where $m_{PL} \simeq 10^{18}$ GeV, and the density of baryon number is about 10^{-9} times the photon number density, which is of order T^3, so that $n_{B0} \approx 10^{-9}m^3$. Putting this together, we find the present ratio of the new charged particles to baryons:

$$(n/n_B)_\infty \approx 10^{12}\,\frac{m}{m_{PL}\mathcal{N}} \approx 10^{-6}\,\frac{m\,(\text{GeV})}{\mathcal{N}}\,. \tag{28.1.13}$$

These new charged particles would experience the same condensations into galaxies, stars, and planets as ordinary baryons, so this would be the ratio observed today on earth. But experiments[4d] that apply mass spectroscopy to samples of water that have been strongly enriched in heavy water-like molecules by electrolysis have set limits of about $10^{-21}n_B$ on the number density of new charged particles with 6 GeV $< m <$ 330 GeV in terrestrial matter. Thus, even if $\mathcal{N}$ is as large as 1000, these measurements decisively rule out the existence of any new charged untrapped particles in this

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

mass range in the numbers that would have been left over from the early universe.

On the other hand, neutral untrapped particles would be left in intergalactic space. Such particles might well provide the 'missing mass,' that seems to be necessary to account for the gravitational field that governs the motion of galaxies in clusters of galaxies. One of these possible neutral particles is the gravitino, whose cosmological abundance is discussed in Section 28.3. Ellis *et al.*[4c] have extended cosmological considerations to all the new particles required by supersymmetry.

28.2 Supersymmetry and Strong–Electroweak Unification

We shall have to defer a detailed assessment of supersymmetric models of particle physics until we are ready to consider how supersymmetry is broken. In this section we will consider the quantitative application of supersymmetry in one context in which the mechanism for the breakdown of supersymmetry is relatively unimportant, and in which supersymmetry has scored what so far is its greatest empirical success.

If the $SU(3)\times SU(2)\times U(1)$ gauge group of the strong and electroweak interactions is embedded in a simple group G that has the known quarks and leptons (plus perhaps some $SU(3)\times SU(2)\times U(1)$-neutral fermions) as a representation, then, as described in Section 21.5, at energies at or above the scale M_X at which G is spontaneously broken, the $SU(3)\times SU(2)\times U(1)$ coupling constants will be related by

$$g_s^2 = g^2 = \frac{5g'^2}{3} \quad \text{at energies } \geq M_X \,. \tag{28.2.1}$$

At energies far below M_X, these couplings are seriously affected by renormalization corrections. If measured at a scale $\mu < M_X$, the couplings will have values $g_s^2(\mu)$, $g^2(\mu)$, $g'^2(\mu)$, governed by one-loop renormalization group equations

$$\mu\frac{d}{d\mu}g'(\mu) = \beta_1\Big(g'(\mu)\Big)\,, \quad \mu\frac{d}{d\mu}g(\mu) = \beta_2\Big(g(\mu)\Big)\,, \quad \mu\frac{d}{d\mu}g_s(\mu) = \beta_3\Big(g_s(\mu)\Big)\,, \tag{28.2.2}$$

with initial conditions at M_X satisfying Eq. (28.2.1). In the original use[5] of these renormalization group equations, discussed in Section 21.5, the beta functions were calculated in one-loop order to be

$$\beta_1 = \frac{5n_g g'^3}{36\pi^2}\,, \tag{28.2.3}$$

$$\beta_2 = \frac{g^3}{4\pi^2}\left(-\frac{11}{6} + \frac{n_g}{3}\right)\,, \tag{28.2.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$$\beta_3 = \frac{g_s^3}{4\pi^2}\left(-\frac{11}{4} + \frac{n_g}{3}\right), \tag{28.2.5}$$

where n_g is the number of generations of quarks and leptons and the relatively small contributions of scalar fields are here neglected. Since M_X will turn out to be many orders of magnitude larger than the energies accessible with today's accelerators, it seems reasonable to suppose that supersymmetry is unbroken over most of the range below M_X, in which case all of the new fields discussed in the previous section need to be included in calculations of the beta functions in Eq. (28.2.1). These new fields introduce three major changes in the calculations of the beta functions:

1. For every gauge boson, there is a Majorana gaugino with the same $SU(3) \times SU(2) \times U(1)$ quantum numbers. Eq. (17.5.41) shows that the ratio of the contribution to the beta function for any gauge coupling of a Dirac fermion that furnishes a contribution of the gauge group with generators t_A to the contribution of the corresponding gauge boson is $-4C_2/11C_1$, where according to Eqs. (17.5.33) and (17.5.34) the ratio of C_1 and C_2 is given by:

$$\sum_{AB} C_{CAB}C_{DBA} = -(C_1/C_2)\mathrm{Tr}\,(t_C t_D)\,. \tag{28.2.6}$$

For the adjoint representation, $(t_C)_{AB} = iC_{ABC}$, so $C_1 = C_2$, and so a Dirac fermion in the adjoint representation makes a contribution that is $-4/11$ that of the gauge bosons. But the gauginos are Majorana fermions, so their contribution is $-2/11$ that of the gauge boson. Thus the term $11/6$ and $11/4$ in Eqs. (28.2.4) and (28.2.5) are reduced by a factor $9/11$ to $9/6$ and $9/4$, respectively.

2. For every left-handed quark, lepton, antiquark, or antilepton field, there is a complex scalar field with the same $SU(3) \times SU(2) \times U(1)$ quantum numbers. Following the same method as in Section 17.5, it is not hard to calculate that the contribution of a complex scalar field belonging to a representation of a gauge group with generators t_A to the beta function for the gauge coupling g_i is

$$[\beta_i(g_i)]_{\text{scalar}} = \frac{g_i^3 C_{2i}}{48\pi^2}\,, \tag{28.2.7}$$

where $\mathrm{Tr}\,(t_A t_B) = g_i^2 C_{2i}\delta_{AB}$. This is $1/4$ the contribution of a Dirac spinor field in the same representation, given by Eq. (18.7.2), and hence $1/2$ the contribution of each left-handed spinor field (including the complex conjugates of the right-handed components of the Dirac fields). Thus the coefficients of n_g in Eqs. (28.2.3)–(28.2.5) should be increased by a factor $3/2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

3. The decrease by a factor 9/11 of the negative gauge boson terms in the beta function and the increase by a factor 3/2 of the positive squark and slepton terms both lead to a general decrease in the rate at which the three gauge coupling constants diverge below M_X from the ratios (28.2.1). This will increase our estimate of M_X, but, as we shall see, in itself it would have no affect on the prediction of the electroweak mixing parameter $\sin^2\theta$. But these changes do enhance the relative contribution of the Higgs scalars, which was neglected in Eqs. (28.2.3)–(28.2.5), and which is now also accompanied with the larger contribution of the accompanying higgsinos. With n_s of the superfields (H_1^0, H_1^-) or (H_2^+, H_2^0) discussed in the previous section, the constant C_{2i} in Eq. (28.2.7) is $[(1/2)^2 + (-1/2)^2]n_s = n_s/2$ for $SU(2)$, and is $2n_s(\pm 1/2)^2 = n_s/2$ for $U(1)$. According to Eq. (28.2.7), the scalar components of these superfields make a contribution to β_1 equal to $n_s g'^3/96\pi^2$ and a contribution to β_2 also equal to $n_s g^3/96\pi^2$. As we have seen, the Majorana higgsinos contribute twice as much to the beta functions as complex scalars with the same quantum numbers, so the superfields (H_1^0, H_1^-) or (H_2^+, H_2^0) make a total contribution to β_1 and β_2 that is 3/2 the contribution of the Higgs scalars, and hence equal to $n_s g'^3/32\pi^2$ and $n_s g^3/32\pi^2$, respectively.

Making all these changes in the beta functions, we now have

$$\beta_1 = \frac{g'^3}{4\pi^2}\left(\frac{5n_g}{6} + \frac{n_s}{8}\right), \tag{28.2.8}$$

$$\beta_2 = \frac{g^3}{4\pi^2}\left(-\frac{9}{6} + \frac{n_g}{2} + \frac{n_s}{8}\right), \tag{28.2.9}$$

$$\beta_3 = \frac{g_s^3}{4\pi^2}\left(-\frac{9}{4} + \frac{n_g}{2}\right). \tag{28.2.10}$$

The solutions of the renormalization group equations (28.2.2) are then

$$\frac{1}{g'^2(\mu)} = \frac{1}{g'^2(M_X)} + \frac{1}{2\pi^2}\left(\frac{5n_g}{6} + \frac{n_s}{8}\right)\ln\left(\frac{M_X}{\mu}\right), \tag{28.2.11}$$

$$\frac{1}{g^2(\mu)} = \frac{1}{g^2(M_X)} + \frac{1}{2\pi^2}\left(-\frac{3}{2} + \frac{n_g}{2} + \frac{n_s}{8}\right)\ln\left(\frac{M_X}{\mu}\right), \tag{28.2.12}$$

$$\frac{1}{g_s^2(\mu)} = \frac{1}{g_s^2(M_X)} + \frac{1}{2\pi^2}\left(-\frac{9}{4} + \frac{n_g}{2}\right)\ln\left(\frac{M_X}{\mu}\right). \tag{28.2.13}$$

It is convenient to take $\mu = m_Z$, so that $SU(2) \times U(1)$ can be regarded as unbroken over almost all of the range of energies in which we use the formulas (28.2.11)–(28.2.13). Using Eq. (28.2.1), the difference between Eqs. (28.2.12) and (28.2.13) gives

$$\frac{1}{g^2(m_Z)} - \frac{1}{g_s^2(m_Z)} = \frac{1}{2\pi^2}\left(\frac{3}{4} + \frac{n_s}{8}\right)\ln\left(\frac{M_X}{m_Z}\right), \tag{28.2.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

Table 28.1. Values of the electroweak mixing parameter $\sin^2\theta$ and unification mass M_X given by Eqs. (28.2.17) and (28.2.18), as functions of the number n_s of left-chiral superfield doublets (H_1^0, H_1^-) or (H_2^+, H_2^0).

n_s	$\sin^2\theta$	M_X (GeV)
0	0.203	8.7×10^{17}
2	0.231	2.2×10^{16}
4	0.253	1.1×10^{15}

while the difference between Eq. (28.2.12) and 3/5 of Eq. (28.2.11) gives

$$\frac{1}{g^2(m_Z)} - \frac{3}{5g'^2(m_Z)} = \frac{1}{2\pi^2}\left(-\frac{3}{2} + \frac{n_s}{20}\right)\ln\left(\frac{M_X}{m_Z}\right) . \tag{28.2.15}$$

Eq. (21.3.19) allows us to express the electroweak couplings in terms of the electroweak mixing angle θ and the positron charge e:

$$g(m_Z) = -e(m_Z)/\sin\theta , \qquad g'(m_Z) = -e(m_Z)/\cos\theta . \tag{28.2.16}$$

We can then solve for the unknowns $\ln(M_X/m_Z)$ and $\sin^2\theta$ in terms of input parameters $e(m_Z)$ and $g_s(m_Z)$:

$$\sin^2\theta = \frac{18 + 3n_s + (e^2(m_Z)/g_s^2(m_Z))(60 - 2n_s)}{108 + 6n_s} , \tag{28.2.17}$$

$$\ln\left(\frac{M_X}{m_Z}\right) = \left(\frac{8\pi^2}{e^2(m_Z)}\right)\left(\frac{1 - (8e^2(m_Z)/3g_s^2(m_Z))}{18 + n_s}\right) . \tag{28.2.18}$$

For $n_s = 0$ Eq. (28.2.17) gives the same result (21.5.15) for $\sin^2\theta$ as was originally calculated (ignoring the small contribution of Higgs scalars) in non-supersymmetric theories, but the value (28.2.18) of $\ln(M_X/m_Z)$ is larger than the original result (21.5.16) by a factor 11/9, which as we have seen arises from gaugino contributions to the beta functions.

Using the same input parameters $e^2(m_Z)/4\pi = (128)^{-1}$, $g_s^2(m_Z)/4\pi = 0.118$, $m_Z = 91.19$ GeV as in Section 21.5 now gives the numerical results shown in Table 28.1. As discussed in the previous section, the necessity of cancelling anomalies in the electroweak currents requires equal numbers of (H_1^0, H_1^-) and (H_2^+, H_2^0) doublets, so we consider only even values of the number n_s of these superfields.

Remarkably, the value $n_s = 2$ for the simplest plausible theory yields a value[6] $\sin^2\theta = 0.231$ which is in perfect agreement with the experimentally

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

observed value, $\sin^2\theta = 0.23$. The value of M_X is 20 times greater[7] than calculated in this way in non-supersymmetric theories, leading to a decrease by a factor 20^{-4} in the rate for proton decay processes like $p \to \pi^0 + e^+$, thus removing a conflict with the experimental non-observation of such processes. (Proton decay is discussed in more detail in Section 28.7.) This increase in the value of M_X brings it closer to the energy scale $\approx 10^{18}$ GeV at which gravitation has the same strength as other interactions. It may be that this remaining gap may be filled by a change in gravitational interactions at very high energies.[7a]

A value $n_s = 4$ would give a value for $\sin^2\theta$ in serious disagreement with experiment, and a value of M_X low enough to revive the conflict with expectations for proton decay. This makes a strong case for having just one of each superfield (H_1^0, H_1^-) and (H_2^+, H_2^0).

Unlike the calculated values of $\sin^2\theta$ and M_X, the calculated value of the common gauge coupling (28.2.1) at M_X does depend on the number of generations as well as the number of scalar doublets. With $n_g = 3$ and $n_s = 2$ and our previous input parameters, Eq. (28.2.13) gives

$$\frac{g^2(M_X)}{4\pi} = \frac{g_s^2(M_X)}{4\pi} = \frac{1}{17.5} . \tag{28.2.19}$$

28.3 Where is Supersymmetry Broken?

Supersymmetry if valid at all is certainly not apparent in the menu of known particles, so any consideration of the implications of supersymmetry at ordinary energies requires us to make some assumption about the mechanism of supersymmetry breaking. It would be simplest to suppose that supersymmetry is broken like $SU(2) \times U(1)$, by effects occurring in the tree approximation of the supersymmetric standard model. This possibility may be definitely ruled out.

One argument against a tree-approximation breakdown of supersymmetry is based on the mass sum rule (27.5.11), which holds separately for each value of the unbroken conserved quantities color and electric charge. In the color-triplet sector with electric charge $-e/3$ the only known fermions are the d, s, and b quarks, for which

$$m_d^2 + m_s^2 + m_b^2 \simeq (5 \text{ GeV})^2 . \tag{28.3.1}$$

According to the sum rule, if there are no other fermions with this color and charge, then the sum of all squared masses for bosons (counting each spin state separately) with the same color and charge must equal about $2(5 \text{ GeV})^2$. In particular, each of the squarks with this color and charge must have a mass no greater than 7 GeV. The existence of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

such light squarks is definitely ruled out experimentally; they would have shown up, for instance, as a contribution to the rate for electron–positron annihilation into hadrons at energies where this process has been studied very thoroughly.

This argument could be invalidated if there were a heavy fourth generation of quarks. There is another argument, due to Dimopoulos and Georgi,[3] which would apply however many heavy quarks there are and which yields an even stronger upper bound on the mass of the lightest squark. The unbroken conservation of charge and color tells us that the only non-zero D_{A0}-terms in the supersymmetric standard model are for the generators y of $U(1)$ and t_3 of $SU(2)$, which we shall call D_1 and D_2, respectively. The values of these generators are $y = -g'/6$ and $t_3 = +g/2$ for the left-handed quarks of charge $2e/3$; $y = -g'/6$ and $t_3 = -g/2$ for the left-handed quarks of charge $-e/3$; $y = 2g'/3$ and $t_3 = 0$ for the right-handed quarks of charge $2e/3$; and $y = -g'/3$ and $t_3 = 0$ for the right-handed quarks of charge $-e/3$. Also, the squark fields are color-triplets, and therefore cannot have vacuum expectation values. According to Eq. (27.5.4), the mass-squared matrix of the charge $2e/3$ color-triplet (not antitriplet) squarks is

$$M^2_{0U} = \begin{bmatrix} \mathscr{M}^*_U \mathscr{M}_U - g'D_1/6 + gD_2/2 & \mathscr{F}^*_U \\ \mathscr{F}_U & \mathscr{M}_U \mathscr{M}^*_U + 2g'D_1/3 \end{bmatrix}, \quad (28.3.2)$$

while the mass-squared matrix of the charge $-e/3$ color-triplet squarks is

$$M^2_{0D} = \begin{bmatrix} \mathscr{M}^*_D \mathscr{M}_D - g'D_1/6 - gD_2/2 & \mathscr{F}^*_D \\ \mathscr{F}_D & \mathscr{M}_D \mathscr{M}^*_D - g'D_1/3 \end{bmatrix}. \quad (28.3.3)$$

Also, Eq. (27.5.6) gives the mass-squared matrices of the quarks of charge $2e/3$ and $-e/3$ here as just $\mathscr{M}^*_U \mathscr{M}_U$ and $\mathscr{M}^*_D \mathscr{M}_D$, respectively, with no mixing with the gauginos.

Now let v_u and v_d be the normalized eigenvectors of the quark mass-squared matrices $\mathscr{M}^*_U \mathscr{M}_U$ and $\mathscr{M}^*_D \mathscr{M}_D$ corresponding to the quarks u and d of lowest mass, and consider the expectation values of the corresponding squark mass-squared matrices

$$\begin{bmatrix} 0 \\ v^*_u \end{bmatrix}^\dagger M^2_{0U} \begin{bmatrix} 0 \\ v^*_u \end{bmatrix} = m^2_u + \frac{2g'D_1}{3}, \quad (28.3.4)$$

$$\begin{bmatrix} 0 \\ v^*_d \end{bmatrix}^\dagger M^2_{0D} \begin{bmatrix} 0 \\ v^*_d \end{bmatrix} = m^2_d - \frac{g'D_1}{3}. \quad (28.3.5)$$

These expectation values are weighted averages of the squared masses of the squarks of charge $2e/3$ and $-e/3$, respectively, so at least one squark of charge $2e/3$ must have a squared mass less than $m^2_u + 2g'D_1/3$, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

at least one squark of charge $-e/3$ must have a squared mass less than $m_d^2 - g'D_1/3$. Thus, depending on the sign of D_1, *there must be either a squark of charge* $2e/3$ *lighter than the u quark, or a squark of charge* $-e/3$ *lighter than the d quark.*

Needless to say, the existence of a charged color-triplet scalar this light would radically change strong-interaction phenomenology. Like the u and d quarks, this colored scalar would appear as an ingredient of hadrons with a 'constituent' mass of a few hundred MeV, which is certainly not seen. Since this scalar is electrically charged, it would also be created in pairs in e^+–e^- annihilation at energies above a few hundred MeV, making a contribution to the annihilation cross-section that would destroy the excellent agreement between theory and experiment for this cross-section. Even worse, since the u and d quarks are so light, and D_1 is expected to be of the order of the supersymmetry-breaking scale, Eqs. (28.3.4) and (28.3.5) indicate that one of the squarks would have a negative squared mass, meaning that this squark field would have to develop a non-vanishing expectation value, breaking both color and charge conservation. We are forced to reject the simple picture of supersymmetry broken spontaneously in the tree approximation in a supersymmetric version of the standard model.

One way out of this conclusion would be to add another $U(1)$ gauge superfield to the theory. If all the quark superfields carry the same value $\tilde{g}$ of this new $U(1)$ generator, then the corresponding D-term $\tilde{D}$ would make an additive contribution $\tilde{g}\tilde{D}$ to the right-hand sides of both Eqs. (28.3.4) and (28.3.5). If this term were sufficiently large, then it could give a large positive value to all of the squark squared masses, avoiding all the problems mentioned above. But there is no sign of such a new neutral gauge boson at accessible energies, and in any case we would still have an upper bound of 7 GeV on the masses of all the squarks of charge $-e/3$.

It is not necessarily a bad thing that we have to look for the breaking of supersymmetry elsewhere than in the tree approximation of the supersymmetric standard model. If supersymmetry were broken in this approximation, then the characteristic mass that sets the scale of supersymmetry breaking would be some mass parameter in the Lagrangian, which would in turn set the scale of all other masses in the standard model. We would then still be confronted with the hierarchy problem: why is this mass scale so much less than 10^{16}–10^{18} GeV?

There is one known way to explain such large mass ratios. If supersymmetry is not spontaneously broken in the tree approximation in whatever field theory unifies all the interactions at some high mass scale M_X, then as shown in Section 27.6, it will not be broken in any order of perturbation theory. But it can be broken by non-perturbative effects. In particular, if there is some gauge field with an asymptotically free gauge coupling

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$\mathcal{G}(\mu)$ at renormalization scale μ, and if $\mathcal{G}^2(\mu)/8\pi^2$ is substantially less than unity for $\mu \approx M_X$, then as discussed in Section 18.3, this gauge interaction will become strong at an energy of order $M_S = M_X \exp(-8\pi^2 b/\mathcal{G}^2(M_X))$, where b is a number of order unity. It is not necessary for $\mathcal{G}^2(M_X)/8\pi^2$ to be very small in order to have M_S many orders of magnitude less than M_X. We will see in Section 29.4 that supersymmetry can indeed be broken in just this way, by a gauge coupling that has become strong at some energy $M_S \ll M_X$. Indeed, this is just what happens to chiral symmetry in quantum chromodynamics; there is no mystery why the proton mass (or at least its main part, due to the dynamical breakdown of chiral symmetry, and not to the tiny masses of the u and d quarks) is so much smaller than the unification scale M_X. Alternatively, the forces that are strong at energy M_S may produce a potential for scalar fields, whose vacuum expectation value then breaks supersymmetry.

There are no signs of any new strong interaction of the known quarks and leptons, so we have to assume that the observed particles of the standard model are neutral with respect to the strong force that breaks supersymmetry. Supersymmetry breaking therefore occurs in a 'hidden sector' of particles that do feel this new strong force. The remaining question, then, is what is the mechanism by which supersymmetry breaking in this hidden sector is communicated to the known particles of the standard model? As we shall see, most of our expectations for the phenomenological implications of supersymmetry depend on the answer to this question, rather than on the details of the breakdown of supersymmetry itself.

Of course, the mechanism for communicating supersymmetry breaking to observed particles must be some sort of interaction that is felt by these particles. There are two leading candidates. One mechanism is provided by the $SU(3) \times SU(2) \times U(1)$ gauge interactions themselves, to be discussed in Section 28.6. The other is gravitation, or rather the auxiliary fields that are superpartners of the gravitational field, to be discussed in Sections 31.4 and 31.7.

Without going into details here, we can make a crude estimate of the supersymmetry-breaking scale M_S for these two possibilities. For gauge-mediated supersymmetry breaking, we expect that the mass splitting between the observed quarks, leptons, and gauge bosons and their superpartners would be of order $g_s^2/16\pi^2$ or $g'^2/16\pi^2$ or $g^2/16\pi^2$ (where g_s, g, and g' are the $SU(3)$, $SU(2)$, and $U(1)$ gauge couplings), depending on which quantum numbers are carried by the supermultiplet in question. (This guess is verified in Section 28.6.) Hence if the squarks, sleptons, and gauginos have masses in the range of 100 GeV to 10 TeV, as argued at the end of Section 28.1, then the supersymmetry-breaking scale M_S would be higher by two or three orders of magnitude — say, of order 100 TeV. On the other hand, if it is gravitation that serves as the mediator of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

supersymmetry breaking, then on dimensional grounds we would expect the mass splittings Δm between the observed particles and their superpartners to be of order $\sqrt{G}M_S^2$, or perhaps of order GM_S^3. (Results of both sorts will be encountered in models described in Section 31.7.) If the squarks, sleptons, and gauginos have masses in the range of 100 GeV to 10 TeV, then M_S would be of order 10^{11} GeV for $\Delta m \approx \sqrt{G}M_S^2$, or 10^{13} GeV for $\Delta m \approx GM_S^3$.

The large difference in estimates of the supersymmetry-breaking scale M_S for gauge- and gravitation-mediated supersymmetry breaking makes an important difference in particle phenomenology and cosmology. As already mentioned several times, supersymmetry dictates that the graviton must have a partner of spin 3/2, the gravitino. When supersymmetry is spontaneously broken at a scale M_S, the gravitino acquires a mass m_g of order $\sqrt{G}M_S^2$. (A precise formula will be given in Section 31.3.) For gauge-mediated supersymmetry breaking, this is very small; if $M_S \approx 100$ TeV then $m_g \approx 1$ eV, so the gravitino would be by far the lightest of the new particles required by supersymmetry — that is, the lightest particle with negative R parity (28.1.6). On the other hand, for gravitationally mediated supersymmetry breaking the gravitino mass is just of the same order of magnitude $\sqrt{G}M_S^2$ as the mass splitting between known particles and their superpartners, so the gravitino would have roughly the same mass as the squarks, sleptons, and gauginos. The gravitino then might or might not be the lightest particle with negative R parity, but its interactions with known particles and their superpartners in this case are of gravitational strength, so that gravitinos would play no direct role in experiments on elementary particles.

* * *

There are limits on the number of gravitinos that could survive from the big bang, which set useful constraints on the scale M_S of supersymmetry breaking. At some point in the distant past the temperature T was presumably high enough so that even purely gravitational interactions would have kept gravitinos in thermal equilibrium with other particles, in which case the number density of gravitinos would have been of the order of T^3, roughly the same as the number density of photons. (We are using units in which the Boltzmann constant k_B as well as $\hbar$ and c are equal to unity.) If gravitinos do not annihilate or decay, then the expansion of the universe will lower their number density in the same way as the number density of photons, so even after the gravitinos go out of equilibrium they would be present in numbers comparable to those of photons. More precisely, since the photons but not the gravitinos are heated by the annihilation of other particles, the number density n_{g0} of gravitinos at present would be one or two orders of magnitude less than

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

the number density $n_{\gamma 0}$ of photons in the cosmic microwave radiation background. In order for the mass density $m_g n_{g0}$ of the gravitinos not to exceed the upper bound on the cosmic mass density set by the observed value of the Hubble constant, m_g would have to be less[8] than about 1 keV. As we have seen, this limit is well satisfied in theories of gauge-mediated supersymmetry breaking, where the gravitino is too light for cosmic gravitinos to contribute appreciably to the mass density of the universe. Since some of the fields that break supersymmetry in these theories must interact at least indirectly with the known quark, lepton, and gauge fields in order for the known particles to show the effects of supersymmetry breaking, the interactions of the gravitino with the known particles and their superpartners are suppressed only by powers of gauge and Yukawa coupling constants, so all the superpartners of the quarks, leptons, and gauge bosons would decay quickly into these known particles and gravitinos. Thus these particles also do not provide candidates in these models for the 'missing mass' sought by cosmologists. (It is possible that conservation laws could keep some particles of the supersymmetry-breaking sector stable, in which case they could conceivably serve as the missing mass.)

On the other hand, for supersymmetry breaking that is gravitationally mediated, gravitinos are heavy enough to be unstable (though gravitino annihilation is still negligible), so the above limit need not apply.[9] We will see in Section 31.3 that the coupling of the gravitino to other fields is proportional to $\sqrt{G}$, so on dimensional grounds the decay rate Γ_g of a gravitino at rest is roughly of the order of Gm_g^3. This is to be compared with the rate of expansion of the universe, which at temperature T is of order $\sqrt{GT^4}$. (We are here ignoring factors of order 10–100, including those involving non-gravitational coupling constants and the number of particle species.) When the cosmic temperature drops to the value $T \approx m_g$ at which gravitinos become non-relativistic, the ratio of their decay rate to the expansion rate is of order $\sqrt{G}m_g = m_g/m_{\text{Planck}} \ll 1$, so gravitino decay becomes significant only after this time, when the gravitinos are highly non-relativistic. As we have seen, their number density will be of order T^3, so their energy density will then be of order $m_g T^3$, which is greater than the energy density of order T^4 of the photons and other particles in thermal equilibrium at temperature T, and therefore makes the dominant contribution to the cosmic gravitational field that governs the rate of expansion of the universe. The expansion rate under these conditions is therefore of order $\sqrt{Gm_g T^3}$, and gravitino decay becomes significant when this equals the gravitino decay rate of order Gm_g^3, and therefore at a temperature

$$T_g \approx G^{1/3} m_g^{5/3} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

As we have seen, if these gravitinos did not decay before the present then their mass had better be less than 1 keV, but they can lead to cosmological difficulties even if they did decay before now. After they decay, their energy must go into the energy of photons and other relativistic particles, so the temperature T_g' after decay is related to the temperature T_g calculated above by the energy conservation condition $m_g T_g^3 \approx T_g'^4$, and hence

$$T_g' \approx G^{1/4} m_g^{3/2} .$$

In particular, since $T_g \ll m_g$, we have $T_g' \gg T_g$. If T_g were less than the temperature $T_n \simeq 0.1$ MeV at which cosmological nucleosynthesis can occur, then gravitinos would still be abundant before nucleosynthesis, giving a higher energy density and hence a faster expansion, so that there would be less time for free neutrons to decay before being incorporated into complex nuclei, and hence more helium would be produced when nucleosynthesis occurs. Also, the ratio of the photon and baryon densities would have been subsequently increased by gravitino decay, so this ratio at the time of nucleosynthesis would have been considerably less than is usually estimated from the present cosmic microwave background temperature, and so nuclear reactions would have incorporated neutrons more completely into helium, and less deuterium would be left today. The present agreement between theory and observation for the cosmic helium and deuterium abundances would thus be destroyed. This problem is avoided if $T_g > 0.1$ MeV, but it can also be avoided under the much weaker condition that $T_g' > 0.4$ MeV, because then after the gravitinos decay the temperature would have been high enough to break up the excess helium and give cosmological nucleosynthesis a fresh start as the universe recools. This condition requires that $m_g > 10$ TeV, which is just barely consistent with the upper bound derived in Section 28.1 on the masses of the superpartners of the known quarks, leptons, and gauge bosons, which for gravitationally mediated supersymmetry breaking are of order m_g. This limit on m_g corresponds to a supersymmetry-breaking scale $M_S > 10^{11}$ GeV for $m_g \approx \sqrt{G} M_S^2$ or $M_S > 10^{13}$ GeV for $m_g \approx G M_S^3$.

28.4 The Minimal Supersymmetric Standard Model

In the previous section we identified two different ways that the breakdown of supersymmetry at a high energy scale M_S could be communicated to the known quarks and leptons: through gauge or gravitational superfields. The supersymmetry-breaking terms in the resulting low-energy effective Lagrangian will then be suppressed by powers of gauge couplings or the Newton constant. Most of these terms will therefore be rather small,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

with the exception that, along with factors of gauge couplings or the Newton constant, on dimensional grounds the mass terms and other superrenormalizable terms in the effective Lagrangian will be proportional to one or more factors of the supersymmetry-breaking scale M_S, which is quite large compared with known particle masses. We can conclude then that, to a fair approximation for gauge-mediated supersymmetry breaking and a very good approximation for gravitationally mediated supersymmetry breaking, the main effect of supersymmetry breaking will be in the superrenormalizable terms of the effective Lagrangian of the supersymmetric standard model. This version of the standard model,[10] which is supersymmetric except for superrenormalizable terms, is usually known as the *minimal supersymmetric standard model.*

With R parity or $B-L$ conserved, the most general superrenormalizable Lagrangian density allowed by $SU(3)\times SU(2)\times U(1)$ gauge symmetry takes the form

$$\begin{aligned}
\mathscr{L}_{SR} = &-\sum_{ij} M^{2\,Q}_{ij}\left(\mathscr{Q}^\dagger_i \mathscr{Q}_j\right) - \sum_{ij} M^{2\,\bar{U}}_{ij}\left(\bar{\mathscr{U}}^\dagger_i \bar{\mathscr{U}}_j\right) - \sum_{ij} M^{2\,\bar{D}}_{ij}\left(\bar{\mathscr{D}}^\dagger_i \bar{\mathscr{D}}_j\right) \\
&-\sum_{ij} M^{2\,L}_{ij}\left(\mathscr{L}^\dagger_i \mathscr{L}_j\right) - \sum_{ij} M^{2\,\bar{E}}_{ij}\left(\bar{\mathscr{E}}^\dagger_i \bar{\mathscr{E}}_j\right) \\
&-\left(\overline{\lambda_3}\, m_{\text{gluino}}\, \lambda_3\right) - \left(\overline{\lambda_2}\, m_{\text{wino}}\, \lambda_2\right) - \left(\overline{\lambda_1}\, m_{\text{bino}}\, \lambda_1\right) \\
&-\sum_{ij} A^D_{ij}\, h^D_{ij}\left(\mathscr{Q}^{\mathrm{T}}_i\, e\, \mathscr{H}_1\right)\bar{\mathscr{D}}_j - \sum_{ij} A^E_{ij}\, h^E_{ij}\left(\mathscr{L}^{\mathrm{T}}_i\, e\, \mathscr{H}_1\right)\bar{\mathscr{E}}_j \\
&-\sum_{ij} A^U_{ij}\, h^U_{ij}\left(\mathscr{Q}^{\mathrm{T}}_i\, e\, \mathscr{H}_2\right)\bar{\mathscr{U}}_j - \sum_{ij} C^D_{ij}\, h^D_{ij}\left(\mathscr{Q}^{\mathrm{T}}_i \mathscr{H}^*_2\right)\bar{\mathscr{D}}_j \\
&-\sum_{ij} C^E_{ij}\, h^E_{ij}\left(\mathscr{L}^{\mathrm{T}}_i \mathscr{H}^*_2\right)\bar{\mathscr{E}}_j - \sum_{ij} C^U_{ij}\, h^U_{ij}\left(\mathscr{Q}^{\mathrm{T}}_i \mathscr{H}^*_1\right)\bar{\mathscr{U}}_j \\
&-B\mu\left(\mathscr{H}^{\mathrm{T}}_2 e \mathscr{H}_1\right) + \text{H.c.}
\end{aligned} \tag{28.4.1}$$

Script letters are used here to denote the scalar components of left-chiral superfields. Sums over $SU(2)$ and color indices are understood, as necessary for invariance under $SU(3)\times SU(2)\times U(1)$, with e the usual antisymmetric 2×2 matrix $i\sigma_2$. All coefficients may be complex, and the gaugino masses may involve terms proportional to the γ_5 as well as the unit matrix.

We follow the custom here of writing the coefficients of the terms involving scalar fields but not their adjoints as equal to the coefficients of the corresponding supersymmetric $\mathscr{F}$-terms in Eq. (28.1.7) times factors A^D_{ij}, A^E_{ij}, A^U_{ij}, and B. This is motivated by the consideration that the smallness of the Yukawa couplings of the light quarks in Eq. (28.1.7) reflects a number of approximate chiral symmetries that if extended to the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

whole supermultiplet would also make the corresponding trilinear terms in Eq. (28.4.1) small, while the appearance of the μ-term in Eq. (28.1.7) violates a possible Peccei–Quinn[4] symmetry that if approximately valid would make both μ and $B\mu$ small. Similar considerations suggest the form in which we write the coefficients of the terms that involve both scalars and their complex conjugates. Also, in Section 31.4 we will describe contributions to Ah and $B\mu$ that really are proportional to h and μ, respectively. However, here we are leaving it an open question whether the Ah, Ch, and $B\mu$ coefficients in Eq. (28.4.1) are necessarily small when the corresponding h and μ coefficients in Eq. (28.1.7) are small.

The Ch-terms in Eq. (28.4.1) have generally been omitted in discussions of the minimal supersymmetric standard model. This is partly because, as discussed in Section 27.7, terms like these that involve the ϕ-components of left-chiral scalar superfields and also their complex conjugates can potentially produce quadratic divergences and thereby raise fine-tuning problems. But we saw in Section 27.7 that the quadratic divergences occur only in 'tadpole' graphs in which a scalar field line disappears into the vacuum, and in the minimum supersymmetric standard model there are no scalars that are neutral under all gauge symmetries, and hence no scalar tadpoles. The Ch-terms are absent in the theories of gravitationally mediated supersymmetry breaking discussed in Section 31.6, and small in the theories of gauge-mediated supersymmetry breaking described in Section 28.6, but there is no reason to suppose that this will always be the case.

Even though they are not supersymmetric, superrenormalizable interactions like those in Eq. (28.4.1) are shown in Section 27.7 to not produce supersymmetry-violating ultraviolet-divergent corrections to the coefficients of the supersymmetric $d = 4$ interactions. The condition of supersymmetry that is imposed on the dimensionless couplings of the minimal supersymmetric standard model therefore does not get in the way of the cancellation of ultraviolet divergences by renormalization of coupling constants. It was this property, rather than any theory of supersymmetry breaking in a high energy hidden sector, that motivated the introduction of the minimal supersymmetric standard model in Reference 10.

The best reason today for studying the implications of the supersymmetric standard model is that, as already mentioned, theories in which supersymmetry is spontaneously broken at a high energy scale are naturally described by the minimal supersymmetric standard model at much lower energies. We can explore the phenomenological implications of the supersymmetric standard model, and be reasonably confident that the results will be relevant whatever detailed model of supersymmetry breaking and its mediation turns out to be correct.

Even without the Ch-terms, if all of the other coefficients in the Lagrangian are constrained only by gauge symmetries and R parity

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

conservation, the minimal supersymmetric standard model will contain over 100 free parameters.[11] Here 'minimal' means nothing more than that the theory contains only a minimal menu of superfields. Sometimes the term 'minimal supersymmetric standard model' is reserved for models that also satisfy restrictions on the coefficients of the superrenormalizable terms, motivated either by some underlying theory or by empirical constraints. For instance, the minimal supersymmetric standard model is sometimes optimistically assumed to satisfy the universality conditions

$$\begin{aligned} &M_{ij}^{2\,Q} = M_{ij}^{2\,\bar{D}} = M_{ij}^{2\,\bar{U}} = M_{ij}^{2\,L} = M_{ij}^{2\,\bar{E}} = M^2\delta_{ij}\,, \\ &m_{\text{gluino}} = m_{\text{wino}} = m_{\text{bino}}\,, \\ &A_{ij}^{D} = A_{ij}^{E} = A_{ij}^{U} = A\,, \qquad C_{ij}^{D} = C_{ij}^{E} = C_{ij}^{U} = 0\,. \end{aligned} \tag{28.4.2}$$

Often these conditions are imposed at the scale $M_X \approx 10^{16}$ GeV of coupling constant unification, with corrections produced only by the renormalization group flow to lower energies. We will not be making such assumptions here.

In analyzing the phenomenological implications of the minimal supersymmetric standard model, we must deal not only with the search for new particles, but also with two classes of severe empirical constraint on processes involving known particles: the experimental upper bounds on various flavor non-conserving processes, and on various modes of CP non-conservation.

Flavor Changing Processes

We saw in Section 21.3 that there is an automatic suppression of flavor-changing processes like K^0–$\bar{K}^0$ oscillations and $K^0 \to \mu^+\mu^-$ in the non-supersymmetric standard model. This is due to the feature of this theory, that it is only the mass splittings of the quarks that prevent them from being defined so that each flavor is separately conserved, so the amplitude of these flavor-changing process must be proportional to several factors of small quark masses. Also, in this theory lepton flavor is automatically conserved, so that processes like $\mu \to e\gamma$ are absolutely forbidden. These satisfactory results are put at risk by the presence of squarks and sleptons in supersymmetric extensions of the standard model, because there is in general no reason to expect that the squark and slepton mass matrices will be diagonal in the same basis as the quark and lepton mass matrices. This does not introduce flavor changing in the interaction of these particles with gauge bosons, which are flavor-independent, but it can produce flavor-changing transitions in which squarks or sleptons turn into quarks or leptons with the emission or absorption of gauginos. Of course, there

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

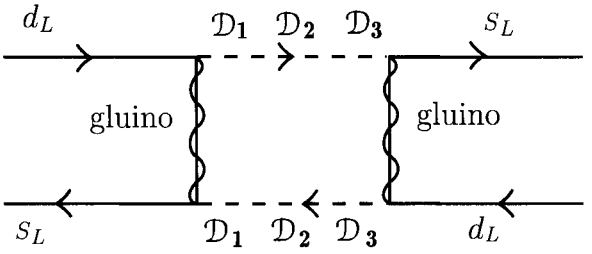

Figure 28.1. A one-loop diagram that can contribute to the $\Delta S = 2$ effective interaction $(\bar{s_L}\gamma^\mu d_L)(\bar{s_L}\gamma_\mu d_L)$ in the supersymmetric standard model. Here solid lines are quarks; dashed lines are squarks; and combined solid and wavy lines are gluinos.

is no problem if the squarks and sleptons are degenerate, in which case their mass matrices are diagonal in any basis.

The most stringent limits on squark mass splittings and/or mixing angles are set by measurements of K^0–$\bar{K}^0$ transitions.[12] These transitions are produced by operators in the effective low-energy Lagrangian density like $(\bar{s_L}\gamma^\mu d_L)(\bar{s_L}\gamma_\mu d_L)$, which can be produced by diagrams like Figure 28.1. The superpartners of the quarks d_L and s_L are in general linear combinations $\sum_i V_{di}\mathscr{D}_i$ and $\sum_i V_{si}\mathscr{D}_i$ of squarks $\mathscr{D}_i$ of definite mass, where V_{ji} is a 3×3 unitary matrix, so the two squark propagators in this diagram contribute a factor

$$\sum_i \frac{V_{di}V_{si}^*}{k^2 + M_i^2 - i\epsilon} \times \sum_j \frac{V_{dj}V_{sj}^*}{k^2 + M_j^2 - i\epsilon},$$

where k is the four-momentum circulating in the loop. Because V_{ji} is unitary, this vanishes if the three squark masses M_i are all equal. If the squark square masses differ from some common value M^2_{squark} by relatively small amounts ΔM_i^2, then this becomes

$$\left(\frac{1}{k^2 + M^2_{\text{squark}} - i\epsilon}\right)^4 \left(\sum_i V_{di}V_{si}^*\Delta M_i^2\right)^2.$$

The amplitude for $d_L\bar{s_L} \to s_L\bar{d_L}$ has dimensionality mass^{-2}, so after multiplying by the gluino propagators and four factors of the strong coupling g_s, and integrating over k, we must get an amplitude proportional to

$$\frac{g_s^4}{\tilde{M}^6}\left(\sum_i V_{di}V_{si}^*\,\Delta M_i^2\right)^2, \tag{28.4.3}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

where $\tilde{M}$ is the larger of M_{squark} and m_{gluino}. This may be compared with the result for this amplitude in the non-supersymmetric standard model, which is produced by W exchange, as shown in Figure 28.2. Ignoring the third generation of quarks, which has only small transition amplitudes to the first two generations, the amplitudes for $d \to u$, $d \to c$, $s \to u$ and $s \to c$ by W^- emission are, respectively, $\cos\theta_c$, $-\sin\theta_c$, $\sin\theta_c$, and $\cos\theta_c$, where θ_c is the Cabibbo angle defined in Section 21.3. Hence in place of the squark propagators here we have the quark propagators

$$\sin\theta_c \cos\theta_c \left(\frac{i\not{k} + m_u}{k^2 + m_u^2 - i\epsilon} - \frac{i\not{k} + m_c}{k^2 + m_c^2 - i\epsilon} \right),$$

and in place of the strong coupling g_s here we have the $SU(2)$ coupling g. Thus in the non-supersymmetric standard model the amplitude for $d_L\overline{s_L} \to s_L\overline{d_L}$ is proportional to

$$\frac{g^4 \sin^2\theta_c \cos^2\theta_c}{m_W^4} \left(m_c - m_u\right)^2, \tag{28.4.4}$$

with a proportionality coefficient of the same order as that in Eq. (28.4.3). With a plausible guess on how to calculate the K^0–$\bar{K}^0$ transition amplitude from the amplitude for $d_L\overline{s_L} \to s_L\overline{d_L}$, the amplitude corresponding to Figure 28.2 is known to give a result in good agreement with experiment. (Indeed, Gaillard and Lee[13] used this calculation to predict that $m_c \approx 1.5$ GeV, before the c quark was discovered.) It therefore seems reasonable to require that the squark exchange result (28.4.3) should be less than the quark exchange result (28.4.4). This yields the condition

$$\left| \sum_i V_{di} V_{si}^* \frac{\Delta M_i^2}{\tilde{M}^2} \right| < \frac{g^2 \sin\theta_c \cos\theta_c}{g_s^2} \frac{(m_c - m_u)\tilde{M}}{m_W^2}. \tag{28.4.5}$$

Taking $g^2/4\pi = 0.036$, $g_s^2/4\pi = 0.118$, $\sin\theta_c = 0.22$, $m_W = 80.4$ GeV, $m_c = 1.5$ GeV, and $m_u \ll m_c$, we find

$$\left| \sum_i V_{di} V_{si}^* \frac{\Delta M_i^2}{\tilde{M}^2} \right| < 1.5 \times 10^{-3} \times (\tilde{M}/100\,\text{GeV}). \tag{28.4.6}$$

The squark masses are unlikely to be much less than m_{gluino}, so we can conclude that either the squark masses are split by no more than about one part in 10^3, or the non-diagonal terms of the mixing matrix V_{ji} are less than about 10^{-3}, or the squarks are heavier than about 10 TeV, or we have some combination of nearly degenerate squarks, nearly zero mixing angles, and heavy squarks. In itself, this result only constrains the superpartners $\mathscr{D}_i$ of the left-handed quarks of charge $-e/3$, but similar limits on the masses and mixing angles of the $\bar{\mathscr{D}}_i$ squarks can be obtained by considering the amplitude for $d_R\overline{s_R} \to s_R\overline{d_R}$. We can also obtain somewhat weaker

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

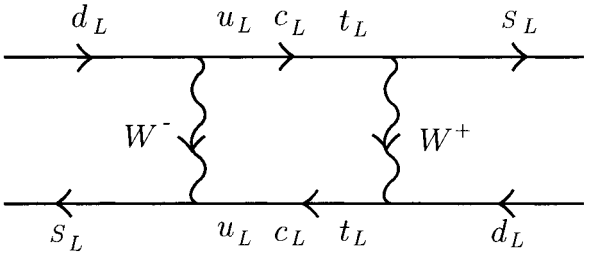

Figure 28.2. A one-loop diagram that can contribute to the $\Delta S = 2$ effective interaction $(\bar{s_L}\gamma^\mu d_L)(\bar{d_L}\gamma_\mu s_L)$ in both the supersymmetric and non-supersymmetric standard models. Here solid lines are quarks and wavy lines are $W^\pm$ bosons.

limits on the masses and mixing angles of the $\mathscr{U}_i$ squarks by considering the amplitudes produced by wino exchange rather than gluino exchange. It should be noted, however, that these arguments put no constraints on the differences between the masses of squarks of different charge, or on the differences between the masses of the superpartners $\mathscr{Q}_i$ and $\bar{\mathscr{Q}}_i$ of left-handed quarks and antiquarks.

Just as for the squarks, the sleptons of definite mass are expected to be non-diagonal linear combinations of the superpartners of the leptons. This leads to the decay process $\mu \to e + \gamma$ through diagrams like Figure 28.3. The experimental upper bound 4.9×10^{-11} on the branching ratio of this process then sets a limit of about 10^{-3} on the fractional mass splitting of sleptons of the same charge but different generations for generic mixing angles, or on the mixing angles for non-degenerate sleptons.[14]

Attempts have been made to explain the degeneracy of the squarks and sleptons in terms of a gauge symmetry connecting the different generations.[14a] In Section 28.6 we will describe an approach to supersymmetry breaking in which this degeneracy appears without needing to impose such symmetries.

CP violation

The second important class of constraint provided by experimental information about known particles has to do with CP-violating effects, such as the electric dipole moments of the neutron and electron.[15] In Section 21.3 we saw that these effects are rather weak in the non-supersymmetric standard model with only one scalar doublet, apart from a potential problem with the parameter θ of quantum chromodynamics, discussed in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

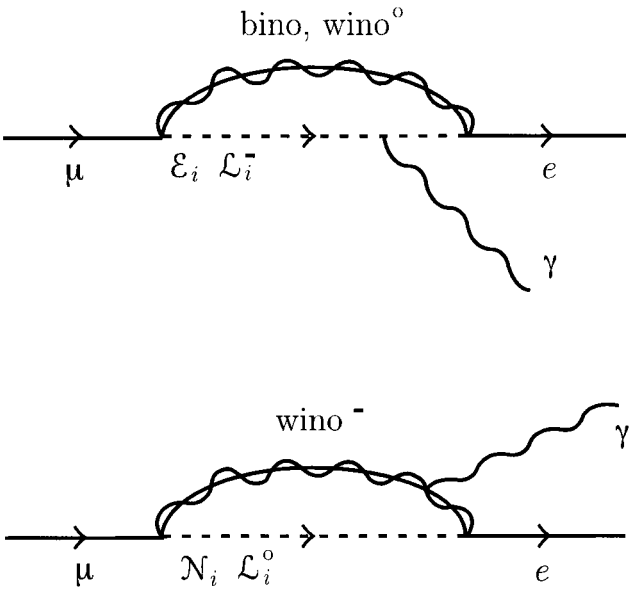

Figure 28.3. One-loop diagrams for the process $\mu \to e + \gamma$. Here solid lines are leptons; dashed lines are sleptons; combined solid and wavy lines are gauginos; and wavy lines are photons.

Section 23.6. This is because all CP-violating phases in the mass matrix of quarks and leptons and their interaction with gauge bosons could be absorbed into the definition of the quark and lepton fields if there were only two generations of quarks and leptons, and although there is a third generation, its mixing with the first two generations is (for mysterious reasons) quite weak. (This argument does not apply to processes that directly involve quarks of the third generation, such as B^0–$\bar{B}^0$ mixing, to be measured in the planned 'B factories.') The electric dipole moment of the neutron in this simple non-supersymmetric version of the standard model is consequently expected[16] to be less than about $10^{-30}\,e$ cm, well below the experimental upper bound, $6.3 \times 10^{-26}\,e$ cm.[16a]

In contrast, the over 100 parameters of the minimum supersymmetric standard model in its most general form involve dozens of CP-violating relative phases. After integrating out the heavy superpartners of the known particles, these phases produce a number of CP-violating effective interactions to be added to the Lagrangian of the standard model. Those of minimum dimensionality, which on dimensional grounds are likely to be the most important, include electric dipole moments of the quarks and leptons,[17] similar CP-violating 'chromoelectric' dipole moments

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

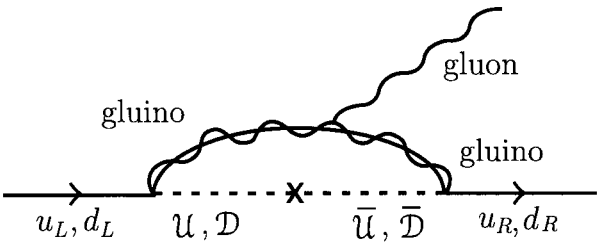

Figure 28.4. A one-loop diagram for the chromoelectric dipole moment of the u or d quarks. Here solid lines are quarks; dashed lines are squarks; combined solid and wavy lines are gluinos; and the wavy line is a gluon. The X represents the insertion of a bilinear interaction arising from a trilinear scalar field interaction combined with the spontaneous breakdown of $SU(2) \times U(1)$. There are also diagrams in which the gluon line is attached to one of the internal squark lines instead of to the gluino line.

contributing to the interactions of gluons with quarks,[18] a CP-violating purely gluonic interaction,[19] and a CP-violating interaction of the lightest Higgs scalar with leptons.[20]

To take one example, consider the quark chromoelectric dipole moments, which in some models make the largest contribution to the electric dipole moment of the neutron. The CP-violating chromoelectric dipole moment operator is $(\bar{q}\gamma_5[\gamma_\mu, \gamma_\nu]\lambda_a q)f_a^{\mu\nu}$ (where q is a u or d color-triplet quark field, $f_a^{\mu\nu}$ is the $SU(3)$ field-strength tensor, and λ_a are the 3×3 generators of $SU(3)$). Since $\gamma_5[\gamma_\mu, \gamma_\nu]$ has matrix elements only between $\bar{q}_L$ and q_R or between $\bar{q}_R$ and q_L, in order for a one-loop graphs to contribute to the chromoelectric dipole moment an external left-handed u or d quark line must emit an internal gluino line and turn into a $\mathcal{U}$ or $\mathcal{D}$ squark line, which next turns into a $\bar{\mathcal{U}}^*$ or $\bar{\mathcal{D}}^*$ squark line, and then into a right-handed u or d quark line by absorbing the internal gluino line, with the external gluon line attached either to the internal gluino line or to one of the internal squark lines. (See Figure 28.4.)

To calculate this, we need to know the mixing of the scalar components $\mathcal{U}_i$ (or $\mathcal{D}_i$) of the left-chiral quark superfields Q_i with the complex conjugates $\bar{\mathcal{U}}_j^*$ (or $\bar{\mathcal{D}}_j^*$) of the scalar components of the left-chiral antiquark superfields $\bar{U}_j$ (or $\bar{D}_j$), produced by the spontaneous breakdown of $SU(2) \times U(1)$, and represented by the X in Figure 28.4. Part of this mixing arises from a contribution of the supersymmetric $\mathcal{F}$-term interaction (28.1.7) to the last term in Eq. (26.4.7):

$$\mathcal{L}_{\mathcal{U}\bar{\mathcal{U}}\mathcal{H}} = -\left|\sum_{ij} h_{ij}^U \mathcal{U}_i \bar{\mathcal{U}}_j + \mu \mathcal{H}_1^0\right|^2 - \left|\sum_{ij} h_{ij}^D \mathcal{D}_i \bar{\mathcal{D}}_j + \mu \mathcal{H}_2^0\right|^2 . \qquad (28.4.7)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

There is also a contribution from the A- and C-terms in Eq. (28.4.1):

$$\mathscr{L}'_{\mathscr{Q}\bar{\mathscr{Q}}\mathscr{H}} = -\sum_{ij} h^D_{ij}\mathscr{D}_i\bar{\mathscr{D}}_j\left[-A^D_{ij}\,\mathscr{H}^0_1 + C^D_{ij}\,\mathscr{H}^{0*}_2\right] - \sum_{ij} h^U_{ij}\mathscr{U}_i\bar{\mathscr{U}}_j\left[A^U_{ij}\,\mathscr{H}^0_2 + C^U_{ij}\,\mathscr{H}^{0*}_1\right] - \text{H.c.} \tag{28.4.8}$$

Replacing the neutral Higgs scalar fields with their expectation values gives the quadratic terms

$$\mathscr{L}_{\mathscr{Q}\bar{\mathscr{Q}}} = -2\,\text{Re}\sum_{ij} m^U_{ij}\mathscr{U}_i\bar{\mathscr{U}}_j\left(\mu^*\cot\beta + A^U_{ij} + C^U_{ij}\cot\beta\right) - 2\,\text{Re}\sum_{ij} m^D_{ij}\mathscr{D}_i\bar{\mathscr{D}}_j\left(\mu^*(\tan\beta)^* + A^D_{ij} - C^D_{ij}(\tan\beta)^*\right), \tag{28.4.9}$$

where $m^U_{ij} = \langle\mathscr{H}^0_2\rangle h^U_{ij}$ and $m^D_{ij} = -\langle\mathscr{H}^0_1\rangle h^D_{ij}$ are the mass matrices of the quarks of charge $2e/3$ and $-e/3$, and

$$\tan\beta \equiv \langle\mathscr{H}^0_2\rangle/\langle\mathscr{H}^0_1\rangle^* . \tag{28.4.10}$$

Neglecting Cabibbo mixing, and for definiteness taking the As and Cs diagonal, Figure 28.4 makes contributions to the chromoelectric dipole moment of the u and d quarks of the form

$$d^{ce}_u = \frac{g_s^3}{16\pi^2}\,\text{Im}\left[m_u\,A'_u I(m_{\mathscr{U}}, m_{\bar{\mathscr{U}}}, m_{\text{gluino}})\right], \tag{28.4.11}$$

$$d^{ce}_d = \frac{g_s^3}{16\pi^2}\,\text{Im}\left[m_d\,A'_d I(m_{\mathscr{D}}, m_{\bar{\mathscr{D}}}, m_{\text{gluino}})\right], \tag{28.4.12}$$

where

$$A'_u \equiv (\mu^* + C_u)\cot\beta + A_u\,, \qquad A'_d = (\mu^* - C_d)(\tan\beta)^* + A_d\,, \tag{28.4.13}$$

and I is a complicated dimensionless function of its arguments arising from the integration over the virtual four-momentum. For $m_{\mathscr{Q}} \simeq m_{\bar{\mathscr{Q}}}$ and the gluino field defined to make m_{gluino} as well as $m_{\mathscr{Q}}$ real, the function I takes the form

$$I(m_{\mathscr{Q}}, m_{\mathscr{Q}}, m_{\text{gluino}}) = m^{-3}_{\text{gluino}}\,J\left(\frac{m^2_{\text{gluino}}}{m^2_{\mathscr{Q}} - m^2_{\text{gluino}}}\right), \tag{28.4.14}$$

where[21]

$$J(z) = 2\left(-z^4 + \frac{4}{3}z^3 + z^2\right)\ln\left(\frac{1+z}{z}\right) + 2z^3 - \frac{11}{3}z^2 . \tag{28.4.15}$$

The hard part of this sort of calculation always lies in estimating the contribution of an operator like the chromoelectric dipole interaction to hadronic matrix elements such as the electric dipole moment of the neutron. We can anticipate that there will be renormalization group

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

corrections required because this operator is to be used at energies of the order of the neutron mass, rather than the masses of the squarks and gluino. More important is simply getting dimensional factors and factors of 4π right. For this purpose, it is usual to use a counting rule[22] known as 'naive dimensional analysis.' A connected graph with V_i vertices of type i and I internal lines will have a number L of loops given by $L = I - \sum_i V_i + 1$. If there are N_i lines attached to a vertex of type i and N external lines of the whole graph, then $2I + N = \sum_i V_i N_i$, so

$$L = 1 - \frac{N}{2} + \sum_i V_i \left(\frac{N_i}{2} - 1 \right) .$$

We expect a factor of order $1/16\pi^2$ for each loop, so the coefficient of an operator $\mathcal{O}$ in the low-energy effective Lagrangian with N field factors will contain an over-all factor

$$(4\pi)^{N-2} \prod_i (4\pi)^{(2-N_i)V_i} .$$

If the operator $\mathcal{O}$ has dimensionality d and the interactions $\mathcal{O}_i$ of type i have dimensionality d_i, then the coefficient of $\mathcal{O}$ will have dimensionality $4 - d - \sum_i (4 - d_i)$, so this coefficient will also have a factor $M^{4-d} \prod_i M^{d_i - 4}$, where M is some scale that is typical of hadronic physics, such as the nucleon mass or the energy $2\pi F_\pi \simeq 1200$ MeV, where the low-energy expansions discussed in Section 19.5 begin to break down. Finally, the contribution to the coefficient of $\mathcal{O}$ from some graph will of course be proportional to the couplings of all the operators $\mathcal{O}_i$ associated with the vertices in the graph. These remarks can be conveniently summarized by defining a 'reduced coupling': the reduced coupling associated with any operator $\mathcal{O}_i$ having N_i field factors, dimensionality d_i, and coupling constant g_i is

$$g_i^{\text{reduced}} \equiv g_i (4\pi)^{2-N_i} M^{\mathscr{D}_i - 4} . \tag{28.4.16}$$

The above estimates suggest the rule of naive dimensional analysis: the reduced coupling of any operator $\mathcal{O}$ in the effective hadronic Lagrangian is roughly equal to the product of the reduced couplings of the interactions that contribute to this effective coupling.

The neutron electric dipole moment is the coefficient of an operator with one photon and two neutron fields and dimensionality 5, so its reduced coupling is $M d_n^e / 4\pi$. Likewise, the quark chromoelectric magnetic moment has reduced coupling $M d_q^{ce} / 4\pi$. In addition to one factor of this reduced coupling, the reduced coupling of the neutron electric dipole operator must have a factor of the reduced coupling $e/4\pi$ of the electromagnetic coupling and an indeterminate number of factors of the reduced strong coupling $g_s/4\pi$, which *at the low-energy scale* M are not very different

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

from unity and will be ignored. Taking the d quark contribution as representative of the contributions of both u and d quarks, the result then is that

$$d_n^e \approx \frac{e\, d_d^{ce}}{4\pi} \approx e \left(\frac{g_s}{4\pi}\right)^3 \mathrm{Im}\left[m_d\, A_d'\right] I(m_{\mathscr{D}}, m_{\bar{\mathscr{D}}}, m_{\mathrm{gluino}})\,. \tag{28.4.17}$$

Further simplifying by setting $m_{\mathrm{gluino}} \simeq m_{\mathscr{D}} \simeq m_{\bar{\mathscr{D}}}$, so that $J = 7/18$, and taking $g_s^2/4\pi$ at the scale of squark and gluino masses to have the same value 0.12 as at m_Z, and $|m_d| \approx 7$ MeV, we have then

$$|d_n^e| \approx 0.5 \times 10^{-23}\, e\,\mathrm{cm}\, \frac{|A_d'|\,|\sin\varphi| \times (100\,\mathrm{GeV})^2}{m_{\mathrm{gluino}}^3}\,, \tag{28.4.18}$$

where φ is the phase of A_d', with the convention that the gluino, quark, and squark masses are taken real. The contribution of the electric dipole moment of the quark is somewhat larger, while the contribution of the purely gluonic CP-odd operator is considerably smaller.[23]

To avoid conflict with the experimental upper bound of $0.97 \times 10^{-25}\, e$ cm, either the phases associated with CP violation in the supersymmetric standard model must be less than about 10^{-2}, or some of the new particles of this model must be heavier than about 1 TeV. Similar conclusions have been reached from calculations of the electric dipole moments of atoms and molecules.[23] Even more stringent conditions on CP-violating phases have been derived[24] by considering the contribution of Figure 28.1 to the one precisely measured CP-violating effect, the imaginary part of the amplitude for K^0–$\bar{K}^0$ oscillation.

28.5 The Sector of Zero Baryon and Lepton Number

Despite the large number of parameters of the supersymmetric standard model, in some contexts it is surprisingly predictive. This is true in particular when we consider the scalar fields whose vacuum expectation values spontaneously break the $SU(2) \times U(1)$ gauge symmetry. In this section we will consider these scalars, along with other fields of zero baryon and lepton number: neutral scalars odd under charge conjugation, charged scalars, and the fermionic superpartners of these scalars and of the $W^{\pm}$ and Z^0.

It is a crucial requirement for supersymmetric versions of the standard model that they should contain scalar doublet 'Higgs' superfields with the right mass and interaction parameters to account for the breakdown of the $SU(2) \times U(1)$ gauge group of the electromagnetic and weak interactions. We saw in Section 28.1 that at least two left-chiral scalar doublets are needed to give mass to the quarks of both charge $2e/3$ and $-e/3$ and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

the charged leptons, while we found in Section 28.2 that two doublets are just what is needed to bring the $SU(3)$, $SU(2)$, and $U(1)$ gauge couplings together at some very high energy. We have therefore assumed that there are two left-chiral scalar $SU(2)$ doublets

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix}, \qquad H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix}. \tag{28.5.1}$$

These have $SU(2)$ and $U(1)$ D-terms (27.4.7) given (assuming zero Fayet–Iliopoulos constant $\xi_{U(1)}$) by

$$\vec{D} = \frac{g}{2}\left(\mathscr{H}_1^\dagger \vec{\tau}\, \mathscr{H}_1\right) + \frac{g}{2}\left(\mathscr{H}_2^\dagger \vec{\tau}\, \mathscr{H}_2\right), \tag{28.5.2}$$

$$D_y = \frac{g'}{2}\left(\mathscr{H}_1^\dagger \mathscr{H}_1\right) - \frac{g'}{2}\left(\mathscr{H}_2^\dagger \mathscr{H}_2\right), \tag{28.5.3}$$

where $\mathscr{H}_{1,2}$ are the scalar components of the superfield doublets $H_{1,2}$ and τ_r are the Pauli matrices, with $\tau_r^2 = 1$. As shown in Eq. (27.4.9), in renormalizable theories this gives the D-term contribution to the scalar field potential

$$\begin{aligned} V_D &= \frac{1}{2}\vec{D}^2 + \frac{1}{2}D_y^2 \\ &= \frac{g^2}{8}\left[\left(\mathscr{H}_1^\dagger \vec{\tau}\, \mathscr{H}_1\right) + \left(\mathscr{H}_2^\dagger \vec{\tau}\, \mathscr{H}_2\right)\right]^2 + \frac{g'^2}{8}\left[\left(\mathscr{H}_1^\dagger \mathscr{H}_1\right) - \left(\mathscr{H}_2^\dagger \mathscr{H}_2\right)\right]^2. \end{aligned} \tag{28.5.4}$$

This can be put in a more convenient form by using the relation

$$(\vec{\tau})_{i\ell} \cdot (\vec{\tau})_{kj} = 2\delta_{ij}\delta_{k\ell} - \delta_{i\ell}\delta_{kj}. \tag{28.5.5}$$

(To prove this, use rotational invariance to show that $\delta_{ij}\delta_{k\ell}$ may be expressed as a linear combination of $(\vec{\tau})_{i\ell} \cdot (\vec{\tau})_{kj}$ and $\delta_{i\ell}\delta_{kj}$, and calculate the coefficients by taking traces on the indices i, j and i, ℓ.) In this way, we may rewrite the D-term part of the scalar field potential as

$$V_D = \frac{g^2}{2}\left|\left(\mathscr{H}_1^\dagger \mathscr{H}_2\right)\right|^2 + \frac{g^2 + g'^2}{8}\left[\left(\mathscr{H}_1^\dagger \mathscr{H}_1\right) - \left(\mathscr{H}_2^\dagger \mathscr{H}_2\right)\right]^2. \tag{28.5.6}$$

As mentioned in Section 28.1, there is just one possible renormalizable term in the superpotential for these two left-chiral doublets, of the form

$$f(H_1, H_2) = \mu\left(H_1^{\mathrm{T}} e H_2\right), \tag{28.5.7}$$

where μ is a constant with the dimensions of a mass and e is the antisymmetric matrix $i\tau_2$. According to Eq. (27.4.9), this gives an additional

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

contribution to the scalar field potential

$$V_\mu = \sum_r \left| \frac{\partial f(\mathscr{H}_1, \mathscr{H}_2)}{\partial \mathscr{H}_{1r}} \right|^2 + \sum_r \left| \frac{\partial f(\mathscr{H}_1, \mathscr{H}_2)}{\partial \mathscr{H}_{2r}} \right|^2$$
$$= |\mu|^2 \left[\left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) + \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) \right] . \tag{28.5.8}$$

For $\mu \neq 0$ the potential $V_D + V_\mu$ evidently has a minimum value of zero, reached at the unique point $\mathscr{H}_1 = \mathscr{H}_2 = 0$. With just these terms in the potential, $SU(2) \times U(1)$ as well as supersymmetry is not spontaneously broken. (The case $\mu = 0$ is not much better; there is a continuous infinity of vacuum states with supersymmetry unbroken and with $SU(2) \times U(1)$ broken down to electromagnetic gauge invariance with all possible strengths, including zero.) This is one more example of the general difficulty, already seen in Section 28.3, of formulating realistic theories in which supersymmetry is spontaneously broken within the standard model.

Under the assumption of the previous section that supersymmetry is violated in the effective Lagrangian only by superrenormalizable terms, the most general such supersymmetry-breaking term involving the scalar doublets is of the form

$$V_m = m_1^2 \left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) + m_2^2 \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) + \mathrm{Re} \left\{ B\mu \left(\mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2 \right) \right\} ,$$

where m_1^2 and m_2^2 are real parameters (not necessarily positive) and $B\mu$ is a parameter of arbitrary phase. We will adjust the *over-all* phase of the superfields H_1 and H_2 so that $B\mu$ is real and positive, and so

$$V_m = m_1^2 \left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) + m_2^2 \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) + B\mu \, \mathrm{Re} \left(\mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2 \right) . \tag{28.5.9}$$

The total scalar potential in the tree approximation is then

$$\begin{aligned} V &= V_D + V_\mu + V_m \\ &= \frac{g^2}{2} \left| \left(\mathscr{H}_1^\dagger \mathscr{H}_2 \right) \right|^2 + \frac{g^2 + g'^2}{8} \left[\left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) - \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) \right]^2 \\ &\quad + (m_1^2 + |\mu|^2) \left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) + (m_2^2 + |\mu|^2) \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) \\ &\quad + B\mu \, \mathrm{Re} \left(\mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2 \right) . \end{aligned} \tag{28.5.10}$$

Note in particular that μ^2, m_1^2, and m_2^2 only appear in the combinations $m_1^2 + |\mu|^2$ and $m_2^2 + |\mu|^2$.

There is one condition on the supersymmetry-breaking parameters m_i^2, derived from the requirement that the potential should be bounded below. For scalar fields going to infinity in generic directions, the potential is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

dominated by the quartic terms V_D, which are positive. There are special directions in which V_D vanishes: those for which (up to an $SU(2) \times U(1)$ gauge transformation)

$$\mathscr{H}_1 = \begin{pmatrix} \phi \\ 0 \end{pmatrix}, \qquad \mathscr{H}_2 = \begin{pmatrix} 0 \\ \phi \end{pmatrix},$$

with ϕ an arbitrary complex quantity. For such directions, $V = (2|\mu|^2 + m_1^2 + m_2^2)|\phi|^2 - B\mu\phi^2$, so (since $B\mu$ has been defined to be positive) in order for this not to go to $-\infty$ as $\phi \to +\infty$, it is necessary that

$$2|\mu|^2 + m_1^2 + m_2^2 \geq B\mu\,. \tag{28.5.11}$$

We wish to look for a minimum of the potential at which electromagnetic gauge invariance is not broken, so let us consider the behavior of the potential as a function of the neutral scalar fields, with the charged scalar fields set equal to zero. In this case Eq. (28.5.10) gives the potential of the neutral scalars as

$$\begin{aligned} V^{\mathrm{N}} = {} & \frac{g^2 + g'^2}{8}\left[\left|\mathscr{H}_1^0\right|^2 - \left|\mathscr{H}_2^0\right|^2\right]^2 + (m_1^2 + |\mu|^2)\left|\mathscr{H}_1^0\right|^2 \\ & + (m_2^2 + |\mu|^2)\left|\mathscr{H}_2^0\right|^2 - B\mu\,\mathrm{Re}\left(\mathscr{H}_1^0\mathscr{H}_2^0\right). \end{aligned} \tag{28.5.12}$$

To find a stationary point, we expand V^{N} around the constant values $\mathscr{H}_i^0 = v_i$, writing

$$\mathscr{H}_i^0 = v_i + \varphi_i\,. \tag{28.5.13}$$

To second order in the φ_i, Eq. (28.5.12) gives

$$\begin{aligned} V_{\mathrm{quad}}^{\mathrm{N}} = {} & \frac{g^2 + g'^2}{4}(|v_1|^2 - |v_2|^2)\left[2\mathrm{Re}\,(v_1^*\varphi_1 - v_2^*\varphi_2) + |\varphi_1|^2 - |\varphi_2|^2\right] \\ & + \frac{g^2 + g'^2}{2}\left[\mathrm{Re}\,(v_1^*\varphi_1 - v_2^*\varphi_2)\right]^2 + (m_1^2 + |\mu|^2)\left(2\mathrm{Re}\,v_1^*\varphi_1 + |\varphi_1|^2\right) \\ & + (m_2^2 + |\mu|^2)\left(2\mathrm{Re}\,v_2^*\varphi_2 + |\varphi_2|^2\right) - B\mu\,\mathrm{Re}\left(v_1\varphi_2 + v_2\varphi_1 + \varphi_1\varphi_2\right) \\ & + \text{constant}\,. \end{aligned} \tag{28.5.14}$$

For the v_i to be equilibrium values of the fields, the terms of first order in the φ_i must vanish

$$\left(m_1^2 + |\mu|^2\right)v_1^* + \frac{g^2 + g'^2}{4}\left(|v_1|^2 - |v_2|^2\right)v_1^* - \frac{1}{2}B\mu\,v_2 = 0\,, \tag{28.5.15}$$

$$\left(m_2^2 + |\mu|^2\right)v_2^* + \frac{g^2 + g'^2}{4}\left(|v_2|^2 - |v_1|^2\right)v_2^* - \frac{1}{2}B\mu\,v_1 = 0\,. \tag{28.5.16}$$

Without changing the over-all phase of the φ_i, we may adjust their relative phases so that v_1 is real. Then Eqs. (28.5.15) and (28.5.16) show that v_2 is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

also real, so that these equations become

$$\left(m_1^2+|\mu|^2\right)v_1+\frac{g^2+g'^2}{4}\left(v_1^2-v_2^2\right)v_1-\frac{1}{2}B\mu v_2=0\,, \quad (28.5.17)$$

$$\left(m_2^2+|\mu|^2\right)v_2+\frac{g^2+g'^2}{4}\left(v_2^2-v_1^2\right)v_2-\frac{1}{2}B\mu v_1=0\,. \quad (28.5.18)$$

These conditions may be used to express the mass parameters in the potential in terms of the convenient quantities

$$\tan\beta\equiv v_2/v_1\,, \quad (28.5.19)$$

$$m_Z^2=\tfrac{1}{2}(g^2+g'^2)(v_1^2+v_2^2)\,, \quad (28.5.20)$$

and

$$m_A^2\equiv 2|\mu|^2+m_1^2+m_2^2\,. \quad (28.5.21)$$

(The parameter m_Z is the mass of the Z vector boson.* We will soon see that m_A is the mass of one of the physical scalars.) Multiplying Eqs. (28.5.17) and (28.5.18), respectively, by v_2 and v_1 and taking the sum and difference gives

$$B\mu=m_A^2\sin 2\beta\,, \quad (28.5.22)$$

and

$$m_1^2-m_2^2=-(m_A^2+m_Z^2)\cos 2\beta\,, \quad (28.5.23)$$

which with Eq. (28.5.21) yields

$$m_1^2+|\mu|^2=\tfrac{1}{2}m_A^2-\tfrac{1}{2}(m_A^2+m_Z^2)\cos 2\beta\,,\ \ m_2^2+|\mu|^2=\tfrac{1}{2}m_A^2+\tfrac{1}{2}(m_A^2+m_Z^2)\cos 2\beta\,. \quad (28.5.24)$$

With linear terms cancelling, the quadratic part (28.5.14) of the neutral scalar potential may then be written

$$\begin{aligned}V_{\text{quad}}^{\text{N}}&=\frac{g^2+g'^2}{4}(v_1^2-v_2^2)\left[|\varphi_1|^2-|\varphi_2|^2\right]+\frac{g^2+g'^2}{2}\left[\text{Re}\,(v_1\varphi_1-v_2\varphi_2)\right]^2\\&\quad+(m_1^2+|\mu|^2)\,|\varphi_1|^2+(m_2^2+|\mu|^2)\,|\varphi_2|^2-B\mu\,\text{Re}\left(\varphi_1\varphi_2\right)+\text{constant}\\&=\tfrac{1}{2}m_Z^2\cos 2\beta\left[|\varphi_1|^2-|\varphi_2|^2\right]+m_Z^2\left[\text{Re}\,(\cos\beta\varphi_1-\sin\beta\varphi_2)\right]^2\\&\quad+\tfrac{1}{2}m_A^2\left(|\varphi_1|^2+|\varphi_2|^2\right)-\tfrac{1}{2}(m_A^2+m_Z^2)\cos 2\beta\left[|\varphi_1|^2-|\varphi_2|^2\right]\\&\quad-m_A^2\sin 2\beta\,\text{Re}\left(\varphi_1\varphi_2\right)+\text{constant}\,. \quad (28.5.25)\end{aligned}$$

* There is a difference of a factor of 2 between the formulas for m_Z^2 given by Eqs. (28.5.20) and (21.3.30), due to the fact that the scalar fields are normalized differently here and in Section 21.3.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

We see from Eq. (28.5.25) that the real and imaginary parts of the φ_i are decoupled. (This is because the potential (28.5.12) is invariant under a charge conjugation or CP transformation $\varphi_i \to \varphi_i^*$.) The mass-squared matrix of the imaginary parts of the φ_i is

$$M^2_{\mathrm{Im}\,\varphi} = \begin{pmatrix} \frac{1}{2}m_A^2\,(1-\cos 2\beta) & \frac{1}{2}m_A^2\,\sin 2\beta \\ \frac{1}{2}m_A^2\,\sin 2\beta & \frac{1}{2}m_A^2\,(1+\cos 2\beta) \end{pmatrix} . \tag{28.5.26}$$

The determinant vanishes, so one eigenvalue is zero, and the other equals the trace, which is just m_A^2. The zero-mass scalar is of course the neutral Goldstone boson associated with the spontaneous breakdown of $SU(2) \times U(1)$ to electromagnetic gauge invariance, and as discussed in Chapter 21 it is eliminated by the Higgs mechanism. As promised, m_A is the mass of one of the physical scalars, the non-Goldstone boson with C negative. This shows that for the field value $\varphi_i = v_i$ to be at least a local minimum of the potential, the parameter m_A^2 defined by Eq. (28.5.21) must be positive. The condition (28.5.11) for good behavior at large field strengths shows that Eq. (28.5.22) has a solution here for β in the range $0 \le \beta \le \pi/2$.

In particular, if $B\mu = 0$ and $0 < \beta < \pi/2$, then Eq. (28.5.22) shows that $m_A = 0$. In this case, the particle A is the Goldstone boson of a $U(1)$ Peccei–Quinn symmetry[4] of the potential (28.5.12) under an equal phase change of $\mathscr{H}_1^0$ and $\mathscr{H}_2^0$, which for $v_1 \neq 0$ and $v_2 \neq 0$ is spontaneously broken without leaving any combination of this and the electroweak $U(1)$ symmetry unbroken. This is the original version of the axion,[25] which as we saw in Section 23.6 acquires only a small mass from the Yukawa interactions of the scalars with the quarks and is experimentally ruled out. Thus we can conclude that $B\mu$ definitely does not vanish.

The elements of the mass-squared matrix for the real scalars are given by Eq. (28.5.25) as

$$\begin{aligned} (M^2_{\mathrm{Re}\,\varphi})_{11} &= \tfrac{1}{2}m_A^2(1-\cos 2\beta) + \tfrac{1}{2}m_Z^2(1+\cos 2\beta)\,, \\ (M^2_{\mathrm{Re}\,\varphi})_{12} &= (M^2_{\mathrm{Re}\,\varphi})_{21} = -\tfrac{1}{2}(m_A^2+m_Z^2)\sin 2\beta\,, \\ (M^2_{\mathrm{Re}\,\varphi})_{22} &= \tfrac{1}{2}m_A^2(1+\cos 2\beta) + \tfrac{1}{2}m_Z^2(1-\cos 2\beta)\,. \end{aligned} \tag{28.5.27}$$

Solving the secular equation gives the eigenvalues

$$m_H^2 = \frac{1}{2}\left[m_A^2+m_Z^2+\sqrt{(m_A^2+m_Z^2)^2-4m_A^2m_Z^2\cos^2 2\beta}\right], \tag{28.5.28}$$

$$m_h^2 = \frac{1}{2}\left[m_A^2+m_Z^2-\sqrt{(m_A^2+m_Z^2)^2-4m_A^2m_Z^2\cos^2 2\beta}\right]. \tag{28.5.29}$$

To calculate the masses of the charged scalars, we evaluate the potential

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

V with the neutral scalars set equal to their vacuum expectation values:

$$\mathscr{H}_1 = \begin{pmatrix} v_1 \\ \mathscr{H}_1^- \end{pmatrix}, \qquad \mathscr{H}_2 = \begin{pmatrix} \mathscr{H}_2^+ \\ v_2 \end{pmatrix}. \tag{28.5.30}$$

Using this in Eq. (28.5.10) then gives the quadratic part of the charged scalar potential

$$V_{\text{quad}}^{\text{C}} = \frac{g^2}{2}\left|v_2(\mathscr{H}_1^-)^* + v_1\mathscr{H}_2^+\right|^2 + \frac{g^2+g'^2}{4}(v_1^2 - v_2^2)\left(|\mathscr{H}_1^-|^2 - |\mathscr{H}_2^+|^2\right)$$
$$+(m_1^2 + |\mu|^2)|\mathscr{H}_1^-|^2 + (m_2^2 + |\mu|^2)|\mathscr{H}_2^+|^2 + B\mu\mathscr{H}_1^-\mathscr{H}_2^+ . \tag{28.5.31}$$

Using Eqs. (28.5.22) and (28.5.24), this may be written

$$V_{\text{quad}}^{\text{C}} = \frac{1}{2}(m_W^2 + m_A^2)\Big[|\mathscr{H}_1^-|^2(1-\cos 2\beta) + |\mathscr{H}_2^+|^2(1+\cos 2\beta)$$
$$+2\sin 2\beta\,\mathscr{H}_1^-\mathscr{H}_2^+\Big], \tag{28.5.32}$$

where m_W is the charged gauge boson mass:

$$m_W^2 = \frac{1}{2}g^2\left(|v_1|^2 + |v_2|^2\right). \tag{28.5.33}$$

The charged scalar mass matrix is then

$$M_{\text{C}}^2 = \frac{1}{2}(m_W^2 + m_A^2)\begin{pmatrix} 1-\cos 2\beta & \sin 2\beta \\ \sin 2\beta & 1+\cos 2\beta \end{pmatrix}. \tag{28.5.34}$$

This has determinant zero, so it has one eigenvalue equal to zero and the other eigenvalue equal to the trace

$$m_C^2 = m_W^2 + m_A^2 . \tag{28.5.35}$$

The zero-mass charged scalar is of course the other Goldstone boson associated with the spontaneous breakdown of $SU(2) \times U(1)$, and like the neutral Goldstone boson found earlier it is eliminated by the Higgs mechanism.

Even without knowing the parameters m_A and β, these results tell us a lot about the relative magnitudes of the scalar boson masses. We can rewrite Eqs. (28.5.28) and (28.5.29) in the form

$$m_H^2 = \frac{1}{2}\left[m_A^2 + m_Z^2 + \sqrt{(m_A^2 - m_Z^2)^2 + 4m_A^2 m_Z^2 \sin^2 2\beta}\right], \tag{28.5.36}$$

$$m_h^2 = \frac{1}{2}\left[m_A^2 + m_Z^2 - \sqrt{(m_A^2 - m_Z^2)^2 + 4m_A^2 m_Z^2 \sin^2 2\beta}\right]. \tag{28.5.37}$$

We see that *the heavier neutral scalar mass m_H is larger than the larger of m_Z and m_A, while the lighter neutral scalar mass m_h is smaller than the smaller of m_Z and m_A.* If the large ratio of top to bottom quark masses is due to a large ratio $v_2/v_1 = \tan\beta$ of scalar field vacuum expectation values, rather than a large ratio of Yukawa couplings, then we expect

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

β to be near $\pi/2$, in which case these inequalities become approximate equalities. Further, Eq. (28.5.35) shows that *the charged scalar mass is greater than both* m_A *and* m_W.

These results are quantitatively modified by various radiative corrections *within* the standard model (as opposed to radiative effects that produce the input parameters m_i^2 in theories of gauge-mediated supersymmetry breaking). The most important corrections arise from the presence in the scalar potential V of terms arising from graphs consisting of a single loop of top or bottom quarks interacting any number of times with external scalar field lines. This is because the top and bottom quarks have by far the strongest couplings to $\mathscr{H}_2$ and $\mathscr{H}_1$, respectively. (It is prudent to include bottom as well as top quark loops here because, as mentioned above, the larger mass of the top quark may be due to a large ratio v_2/v_1, rather than to a large ratio of Yukawa couplings, but even in this case we will see that the dominant corrections are due to top quark loops.)

Let us first consider the neutral scalars, at least one of which would be lighter than the Z boson in the absence of radiative corrections. The effect of these top and bottom loops is to contribute a term of form $U_t(|\mathscr{H}_2^0|^2) + U_b(|\mathscr{H}_1^0|^2)$ to $V^{\rm N}$. We will absorb any terms in U_b or U_t that are linear in $|\mathscr{H}_1^0|^2 - v_1^2$ or $|\mathscr{H}_2^0|^2 - v_2^2$ into the input parameters m_1^2 and m_2^2, so that

$$U_b'(v_1^2) = U_t'(v_2^2) = 0\,. \tag{28.5.38}$$

Our earlier results (28.5.24) and (28.5.22) for $m_1^2 + |\mu|^2$, $m_2^2 + |\mu|^2$ and $B\mu$ are then unchanged. Also, the mass matrix for the C-odd neutral scalars is still given by Eq. (28.5.26). On the other hand, the elements of the mass-squared matrix for the C-even neutral scalars are now given by

$$\begin{aligned}
(M^2_{{\rm Re}\,\varphi})_{11} &= \tfrac{1}{2}m_A^2(1-\cos 2\beta) + \tfrac{1}{2}m_Z^2(1+\cos 2\beta) + \Delta_b\,,\\
(M^2_{{\rm Re}\,\varphi})_{12} &= (M^2_{{\rm Re}\,\varphi})_{21} = -\tfrac{1}{2}(m_A^2 + m_Z^2)\sin 2\beta\,,\\
(M^2_{{\rm Re}\,\varphi})_{22} &= \tfrac{1}{2}m_A^2(1+\cos 2\beta) + \tfrac{1}{2}m_Z^2(1-\cos 2\beta) + \Delta_t\,,
\end{aligned} \tag{28.5.39}$$

where

$$\Delta_b = 2v_1^2\,U_b''(v_1^2)\,, \qquad \Delta_t = 2v_2^2\,U_t''(v_2^2)\,. \tag{28.5.40}$$

The solutions of the secular equation are then

$$\begin{aligned}
m_H^2 = \frac{1}{2}\Bigg[& m_A^2 + m_Z^2 + \Delta_t + \Delta_b \\
& + \sqrt{\Big((m_A^2 - m_Z^2)\cos 2\beta + \Delta_t - \Delta_b\Big)^2 + \Big(m_A^2 + m_Z^2\Big)^2 \sin^2 2\beta}\,\Bigg]\,,
\end{aligned} \tag{28.5.41}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$$m_h^2 = \frac{1}{2}\Bigg[m_A^2 + m_Z^2 + \Delta_t + \Delta_b - \sqrt{\left((m_A^2 - m_Z^2)\cos 2\beta + \Delta_t - \Delta_b\right)^2 + \left(m_A^2 + m_Z^2\right)^2 \sin^2 2\beta}\Bigg] . \tag{28.5.42}$$

In considering searches for these particles, it is important to note that the light Higgs mass m_h increases as the unknown mass m_A increases, reaching a finite upper bound for $m_A \to \infty$

$$m_h \leq m_h(m_A \to \infty) = m_Z^2 \cos^2 2\beta + \Delta_t \sin^2 \beta + \Delta_b \cos^2 \beta . \tag{28.5.43}$$

To calculate Δ_b and Δ_t, we recall from Section 16.2 that the potentials U_b and U_t are given by

$$U_b(|\mathscr{H}_1^0|^2) = -\frac{3}{16\pi^2}\left|\lambda_b \mathscr{H}_1^0\right|^4 \left[\ln \frac{\left|\lambda_b \mathscr{H}_1^0\right|^2}{M_{sb}^2} - \frac{3}{2}\right] + \text{linear terms} , \tag{28.5.44}$$

$$U_t(|\mathscr{H}_2^0|^2) = -\frac{3}{16\pi^2}\left|\lambda_t \mathscr{H}_2^0\right|^4 \left[\ln \frac{\left|\lambda_t \mathscr{H}_2^0\right|^2}{M_{st}^2} - \frac{3}{2}\right] + \text{linear terms} , \tag{28.5.45}$$

where $\lambda_t = m_t/v_2$ and $\lambda_b = m_b/v_1$ are the Yukawa couplings of the top and bottom quarks; M_{st} and M_{sb} are the masses of the stop and sbottom (the scalar superpartners of the top and bottom quarks); these masses and the terms $-3/2$ inside the square brackets are chosen to satisfy the condition that supersymmetry-breaking corrections due to stop and sbottom loops would cancel the corrections due to top and bottom loops if the masses were equal; and the 'linear terms' are linear in $|\mathscr{H}_2^0|^2$ or $|\mathscr{H}_1^0|^2$, with coefficients adjusted to satisfy Eq. (28.5.38). (The factor 3 takes account of the three quark colors.) Then Eq. (28.5.40) gives

$$\Delta_b = -\frac{3}{4\pi^2}|\lambda_b|^4 v_1^2 \ln\left(\frac{\lambda_b v_1^2}{M_{sb}^2}\right) = \frac{3\sqrt{2}\, m_b^4\, G_F}{2\pi^2 \cos^2\beta} \ln\left(\frac{M_{sb}^2}{m_b^2}\right) , \tag{28.5.46}$$

$$\Delta_t = -\frac{3}{4\pi^2}|\lambda_t|^4 v_2^2 \ln\left(\frac{\lambda_t v_2^2}{M_{st}^2}\right) = \frac{3\sqrt{2}\, m_t^4\, G_F}{2\pi^2 \sin^2\beta} \ln\left(\frac{M_{st}^2}{m_t^2}\right) , \tag{28.5.47}$$

where $G_F = 1.17 \times 10^{-5}\ \text{GeV}^{-2}$ is the Fermi coupling constant, given by Eq. (21.3.34) as $G_F = g^2/4\sqrt{2}m_W^2$. Taking $m_b = 4.3$ GeV, $m_t = 180$ GeV, $M_{st} \sim M_{sb} \sim 1$ TeV, and $m_Z = 91.2$ GeV gives $\Delta_b \sim 1.1 \times 10^{-6}\, m_Z^2/\cos^2\beta$ and $\Delta_t \sim 1.1\, m_Z^2/\sin^2\beta$. We see that even if $\tan\beta$ is as large as m_t/m_b, the top-quark correction Δ_t will still be much larger than Δ_b.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

The effect of Δ_t is to *increase* both m_H and m_h. Taking this and other radiative corrections into account,[26] for $\tan\beta > 10$ the upper bound (28.5.43) on the lightest neutral scalar mass is raised by radiative corrections from just below m_Z to between 100 GeV and 110 GeV for stop masses between 300 GeV and 1 TeV. For comparison, an experimental *lower* bound[27] of 62.5 GeV on m_h, m_H, and m_A is set by the absence of an hA or HA final state in e^+e^- collisions at 130 to 172 GeV. Also, calculations of radiative corrections involving Higgs scalars are consistent with precision measurements of electroweak phenomena for m_h in the range[27a] of 27 to 140 GeV.

Radiative corrections are less important for the charged scalars. Since the attachment of charged scalar lines allows transitions between top and bottom quarks, the correction to the scalar field potential here takes a more general form, constrained by $SU(2) \times U(1)$ to be

$$\Delta V = U(\mathscr{H}_2^\dagger \mathscr{H}_2\,,\ \mathscr{H}_1^\dagger \mathscr{H}_1\,,\ \mathscr{H}_2^\dagger \mathscr{H}_1\,,\ \mathscr{H}_1^\dagger \mathscr{H}_2\,,\ \mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2)\,. \qquad (28.5.48)$$

(Quark loops do not actually produce any dependence on $\mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2$.) Every appearance of a $\mathscr{H}_1$ or $\mathscr{H}_2$ doublet is accompanied with a factor λ_b or λ_t, respectively, so terms involving $\mathscr{H}_1$ will be suppressed, as we have already seen in calculating the neutral scalar masses. To a good approximation, then, the correction to the effective potential is of the form

$$\Delta V \simeq U(\mathscr{H}_2^\dagger \mathscr{H}_2\,,\ 0\,,\ 0\,,\ 0) = U(|v_2+\varphi_2|^2 + |\mathscr{H}_2^-|^2\,,\ 0\,,\ 0\,,\ 0)\,. \qquad (28.5.49)$$

By going back to the case where the charged fields vanish, we see that the function U must be just the same as what was earlier called U_t. Any terms in U of first order in an expansion in powers of $|v_2+\varphi_2|^2 + |\mathscr{H}_2^-|^2 - v_2^2$ would simply serve to redefine the constant m_2^2, and are eliminated by the convention (28.5.38). Terms in U of second order in $|v_2+\varphi_2|^2+|\mathscr{H}_2^-|^2-v_2^2$ are genuine radiative corrections, but although they contain terms of second order in $|\varphi_2|^2$ which do affect the neutral scalar masses, they contain no terms of second order in $|\mathscr{H}_2^-|^2$ which could shift the non-Goldstone charged scalar mass. Fortunately, radiative corrections are not needed to avoid a conflict with experiment, because in the absence of an upper bound on m_A there is no theoretical upper bound on the charged scalar mass (28.5.35). An experimental lower bound[28] $m_C \geq 59$ GeV is provided by the non-observation of the process $e^+e^- \to \mathscr{H}^+\mathscr{H}^-$ at 181 to 184 GeV. There is a much more stringent lower bound on m_C provided by the rate of the process $b \to s\gamma$ (measured in decays like $B \to K^*\gamma$), which can occur by transitions to an intermediate $\mathscr{H}^- u$ or $\mathscr{H}^- c$ state, with the photon radiated from the virtual quark or $\mathscr{H}^-$. The present agreement between theory and experiment for this process sets a lower bound[29] of about 150 GeV on m_C (and higher for $\tan\beta < 1$). Using Eq. (28.5.35), this gives an important lower bound $m_A > 125$ GeV.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

There are two conditions on the masses m_i^2 that must be satisfied by any model of supersymmetry breaking in order to give a successful account of electroweak symmetry breaking. One of them is provided by the requirement that the potential is bounded below, which as we have seen requires that

$$2|\mu|^2 + m_1^2 + m_2^2 > B\mu \,.$$

Since $B\mu$ is defined to be positive, this ensures that the C-odd neutral scalar squared mass (28.5.21) is positive. The other condition is provided by Eqs. (28.5.22) and (28.5.24), which for arbitrary values of β require that

$$4\,(m_1^2 + |\mu|^2)\,(m_2^2 + |\mu|^2) \leq (B\mu)^2 \,. \tag{28.5.50}$$

We can easily see from Eq. (28.5.10) that this condition ensures that the matrix of second derivatives of the potential has a negative eigenvalue at $\mathscr{H}_1 = \mathscr{H}_2 = 0$, so this $SU(2) \times U(1)$-invariant point is one of unstable equilibrium, and therefore $SU(2) \times U(1)$ must be spontaneously broken. If β is very near $\pi/2$, then Eq. (28.5.24) tells us that this condition is satisfied by having $m_1^2 + |\mu|^2$ positive and $m_2^2 + |\mu|^2$ *negative*. As we will see in the next section, the renormalization group flow of the parameters in the scalar field Lagrangian provides a mechanism for driving $m_2^2 + |\mu|^2$ negative.

Even with a minimal set of superfields, in supersymmetry theories there are several pairs of particles with different $SU(2) \times U(1)$ transformation properties, but the same charge, color, and baryon and lepton number, which become mixed when $SU(2) \times U(1)$ is spontaneously broken. We have already seen an example of this in the previous section, in which we had to deal with the mixing of the scalar superpartners of the left-handed quarks with the complex conjugates of the scalar superpartners of the left-handed antiquarks. A similar mixing occurs between the higgsinos and gauginos, both charged and neutral; the particles of definite mass are not higgsinos or gauginos, but mixtures known as *charginos* and *neutralinos*. Let us consider the charginos, which provide a useful bound on μ. According to Eq. (27.4.8), there are off-diagonal supersymmetric mass terms in the Lagrangian density

$$-\mathrm{Re}\left[\mu\left(h_{1L}^{-\mathrm{T}}\epsilon h_{2L}^{+}\right) + i\sqrt{2}m_W\cos\beta\left(w_L^{-\mathrm{T}}\epsilon h_{2L}^{+}\right) + i\sqrt{2}m_W\sin\beta\left(w_L^{+\mathrm{T}}\epsilon h_{1L}^{-}\right)\right] ,$$

To this, we should add a term for the wino mass generated by gauge interactions with the supersymmetry-breaking sector

$$-m_{\mathrm{wino}}\mathrm{Re}\left(w_L^{+\mathrm{T}}\epsilon w_L^{-}\right) .$$

The squared chargino masses are then the eigenvalues of the matrix

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$\mathscr{M}_C^\dagger \mathscr{M}_C$, where

$$\mathscr{M}_C = \begin{pmatrix} m_{\text{wino}} & i\sqrt{2}m_W \sin\beta \\ i\sqrt{2}m_W \cos\beta & \mu \end{pmatrix} . \tag{28.5.51}$$

These two eigenvalues are

$$m^2_{\text{chargino}} = \frac{1}{2}\Big[m^2_{\text{wino}} + 2m_W^2 + |\mu|^2 \pm \Big((m^2_{\text{wino}} - |\mu|^2)^2 + 4m_W^4 \cos^2 2\beta + 4m_W^2(m^2_{\text{wino}} + |\mu|^2 - 2m_{\text{wino}} \text{Re}\,\mu \sin 2\beta)\Big)^{1/2}\Big] . \tag{28.5.52}$$

We expect the wino mass m_{wino} to be much larger than m_W. If it is also much larger than $|\mu|$, then the heavier chargino is mostly a wino, with mass m_{wino}, while the lightest chargino is mostly a higgsino, with mass $|\mu|$. In any case, $|\mu|$ is greater than the lightest chargino mass, which the non-appearance of gauginos in e^+–e^- annihilation tells us is greater than about 60 GeV, and probably greater than m_W. The search for neutralinos in e^+–e^- annihilation (with assumptions like (28.4.2)) has set a lower bound of 27 GeV on the mass of the lightest neutralino.[29a]

28.6 Gauge Mediation of Supersymmetry Breaking

In this section we will consider the possibility that the breakdown of supersymmetry is transmitted to the known particles through interactions of the ordinary $SU(3) \times SU(2) \times U(1)$ gauge bosons and their superpartners.[30] It is assumed here that supersymmetry is dynamically broken in a sector of superfields, *not* including the superfields of the observed quarks and leptons, and that some of the chiral superfields in the symmetry-breaking sector, known as the *messenger superfields*, have non-vanishing $SU(3) \times SU(2) \times U(1)$ quantum numbers. In order for the messenger particles to be able to get large (say, of order 1 TeV) masses without breaking $SU(3) \times SU(2) \times U(1)$, it is necessary for them to furnish a real (or pseudoreal) representation of $SU(3) \times SU(2) \times U(1)$, which automatically also means that they do not introduce any new anomalies. Although most treatments of gauge-mediated supersymmetry breaking in the literature also make specific assumptions about the interactions of the messenger superfields with the other superfields responsible for supersymmetry breaking, the most important predictions of this class of theories in fact do not depend on these assumptions. We shall therefore put off assuming anything about the interactions of the messenger superfields with the other superfields of the symmetry-breaking sector. We will, however, make another assumption about the $SU(3) \times SU(2) \times U(1)$ properties of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

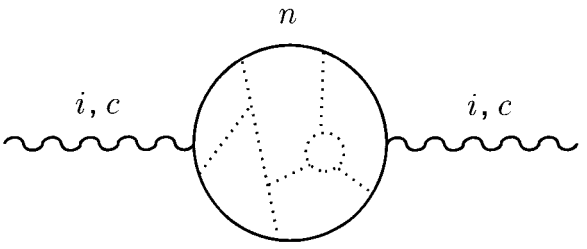

Figure 28.5. A diagram of the sort that introduces a breakdown of supersymmetry into the propagator of the gauge superfields. Here wavy lines are any component fields of the gauge superfields; solid lines are component fields of the messenger superfields; and dotted lines are component fields of the $SU(3) \times SU(2) \times U(1)$-neutral superfields of the supersymmetry-breaking sector.

messenger superfields, which has a strong phenomenological motivation. In order for the messenger particles not to interfere with the unification of couplings discussed in Section 28.2, we assume that they have the same ratios for the total traces of all squared $SU(3) \times SU(2) \times U(1)$ gauge generators as the ordinary quarks and leptons. This condition will automatically be satisfied if the messenger superfields (together perhaps with some $SU(3) \times SU(2) \times U(1)$-neutral chiral superfields) furnish a complete representation of some simple group G that contains $SU(3) \times SU(2) \times U(1)$, of which the quarks and leptons (again, together perhaps with some $SU(3) \times SU(2) \times U(1)$-neutral chiral superfields) also form a complete representation. (For instance, these left-chiral superfields might form N $SU(2)$ singlet $SU(3)$ triplets with charge $e/3$ and N $SU(2)$ doublet $SU(3)$ singlets with charges 0 and $-e$, which together form N representations $\mathbf{5}$ of $SU(5)$, together with an equal number of left-chiral superfields in the complex-conjugate representations of $SU(3) \times SU(2) \times U(1)$, which form N representations $\bar{\mathbf{5}}$.) However, for our present purposes we will neither need to assume that G is an actual symmetry group of the theory, nor adopt any particular choice of G or of the representations furnished by the messenger particles.

This interaction of the messenger superfields both with the other chiral and/or gauge superfields of the supersymmetry-breaking sector and with the $SU(3) \times SU(2) \times U(1)$ gauge superfields can be expected to produce a breakdown of supersymmetry in the propagators of the component fields of the $SU(3) \times SU(2) \times U(1)$ gauge superfields. To lowest order in the $SU(3) \times SU(2) \times U(1)$ couplings, the leading contribution to the propagators comes from the diagram shown in Figure 28.5, in which a pair of gauge, gaugino, or auxiliary D-field lines is attached to a loop

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

of messenger fields, which also may have any number of interactions with $SU(3) \times SU(2) \times U(1)$-neutral fields of the supersymmetry-breaking sector. The supersymmetry-breaking corrections Δ_{ic} to the propagators of the gauge superfields (with $i = 1, 2, 3$ for $SU(3)$, $SU(2)$, and $U(1)$, and $c = V, \lambda, D$ labelling the different components of each gauge superfield) therefore have the forms

$$\begin{aligned} \Delta_{3c}(q) &= (g_s^2/16\pi^2) \sum_n T_{3n} \Pi_{cn}(q) , \\ \Delta_{2c}(q) &= (g^2/16\pi^2) \sum_n T_{2n} \Pi_{cn}(q) , \\ \Delta_{1c}(q) &= (g'^2/16\pi^2) \sum_n T_{1n} \Pi_{cn}(q) , \end{aligned} \tag{28.6.1}$$

where n labels the different messenger superfields; $\Pi_{cn}(q)$ are more-or-less complicated functions of the four-momentum q; T_{3n} and T_{2n} are the traces of the squares of any generator of $SU(3)$ and $SU(2)$, respectively, in the representations furnished by the nth messenger superfield (normalized so that in the defining representations $T_3 = T_2 = 1/2$); and T_{1n} is the sum of the squares of the electroweak hypercharges of the nth messenger superfield. One immediate consequence is that the gauginos acquire masses of the same form:*

$$\begin{aligned} m_{\text{gluino}} &= (g_s^2/16\pi^2) \sum_n T_{3n} M_{gn} , \\ m_{\text{wino}} &= (g^2/16\pi^2) \sum_n T_{2n} M_{gn} , \\ m_{\text{bino}} &= (g'^2/16\pi^2) \sum_n T_{1n} M_{gn} , \end{aligned} \tag{28.6.2}$$

where the M_{gn} are masses that characterize the different messenger superfields. As already mentioned, in order to preserve the unification of couplings at very high energy we assume that the sums of the T_n have the same ratios as for the observed quarks and leptons:

$$\sum_n T_{3n} = \sum_n T_{2n} = \sum_n 3T_{1n}/5 \equiv T . \tag{28.6.3}$$

The breakdown of supersymmetry in these propagators is then communicated to the squarks and sleptons of the supersymmetric standard model through the diagrams shown in Figure 28.6, in which an

* Recall that the bino is the superpartner of the $U(1)$ gauge field B_μ that appears in the Lagrangian of the standard model. We are not yet taking $SU(2) \times U(1)$ breaking into account, so the gaugino, squark, and slepton masses calculated here should be understood as parameters appearing in the $SU(3) \times SU(2) \times U(1)$-invariant effective Lagrangian of the standard mode.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

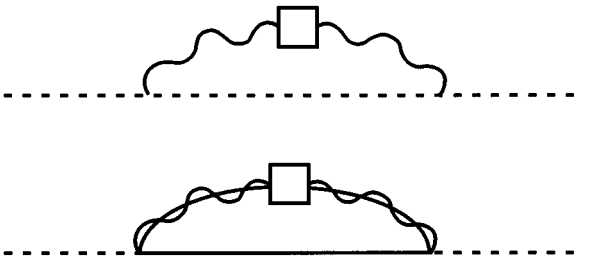

Figure 28.6. Diagrams that communicate supersymmetry breaking to the squarks and sleptons. Here dashed lines are squarks or sleptons; wavy lines are $SU(3) \times SU(2) \times U(1)$ gauge bosons or auxiliary D fields; solid lines are quarks or leptons; combined solid and wavy lines are $SU(3) \times SU(2) \times U(1)$ gauginos; and the squares represent insertions of the supersymmetry-breaking propagator correction shown in Figure 28.5.

$SU(3) \times SU(2) \times U(1)$ gauge boson or gaugino or auxiliary D field is emitted and reabsorbed by the squark or slepton. We are calculating the effective low-energy theory in which $SU(3) \times SU(2) \times U(1)$-breaking is not yet taken into account, so there is no mixing among the $SU(3)$, $SU(2)$, and $U(1)$ propagators, and each propagator acts like a unit matrix on the gauge indices. Thus the squared mass given to any squark or slepton will be proportional to a sum over the squares of all the $SU(3) \times SU(2) \times U(1)$ generators (including coupling constants) in the representation furnished by that squark or slepton. The sums of the squared $SU(2)$ and $SU(3)$ generators in the defining representations are

$$\sum_{a=1}^{3} \left(g\sigma_a/2\right)^2 = \frac{3g^2}{4} \cdot 1\,, \qquad \sum_{\alpha=1}^{8} \left(g_s\lambda_\alpha/2\right)^2 = \frac{4g_s^2}{3} \cdot 1\,,$$

where σ_a are the Pauli isospin matrices (5.4.18) and λ_α are the Gell-Mann matrices (19.7.2). For $U(1)$ the generator is just the weak hypercharge (21.3.7), including a factor g'. The squared squark and slepton masses therefore have the forms

$$\begin{aligned}
M_Q^2 &= 2\sum_n M_{sn}^2 \left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 T_{3n} + \frac{3}{4}\left(\frac{g^2}{16\pi^2}\right)^2 T_{2n} + \left(\frac{1}{6}\right)^2 \left(\frac{g'^2}{16\pi^2}\right)^2 T_{1n}\right] \\
M_U^2 &= 2\sum_n M_{sn}^2 \left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 T_{3n} + \left(\frac{2}{3}\right)^2 \left(\frac{g'^2}{16\pi^2}\right)^2 T_{1n}\right], \\
M_D^2 &= 2\sum_n M_{sn}^2 \left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 T_{3n} + \left(-\frac{1}{3}\right)^2 \left(\frac{g'^2}{16\pi^2}\right)^2 T_{1n}\right],
\end{aligned} \tag{28.6.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$$M_L^2 = 2\sum_n M_{sn}^2 \left[\frac{3}{4} \left(\frac{g^2}{16\pi^2} \right)^2 T_{2n} + \left(\frac{1}{2} \right)^2 \left(\frac{g'^2}{16\pi^2} \right)^2 T_{1n} \right] ,$$

$$M_{\bar{E}}^2 = 2\sum_n M_{sn}^2 \left(\frac{g'^2}{16\pi^2} \right)^2 T_{1n} ,$$

where Q, $\bar{U}$, $\bar{D}$, L, and $\bar{E}$ are the scalar superpartners of the left-handed quark doublets, the left-handed antiquarks of charge $-2e/3$ and $+e/3$, the left-handed lepton doublets, and the left-handed charged antileptons, and M_{sn} are some new masses that characterize the nth messenger superfields. (The factor 2 is extracted from M_{sn}^2 for future convenience.) The squark and slepton masses that are produced in this way are automatically the same in all three generations, thus avoiding the problem with flavor-changing processes discussed in Section 28.4.

We expect all M_{gn} and M_{sn} to be roughly of the same order of magnitude, so that the gluino and squarks will have comparable masses, while the wino, bino, and sleptons will be much lighter, with masses suppressed by squares of electroweak coupling constants.

We can go considerably further than this with some reasonable dynamical assumptions. Suppose that the effects of supersymmetry breaking on the messenger superfields may be modeled by including these superfields along with a set of $SU(3) \times SU(2) \times U(1)$-neutral chiral superfields S_n (not necessarily all distinct) in a superpotential

$$f(\Phi, \bar{\Phi}, S) = \sum_n \lambda_n S_n \Phi_n \bar{\Phi}_n , \tag{28.6.5}$$

where $\bar{\Phi}_n$ and Φ_n are left-chiral messenger superfields in complex-conjugate representations of $SU(3) \times SU(2) \times U(1)$ and λ_n are a set of coupling coefficients. (Here and below we suppress the $SU(3) \times SU(2)$ indices that are summed over in calculating scalar products like $\Phi_n \bar{\Phi}_n$.) The superfields S_n are supposed to have non-vanishing vacuum expectation values $\mathscr{S}_n$ and $\mathscr{F}_n$ for their scalar and auxiliary components, respectively. It is the non-zero values of $\mathscr{F}_n$ that in these models introduce supersymmetry breaking in the masses of the Φ_n and $\bar{\Phi}_n$ particles. Section 26.4 shows that, with gauge couplings neglected, the squared masses of the spinor components of Φ_n (and $\bar{\Phi}_n$) are the eigenvalues of the matrix $\mathscr{M}_n^\dagger \mathscr{M}_n$, with $\mathscr{M}_n$ defined by Eq. (26.4.11), which gives

$$\mathscr{M}_n = \begin{pmatrix} 0 & \lambda_n \mathscr{S}_n \\ \lambda_n \mathscr{S}_n & 0 \end{pmatrix} ,$$

so that the messenger fermions have masses $|\lambda_n \mathscr{S}_n|$. To find the mass terms for the scalar components ϕ_n and $\bar{\phi}_n$ of the Φ_n and $\bar{\Phi}_n$ superfields, we note that integrating out the auxiliary fields of the Φ_n and $\bar{\Phi}_n$ yields a

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

potential

$$\sum_n \left|\frac{\partial f(\phi,\bar\phi,\mathcal{S})}{\partial\phi_n}\right|^2 + \sum_n \left|\frac{\partial f(\phi,\bar\phi,\mathcal{S})}{\partial\bar\phi_n}\right|^2 = \sum_n |\lambda_n \mathcal{S}_n|^2 \left[|\phi_n|^2 + |\bar\phi_n|^2\right] ,$$

to which we now must add the contribution of the auxiliary component of the S_n, given by the second term of Eq. (26.4.4) as:

$$2\mathrm{Re}\sum_n \left[\lambda_n \mathcal{F}_n \frac{\partial f(\phi,\bar\phi,\mathcal{S})}{\partial \mathcal{S}_n}\right] = 2\mathrm{Re}\sum_n \left[\mathcal{F}_n \lambda_n \phi_n \bar\phi_n\right] .$$

The complex scalar fields of definite mass are then $(\phi_n \pm e^{-i\alpha_n}\bar\phi_n)/\sqrt{2}$, where α_n is the phase of $\lambda_n \mathcal{F}_n$, with squared masses $|\lambda_n \mathcal{S}_n|^2 \pm |\lambda_n \mathcal{F}_n|$. (Note that this pattern, of a pair of complex scalars with squared masses equidistant above and below a Majorana fermion squared mass, is just what we would expect from the sum rule (27.5.11).) Since these squared masses must be positive, it follows that

$$|\mathcal{F}_n| \le |\lambda_n||\mathcal{S}_n|^2 . \tag{28.6.6}$$

The gaugino masses in models based on Eq. (28.6.5) are given by diagrams of the form shown in Figure 28.5, but now with just a single loop, not including what are shown as dotted lines in Figure 28.5. A detailed calculation gives the coefficients M_{gn} in Eq. (28.6.2) as[31]

$$M_{gn} = \frac{|\mathcal{F}_n|}{|\mathcal{S}_n|}\, g\left(\frac{|\mathcal{F}_n|}{|\lambda_n||\mathcal{S}_n|^2}\right) , \tag{28.6.7}$$

where

$$\begin{aligned} g(x) &= \frac{1}{2x^2}\left[(1+x)\ln(1+x) + (1-x)\ln(1-x)\right] \\ &= 1 + \frac{x^2}{6} + \frac{x^4}{15} + \frac{x^6}{28} + \cdots . \end{aligned} \tag{28.6.8}$$

The masses of the squarks and sleptons are given by the diagrams of Figure 28.6, which now involve just two loops. Another detailed calculation gives the mass parameters M_{sn}^2 in Eq. (28.6.4) as[31]

$$M_{sn}^2 = \frac{|\mathcal{F}_n|^2}{|\mathcal{S}_n^2|} f\left(\frac{|\mathcal{F}_n|}{|\lambda_n||\mathcal{S}_n|^2}\right) , \tag{28.6.9}$$

where

$$\begin{aligned} f(x) &= \frac{1+x}{x^2}\left[\ln(1+x) - 2\,\mathrm{Li}_2\left(\frac{x}{1+x}\right) + \frac{1}{2}\mathrm{Li}_2\left(\frac{2x}{1+x}\right)\right] + x \to -x \\ &= 1 + \frac{1}{36}x^2 - \frac{11}{450}x^4 - \frac{319}{11760}x^6 + \cdots , \end{aligned} \tag{28.6.10}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

with Li_2 the dilogarithm

$$\mathrm{Li}_2(x) \equiv -\int_0^x \frac{\ln(1-t)}{t} dt \,. \tag{28.6.11}$$

In particular, if (as is usually assumed) the various S_n are all the same, and if $|\mathscr{F}| \ll |\lambda_n||\mathscr{S}|^2$ for all n, then f and g in Eqs. (28.6.9) and (28.6.7) may be set equal to 1, so that

$$M_{gn} = M_{sn} = |\mathscr{F}|/|\mathscr{S}| \equiv M \,. \tag{28.6.12}$$

Using Eq. (28.6.3), we can express the gaugino masses (28.6.2) as

$$\begin{aligned} m_{\text{wino}} &= (g^2/16\pi^2)TM \,, \\ m_{\text{bino}} &= (5/3)(g'^2/16\pi^2)TM \,, \\ m_{\text{gluino}} &= (g_s^2/16\pi^2)TM \,, \end{aligned} \tag{28.6.13}$$

while the squared squark and slepton masses (28.6.4) become

$$\begin{aligned} M_Q^2 &= 2TM^2\left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 + \frac{3}{4}\left(\frac{g^2}{16\pi^2}\right)^2 + \frac{5}{3}\left(\frac{1}{6}\right)^2\left(\frac{g'^2}{16\pi^2}\right)^2\right] \,, \\ M_{\bar{U}}^2 &= 2TM^2\left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 + \frac{5}{3}\left(\frac{2}{3}\right)^2\left(\frac{g'^2}{16\pi^2}\right)^2\right] \,, \\ M_{\bar{D}}^2 &= 2TM^2\left[\frac{4}{3}\left(\frac{g_s^2}{16\pi^2}\right)^2 + \frac{5}{3}\left(-\frac{1}{3}\right)^2\left(\frac{g'^2}{16\pi^2}\right)^2\right] \,, \\ M_L^2 &= 2TM^2\left[\frac{3}{4}\left(\frac{g^2}{16\pi^2}\right)^2 + \frac{5}{3}\left(\frac{1}{2}\right)^2\left(\frac{g'^2}{16\pi^2}\right)^2\right] \,, \\ M_{\bar{E}}^2 &= 2TM^2\frac{5}{3}\left(\frac{g'^2}{16\pi^2}\right)^2 \,. \end{aligned} \tag{28.6.14}$$

There is no special reason to expect that $|\mathscr{F}| \ll |\lambda_n\mathscr{S}|^2$, but this assumption is not actually very restrictive, because Eq. (28.6.6) already requires that $|\mathscr{F}| \leq |\lambda_n||\mathscr{S}|^2$, and it turns out that the functions $f(x)$ and $g(x)$ do not differ much from unity for $x < 1$ unless x is very close to 1.

The extreme simplicity of the results (28.6.13) and (28.6.14) has been explained by Giudice and Rattazzi[32] by using Seiberg's arguments of holomorphy,[33] described in Section 27.6. Suppose we introduce a superpotential for the messenger superfields like (28.6.5), but with a single external singlet superfield S:

$$f(S,\Phi) = S\sum_n \lambda_n \Phi_n \bar{\Phi}_n \,. \tag{28.6.15}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

Among other effects the kinematic term for the gauge superfield V_i (with $i = 3, 2, 1$ for $SU(3)$, $SU(2)$, and $U(1)$) in the Wilsonian effective Lagrangian at a renormalization scale μ will now take the form

$$\mathscr{L}_{\text{gauge},\mu} = \text{Re}\left[\sum_i N_i(S,\mu)\sum_{\alpha\beta}\left(W_{iL\alpha}\epsilon_{\alpha\beta}W_{iL\beta}\right)\right]_{\mathscr{F}}, \tag{28.6.16}$$

with some functions $N_i(S,\mu)$ replacing the factors $1/2g_i^2(\mu)$ in Eq. (27.3.22). (The θ-term is dropped here because it has no effect in perturbation theory. A sum is implied over indices on $W_{iL\alpha}$ that label different members of the adjoint representations of $SU(3)$ and $SU(2)$, which are not explicitly shown.) The gauge coupling constant is now given by setting the superfield S equal to the expectation value $\mathscr{S}$ of its scalar component

$$\frac{1}{2g_i^2(\mu)} = N_i(\mathscr{S},\mu)\,. \tag{28.6.17}$$

Also, recalling that $W_{iL\alpha} = \lambda_{iL\alpha} + O(\theta)$ and using Eq. (27.2.11), the terms in the Lagrangian density (28.6.16) of second order in the gaugino fields are, up to derivative terms,

$$-2\sum_i \text{Re}\left[N_i(\mathscr{S},\mu)\left(\bar{\lambda}_{iR}\,\not{\partial}\lambda_{iR}\right) + [N_i(S,\mu)]_{\mathscr{F}}\left(\lambda_{iL}^{\mathrm{T}}\epsilon\lambda_{iL}\right)\right]\,.$$

This gives the gaugino masses

$$m_{gi}(\mu) = \left|\frac{[N_i(S,\mu)]_{\mathscr{F}}}{2N_i(\mathscr{S},\mu)}\right| = g_i^2(\mu)\left|[N_i(S,\mu)]_{\mathscr{F}}\right|\,. \tag{28.6.18}$$

Now let us consider the behavior of $N_i(\mathscr{S},\mu)$ as a function of a real positive $\mathscr{S}$, with the phases of the messenger superfields adjusted so that all λ_n are real and positive. Suppose we fix the values of the gauge couplings $g_i(\mu)$ at some scale $\mu = K$ above all the messenger particle masses. Taking account of the change in the constants b_i in the renormalization group equations $\mu dg_i(\mu)/d\mu = b_i g_i^3$ as μ passes through the various messenger masses, the solution of this equation when μ is *below* all the messenger particle masses takes the form

$$\frac{1}{g_i^2(\mu)} = \frac{1}{g_i^2(K)} - 2b_i^{(0)}\ln\left(\frac{M_1}{K}\right) - 2b_i^{(1)}\ln\left(\frac{M_2}{M_1}\right) - \cdots - 2b_i^{(N)}\ln\left(\frac{\mu}{M_N}\right)\,,$$

where we label the messenger particles so that their masses $M_n = \lambda_n\mathscr{S}$ satisfy

$$M_1 > M_2 > \cdots > M_N\,,$$

and $b_i^{(n)}$ is calculated taking account only of particles with masses less than M_n. Since all M_n are proportional to $\mathscr{S}$, we see that $N_i(\mathscr{S},\mu)$ has the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$\mathscr{S}$ dependence

$$N_i(\mathscr{S},\mu) = -b_i^{\text{messenger}} \ln \mathscr{S} + \mathscr{S}\text{-independent terms}\,, \tag{28.6.19}$$

where $b_i^{\text{messenger}} = b_i^{(0)} - b_i^{(N)}$ is the contribution to b_i of all the messenger superfields. According to Eq. (27.9.45) (with $C_{i1} = 0$ and $C_{i2}^f = C_{i2}^s = \sum_n T_{in}$), this is

$$b_i^{\text{messenger}} = \frac{1}{16\pi^2} \sum_n T_{in}\,. \tag{28.6.20}$$

Since $N_i(S,\mu)$ is required by supersymmetry to be a holomorphic function of S, we see that

$$N_i(S,\mu) = -\frac{1}{16\pi^2} \sum_n T_{in} \ln S + S\text{-independent terms}\,. \tag{28.6.21}$$

Expanding around $S = \mathscr{S}$, to first order in $\mathscr{F}$ we have $[\ln \mathscr{S}]_{\mathscr{F}} = \mathscr{F}/\mathscr{S}$ and so Eq. (28.6.18) gives the gaugino masses

$$m_{gi}(\mu) = \frac{g_i^2(\mu)}{16\pi^2} \sum_n T_{in} \left|\frac{\mathscr{F}}{\mathscr{S}}\right|\,. \tag{28.6.22}$$

Using Eq. (28.6.3), we see that this is the same as our previous result (28.6.13). Eq. (28.6.14) for the squark and slepton masses was obtained by Giudice and Rattazzi in a similar way, by studying the kinematic terms for the quark and lepton superfields instead of the gauge superfields.

Incidentally, Eqs. (28.6.13) and (28.6.14) could be obtained (in general with different values of M in Eq. (28.6.13) and Eq. (28.6.14)) without a specific dynamical assumption like Eq. (28.6.5), if we supposed that the supersymmetry-breaking sector respected invariance under some grand unified group G that had both the known quarks and leptons and the messenger fields as complete representations. In this case the coefficients M_{gn} and M_{sn} in Eqs. (28.6.2) and (28.6.4) would have values $M_g(d)$ and $M_s(d)$, respectively, that depend only on the irreducible representation d of G to which the nth messenger fields belong. The sums of T_{in} over n belonging to any irreducible representation d of G have the same ratios as the sums in Eq. (28.6.3), so $\sum_{n\in d} T_{in} = k_i T(d)$, where $k_3 = k_2 = 1$, $k_1 = 5/3$, and therefore

$$\sum_n T_{in} M_{gn} = \sum_d M_g(d) \sum_{n\in d} T_{in} = k_i M_g\,,$$

where $M_g = \sum_d M_g(d) T(d)$. Likewise

$$\sum_n T_{in} M_{sn}^2 = \sum_d M_s^2(d) \sum_{n\in d} T_{in} = k_i M_s^2\,,$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

where $M_s^2 = \sum_d M_s^2(d)T(d)$. Eqs. (28.6.2) and (28.6.4) would then yield Eqs. (28.6.13) and (28.6.14), except with M_g in place of TM and M_s^2 in place of $2TM^2$. But the assumption that M_{gn} and M_{sn} respect invariance under G is implausible if the messenger mass scale is far below the grand unified scale, since whatever we assume about a grand unified gauge group, the $SU(3) \times SU(2) \times U(1)$ gauge interactions would make coupling constants like λ_n run differently for Φ_n with different $SU(3) \times SU(2) \times U(1)$ quantum numbers in the same representation of the grand unified group.

These results are subject to various radiative corrections, of which the most important is that we must use values of g_s, g, and g' renormalized at a scale comparable to the mass being calculated. Indeed, the gaugino mass ratios given in Eq. (28.6.13) could also be derived under quite different assumptions: that all gaugino masses are equal at the grand unification scale where the coupling constants are related by $g_s^2 = g^2 = 5g'^2/3$ and become different at lower energies as described by the renormalization group equations.

For an illustrative example of numerical results, suppose that the messenger superfields form an $SU(2)$ singlet $SU(3)$ triplet with charge $e/3$ and an $SU(2)$ doublet $SU(3)$ singlet with charges 0 and $-e$, together with left-chiral superfields in the complex-conjugate representations of $SU(3) \times SU(2) \times U(1)$. Then, as already mentioned, Eq. (28.6.3) is satisfied with $T = 2 \times 1/2 = 1$, so by using the correct values of the gauge couplings, the masses of the squarks, gluino, L sleptons, wino, E sleptons, and bino are calculated to be in the ratios[34] 11.6 :: 7.0 :: 2.5 :: 2 :: 1.1 :: 1.0. In larger representations of G we could have $T \gg 1$, in which case the gluino would be the heaviest of these particles, and the sleptons the lightest ones.

There are also radiative corrections aside from the running of the gauge couplings. According to one calculation,[34] in the model with an $SU(2)$ singlet $SU(3)$ triplet with charge $e/3$ and an $SU(2)$ doublet $SU(3)$ singlet with charges 0 and $-e$, together with left-chiral superfields in the complex-conjugate representations of $SU(3) \times SU(2) \times U(1)$, radiative corrections give the ratios of the masses of the squarks, gluino, L sleptons, wino, E sleptons, and bino as 9.3 :: 6.4 :: 2.6 :: 1.9 :: 1.35 :: 1.0.

As we saw in the previous section, the wino and bino may mix with the charged and neutral higgsinos, so the wino and bino masses that have been calculated here must be regarded as inputs to a calculation of physical masses of mixtures known as *charginos* and *neutralinos*, rather than physical masses themselves.

Now we shall consider the masses of the Higgs scalars in these models. If we were to consider only the two-loop diagrams by which these scalars get masses through gauge interactions with the supersymmetry-breaking sector, then since (apart from signs) they have the same $SU(3) \times SU(2) \times$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$U(1)$ quantum numbers as the left-handed lepton doublets, their masses would be given by a formula like the fourth of Eqs. (28.6.4)

$$[m_1^2]_{2\text{ loop}} = [m_2^2]_{2\text{ loop}} = M_L^2$$
$$= \sum_n M_{sn}^2 \left[\frac{3}{4} \left(\frac{g^2}{16\pi^2} \right)^2 T_{2n} + \left(\frac{1}{2} \right)^2 \left(\frac{g'^2}{16\pi^2} \right)^2 T_{1n} \right]^2 . \quad (28.6.23)$$

If this were the whole story, then it would be impossible to meet the condition for $SU(2) \times U(1)$ breaking found in the previous section that (unless $\tan\beta$ is very close to unity) one of $m_1^2 + |\mu|^2$ and $m_2^2 + |\mu|^2$ must be negative. Fortunately, the large masses of the top quark and squark produce a negative contribution to m_2^2 that leads naturally to a spontaneous breakdown of electroweak symmetry. The couplings of the Higgs doublets to the third-generation quark superfields are described by the superpotential

$$f_{3\text{rd gen}} = \lambda_b \left(H_1^{\mathrm{T}} e Q \right) \bar{B} + \lambda_t \left(H_2^{\mathrm{T}} e Q \right) \bar{T} , \quad (28.6.24)$$

where Q is the $SU(2)$ quark doublet left-chiral superfield (T, B), while $\bar{T}$ and $\bar{B}$ are the left-chiral superfields of the left-handed top and bottom antiquarks, and λ_t and λ_b are the Yukawa couplings, related to the t and b quark masses by $m_t = \lambda_t v_2$ and $m_b = \lambda_b v_1$. The last term of Eq. (26.4.7) then gives the terms in the potential involving an interaction between the squark fields and Higgs fields as

$$V_{sq\,H} = \left| \lambda_b \mathscr{H}_1^- \bar{\mathscr{B}} + \lambda_t \mathscr{H}_2^0 \bar{\mathscr{T}} \right|^2 + \left| \lambda_b \mathscr{H}_1^0 \bar{\mathscr{B}} + \lambda_t \mathscr{H}_2^+ \bar{\mathscr{T}} \right|^2$$
$$+ \left| \lambda_b \right|^2 \left| \mathscr{H}_1^0 \mathscr{B} - \mathscr{H}_1^- \mathscr{T} \right| + \left| \lambda_t \right|^2 \left| \mathscr{H}_2^+ \mathscr{B} - \mathscr{H}_2^0 \mathscr{T} \right|^2 , \quad (28.6.25)$$

where script letters denote the scalar field components of superfields. The squark loop contributions to the potential of $\mathscr{H}_1$ and $\mathscr{H}_2$ are then

$$V_H^{\text{squark loop}} = 3\langle \mathscr{S}\mathscr{S}^* \rangle \left[2|\lambda_b|^2 \left(\mathscr{H}_1^\dagger \mathscr{H}_1 \right) + 2|\lambda_t|^2 \left(\mathscr{H}_2^\dagger \mathscr{H}_2 \right) \right] , \quad (28.6.26)$$

where $\langle \mathscr{S}\mathscr{S}^* \rangle$ is the vacuum expectation value of the product of any one of the squark fields and its complex conjugate at the same spacetime point. (In taking this to be the same for all squark types, we are here using Eq. (28.6.4), which tells us that the squark masses M_Q do not vary much among $\mathscr{T}$, $\mathscr{B}$, $\bar{\mathscr{T}}$, and $\bar{\mathscr{B}}$ squarks. The factor 3 in Eq. (28.6.26) takes account of the three colors of each squark type.) The vacuum expectation value $\langle \mathscr{S}\mathscr{S}^* \rangle$ is given in lowest order by

$$\langle \mathscr{S}\mathscr{S}^* \rangle \equiv \langle \mathscr{S}(x)\, \mathscr{S}^*(x) \rangle_{\text{VAC}} = \frac{-i}{(2\pi)^4} \int \frac{d^4 p}{p^2 + M_Q^2 - i\epsilon} . \quad (28.6.27)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

This of course is divergent, but the contribution to supersymmetry-breaking coefficients would be cancelled by quark loops if the squarks had the zero bare mass of the quarks, so the effect of quark loops is to subtract from (28.6.27) the same expression with M_Q replaced with zero, which with a Wick rotation becomes

$$\begin{aligned}\langle \mathscr{S}\mathscr{S}^* \rangle &\to \frac{M_Q^2\, i}{(2\pi)^4}\int \frac{d^4p}{(p^2+M_Q^2-i\epsilon)(p^2-i\epsilon)} \\ &= -\frac{M_Q^2}{(2\pi)^4}\int_0^{M^2}\frac{\pi^2 dp^2}{p^2+M_Q^2} \simeq -\frac{M_Q^2}{16\pi^2}\ln\left(\frac{M^2}{M_Q^2}\right)\,.\end{aligned}$$

We have inserted an ultraviolet cut-off at the messenger mass M, because at momenta above M the squark mass must be replaced with a momentum-dependent mass, which goes to the supersymmetric value zero at very high momenta. Making this substitution in Eq. (28.6.26) gives a net contribution to the potential due to squark and quark loops:

$$(V_m)^{3\text{ loop}} = \frac{3M_Q^2}{16\pi^2}\ln\left(\frac{M^2}{M_Q^2}\right)\left[2|\lambda_b|^2\left(\mathscr{H}_1^\dagger\mathscr{H}_1\right)+2|\lambda_t|^2\left(\mathscr{H}_2^\dagger\mathscr{H}_2\right)\right]\,. \tag{28.6.28}$$

(This is a three-loop contribution, because the squared squark masses are given by two-loop diagrams. There are also terms in the potential that are quadratic in both the Higgs and squark superfields, arising from products of the Higgs and squark terms in the squares of the $SU(2)\times U(1)$ gauge field D-components. These do not make a three-loop contribution to the Higgs masses because the sum of each $SU(2)\times U(1)$ quantum number of the squarks vanishes, so that their contributions to the $SU(2)\times U(1)$ gauge field D-components have zero vacuum expectation value.) Comparing Eq. (28.6.28) with Eq. (28.5.9) and adding the two-loop contribution (28.6.23) to the masses, we see then that

$$m_1^2 \simeq M_L^2 - \frac{3M_Q^2|\lambda_b|^2}{8\pi^2}\ln\left(\frac{M^2}{M_Q^2}\right)\,, \tag{28.6.29}$$

$$m_2^2 \simeq M_L^2 - \frac{3M_Q^2|\lambda_t|^2}{8\pi^2}\ln\left(\frac{M^2}{M_Q^2}\right)\,. \tag{28.6.30}$$

Using Eq. (28.6.14) and $|\lambda_t| = m_t/v_2 = m_t(2\sqrt{2}G_F)^{1/2}/\sin\beta$, we may write

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

(28.6.30) as

$$m_2^2 \simeq 2TM^2\left[\frac{3}{4}\left(\frac{g^2}{16\pi^2}\right)^2+\frac{5}{12}\left(\frac{g'^2}{16\pi^2}\right)^2 -\frac{\sqrt{2}G_F m_t^2}{\pi^2\sin^2\beta}\left(\frac{g_s^2}{16\pi^2}\right)^2\ln\left(\frac{3}{8T(g_s^2/16\pi^2)}\right)\right]. \tag{28.6.31}$$

For $T = 1$, $g_s^2/4\pi = 0.118$, $g^2/4\pi = 0.0340$, $g'^2/4\pi = 0.0101$, and $m_t = 180$ GeV, this is

$$m_2^2 \simeq M_L^2\left[1-\frac{3.06}{\sin^2\beta}\right], \tag{28.6.32}$$

which is negative for all values of β, thus providing a natural mechanism for the spontaneous breaking of the electroweak gauge symmetry. Also, $M_L^2 = (0.91\times 10^{-4})M^2/8\pi^2$. Unless $\tan\beta$ is huge, we have $|\lambda_b| \ll |\lambda_t|$, so Eq. (28.6.27) gives

$$m_1^2 \simeq M_L^2 . \tag{28.6.33}$$

The aspect of electroweak phenomenology for which the predictions of gauge-mediated supersymmetry models are most uncertain and unsatisfactory has to do with the parameter μ in the supersymmetry-preserving term $\mu[(H_1^{\mathrm{T}}eH_2)]_{\mathscr{F}}$ and the related supersymmetry-breaking term $B\mu$ in the Lagrangian density. These are related because the interactions of the Higgs superfields with the gauge, lepton, and quark superfields are invariant under a symmetry

$$\begin{aligned} &H_1 \to e^{i\varphi}H_1 , && H_2 \to e^{i\varphi}H_2 , \\ &Q \to e^{-i\varphi}Q , && V_i \to V_i , \\ &\bar{D} \to \bar{D} , && \bar{U} \to \bar{U} , \end{aligned} \tag{28.6.34}$$

which in the absence of a superpotential term $\mu(H_1^{\mathrm{T}}eH_2)$ would forbid radiative corrections from producing a term $B\mu\,\mathrm{Re}\,(\mathscr{H}_1^{\mathrm{T}}e\mathscr{H}_2)$ in the scalar field potential.

It is not possible for $B\mu$ to vanish because then Eq. (28.5.22) and the fact that (as we have seen) $m_A \neq 0$ would imply that $\sin 2\beta = 0$, or in other words either $v_1 = 0$ or $v_2 = 0$, which would imply that either all the charge $-e/3$ quarks and charged leptons are massless or all the charge $+2e/3$ quarks are massless. (If $B\mu = 0$ and $\mu = 0$ then this problem persists to all orders, because the appearance of non-zero vacuum expectation values for both v_1 and v_2 would imply that symmetry under any combination of the transformation (28.6.34) and electroweak $U(1)$ gauge transformations is spontaneously broken, so the C-odd neutral scalar would be a Goldstone boson with $m_A = 0$.) It is natural to try to account for a non-zero value

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

of $B\mu$ as a radiative correction in a theory in which symmetry under the transformation (28.6.34) is explicitly broken by a supersymmetric term $\mu[(H_1^{\mathrm{T}} e H_2)]_{\mathscr{F}}$ in the Lagrangian density. This gives a very small value of $B\mu$ at the messenger scale,[34] although renormalization group effects greatly increase $B\mu$ at lower energies. According to Eq. (28.5.22), a relatively small $B\mu$ would fit well with the idea that the large mass of the top quark arises from a large value of $\tan\beta$. In any case, as already mentioned, the experimental lower bound on the chargino mass tells us that $|\mu|$ is at least about 60 GeV.

The trouble is that the appearance of a non-zero value of μ resurrects the hierarchy problem that supersymmetry was supposed to solve: instead of asking why the Higgs mass terms in the Lagrangian density are so much smaller than the Planck mass or the mass at which gauge couplings are unified, we now have to ask why μ is this small?

The hierarchy problem would be put to rest if the Higgs superfields interact with the supersymmetry-breaking sector in such a way that a term $\mu[(H_1^{\mathrm{T}} e H_2)]_{\mathscr{F}}$ is forbidden by some symmetry, but appears when that symmetry is spontaneously broken. A massless Goldstone boson may be avoided if the symmetry is discrete rather than continuous. The simplest possibility is simply to extend the symmetry transformation (28.6.34) to include a transformation

$$S \to e^{-2i\varphi} S \,,$$

which would allow a term in the superpotential of the form

$$\lambda' S(\mathscr{H}_1^{\mathrm{T}} e \mathscr{H}_2) \,.$$

We can avoid a continuous symmetry by also including a term S^3 in the superpotential, so that the Lagrangian is invariant only under transformations with φ a multiple of $2\pi/3$, which is still enough to forbid a non-zero bare value of μ. In this case the appearance of non-zero vacuum values $\mathscr{S}$ and $\mathscr{F}$ for the scalar and auxiliary components of S yields

$$B\mu = |\lambda' \mathscr{F}| \,, \qquad \mu = |\lambda' \mathscr{S}| \,.$$

This would result in B having a very large $|\lambda'|$-independent value M, given by (28.6.12), which is greater by a factor of order $(g_s^2/16\pi)^{-1} \simeq 100$ than the squark or gluino masses. But then, since Eqs. (28.6.32) and (28.6.33) give $m_1^2 + m_2^2 < 0$, the stability condition (28.5.11) would require that $|\mu| \geq M/2$, and hence Eq. (28.5.22) would require that m_A is also much greater than the squark and slepton masses. This is ruled out by the relation (28.5.23) and the estimates (28.6.33) and (28.6.32) of m_1^2 and m_2^2 unless $\tan\beta$ is very close to unity.

We will see in Section 31.6 that theories of gravity-mediated supersymmetry breaking naturally yield acceptable values of $B\mu$ and μ. Such

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

theories are characterized by a very high energy scale of supersymmetry breaking, in various versions of order either 10^{11} GeV or 10^{13} GeV. There have been several suggestions[35] on ways to obtain acceptable values of $B\mu$ and μ in theories in which supersymmetry is broken at relatively low energy, such as theories of gauge-mediated supersymmetry breaking, but none of them is particularly compelling. Also, since we do not know where μ comes from, we do not have any reason to suppose that it is real, so theories of gauge-mediated supersymmetry breaking are at risk of yielding too much CP violation, just as in the more general framework of Section 28.4.

Like the scalar squared masses m_1^2 and m_2^2, the parameters A_{ij} and C_{ij} in Eq. (28.4.1) are given by two-loop diagrams. However, they have the dimensions of mass rather than mass squared, and they are much less than the scalar and gaugino masses, so they make a relatively unimportant contribution to supersymmetry breaking.

As in any model with supersymmetry broken at energies much less than 10^{10} GeV, the lightest R-odd particle in all models based on gauge-mediated supersymmetry breaking is the gravitino. As we will see in Section 31.3, the gravitino mass is of the order of $\sqrt{G}$ times the squared energy F characteristic of supersymmetry breaking, defined so that the vacuum energy is $F^2/2$. Where supersymmetry is broken by the $\mathscr{F}$-terms $\mathscr{F}_{n0}$ of the $SU(3) \times SU(2) \times U(1)$-neutral chiral superfields S_n, we have $F^2 = \sum_n |\mathscr{F}_{n0}|^2$. If there are no large dimensionless quantities in the Lagrangian for the S_n, then squark masses in this model are of the order of $g_s^2\sqrt{F}/16\pi^2 \approx 10^{-2}\sqrt{F}$, so in order for this to be less than the naturalness bound of 10^4 GeV, we must have $\sqrt{F} < 10^6$ GeV, which gives a gravitino mass less than 1 keV. Gravitational couplings are so weak at accessible energies that it is only the helicity $\pm 1/2$ states of the gravitino that can actually be produced, and these behave just like goldstino states. As shown by Eq. (27.5.12), in the models discussed here the goldstino field appears with coefficient $i\sqrt{2}\mathscr{F}_n$ in the fermionic component ψ_n of S_n. Goldstinos are emitted in the decay of the R-odd sparticles into the corresponding R-even particles of the standard model through radiative corrections, with goldstinos emerging from vertices connecting internal ψ_n and $\mathscr{S}_n$ lines. According to Eq. (29.2.10), the goldstino emission amplitudes are inversely proportional to F, which makes these decays comparatively slow, although perhaps fast enough to be observed.

Because the decay of R-odd particles into goldstinos is slow, it is phenomenologically important in these models to identify the next-to-lightest R-odd particle, into which all the heavier R-odd particles will decay before it in turn decays into a goldstino. As we have seen, the next-to-lightest R-odd particle is usually a slepton, wino, or bino. (It is possible in models with a messenger superfield having the same $SU(3) \times SU(2) \times U(1)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

quantum numbers as the Higgs doublets that the mixing between these superfields would lower the mass of the doublet messengers so much that the lightest R-odd particle would be the gluino.[35a]) Detailed calculations including the effects of $SU(2) \times U(1)$ breaking indicate that here is a large region of parameter space in which the next-to-lightest R-odd particle is one of the two tau sleptons.[35b]

28.7 Baryon and Lepton Non-Conservation

The extra particles in supersymmetric models provide several new mechanisms for baryon and lepton non-conservation. We saw in Section 28.1 that there are various baryon- and lepton-number non-conserving supersymmetric operators (28.1.2) and (28.1.3) of dimensionality four that can be included in a renormalizable $SU(3) \times SU(2) \times U(1)$-invariant theory and that would lead to processes like proton decay at a catastrophic rate. These terms can be excluded from the Lagrangian by imposing R parity conservation (or, equivalently, invariance under a change of sign of all quark and lepton chiral superfields) but this does not exclude various $SU(3) \times SU(2) \times U(1)$-invariant but baryon- and lepton-number non-conserving operators of dimensionality $d > 4$. As discussed in Section 21.3, if there is an underlying mechanism for baryon and lepton non-conservation characterized by some high mass scale M, then these operators will appear in the effective Lagrangian of the standard model with coefficients proportional to M^{4-d}. With only the fields of the non-supersymmetric standard model, operators that can violate baryon conservation have a minimum dimensionality six,[36] and thus give baryon non-conserving amplitudes proportional to M^{-2}. The new fields required by supersymmetry lead to two important changes in estimates of baryon non-conserving processes like proton and bound neutron decay. As we saw in Section 28.2, the change in the renormalization group equations gives larger estimates of M, which decreases the effect of the dimension six operators. At the same time, these new fields allow the construction of new operators of dimensionality five, which give baryon non-conserving amplitudes proportional to M^{-1} and are therefore likely to make the dominant contribution to proton and bound neutron decay.[37]

The supersymmetric operators of dimensionality five that can be formed out of chiral superfields (generically called Φ) are of the forms $(\Phi^* \Phi \Phi)_D$ and $(\Phi \Phi \Phi \Phi)_{\mathscr{F}}$, and their complex conjugates. (We do not consider operators that include derivatives or gauge fields, because they do not offer any additional possibility of baryon or lepton non-conservation.) In the notation of Section 28.1, the $SU(3) \times SU(2) \times U(1)$-invariant operators

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

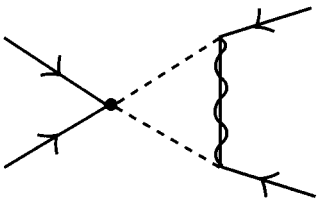

Figure 28.7. A diagram that can produce a four-fermion interaction among quarks and/or leptons that violates baryon and lepton number conservation. Here solid lines are quarks and/or leptons; dashed lines are squarks and/or sleptons; the combined solid and wavy line is a gaugino; and the dot is a vertex arising directly from the $\mathscr{F}$-term interactions (28.7.3).

of dimensionality five that also conserve R parity are

$$(LLH_2H_2)_{\mathscr{F}}\,, \tag{28.7.1}$$

$$(L\bar{E}H_2^*)_D\,,\quad (Q\bar{D}H_2^*)_D\,,\quad (Q\bar{U}H_1^*)_D\,,\quad (QQ\bar{U}\bar{D})_{\mathscr{F}}\,,\quad (Q\bar{U}L\bar{E})_{\mathscr{F}}\,, \tag{28.7.2}$$

and

$$(QQQL)_{\mathscr{F}}\,,\qquad (\bar{U}\bar{U}\bar{D}\bar{E})_{\mathscr{F}}\,, \tag{28.7.3}$$

with obvious contraction of indices as dictated by $SU(3)$ and $SU(2)$ conservation. The interaction (28.7.1) is the supersymmetric version of the dimensionality five operator which would provide small neutrino masses in some theories.[38] The interactions (28.7.2) only provide small corrections to processes that already occur in the renormalizable terms of the supersymmetric standard model. It is the interactions (28.7.3) that provide new mechanisms for baryon as well as lepton non-conservation.

According to Eq. (26.4.4), the quarks and leptons enter in the interactions (28.7.3) through terms involving a pair of quark and/or lepton fields and a pair of squark and/or slepton fields. In order to generate reactions among quarks and leptons alone, it is necessary for the pair of squarks and/or sleptons to be converted into a pair of quarks and/or leptons by exchanging a gaugino in the one-loop diagram shown in Figure 28.7. This will produce effective four-fermion $qqq\ell$ interactions with $d = 6$ among three quarks and a lepton. The coupling g_6 of these interactions will be proportional to the square of the gauge coupling g or g' or g_s of the gaugino, to the supersymmetry-breaking mass of the gaugino, to the inverse square of the larger of the gaugino and squark or slepton masses (which is needed to give a coupling of the right dimensionality), and to a factor of order $1/8\pi^2$ from the loop integration.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

It might be thought that the greater strength of the gluino coupling would make gluino exchange the dominant contribution to g_6. (Indeed, in gauge-mediated supersymmetry-breaking theories, for moderate values of the trace (28.6.3) and with $g \approx g'$, Eqs. (28.6.13)–(28.6.14) give

$$m_{\text{gluino}} \approx m_{\text{squark}} \approx \frac{g_s^2}{16\pi^2} M_* \,,$$

$$m_{\text{wino}} \approx m_{\text{slepton}} \approx m_{\text{bino}} \approx \frac{g^2}{16\pi^2} M_* \,,$$

where M_* is a mass characterizing the messenger sector. Therefore in such theories individual gluino exchange diagrams give contributions proportional to g_s^2/m_{gluino}, while wino or bino exchange (or, more precisely, chargino or neutralino exchange) makes a contribution proportional to $g^2 m_{\text{wino}}/m_{\text{squark}}^2$, which is smaller by a factor roughly equal to $m_{\text{wino}} g^2/m_{\text{gluino}} g_s^2 \approx g^4/g_s^4$.) However, there is a cancellation among the diagrams for gluino exchange that strongly suppresses their contribution. This was originally shown using a Fierz identity among four-fermion operators,[39] but the same result can be obtained with no equations at all. To conserve color, the coefficients of the operators $(QQQL)_{\mathcal{F}}$ and $(\bar{U}\bar{U}\bar{D}\bar{E})_{\mathcal{F}}$ must be totally antisymmetric in the colors of the three quark or antiquark superfields, and since these superfields are bosonic, they must then also be antisymmetric in their flavors as well. Gluino interactions are flavor-independent, so if we can neglect the flavor dependence of the squark masses then the coefficients of the four-fermion $d = 6$ operators will also be totally antisymmetric in flavor as well as color. Fermi statistics then requires the coefficients of these operators to be also totally antisymmetric in the spin indices of the quark or antiquark fields. But the three quark or antiquark fields in the $d = 6$ operators derived by gaugino exchange from $QQQL_{\mathcal{F}}$ or $(\bar{U}\bar{U}\bar{D}\bar{E})_{\mathcal{F}}$ are all left-handed and therefore have only two independent spin indices, so no coefficient can be antisymmetric in all three spins. The contribution of gluino exchange to the $d = 6$ operators would therefore vanish if the squark masses were all equal and so this contribution is suppressed by the fractional differences among the masses of the different squarks. According to Eq. (28.6.4), in theories of gauge-mediated supersymmetry breaking the fractional differences between the $\bar{U}$ and $\bar{D}$ squark masses are of order g'^4/g_s^4, so gluino exchange between the antisquarks in the $(\bar{U}\bar{U}\bar{D}\bar{E})_{\mathcal{F}}$ operator generates a dimension six four-fermion operator with a coefficient of the same order as that generated by bino exchange. However, since gluinos conserve flavor this operator like the $(\bar{U}\bar{U}\bar{D}\bar{E})_{\mathcal{F}}$ operator must be totally antisymmetric in antiquark flavors, so that it must involve c or t quarks and therefore cannot contribute directly to proton or bound neutron decay. On the other hand, Eq. (28.6.4) indicates that the fractional mass differences among the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

Q quarks of different flavors is much less than of order g^4/g_s^4, so that gluino exchange between the squarks in the $(QQQL)_{\mathscr{F}}$ operator makes a contribution to g_5 that is much smaller than wino or bino exchange. We conclude that at least in the theories of gauge-mediated supersymmetry breaking, gluino exchange makes a smaller contribution to proton or bound neutron decay than wino or bino exchange. In other models gluino exchange may make a contribution comparable to that of other processes.[40]

For $g \approx g'$ and $m_{\text{wino}} \approx m_{\text{bino}}$, the contribution of wino or bino exchange to the dimension six operators is of order

$$g_6 \approx \frac{g^2\, g_5\, m_{\text{wino}}}{8\pi^2 m_{\text{squark}}^2}\,, \tag{28.7.4}$$

where g_5 is a typical value of the couplings of the effective $d = 5$ interactions (28.7.3). If the wino and squark masses have the same ratio (g^2/g_s^2) as in theories of gauge-mediated supersymmetry breaking, then this gives

$$g_6 \approx \frac{g^4\, g_5}{8\pi^2 g_s^2 m_{\text{squark}}}\,. \tag{28.7.5}$$

The four-fermion $qqq\ell$-terms of dimensionality six in the effective Lagrangian are the same as those that had been supposed to generate processes like proton decay in non-supersymmetric theories.[36] They produce proton and bound neutron decay at a rate which on dimensional grounds must be of the form

$$\Gamma_N = c_N\, m_N^5\, g_6^2\,, \tag{28.7.6}$$

where c_N is a pure number that must be calculated by non-perturbative calculations in quantum chromodynamics. Much work has gone into these calculations, with results[41] generally in the range $c_N \approx 3 \times 10^{-3 \pm 0.7}$.

To estimate g_5, we note that it is not possible to produce $\mathscr{F}$-terms like (28.7.3) that involve only left-chiral superfields by the tree-approximation exchange of gauge supermultiplets, which always interact with both left-chiral superfields and their right-chiral complex conjugates. Thus the interactions (28.7.3) arise in the tree approximation only from the exchange of particles of chiral superfields, and therefore g_5 is of the order of g_T^2/M_T^2, where g_T is a typical baryon and lepton non-conserving coupling of some superheavy left-chiral superfield of mass M_T to the quark and lepton superfields. To produce the interactions (28.7.3), these superheavy particles must be color triplets or antitriplets, and $SU(2)$ triplets or singlets. Whatever gauge group unifies the strong and electroweak interactions presumably dictates some relation between the interactions of the superheavy color triplet T and the familiar color singlet H_1 and H_2. Then g_T will be of the same order as the Yukawa coupling constants in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

the interactions (28.1.2) and (28.1.3) that give mass to the known quarks and leptons, and which are equal to quark or lepton masses divided by the vacuum expectation values of order $G_F^{-1/2} \simeq 300$ GeV of $\mathscr{H}_1^0$ or $\mathscr{H}_2^0$. We therefore take

$$g_5 \approx \frac{G_F m_f^2}{M_T} , \tag{28.7.7}$$

where m_f is some typical quark or lepton mass. We have seen that the dimension five operators are antisymmetric in quark flavors, so as a compromise between the masses of the s quark and the u or d quarks, we will take $m_f = 30$ MeV. Combining Eqs. (28.7.5)–(28.7.7) and taking $M_T = 2 \times 10^{16}$ GeV (as suggested by the results of Section 28.2), $c_N = 0.003$, $g_s^2/4\pi = 0.118$, $g^2/4\pi = 1/(0.23 \times 137)$, and $m_{\text{squark}} = 1$ TeV, we find a proton (or bound neutron) lifetime Γ_N^{-1} of about 2×10^{31} years.[42] This is not very different from experimental lower bounds on the partial lifetimes of what are expected to be the leading modes of proton decay, variously quoted as ranging from 10^{31} to 5×10^{32} years. At the time of writing the most stringent bounds are set by the non-observation of proton decay in the large Super Kamiokande neutrino detector in Japan:[42a] the partial lifetimes for the decays $p \to e^+\pi^0$ and $p \to \bar{\nu}K^+$ are greater than 2.1×10^{33} years and 5.5×10^{32} years, respectively. There is an uncertainty of a factor of at least 100 in the above estimate of the theoretical lifetime from the uncertainty in the squark mass alone, so it is too soon to say that there is any discrepancy between experiment and theoretical expectations. On the other hand, supersymmetry raises the possibility that baryon non-conservation may be discovered soon.

We can also say a little on general grounds about the expected branching ratios for various proton and bound neutron decay modes. As we have mentioned, the dimension five operators (28.7.3) must be totally antisymmetric in the flavor of the quark superfields, so the only operators that concern us here are of the forms $(U_i D_j D_k N_\ell)_{\mathscr{F}}$, $(D_i U_j U_k E_\ell)_{\mathscr{F}}$, and $(\bar{D}_i \bar{U}_j \bar{U}_k \bar{E}_\ell)_{\mathscr{F}}$, where i, j, k, ℓ are generation indices, and in each case $j \neq k$. The exchange of neutral winos or binos then produces $d = 6$ four-fermion operators of the forms $u_i d_j d_k \nu_\ell$, $d_i u_j u_k e_\ell$, and $\bar{d}_i \bar{u}_j \bar{u}_k \bar{e}_\ell$, with $j \neq k$ and i and ℓ arbitrary, while charged wino exchange produces the same four-fermion operators with $i \neq j$ and k and ℓ arbitrary. The only quarks light enough to be involved in proton decay are u, s, and d; ignoring all the others and the small mixing angles in the third generation, we have

$$u_1 = u , \quad d_1 = d \cos\theta_c + s \sin\theta_c , \quad d_2 = -d \sin\theta_c + s \cos\theta_c ,$$

where θ_c is the Cabibbo angle, while u_2, u_3, and d_3 can be ignored. The four-fermion operators that can be produced by wino or bino exchange and can contribute to proton or bound neutron decay thus are

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

$u\,d\,s\,\nu_\ell \cos(2\theta_c)$, $u\,d\,d\,\nu_\ell \sin(2\theta_c)$, $u\,u\,s\,e_\ell \cos\theta_c$, and $u\,u\,d\,e_\ell \sin\theta_c$, plus others with quarks and leptons replaced with antiquarks and antileptons. All other things being equal, the dominant decay modes are therefore $p \to K^+\bar{\nu}$, $n \to K^0\bar{\nu}$, $p \to K^0 e^+$, and $p \to K^0\mu^+$, while the rates for the decay modes $p \to \pi^+\bar{\nu}$, $n \to \pi^0\bar{\nu}$, $p \to \pi^0 e^+$, $p \to \pi^0\mu^+$, and $n \to \pi^- e^+$ are suppressed by factors $\sin^2\theta_c = 0.05$, though also enhanced somewhat by the greater phase space available.

These considerations do not lead to definite predictions of branching ratios, because in addition to all the factors mentioned above, the coefficients generically called g_5 of the operators (28.7.3) may have a strong dependence on the superfield flavors appearing in these operators. To go further, one needs a specific theory for the generation of the dimension five operators. Most of the authors of Reference 42 concluded on the basis of a supersymmetric version of $SU(5)$ theories that proton and bound neutron decay would be dominated by the processes $p \to K^+\bar{\nu}$ and $n \to K^0\bar{\nu}$, but for a model based on $SO(10)$ charged lepton modes can become prominent.[43] Also, in some models higgsino exchange can compete with wino and bino exchange,[44] increasing the rate of $p \to K^+\bar{\nu}$. It seems a good idea in searches for baryon non-conservation to keep an open mind as to the decay modes to be expected in proton or bound neutron decay.

Of course, it is possible that all of these baryon non-conserving processes are prohibited by some sort of conservation law. As mentioned in Section 28.1, string theory argues against baryon conservation being a fundamental global continuous symmetry, but the baryon non-conserving operators (28.7.3) may be forbidden by a $\mathbb{Z}_3$ multiplicative symmetry known as baryon parity,[45] under which the Q superfield is neutral; the H_2 and $\bar{D}$ superfields are multiplied by the phase $\exp(i\pi/3)$, and the L, H_1, $\bar{U}$, and $\bar{E}$ superfields are multiplied by the opposite phase $\exp(-i\pi/3)$. This symmetry is designed to allow the fundamental Yukawa couplings (28.1.2) and (28.1.3) as well as the μ-term (28.5.7) and the lepton non-conserving terms (28.1.4) and (28.7.1), but it rules out the dimension four baryon non-conserving term (28.1.5) and the dimension five baryon non-conserving terms (28.7.3). This symmetry is spontaneously broken by the appearance of vacuum expectation values of $\mathscr{H}_1^0$ and $\mathscr{H}_2^0$ (and perhaps the sneutrino fields $\mathscr{N}$), and with no conservation law for R parity, there would be nothing to keep the lightest supersymmetric particle stable.

Problems

1. Suppose that the interactions (28.1.4) and (28.1.5) were actually present in the Lagrangian of a supersymmetric version of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

standard model. Make a rough estimate of how heavy the squarks and sleptons would have to be to avoid a conflict with experimental bounds on the proton lifetime.

2. Suppose that the typical mass m of the gauginos, higgsinos, squarks and sleptons is very much larger than m_Z. Give the renormalization group equations for the running gauge couplings at energies both above and below m. Use the results, together with the unification assumption employed in Section 28.2, to give formulas for $\sin^2\theta$ and the unification scale M in terms of m, m_Z, $e(m_Z)$, $g_s(m_Z)$, and n_s. How large could m be without violating experimental bounds on $\sin^2\theta$ and M?

3. Give formulas for the couplings of the quarks and leptons to the lightest CP-even neutral scalar particle in the minimum supersymmetric standard model in terms of the parameters m_A, m_Z, β, G_F, and the quark and lepton masses.

4. Use holomorphy arguments to derive the one-loop formula for the gluino mass in a theory of gauge-mediated supersymmetry breaking, in which the messenger superfields Φ_n and $\bar{\Phi}_n$ get their masses from a term $\sum_n \lambda_n S_n(\bar{\Phi}_n\Phi_n)$ in the superpotential, in terms of the expectation values $\mathscr{S}_n$ and $\mathscr{F}_n$ of the ϕ- and $\mathscr{F}$-components of the singlet superfields S_n, in the limit where $|\mathscr{F}_n| \ll |\lambda_n||\mathscr{S}_n|^2$.

5. Taking account of the possibility of a small flavor dependence of squark masses, estimate the gluino exchange contribution to baryon and lepton non-conserving four-fermion interactions among the quarks and leptons. Set an upper bound on these contributions, using the bound on squark mass splittings from the rate of $K^0 \to \overline{K}^0$ conversion.

References

1. We will not consider here the possibility of unification at much lower energies raised by I. Antoniadis, *Phys. Lett.* **B246**, 377 (1990); J. Lykken, *Phys. Rev.* **D54**, 3693 (1996), and revived by N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *Phys. Lett.* **B429**, 263 (1998); K. R. Dienes, E. Dudea, and T. Gherghetta, *Phys. Lett.* **B436**, 55 (1998); I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *Phys. Rev. Lett.* **B436**, 257 (1998).

1*a*. S. Weinberg, in *Proceedings of the XVII International Conference on High Energy Physics, London, 1974*, J. R. Smith, ed. (Rutherford Laboratory, Chilton, Didcot, England, 1974); S. Weinberg, *Phys.*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

Rev. **D13**, 974 (1976); E. Gildener and S. Weinberg, *Phys. Rev.* **D13**, 3333 (1976).

1*b*. T. Banks and L. Dixon, *Nucl. Phys.* **B307**, 93 (1988). For a detailed treatment, see J. Polchinski, *String Theory* (Cambridge University Press, Cambridge, 1998): Chapter 18.

2. Additive R-conservation laws were introduced by A. Salam and J. Strathdee, *Nucl. Phys.* **B87**, 85 (1975); P. Fayet, *Nucl. Phys.* **B90**, 104 (1975), reprinted in *Supersymmetry*, S. Ferrara, ed., (North Holland/World Scientific, Amsterdam/Singapore, 1987). The R parity can be defined in terms of an R quantum number as $\exp(i\pi R)$, and may be conserved multiplicatively even if R is not conserved additively; see G. Farrar and P. Fayet, *Phys. Lett.* **76B**, 575 (1978); P. Fayet, in *Unification of the Fundamental Particle Interactions*, S. Ferrara, J. Ellis, and P. van Nieuwenhuizen, eds. (Plenum, New York, 1980); S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Lett.* **112B**, 133 (1982); G. Farrar and S. Weinberg, *Phys. Rev.* **D27**, 1731 (1983), reprinted in *Supersymmetry*.

3. S. Dimopoulos and H. Georgi, *Nucl. Phys.* **B193**, 150 (1981), reprinted in *Supersymmetry*, Ref. 2.

4. R. D. Peccei and H. Quinn, *Phys. Rev. Lett.* **38**, 1440 (1977); *Phys. Rev.* **D16**, 1791 (1977).

4*a*. S. Dimopoulos and G. F. Giudice, *Phys. Lett.* **B357**, 573 (1995); A. Pomerol and D. Tommasini, *Nucl. Phys.* **B466**, 3 (1996); G. Dvali and A. Pomerol, *Phys. Rev. Lett.* **77**, 3728 (1996); *Nucl. Phys.* **B522**, 3 (1998); A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Phys. Lett.* **B388**, 588 (1996); R. N. Mohapatra and A. Riotto, *Phys. Rev.* **D55**, 1 (1997); R.-J. Zhang, *Phys. Lett.* **B402**, 101 (1997); H-P. Nilles and N. Polonsky, *Phys. Lett.* **B412**, 69 (1997); D. E. Kaplan, F. Lepeintre, A. Masiero, A. E. Nelson, and A. Riotto, hep-ph/9806430, to be published; J. Hisano, K. Kurosawa, and Y. Nomura, *Phys. Lett.* **B445**, 316 (1999). This pattern of masses can arise naturally from radiative corrections; see J. L. Feng, C. Kolda, and N. Polonsky, *Nucl. Phys.* **B546**, 3 (1999); J. Bagger, J. L. Feng, and N. Polonsky, hep-ph/9905292, to be published.

4*b*. B. W. Lee and S. Weinberg, *Phys. Rev. Lett.* **39**, 165 (1977); D. A. Dicus, E. W. Kolb, and V. L. Teplitz, *Phys. Rev. Lett.* **39**, 168 (1977).

4*c*. S. Wolfram, *Phys. Lett.* **82B**, 65 (1979); J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. Olive, and M. Srednicki, *Nucl. Phys.* **B238**, 453 (1984).

4*d*. P. F. Smith and J. R. J. Bennett, *Nucl. Phys.* **B149**, 525 (1979).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

5. H. Georgi, H. R. Quinn, and S. Weinberg, *Phys. Rev. Lett.* **33**, 451 (1974).

6. S. Dimopoulos and H. Georgi, Ref. 3; J. Ellis, S. Kelley, and D. V. Nanopoulos, *Phys. Lett.* **B260**, 131 (1991); U. Amaldi, W. de Boer, and H. Furstmann, *Phys. Lett.* **B260**, 447 (1991); C. Giunti, C. W. Kim and U. W. Lee, *Mod. Phys. Lett.* **16**, 1745 (1991); P. Langacker and M.-X. Luo, *Phys. Rev.* **D44**, 817 (1991). For other references and more recent analyses of the data, see P. Langacker and N. Polonsky, *Phys. Rev.* **D47**, 4028 (1993); **D49**, 1454 (1994); L. J. Hall and U. Sarid, *Phys. Rev. Lett.* **70**, 2673 (1993).

7. S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Rev.* **D24**, 1681 (1981), reprinted in *Supersymmetry*, Ref. 2.

7*a*. P. Hořava and E. Witten, *Nucl. Phys.* **B460**, 506 (1996); *ibid.* **B475**, 94 (1996); E. Witten, *Nucl. Phys.* **B471**, 135 (1996); P. Hořava, *Phys. Rev.* **D54**, 7561 (1996).

8. H. Pagels and J. R. Primack, *Phys. Rev. Lett.* **48**, 223 (1982).

9. S. Weinberg, *Phys. Rev. Lett.* **48**, 1303 (1983).

10. S. Dimopoulos and H. Georgi, Ref. 3; N. Sakai, *Z. Phys. C* **11**, 153 (1981). For reviews, see H. E. Haber and G. L. Kane, *Phys. Reports* **117**, 75 (1985); J. A. Bagger, in *QCD and Beyond: Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, University of Colorado, June 1995*, D. E. Soper, ed. (World Scientific, Singapore, 1996); V. Barger, in *Fundamental Particles and Interactions: Proceedings of the FCP Workshop on Fundamental Particles and Interactions, Vanderbilt University, May 1997*, R. S. Panvini, T. J. Weiler, eds. (American Institute of Physics, Woodbury, NY, 1998); J. F. Gunion, in *Quantum Effects in the MSSM – Proceedings of the International Workshop on Quantum Effects in the MSSM, Barcelona, September 1997*, J. Solà, ed. (World Scientific Publishing, Singapore, 1998); S. Dawson, in *Proceedings of the 1997 Theoretical Advanced Study Institute on Supersymmetry, Supergravity, and Supercolliders*, J. Bagger, ed. (World Scientific, Singapore, 1998); S. P. Martin, in *Perspectives on Supersymmetry*, G. L. Kane, ed. (World Scientific, Singapore, 1998); K. R. Dienes and C. Kolda, in *Perspectives on Supersymmetry*, ibid.

11. S. Dimopoulos and D. Sutter, *Nucl. Phys.* **B194**, 65 (1995); H. Haber, *Nucl. Phys. Proc. Suppl.* **62**, 469 (1998).

12. S. Dimopoulos and H. Georgi, Ref. 3; J. Ellis and D. V. Nanopoulos, *Phys. Lett.* **110B**, 44 (1982); J. F. Donoghue, H-P. Nilles, and D. Wyler, *Phys. Lett.* **128B**, 55 (1983). For strong-interaction corrections to these calculations, see J. A. Bagger, K. T. Matchev, and

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

R.-J. Zhang, *Phys. Lett.* **B412**, 77 (1997). Conditions under which limits on flavor changing processes do not constrain squark masses are discussed by R. Barbieri and R. Gatto, *Phys. Lett.* **110B**, 211 (1981); Y. Nir and N. Seiberg, *Phys. Lett.* **B309**. 337 (1993). For a detailed review, see F. Gabbiani, E. Gabrielli, A. Masiero, and L. Silvestrini, *Nucl. Phys.* **B477**, 321 (1996).

13. M. K. Gaillard and B. W. Lee, *Phys. Rev.* **D10**, 897 (1974).

14. J. Ellis and D. V. Nanopoulos, Ref. 12. Detailed results are given by F. Gabbiani and A. Masiero, *Nucl. Phys.* **B322**, 235 (1989); J. S. Hagelin, S. Kelley, and T. Tanaka, *Nucl. Phys.* **B415**, 293 (1994). The most complete treatment is by D. Sutter, Stanford University Ph. D. thesis (unpublished) and S. Dimopoulos and D. Sutter, Ref. 11.

14*a*. M. Dine, R. Leigh, and A. Kagan, *Phys. Rev.* **D48**, 4269 (1993).

15. For more recent reviews, see Y. Grossman, Y. Nir, and R. Rattazzi, in *Heavy Flavours II*, A. J. Buras and M. Lindner, eds. (World Scientific, Singapore, 1998); A. Masiero and L. Silvestrini, in *Perspectives on Supersymmetry*, Ref. 10.

16. J. Ellis and M. K. Gaillard, *Nucl. Phys.* **B150**, 141 (1979); D. V. Nanopoulos, A. Yildiz, and P. H. Cox, *Ann. Phys. (N.Y.)* **127**, 126 (1980); M. B. Gavela, A. Le Yaouanc, L. Oliver, O. Pène, J.-C. Raynal, and T. N. Pham, *Phys. Lett.* **109B**, 215 (1982); B. H. J. McKellar, S. R. Choudhury, X-G. He, and S. Pakvasa, *Phys. Lett.* **B197**, 556 (1987).

16*a*. P. G. Harris *et al.*, *Phys. Rev. Lett.* **82**, 904 (1999).

17. J. Ellis, S. Ferrara, and D. V. Nanopoulos, *Phys. Lett.* **114B**, 231 (1982); J. Polchinski and M. B. Wise, *Phys. Lett.* **125B**, 393 (1983); M. Dugan, B. Grinstein, and L. Hall, *Nucl. Phys.* **B255**, 413 (1985).

18. R. Arnowitt, J. Lopez, and D. Nanopoulos, *Phys. Rev.* **D42**, 2423 (1990); R. Arnowitt, M. Duff, and K. Stelle, *Phys. Rev.* **D43**, 3085 (1991); Y. Kizuri and N. Oshimo, *Phys. Rev.* **D45**, 1806 (1992).

19. S. Weinberg, *Phys. Rev. Lett.* **63**, 2333 (1989); D. Dicus, *Phys. Rev.* **D41**, 999 (1990); J. Dai, H. Dykstra, R. G. Leigh, S. Paban, and D. A. Dicus, *Phys. Lett.* **B237**, 216 (1990); E. Braaten, C. S. Li, and T. C. Yuan, *Phys. Rev. Lett.* **64**, 1709 (1990); A. De Rújula, M. B. Gavela, O. Pène, and F. J. Vegas, *Phys. Lett.* **B245**, 640 (1990); R. Arnowitt, M. J. Duff, and K. S. Stelle, Ref. 18; T. Ibrahim and P. Nath, *Phys. Lett.* **148B**, 98 (1998).

20. K. S. Babu, C. Kolda, J. March-Russell, and F. Wilczek, *Phys. Rev.* **D59**, 016004 (1999).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

21. R. Arnowitt, M. J. Duff, and K. S. Stelle, Ref. 18. In terms of the functions J_1 and J_2 in this reference, the function J here is taken as $2J_1 + \frac{2}{3}J_2$, under the assumption that the graphs with the gluon attached to the squark or gluino lines add constructively.

22. H. Georgi and L. Randall, *Nucl. Phys.* **B276**, 241 (1980); A. Manohar and H. Georgi, *Nucl. Phys.* **B238**, 189 (1984); S. Weinberg, Ref. 19.

23. W. Fischler, S. Paban, and S. Thomas, *Phys. Lett.* **B 289**, 373 (1992).

24. J. Ellis and D. V. Nanopoulos, Ref. 12; F. Gabbiani and A. Masiero, Ref. 14; F. Dine, A. Kagan, and S. Samuel, *Phys. Lett.* **B243** 250 (1990); F. Gabbiani, E. Gabrielli, A. Masiero, and L. Silvestrini, Ref. 12.

25. S. Weinberg, *Phys. Rev. Lett.* **40**, 223 (1978); F. Wilczek, *Phys. Rev. Lett.* **40**, 279 (1978).

26. A. Brignole, J. Ellis, G. Ridolfi, and F. Zwirner, *Phys. Lett.* **B271**, 123 (1991); M. Carena, M. Quiros, and C. E. M. Wagner, *Nucl. Phys.* **B461**, 407 (1996); S. Heinemayer, W. Hollik, and G. Weiglein, hep-ph/9812472, hep-ph/9903404, hep-ph/9903504, to be published. The numerical results for m_h cited here are taken from calculations quoted by S. Dawson, Ref. 10.

27. R. Barate *et al.* (ALEPH collaboration), *Phys. Lett.* **B412**, 173 (1997).

27*a*. M. Grünewald and D. Karlen, in *Proceedings of the XXIX International Conference on High Energy Nuclear Physics*, A. Astbury, D. Axen, and J. Robinson, eds. (TRIUMF, Vancouver, 1999).

28. R. Barate *et al.* (ALEPH collaboration), 1999 CERN preprint EP-99-011, to be published in *Phys. Lett.* A lower bound of 54.5 GeV had been obtained earlier from experiments on e^+–e^- annihilation at 130 to 172 GeV by P. Abreu *et al.* (DELPHI collaboration), *Phys. Lett.* **B420**, 140 (1998).

29. A. J. Buras, M. Misiak, M. Münz, and S. Pokorski, *Nucl. Phys.* **B424**, 374 (1994).

29*a*. R. Barate *et al.* (ALEPH collaboration), 1999 CERN preprint EP-99-014, to be published in *Eur. Phys. J.*

30. M. Dine, W. Fischler, and M. Srednicki, *Nucl. Phys.* **B189**, 575 (1981); S. Dimopoulos and S. Raby, *Nucl. Phys.* **B192**, 353 (1982); M. Dine and W. Fischler, *Phys. Lett.* **110B**, 227 (1982); *Nucl. Phys.* **B204**, 346 (1982); C. Nappi and B. Ovrut, *Phys. Lett.* **113B**, 175 (1982); L. Alvarez-Gaumé, M. Claudson, and M. Wise, *Nucl. Phys.*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

B207, 96 (1982); S. Dimopoulos and S. Raby, *Nucl. Phys.* **B219**, 479 (1983). This class of models was revived by M. Dine and A. E. Nelson, *Phys. Rev.* **D48**, 1277 (1993); **D51**, 1362 (1995); J. Bagger, Ref. 10; M. Dine, A. Nelson, and Y. Shirman, *Phys. Rev.* **D51**, 1362 (1995); M. Dine, A. Nelson, Y. Nir, and Y. Shirman, *Phys. Rev.* **D53**, 2658 (1996). For reviews, see C. Kolda, *Nucl. Phys. Proc. Suppl.* **62**, 266 (1998); G. F. Giudice and R. Rattazzi, hep-ph/9801271, to be published in *Phys. Rep*; S. L. Dubovsky, D. S. Gorbunov, and S. V. Troitsky, hep-ph/9905466, to be published. The phenomenology of these models is described by S. Dimopoulos, S. Thomas, and J. D. Wells, *Nucl. Phys.* **B488**, 39 (1997).

31. S. Dimopoulos, G. F. Giudice, and A. Pomerol, *Phys. Lett.* **389B**, 37 (1997); S. P. Martin, *Phys. Rev.* **D55**, 3177 (1997).

32. G. F. Giudice and R. Rattazi, *Nucl. Phys.* **B511**, 25 (1998). This work has been extended by N. Arkani-Hamed, G. F. Giudice, M. A. Luty, and R. Rattazzi, *Phys. Rev.* **D58**, 115005 (1998).

33. N. Seiberg, *Phys. Lett.* **B318**, 469 (1993).

34. K. S. Babu, C. Kolda, and F. Wilczek, *Phys. Rev. Lett.* **77**, 3070 (1996).

35. J. E. Kim and H-P. Nilles, *Phys. Lett.* **138B**, 150 (1984); J. Ellis, J. F. Gunion, H. E. Haber, L. Roszkowski, and F. Zwirner, *Phys. Rev.* **D39**, 844 (1989); E. J. Chun, J. E. Kim, and H-P. Nilles, *Nucl. Phys.* **B370**, 105 (1992); M. Dine and A. E. Nelson, Ref. 30; M. Dine, A. E. Nelson, Y. Nir, and Y. Shirman, Ref. 30; G. Dvali, G. F. Giudice, and A. Pomerol, *Nucl. Phys.* **B478**, 31 (1996); S. Dimopoulos, G. Dvali, and R. Rattazzi, *Phys. Lett.* **413B**, 336 (1997); H-P. Nilles and N. Polonsky, *Nucl. Phys.* **B484**, 33 (1997); G. Cleaver, M. Cvetič, J. R. Espinosa, L. Everett, and P. Langacker, *Phys. Rev.* **D57**, 2701 (1998); P. Langacker, N. Polonsky, and J. Wang, hep-ph/9905252, to be published; J. E. Kim, hep-ph/9901204, to be published.

35*a*. S. Raby, *Phys. Lett.* **B422**, 158 (1998).

35*b*. D. A. Dicus, B. Dutta, and S. Nandi, *Phys. Rev. Lett.* **78**, 3055 (1997); *Phys. Rev.* **D56**, 5748 (1997).

36. S. Weinberg, *Phys. Rev. Lett.* **43**, 1566 (1979); F. Wilczek and A. Zee, *Phys. Rev. Lett.* **43**, 1571 (1979).

37. S. Weinberg, *Phys. Rev.* **D26**, 287 (1982); N. Sakai and T. Yanagida, *Nucl. Phys.* **B197**, 533 (1982). These articles are reprinted in *Supersymmetry*, Ref. 2.

38. S. Weinberg, Ref. 36.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

39. J. Ellis, J. S. Hagelin, D. V. Nanopoulos, and K. Tamvakis, *Phys. Lett.* **124B**, 484 (1983); V. M. Belyaev and M. I. Vysotsky, *Phys. Lett.* **127B**, 215 (1983).

40. V. Lucas and S. Raby, *Phys. Rev.* **D55**, 6986 (1997).

41. This estimate is taken from a compilation of non-supersymmetric calculations of the total two-body decay proton decay rate in terms of the superheavy gauge boson masses M_X by P. Langacker, in *Proceedings of the 1983 Annual Meeting of the Division of Particles and Fields of the American Physical Society* (American Institute of Physics, New York, 1983): 251. To express the result in terms of g_6, I have assumed that, in effect, the coupling g_6 that was used in these calculations was given by $g_6 = g^2(M_X)/M_X^2$, where $g(M_X)$ has the value appropriate to non-supersymmetric theories, with $g^2(M_X)/4\pi \simeq 1/41$.

42. For more detailed (mostly model-dependent) calculations, including renormalization group corrections to g_5, see S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Lett.* **112B**, 133 (1982); J. Ellis, D. V. Nanopoulos, and S. Rudaz, *Nucl. Phys.* **B202**, 43 (1982); W. Lang, *Nucl. Phys.* **B203**, 277 (1982); J. Ellis, J. S. Hagelin, D. V. Nanopoulos, and K. Tamvakis, Ref. 39; V. M. Belyaev and M. I. Vysotsky, Ref. 39; L. E. Ibáñez and C. Muñoz, *Nucl. Phys.* **B245**, 425 (1984); P. Nath, A. H. Chamseddine, and R. Arnowitt, *Phys. Rev.* **D32**, 2385 (1985); J. Hisano, H. Murayama, and T. Yanagida, *Nucl. Phys.* **B402**, 46 (1993); V. Lucas and S. Raby, Ref. 40. For a review, see P. Nath and R. Arnowitt, *Phys. Atom. Nucl.* **61**, 975 (1997).

42*a*. M. Takita *et al.*, in *Proceedings of the XXIX International Conference on High Energy Nuclear Physics*, Ref. 27a.

43. K. S. Babu, J. C. Pati, and F. Wilczek, *Phys. Lett.* **423B**, 337 (1998).

44. V. Lucas and S. Raby, Ref. 40; T. Goto and T. Nihei, *Phys. Rev.* **D59**, 115009 (1999).

45. L. Ibáñez and G. Ross, *Nucl. Phys.* **B368**, 3 (1991).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.007

29

Beyond Perturbation Theory

Most of the implications of supersymmetry discussed so far have been inferred with the use of perturbation theory. In this chapter we shall consider some results that apply even when non-perturbative effects are taken into account.

29.1 General Aspects of Supersymmetry Breaking

Supersymmetry is not observed in the spectrum of known particles, so it must be broken. We saw in the previous chapter that supersymmetry breaking in the tree approximation of the standard model is experimentally ruled out, and that the large disparity between the electroweak breaking scale and the Planck or grand unification scales suggests that supersymmetry is likely to be broken when some running gauge coupling becomes strong. It is therefore essential for us to explore spontaneous supersymmetry breaking without using perturbation theory.

We saw in Section 26.7 that the supersymmetry of the action implies the existence of a supersymmetry current $S^\mu(x)$. This current is a Majorana spinor in the sense of Eq. (26.A.2):

$$S^\mu(x)^* = -\beta\gamma_5\epsilon S^\mu(x) \; ; \tag{29.1.1}$$

it is conserved,

$$\partial_\mu S^\mu_\beta(x) = 0 \; ; \tag{29.1.2}$$

and the integral of its time component is the supersymmetry generator

$$\int d^3x \; S^0_\beta(x) = Q_\beta \; , \tag{29.1.3}$$

for which the commutator of $-i(\bar{\alpha}Q)$ with any operator gives the change of that operator under a supersymmetry transformation with infinitesimal Majorana spinor parameter α.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

The arguments that led to these results relied on the supersymmetry of the action; nothing depended on whether or not supersymmetry is spontaneously broken, except perhaps for the assumption that the integral (29.1.3) exists. In fact this assumption may be violated in theories with massless fermions, which can have long range effects (or, equivalently, poles at zero four-momentum) that would make this integral not converge. We will see here that such massless fermions are necessary consequences of supersymmetry breaking. In order to avoid the issue of the convergence of this integral even in theories with massless fermions, it is very convenient to work in a space of finite volume V. We can do this while maintaining translation invariance by imposing periodic boundary conditions: all fields are assumed to be unaffected by translations of any spatial coordinate x^i by an amount $V^{1/3}$.

The existence of an operator Q_α that induces supersymmetric transformations on quantum fields would allow us to derive all the consequences of supersymmetry, provided that there exists a supersymmetric vacuum state $|\text{VAC}\rangle$ of zero three-momentum from which multiparticle states may be constructed by acting with field operators. But if $|\text{VAC}\rangle$ is supersymmetric in the sense that $Q_\alpha|\text{VAC}\rangle = 0$, then it follows from the anticommutation relation (25.2.36) that this state has zero energy as well as zero momentum. Conversely, by taking the vacuum expectation value of the positive operator (returning for a moment to two-component notation) $\{\text{Q}_a, \text{Q}_a^\dagger\}$, we see that if the vacuum has zero energy then it must be annihilated by Q_a and $\text{Q}_a^\dagger$, and hence be supersymmetric, while if it is not supersymmetric then its energy must be positive-definite. *The question of whether supersymmetry is or is not spontaneously broken is thus entirely a question of whether the vacuum has a positive-definite or a zero energy.*

The same reasoning led Witten to the conclusion that extended supersymmetry with $N > 1$ two-component spinor generators Q_{ar} and their adjoints cannot be spontaneously broken to extended supersymmetry with fewer generators or to simple supersymmetry, because if any one of the generators does not annihilate the vacuum then the vacuum energy cannot vanish, and it follows from this that none of the generators can annihilate the vacuum.[1] Usually the energy of the vacuum state appears as an ill-defined additive constant in the energies of all states, but here this constant is given a meaning by the appearance of the energy-momentum four-vector in the anticommutation relations of supersymmetry. One of the advantages of working in a finite volume is that it makes it meaningful to talk about the total energy of the vacuum.

Hughes, Liu, and Polchinski have pointed out that there are theories that exhibit a sort of partly broken supersymmetry.[1a] These theories do not have a supersymmetry algebra of the sort described in Chapter 25. Instead, they have an algebra of *currents*, based on anticommutation

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

relations like Eq. (26.7.45):

$$\int d^3x \left\{ S^0_{r\alpha}(x), \bar{S}^\mu_{s\beta}(y) \right\} = -2i\,\delta_{rs}\gamma_\nu \Theta^{\mu\nu}(y) + 2i\gamma^\mu_{\alpha\beta} C_{rs} ,$$

where $\Theta^{\mu\nu}$ is an energy-momentum tensor satisfying the conservation condition $\partial_\mu \Theta^{\mu\nu} = 0$, and C_{rs} is a new ingredient, a constant. For $N = 1$ this constant could be regarded as a term $-\eta^{\mu\nu} C$ in $\Theta^{\mu\nu}$, but this is not possible for extended supersymmetry unless $C_{rs} \propto \delta_{rs}$, which need not be the case. This algebra is not ruled out by the Haag–Lopuszanski–Sohnius theorem proved in Section 25.2 because it could not be a symmetry of the S-matrix. It is true that spontaneously broken symmetries are never symmetries of the S-matrix, but they are usually assumed to be based on algebras or superalgebras that *could* be symmetries of the S-matrix in some phases of some theories. With C_{rs} not proportional to δ_{rs}, the superalgebra of currents is one that could not generate a symmetry of the S-matrix in any phase of any theory. Here we will only consider superalgebras of the sort described in Chapter 25, to which Witten's argument does apply.

Another advantage of working in a finite volume is that all states become discrete and normalizable. An immediate consequence of the fact that Q_α commutes with P_μ is that any state of non-zero energy is paired with another state of the same energy and momentum but opposite statistics. To see this, note that for any three-momentum $\mathbf{p}$ we may find a two-component spinor u_a such that $\sum_{ab} u^*_a \boldsymbol{\sigma}_{ab} \cdot \mathbf{p} u_b = 0$ and $\sum_a |u_a|^2 = 1$. (For $\mathbf{p}$ in the 3-direction, take $u = (1,1)/\sqrt{2}$. For $\mathbf{p}$ in any other direction, apply to this u the spin 1/2 representation of the rotation that takes the 3-direction into the direction of $\mathbf{p}$.) Then, within the space with four-momentum p^μ, the anticommutation relations (25.2.31) and (25.2.32) give

$$\mathrm{Q}^2(p) = p^0 , \tag{29.1.4}$$

where $\mathrm{Q}(p)$ is the Hermitian linear combination of supersymmetry generators:

$$\mathrm{Q}(p) = \sum_a u_a \mathrm{Q}_a + \sum_a u^*_a \mathrm{Q}^*_a . \tag{29.1.5}$$

Acting with $\mathrm{Q}(p)$ on any normalized state $|X\rangle$ with four-momentum p^μ and $p^0 > 0$, we get another normalized state $|Y\rangle = \mathrm{Q}(p)|X\rangle/\sqrt{p^0}$ with opposite statistics and the same four-momentum. Furthermore, $|X\rangle$ is the only state related to $|Y\rangle$ in this way, because if $|Y\rangle = \mathrm{Q}(p)|X\rangle/\sqrt{p^0}$ then, according to Eq. (29.1.4), $|X\rangle = \mathrm{Q}(p)|Y\rangle/\sqrt{p^0}$. The multiplicity of supersymmetry generators and of spin states will usually cause these pairs of fermionic and bosonic states to be joined by other pairs, all with the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

same four-momentum, but for the moment it is enough to know that all states of non-zero energy may at least be grouped in these pairs.

When supersymmetry is broken we do not expect states with a definite number of particles to form supermultiplets with other states of opposite statistics and with the same four-momentum and the same number of particles. The pairing of states then requires the existence of a massless fermion, so that an n-particle state can be paired with a state of the same energy and momentum but opposite statistics, consisting of the same n particles together with a massless fermion of zero energy and momentum. This massless fermion is known as a *goldstino*. To be more precise, any n-particle state is accompanied with two states of the same energy and momentum and opposite statistics, containing an additional zero-momentum goldstino of spin up or down, and with another state of the same energy and momentum and the *same* statistics, containing two additional zero-momentum goldstinos of opposite spin. In particular, when supersymmetry is spontaneously broken the vacuum state has non-zero energy, so it must be paired with a fermionic state of the same energy and zero momentum; more precisely, the vacuum and the state containing two zero-momentum goldstinos are paired with the two states of a single zero-momentum goldstino. It is only when supersymmetry is unbroken that there is a state of zero energy, the vacuum, that can be unpaired.

The pairing of states of non-zero energy provides a valuable diagnostic tool that can in some cases tell us that supersymmetry is not spontaneously broken, even where perturbation theory is not adequate to answer this question. When all interactions are weak, we can rely on perturbation theory to give us a *qualitative* picture of the spectrum. If it turns out that in the tree approximation there are n vacuum states with zero energy, and no massless fermions, then we can be confident that for weak coupling there are no zero-energy fermionic states with which we could pair the n vacuum states, so these unpaired states would have to have precisely zero energy. Then, as we increase the strength of the couplings or vary the parameters of the theory in any other way, states may move from positive to zero energy or vice-versa, but they will generally not suddenly appear or disappear. (There is an exception for changes in the parameters that change the asymptotic behavior of the Lagrangian for large fields; as we will shortly see, this *can* produce or destroy states.) Because each state of non-zero energy is always paired with another of opposite statistics, they can only make the transition from zero to non-zero energy or vice-versa in such pairs, so the number of bosonic zero-energy states minus the number of fermionic zero-energy states does not change as the parameters of the theory are varied, as long as the large-field behavior of the Lagrangian is not changed. This difference is known as the *Witten index*.[2] Formally, this index is $\mathrm{Tr}\,(-1)^F$, where F is the fermion number; the pairing of states

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

discussed earlier insures that this trace can receive no contribution from states of non-zero energy. If the Witten index is non-zero then there must be *some* states of zero energy, and so supersymmetry cannot be broken. In particular, in a theory where the tree approximation gives n zero-energy vacuum states and no zero-energy fermions, the Witten index is n for weak coupling, where the tree approximation can be trusted to give a qualitative picture of the spectrum, and the index remains equal to n as the strength of the coupling is increased, so we can be certain that higher-order effects or even non-perturbative effects do not break supersymmetry.

As an example of the use of the Witten index, consider the Wess–Zumino theory of a single chiral superfield with a cubic polynomial superpotential of the form (26.4.16):

$$f(\phi) = \tfrac{1}{2}m^2\phi^2 + \tfrac{1}{6}g\phi^3 ,$$

where ϕ is the complex scalar field component of the superfield. We saw in Section 26.4 that this model does not exhibit supersymmetry breaking in the tree approximation, but what about higher orders of perturbation theory, and what about non-perturbative effects? Perturbation theory gives a good approximation to the energy spectrum when m is large and g is small; it tells us that in this case there are two bosonic states near zero energy, corresponding to the solutions $\phi = 0$ and $\phi = -2m^2/g$ of the equation $\partial f(\phi)/\partial\phi = 0$, and no fermionic states near zero energy; the lowest-energy fermionic state is a zero-momentum one-fermion state of energy near $|m|$. In typical scalar field theories we would not expect the two bosonic states to have precisely zero energy; even though each has zero energy in the tree approximation, higher-order effects (including tunneling through the barrier between $\phi = 0$ and $\phi = -2m^2/g$) would be expected to mix them and shift their energies away from zero. (It is only in the limit of infinite volume that this barrier becomes impassable.) But in supersymmetric theories these states must have precisely zero energy, because there is no low-energy fermion state with which they could be paired. Thus for m large and g small the Witten index is 2. Because the Witten index is invariant under changes in the parameters of the theory, the Witten index remains equal to 2 even when g is large, where perturbation theory breaks down, and even when m vanishes, where the two potential wells merge. (It is not easy to calculate the Witten index directly in this case, because of the presence in the tree approximation of massless bosons as well as massless fermions.) Since the Witten index is not zero, supersymmetry remains strictly unbroken in the Wess–Zumino model, whatever the values of its parameters.

The same arguments can generally be used in theories with several chiral scalar superfields to show that the Witten index is positive, and that supersymmetry is therefore not spontaneously broken. The O'Raifeartaigh

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

models discussed in Section 26.5 are an exception, because there are flat directions in which the potential remains constant as the fields go to infinity, rather than growing as some power of the fields. These models provide a good illustration of the fact that although the Witten index must be zero for supersymmetry to be broken, a zero Witten index does not necessarily imply that supersymmetry *is* broken. For instance, if we write the superpotential used as an illustration in Section 26.5 in terms of canonically normalized superfields, then it takes the form

$$f(X, Y_1, Y_2) = mY_1(X - a) + gY_2X^2 ,$$

with arbitrary parameters m, g, and a. The potential is then

$$U(x, y_1, y_2) = |g|^2|x|^4 + |m|^2|x - a|^2 + |my_1 + 2gxy_2|^2 ,$$

with lower-case letters denoting the scalar components of the left-chiral superfields. For m and a non-zero and g small, perturbation theory gives a good estimate of the spectrum, and tells us that there is a minimum of the potential near $x = a - 2|g|^2|a|^4/|m|^2$ with $my_1 + 2xy_2 = 0$, at which the vacuum energy is approximately equal to $|ga^2|^2V$. Since this energy appears as an additive constant in all states, there are no zero-energy states, and the Witten index vanishes. The matrix $\mathscr{M}$ of second derivatives of the superpotential (with rows and columns labelled in the order x, y_1, y_2) is here

$$\mathscr{M} = \begin{pmatrix} 2gy_2 & m & 2gx \\ m & 0 & 0 \\ 2gx & 0 & 0 \end{pmatrix} .$$

This has an eigenvector $(0, 2gx, -m)$ with eigenvalue zero, so there is a massless fermion here; this is the goldstino associated with the breaking of supersymmetry. The fermionic state degenerate with the vacuum consists of a single goldstino of zero energy and momentum. (Once again, there are two vacuum-energy fermionic states, with opposite orientations for the goldstino spin, and two vacuum-energy bosonic states: the vacuum, and a state containing two goldstinos with opposite spin.) Now, as $a \to 0$ supersymmetry may become unbroken (and we shall see later that it does), but the Witten index must remain 0; in this case the massless fermion is no longer a goldstino, but continuity demands that its mass remains zero, so it remains paired with the vacuum state. This of course is a general feature of theories in which supersymmetry is restored at isolated values of the parameters; continuity demands that the massless fermion that plays the role of the goldstino when supersymmetry is broken remains massless (though no longer a goldstino) at the value of the parameters where supersymmetry is restored, so that the vacuum remains paired with the massless fermion state, and the Witten index remains zero.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

This model provides a good illustration of why we had to qualify the statement that the Witten index does not change when we vary the parameters of a supersymmetric theory, with the proviso that the parameters must not be changed in such a way as to alter the asymptotic behavior of the Lagrangian density for large fields. Suppose we unflatten the flat direction in this model by adding a small term to the superpotential, so that it now reads

$$f(X, Y_1, Y_2) = mY_1(X - a) + gY_2X^2 + \tfrac{1}{2}\epsilon(Y_1^2 + Y_2^2) ,$$

where ϵ is a small mass parameter. Now there are two solutions of the conditions for supersymmetry to be preserved:

$$0 = \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y_1} = \frac{\partial f}{\partial y_2} .$$

At these solutions, x is at one of the roots of the quadratic equation $2g^2x^2 + m(x - a) = 0$, while y_1 and y_2 are of order $1/\epsilon$: $y_1 = -m(x - a)/\epsilon$ and $y_2 = -gx^2/\epsilon$. We see that the reason that the Witten index can change from 0 to 2 when we turn on the small parameter ϵ is that two new minima of the potential come in from infinite field values.

In deciding whether or not supersymmetry is broken in theories with a vanishing Witten index, it is often useful to use conservation laws to limit the pairings that may occur, and to define a new sort of index. If K is a quantum operator that commutes with the supersymmetry generators Q_α (and hence also with the Hamiltonian), then all states of non-zero energy that have a definite value for K are paired with states of opposite statistics and the same energy and momentum *and* the same value of K. Also, not only is the Witten index $\mathrm{Tr}\,(-1)^F$ independent of the parameters of the theory (as long as they are not varied in a way that varies the large-field asymptotic behavior of the Lagrangian) — so also is the *weighted Witten index*, given by $\mathrm{Tr}\,g(K)(-1)^F$, where $g(K)$ is an arbitrary function of the conserved quantity. To use the conservation law in this way, it is not necessary that it be unbroken when the volume V becomes infinite; it is only necessary that K commute with the supersymmetry generators.

In diagnosing the possibilities of supersymmetry breaking, it is sometimes helpful to work with a linear combination of weighted Witten indices for a number of different conserved quantities. In particular, consider the quantity

$$W_G = \sum_{h \in G} \mathrm{Tr}\left\{h(-1)^F\right\} , \tag{29.1.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

with the sum running over all elements of some symmetry group G. (For compact continuous groups, this sum should be interpreted as an integral over the group volume, with a suitable invariant measure.) Within any irreducible representation other than the identity, the 'characters' $\mathrm{Tr}\, h$ add up to zero when summed over a finite or compact group, so

$$W_G = \sum_f N(f)(-1)^f \,, \tag{29.1.7}$$

where $N(f)$ is the number of times that the identity representation of G appears among states with fermion number f. In other words, W_G is just the Witten index, but evaluated using only G-invariant states. As long as G is conserved, W_G remains independent of the parameters of the theory, and if non-zero indicates that supersymmetry is unbroken.

Conservation laws will be used in this way in Section 29.4 to study the spontaneous breaking of supersymmetry in gauge theories, but a simpler (though academic) example is provided by the O'Raifeartaigh-style model discussed earlier, only now with the parameter a set equal to zero. The superpotential here is

$$f(X, Y_1, Y_2) = mY_1X + gY_2X^2 \,, \tag{29.1.8}$$

yielding a tree-approximation potential

$$U(x, y_1, y_2) = |m|^2|x|^2 + |g|^2|x|^4 + |my_1 + 2gxy_2|^2 \,. \tag{29.1.9}$$

We know that the Witten index vanishes here, because we have seen that it vanished for $a \neq 0$, but is supersymmetry broken? With $a = 0$, there are now field values with $x = y_1 = 0$ where the potential vanishes in the tree approximation, but how can we tell whether effects of higher order in g or even non-perturbative effects give the corresponding states a small energy? To answer this, we note that this superpotential (and hence the Lagrangian density) is invariant under a discrete symmetry K, under which the superfields are transformed by

$$KXK^{-1} = iX \,, \qquad KY_1K^{-1} = -iY_1 \,, \qquad KY_2K^{-1} = -Y_2 \,. \tag{29.1.10}$$

(Note that this symmetry is violated by the term $-maY_1$ in the original superpotential, so none of the results obtained using K will apply to that superpotential.) Since the potential vanishes for $x = y_1 = 0$ and y_2 arbitrary, for small g we may use perturbation theory to tell us that for each y_2 there is a bosonic vacuum state *near* zero energy. For $y_2 = 0$ this vacuum is even under K. For any non-zero value of $|y_2|$ we can take linear combinations of the two zero-energy states with $y_2 = \pm|y_2|$, of which one

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

will be even under K and the other odd. As we have seen, there is also a massless fermion here, the fermionic component of Y_2, but this fermion is odd under K, so it cannot be paired with the even vacuum states. The only other fermions in the theory have tree-approximation masses $|m|$, so for small g they can't be paired with the even vacuum states either. We conclude then that for small g the even vacuum states must have precisely zero energy, and supersymmetry is unbroken. It is not easy to calculate the weighted Witten index here, because there are an infinite number of bosonic zero-energy states that are even under K, consisting of either zero or two Y_2-fermions with zero momentum plus any even number of zero-momentum Y_2 bosons, but it is clear that $\mathrm{Tr}\, K(-)^F > 0$, and since this is independent of g (as long as $g \neq 0$), supersymmetry cannot be broken for any finite g.

29.2 Supersymmetry Current Sum Rules

We now turn to sum rules that yield exact quantitative relations between the vacuum energy and parameters describing the strength of supersymmetry breaking.

Let us again start by assuming that the world is placed in a box of volume V, with periodic boundary conditions to preserve translation invariance. The vacuum expectation value of the anticommutation relations (25.2.36) may then be expressed as a sum over discrete states $|X, \mathrm{Box}\rangle$:

$$\begin{aligned}\sum_X \left\langle \mathrm{VAC}\middle|Q_\alpha\middle|X,\mathrm{Box}\right\rangle \left\langle \mathrm{VAC}\middle|Q_\beta\middle|X,\mathrm{Box}\right\rangle^* \\ +\sum_X \left\langle \mathrm{VAC}\middle|Q^*_\beta\middle|X,\mathrm{Box}\right\rangle \left\langle \mathrm{VAC}\middle|Q^*_\alpha\middle|X,\mathrm{Box}\right\rangle^* \\ = -2i\left(\gamma_\mu\beta\right)_{\alpha\beta}\left\langle \mathrm{VAC}\middle|P^\mu\middle|\mathrm{VAC}\right\rangle ,\end{aligned} \tag{29.2.1}$$

with the label 'Box' indicating that the states are normalized to have Kronecker deltas rather than delta functions as their scalar products. Setting $\beta = \alpha$, summing over α and using Eq. (25.2.37) gives

$$\sum_{X,\alpha}\left|\left\langle \mathrm{VAC}\middle|Q_\alpha\middle|X,\mathrm{Box}\right\rangle\right|^2 = 4\left\langle \mathrm{VAC}\middle|P^0\middle|\mathrm{VAC}\right\rangle . \tag{29.2.2}$$

Because Q_α commutes with the four-momentum, it is only states with zero three-momentum and the same energy as the vacuum that can contribute

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

to this sum. To find the volume dependence of the matrix elements that do contribute in Eq. (29.2.2), we note that a box-normalized state $|X, \text{Box}\rangle$ containing N_X particles is related according to Eq. (3.4.3) to the corresponding continuum-normalized state $|X\rangle$ by

$$\left|X, \text{Box}\right\rangle = \left((2\pi)^3/V\right)^{N_X/2}\left|X\right\rangle . \tag{29.2.3}$$

For states with $\mathbf{p}_X = 0$, the spatial integral of the supersymmetry current time-component S^0_α gives another factor of V, so we conclude that for box-normalized states with $\mathbf{p}_X = 0$

$$\left\langle \text{VAC}\middle|Q_\alpha\middle|X, \text{Box}\right\rangle = (2\pi)^{3N_X/2}V^{1-N_X/2}\left\langle \text{VAC}\middle|S^0_\alpha(0)\middle|X\right\rangle . \tag{29.2.4}$$

Since invariance under 2π-rotations does not allow X to be a zero-particle state, the dominant terms in Eq. (29.2.2) when $V \to \infty$ will be those from one-particle states. In this limit, Eq. (29.2.2) becomes

$$(2\pi)^3 \sum_{X,\alpha}{}^{(0)} \left|\left\langle \text{VAC}\middle|S^0_\alpha(0)\middle|X\right\rangle\right|^2 = 4\,\rho_{\text{VAC}} , \tag{29.2.5}$$

where ρ_{VAC} is the vacuum energy density

$$\rho_{\text{VAC}} \equiv \left\langle \text{VAC}\middle|P^0\middle|\text{VAC}\right\rangle/V , \tag{29.2.6}$$

and the superscript (0) indicates that the sum in Eq. (29.2.5) runs only over one-particle states with zero four-momentum. These of course are the two helicity states of the goldstino.

We see again from Eq. (29.2.5) that if the vacuum energy density does not vanish then the vacuum is not invariant under supersymmetry transformations, but is rather transformed into one-goldstino states. Conversely, Eq. (29.2.2) shows that if the vacuum is not invariant under supersymmetry then according to Eq. (29.2.2) its energy in a finite box cannot vanish, although it is conceivable that supersymmetry transformations might take the vacuum only into multiparticle states, in which case the vacuum energy density would vanish in the limit of large volume.

To evaluate the one-goldstino contribution in Eq. (29.2.5), we use Lorentz invariance to write the matrix element of the supersymmetry current between the vacuum and a one-goldstino state $|\mathbf{p}, \sigma\rangle$, with momentum

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

$\mathbf{p}$ and helicity λ, in the form*

$$\left\langle \mathrm{VAC}\middle|S^\mu(0)\middle|\mathbf{p},\lambda\right\rangle = (2\pi)^{-3/2}\left[\left(\frac{1+\gamma_5}{2}\right)\left(\gamma^\mu F + ip^\mu F'\right) + \left(\frac{1-\gamma_5}{2}\right)\left(\gamma^\mu F^* + ip^\mu F'^*\right)\right] u(\mathbf{p},\lambda)\,, \quad (29.2.7)$$

where $u(\mathbf{p},\lambda)$ are the coefficient functions for a massless Dirac field introduced in Section 5.5, and F and F' are unknown constants. The matrix element (29.2.7) satisfies a conservation condition $p_\mu\left\langle \mathrm{VAC}\middle|S^\mu(0)\middle|\mathbf{p},\lambda\right\rangle = 0$ because $u(\mathbf{p},\lambda)$ satisfies the momentum-space Dirac equation (5.5.42) for zero mass, and p^μ is on the light-cone. The sum over helicities gives

$$\sum_\lambda u(\mathbf{p},\lambda)\,\bar{u}(\mathbf{p},\lambda) = -i\not{p}/2p^0\,.$$

(The Dirac spinor $u(\mathbf{p},\lambda)$ for a massless particle of momentum $\mathbf{p}$ and spin λ is not well defined for $\mathbf{p}\to 0$, but there is no problem if we regard the sum over X in Eq. (29.2.5) as a sum over helicities for a small momentum $\mathbf{p}$ of fixed direction.) After a straightforward calculation, we then have

$$4\,\rho_{\mathrm{VAC}} = \mathrm{Tr}\left\{\left(F\frac{1+\gamma_5}{2} + F^*\frac{1-\gamma_5}{2}\right)\gamma^0\frac{-i\not{p}\beta}{2p^0}(\gamma^0)^\dagger\left(F^*\frac{1+\gamma_5}{2} + F\frac{1-\gamma_5}{2}\right)\right\},$$

and therefore

$$\rho_{\mathrm{VAC}} = |F|^2/2\,. \quad (29.2.8)$$

An alternative proof of this formula will be given at the end of this section.

---

* Lorentz invariance alone would yield this formula with independent coefficients F_L, F'_L and F_R, F'_R for the matrices proportional to $(1+\gamma_5)/2$ and $(1-\gamma_5)/2$, respectively. It is CPT invariance that imposes the relations $F_R = F_L^*$ and $F'_R = F'^*_L$. To see this, we must use the CPT-transformation properties of the supersymmetry current

$$\mathrm{CPT}\,S^\mu(x)(\mathrm{CPT})^{-1} = -\gamma_5 S^\mu(-x)^* = -\beta\epsilon S^\mu(-x)$$

(see Section 5.8) and of the one-particle states

$$\mathrm{CPT}|\mathbf{p},\lambda\rangle = \chi_\lambda|\mathbf{p},-\lambda\rangle$$

with χ_λ a phase factor that depends on how we define the relative phases of the helicity states. We also need the reality properties of the coefficient functions $u(\mathbf{p},\lambda)$. These are related to those of the one-particle states by the definition

$$\langle \mathrm{VAC}|\psi_{\mathrm{REN}}(x)|\mathbf{p},\lambda\rangle = (2\pi)^{-3/2}\exp(ip\cdot x)\,u(\mathbf{p},\lambda)\,,$$

where $\psi_{\mathrm{REN}}(x)$ is a renormalized Majorana field, with CPT-transformation property

$$\mathrm{CPT}\,\psi_{\mathrm{REN}}(x)\,(\mathrm{CPT})^{-1} = \gamma_5\,\psi_{\mathrm{REN}}(-x)^* = \beta\,\epsilon\,\psi_{\mathrm{REN}}(-x)\,.$$

which yields $u(\mathbf{p},\lambda) = \chi_\lambda^*\,\beta\,\epsilon\,u^*(\mathbf{p},-\lambda)$. Putting this together with the CPT-transformation properties of the supersymmetry current and one-particle states yields the relations $F_R = F_L^*$ and $F'_R = F'^*_L$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

The parameter F plays much the same role in the interactions of soft goldstinos as does the parameter F_π (introduced in Section 19.4) in the interactions of soft pions. The matrix element of the supersymmetry current between any two states X and Y may be split into terms that have a one-goldstino pole at $p^\mu = 0$ in the momentum transfer $p \equiv p_X - p_Y$ and terms that do not have this pole

$$\langle X|S^\mu(0)|Y\rangle = \left\{ \left(\frac{1+\gamma_5}{2}\right)[\gamma^\mu F + ip^\mu F'] + \left(\frac{1-\gamma_5}{2}\right)[\gamma^\mu F^* + ip^\mu F'^*] \right\} \times \left(\frac{-i\not{p}}{p^2}\right) M(X \to Y + g) + \langle X|S^\mu(0)|Y\rangle_{\text{no pole}}\,, \quad (29.2.9)$$

where $\bar{u}M(X \to Y + g)$ is the amplitude for emitting a goldstino with four-momentum p and Dirac wave function u, and the subscript 'no pole' denotes the terms in the matrix element that do not have the one-goldstino pole in the four-momentum p. The conservation of the current S^μ tells us that this vanishes if contracted with p_μ, so in the limit $p^\mu \to 0$ the goldstino emission amplitude is**

$$M(X \to Y + g) \to -i\left\{\left(\frac{1+\gamma_5}{2F}\right) + \left(\frac{1-\gamma_5}{2F^*}\right)\right\} p_\mu \langle X|S^\mu(0)|Y\rangle_{\text{no pole}}\,. \quad (29.2.10)$$

There is another sum rule that provides an alternative proof of the existence of goldstinos when supersymmetry is spontaneously broken, and that also relates the parameter F and the vacuum energy density to the D-terms and $\mathscr{F}$-terms that characterize the strength of supersymmetry breaking. (This use of a sum rule is analogous to the second proof in Section 19.2 of the existence of Goldstone bosons when ordinary symmetries are spontaneously broken.) To derive this sum rule, we will now give up the device of a finite volume, and instead avoid the question of the convergence of the integral (29.1.3) by working with a local consequence of the supersymmetry of the action

$$\left[\left(\bar{S}^0(\mathbf{x},t)\,\alpha\right),\, \chi(\mathbf{y},t)\right] = \left[\left(\bar{\alpha}\,S^0(\mathbf{x},t)\right),\, \chi(\mathbf{y},t)\right] = i\,\delta^3(\mathbf{x}-\mathbf{y})\,\delta\chi(\mathbf{x},t) + \cdots\,, \quad (29.2.11)$$

** The amplitude $\langle X|S^\mu(0)|Y\rangle_{\text{no pole}}$ may have poles that go as $1/p\cdot k$ as $p^\mu \to 0$; they would not arise from the goldstino propagator, which is explicitly excluded in this matrix element, but rather from other particle propagators produced by the insertion of the supersymmetry current in external lines of momentum k of the process $X \to Y$. In the limit $p^\mu \to 0$, the contribution of these poles in Eq. (29.2.10) would dominate the amplitudes for emission or absorption of soft goldstinos. But for such poles to arise, the goldstino would have to be emitted in a transition between a pair of *degenerate* particles of opposite statistics, which are not likely to appear in theories in which supersymmetry is spontaneously broken. The interactions of soft goldstinos differ in this respect from the interactions of soft pions, photons, or gravitons.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

where $\chi(x)$ is an arbitrary fermionic or bosonic field, $\delta\chi(x)$ is the change in $\chi(x)$ induced by the supersymmetry transformation with infinitesimal parameter α that leaves the action invariant, and the dots denote terms involving derivatives of $\delta^3(\mathbf{x}-\mathbf{y})$. Let us consider the vacuum expectation value of the anticommutator of an arbitrary left-handed spinor field $\psi_L(x)$ with the covariant conjugate $\bar{S}^\mu(y)$ of the supersymmetry current. By summing over a complete set $|X\rangle$ of intermediate states (including integration over particle momenta), this vacuum expectation value may be put in the form

$$\left\langle \mathrm{VAC}\middle|\left\{\psi_{L\alpha}(x),\,\bar{S}^\mu_\beta(y)\right\}\middle|\mathrm{VAC}\right\rangle = \int d^4p\; e^{ip\cdot(x-y)}\left[G^\mu_{\alpha\beta}(p)+\tilde{G}^\mu_{\alpha\beta}(-p)\right], \tag{29.2.12}$$

where

$$G^\mu_{\alpha\beta}(p) \equiv \sum_X \delta^4(p-p_X)\left\langle \mathrm{VAC}\middle|\psi_{L\alpha}(0)\middle|X\right\rangle\left\langle X\middle|\bar{S}^\mu_\beta(0)\middle|\mathrm{VAC}\right\rangle, \tag{29.2.13}$$

$$\tilde{G}^\mu_{\alpha\beta}(p) = \sum_X \delta^4(p-p_X)\left\langle \mathrm{VAC}\middle|\bar{S}^\mu_\beta(0)\middle|X\right\rangle\left\langle X\middle|\psi_{L\alpha}(0)\middle|\mathrm{VAC}\right\rangle. \tag{29.2.14}$$

Lorentz invariance requires that the matrices $G^\mu(p)$ and $\tilde{G}^\mu(p)$ must take the forms:

$$G^\mu(p) = \theta(p^0)\left(\frac{1+\gamma_5}{2}\right)\Big[\gamma^\mu G^{(1)}(-p^2)+\not{p}p^\mu G^{(2)}(-p^2) + p^\mu G^{(3)}(-p^2)+\not{p}\gamma^\mu G^{(4)}(-p^2)\Big] \tag{29.2.15}$$

and

$$\tilde{G}^\mu(p) = \theta(p^0)\left(\frac{1+\gamma_5}{2}\right)\Big[\gamma^\mu \tilde{G}^{(1)}(-p^2)+\not{p}p^\mu \tilde{G}^{(2)}(-p^2) + p^\mu \tilde{G}^{(3)}(-p^2)+\not{p}\gamma^\mu \tilde{G}^{(4)}(-p^2)\Big]. \tag{29.2.16}$$

The right-hand sides of Eqs. (29.2.15) and (29.2.16) are unaffected if we replace $-p^2$ with m^2, multiply by a factor $\delta(p^2+m^2)$, and integrate over m^2. In this way, Eq. (29.2.12) becomes

$$(2\pi)^{-3}\left\langle \mathrm{VAC}\middle|\left\{\psi_L(x),\,\bar{S}^\mu(y)\right\}\middle|\mathrm{VAC}\right\rangle = \left(\frac{1+\gamma_5}{2}\right)\int_0^\infty dm^2\Big[\gamma^\mu G^{(1)}(m^2) - \not{\partial}\partial^\mu G^{(2)}(m^2) - i\partial^\mu G^{(3)}(m^2) - i\not{\partial}\gamma^\mu G^{(4)}(m^2)\Big]\Delta_+(x-y,m)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

$$+\left(\frac{1+\gamma_5}{2}\right)\int_0^\infty dm^2\left[\gamma^\mu\tilde{G}^{(1)}(m^2)-\not{\partial}\partial^\mu\tilde{G}^{(2)}(m^2)-i\partial^\mu\tilde{G}^{(3)}(m^2)\right.$$
$$\left.-i\not{\partial}\gamma^\mu\tilde{G}^{(4)}(m^2)\right]\Delta_+(y-x,m)\,, \tag{29.2.17}$$

where $\Delta_+(x,m)$ is the standard function

$$\Delta_+(x,m)\equiv(2\pi)^{-3}\int d^4p\;\theta(p^0)\,\delta(p^2+m^2)\,\exp(ip\cdot x)\,. \tag{29.2.18}$$

Causality requires that the anticommutator on the left-hand side of Eq. (29.2.17) should vanish for spacelike separations $x-y$. For such separations, $\Delta_+(x-y,m)$ is an even function of $x-y$, so Eq. (29.2.17) vanishes for all spacelike separations if and only if

$$\begin{aligned} G^{(1)}(m^2)&=-\tilde{G}^{(1)}(m^2)\,, & G^{(2)}(m^2)&=-\tilde{G}^{(2)}(m^2)\,,\\ G^{(3)}(m^2)&=+\tilde{G}^{(3)}(m^2)\,, & G^{(4)}(m^2)&=+\tilde{G}^{(4)}(m^2)\,, \end{aligned} \tag{29.2.19}$$

so that for general $x-y$, Eq. (29.2.17) now reads

$$(2\pi)^{-3}\left\langle \mathrm{VAC}\middle|\left\{\psi_L(x),\bar{S}^\mu(y)\right\}\middle|\mathrm{VAC}\right\rangle=\int_0^\infty dm^2\left[G^{(1)}(m^2)\gamma^\mu-G^{(2)}(m^2)\not{\partial}\partial^\mu\right.$$
$$\left.-iG^{(3)}(m^2)\partial^\mu-iG^{(4)}(m^2)\not{\partial}\gamma^\mu\right]\Delta(x-y,m)\left(\frac{1+\gamma_5}{2}\right)\,, \tag{29.2.20}$$

where, as usual,

$$\Delta(x-y,m)\equiv\Delta_+(x-y,m)-\Delta_+(y-x,m)\,. \tag{29.2.21}$$

We next impose the supersymmetry current conservation condition (29.1.2), which (because $\Box\Delta=m^2\Delta$) yields

$$G^{(1)}(m^2)=m^2G^{(2)}(m^2)\,,\qquad m^2G^{(3)}(m^2)=-m^2G^{(4)}(m^2)\,. \tag{29.2.22}$$

Finally, we must relate these spectral functions to the breaking of supersymmetry. Recall that for $x^0=y^0$,

$$\Delta(\mathbf{x}-\mathbf{y},0,m)=0\,,\qquad \dot{\Delta}(\mathbf{x}-\mathbf{y},0,m)=-i\delta^3(\mathbf{x}-\mathbf{y})\,.$$

Setting $x^0=y^0=t$ and $\mu=0$ in Eq. (29.2.20) thus yields

$$\left\langle \mathrm{VAC}\middle|\left\{\psi_L(\mathbf{x},t),\bar{S}^0(\mathbf{y},t)\right\}\middle|\mathrm{VAC}\right\rangle=(2\pi)^3\delta^3(\mathbf{x}-\mathbf{y})\left(\frac{1+\gamma_5}{2}\right)$$
$$\times\int_0^\infty dm^2\left[G^{(3)}(m^2)+G^{(4)}(m^2)\right]\,. \tag{29.2.23}$$

Contracting on the right with an infinitesimal Majorana fermionic supersymmetry transformation parameter α and using Eq. (29.2.11) then gives

$$i\left\langle\delta\psi_L\right\rangle_{\mathrm{VAC}}=(2\pi)^3\left(\frac{1+\gamma_5}{2}\right)\alpha\int_0^\infty dm^2\left[G^{(3)}(m^2)+G^{(4)}(m^2)\right]\,. \tag{29.2.24}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

But Eq. (29.2.22) shows that the integrand of the integral over m^2 in Eq. (29.2.24) vanishes, except perhaps at $m^2 = 0$, so we can conclude that

$$G^{(3)}(m^2) + G^{(4)}(m^2) = \delta(m^2)\mathcal{G}\,, \tag{29.2.25}$$

with a constant coefficient $\mathcal{G}$ given by

$$i\left\langle \delta\psi_L \right\rangle_{\rm VAC} = (2\pi)^3 \mathcal{G}\alpha_L\,. \tag{29.2.26}$$

As we saw in Sections 26.4 and 27.4, a breakdown of supersymmetry is signaled by the appearance of vacuum expectation values of the changes $\delta\psi$ under supersymmetry transformations of one or more spinor fields ψ. Eqs. (29.2.25) and (29.2.26) show that for any such spinor field, the spectral function $G^{(3)}(m^2) + G^{(4)}(m^2)$ has a delta function singularity at $m^2 = 0$, which could only arise from the appearance of a massless one-particle state $|g\rangle$ in the sums over states in Eqs. (29.2.13) and/or (29.2.14). For the matrix elements $\langle {\rm VAC}|\psi|g\rangle$ or $\langle g|\psi|{\rm VAC}\rangle$ not to vanish, this massless particle must have spin 1/2. This is the goldstino.

Now let's calculate the contribution of a one-goldstino state $|\mathbf{p},\lambda\rangle$ with momentum $\mathbf{p}$ and helicity λ to the spectral functions $G^{(i)}(m^2)$. Lorentz invariance tells us that the matrix element of a general fermion field (not just the renormalized goldstino field) in Eq. (29.2.13) takes the form[†]

$$\left\langle {\rm VAC}\middle|\psi(0)\middle|\mathbf{p},\lambda\right\rangle = \frac{1}{(2\pi)^{3/2}}\left[N\left(\frac{1+\gamma_5}{2}\right) + N^*\left(\frac{1-\gamma_5}{2}\right)\right] u(\mathbf{p},\lambda)\,, \tag{29.2.27}$$

where N is a constant, characterizing the particular fermion field. Also, the delta function in Eq. (29.2.13) is

$$\delta^4(p - p_g) = 2p^0\delta(p^2)\theta(p^0)\delta^3(\mathbf{p} - \mathbf{p}_g)\,.$$

Together with Eq. (29.2.7), this yields

$$\left[G^\mu_{\alpha\beta}(p)\right]_1 = \frac{N}{(2\pi)^3}\left(\frac{1+\gamma_5}{2}\right)\theta(p^0)\delta(p^2)\;\not{p}\left[\gamma^\mu F + ip^\mu F'\right],$$

with the subscript 1 indicating the one-goldstino contribution. Comparing this with Eq. (29.2.15) shows that

$$\begin{aligned} \left[G^{(2)}(m^2)\right]_1 &= i(2\pi)^{-3}NF'\delta(m^2)\,, \qquad \left[G^{(4)}(m^2)\right]_1 = (2\pi)^{-3}NF\delta(m^2)\,, \\ \left[G^{(1)}(m^2)\right]_1 &= \left[G^{(3)}(m^2)\right]_1 = 0\,. \end{aligned} \tag{29.2.28}$$

[†] Again, Lorentz invariance would allow independent coefficients N_L and N_R for $(1+\gamma_5)$ and $(1-\gamma_5)$, respectively. Using the CPT transformation of the one-particle state and the reality property of the coefficient functions discussed in the first footnote of this section, and the CPT transformation of a general fermion field, ${\rm CPT}\psi(x)({\rm CPT})^{-1} = \beta\epsilon\psi(-x)$, we find that $N_L = N_R^*$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Thus Eqs. (29.2.25) and (29.2.26) yield

$$i\left\langle \delta\psi_L \right\rangle_{\text{VAC}} = NF\alpha_L \,. \tag{29.2.29}$$

Evidently if the change in any fermion field under a supersymmetry transformation has a non-zero vacuum expectation value then the N-factor for that field cannot vanish, so there must be a one-goldstino state that contributes to these spectral functions.

To be a little more specific, recall that the fermionic components ψ_{Ln} of a left-chiral scalar superfield Φ_{Ln} obey the supersymmetry transformation rule (26.3.15):

$$\delta\psi_{Ln} = \sqrt{2}\partial_\mu\phi_n\,\gamma^\mu\,\alpha_R + \sqrt{2}\mathscr{F}_n\,\alpha_L \,, \tag{29.2.30}$$

while Eq. (27.3.5) gives the supersymmetry transformation of the gaugino fields as

$$\delta\lambda_A = \left(\tfrac{1}{4}f_{A\mu\nu}\,[\gamma^\nu\,,\,\gamma^\mu] + i\gamma_5 D_A\right)\alpha \,. \tag{29.2.31}$$

Hence the N-factors in the matrix elements of the ψ_n and λ_A between the vacuum and a one-goldstino state are given by

$$N_n = iF^{-1}\sqrt{2}\left\langle \mathscr{F}_n \right\rangle_{\text{VAC}} \,, \qquad N_A = -F^{-1}\left\langle D_A \right\rangle_{\text{VAC}} \,. \tag{29.2.32}$$

Let us see how the results (29.2.32) arise in the tree approximation. For a renormalizable theory of gauge and chiral superfields, Eq. (27.4.30) gives the left-handed part of the fermion mass matrix as

$$\begin{aligned} M_{nm} &= \left(\frac{\partial^2 f(\phi)}{\partial\phi_n\partial\phi_m}\right)_{\phi=\phi_0} \,, \\ M_{nA} &= M_{An} = i\sqrt{2}(t_A\phi_0)_n^* \,, \qquad M_{AB} = 0 \,. \end{aligned} \tag{29.2.33}$$

The vacuum fields ϕ_{n0} are at a minimum of the potential given by Eq. (27.4.9), so

$$0 = \left(\frac{\partial V(\phi)}{\partial\phi_n}\right)_0 = -\sum_m M_{nm}\mathscr{F}_{m0} + \sum_A \left(\phi_0^\dagger t_A\right)_n D_{A0} \,, \tag{29.2.34}$$

where the $\mathscr{F}$s and Ds are given by Eqs. (27.4.6) and (27.4.7)

$$\mathscr{F}_n = -\left(\partial f(\phi)/\partial\phi_n\right)^* \,, \qquad D_A = \xi_A + \sum_{nm}\phi_n^*(t_A)_{nm}\phi_m \,,$$

and the subscript 0 indicates that we are to set $\phi_n = \phi_{n0}$. Furthermore, the gauge invariance of the superpotential requires that

$$\sum_n \mathscr{F}_n\left(t_B\phi\right)_n^* = 0 \,, \tag{29.2.35}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

for all values of ϕ. Therefore the left-handed quark mass matrix M has an eigenvector v with $Mv = 0$, where

$$v_n = \sqrt{2}\mathscr{F}_{n0}\,, \qquad v_A = iD_{A0}\,. \tag{29.2.36}$$

Thus in the tree approximation, if we expand the left-handed fermion fields in renormalized fields for particles of definite mass, then the coefficients of the goldstino field in ψ_{Ln} and in λ_{AL} are proportional to $\sqrt{2}\mathscr{F}_{n0}$ and iD_{A0}, respectively, in agreement with Eq. (29.2.32).

If we normalize the spinor fields ψ_{Ln} and λ_A so that the matrix connecting these fields to the renormalized fields of particles of definite mass is unitary, then

$$\sum_n |N_n|^2 + \sum_A |N_A|^2 = 1\,, \tag{29.2.37}$$

so Eq. (29.2.32) gives the non-perturbative result

$$|F|^2 = 2\sum_n \left|\left\langle \mathscr{F}_n \right\rangle_{\rm VAC}\right|^2 + \sum_A \left|\left\langle D_A \right\rangle_{\rm VAC}\right|^2 . \tag{29.2.38}$$

The result (29.2.38) for $|F|^2$ allows us to express the vacuum expectation density (29.2.8) in terms of the vacuum expectation values of auxiliary fields

$$\rho_{\rm VAC} = \sum_n \left|\left\langle \mathscr{F}_n \right\rangle_{\rm VAC}\right|^2 + \frac{1}{2}\sum_A \left|\left\langle D_A \right\rangle_{\rm VAC}\right|^2 . \tag{29.2.39}$$

This is the non-perturbative generalization of the zeroth-order result (27.4.9). It confirms a result used in Section 27.6, that the conditions $\langle \mathscr{F}_n \rangle = \langle D_A \rangle = 0$ are sufficient as well as necessary for supersymmetry to be unbroken.

* * *

It is instructive to see how Eq. (29.2.8) can be derived without the device of a finite volume. For this purpose, consider the vacuum expectation value of the anticommutator of two supersymmetry currents. By using Lorentz invariance and the vanishing of anticommutators at spacelike separations in the same way as in the previous section, we find

$$\begin{aligned}\left\langle {\rm VAC}\middle|\left\{S^\mu(x),\,\bar{S}^\nu(y)\right\}\middle|{\rm VAC}\right\rangle = -i\int dm^2 \Big[&H^{(1)}(m^2)\gamma^\mu\partial^\nu + H^{(2)}(m^2)\gamma^\nu\partial^\mu \\ &+H^{(3)}(m^2)\,\not{\partial}\partial^\mu\partial^\nu + H^{(4)}(m^2)\,\not{\partial}\eta^{\mu\nu} + H^{(5)}(m^2)\epsilon^{\mu\nu\lambda\rho}\partial_\lambda\gamma_\rho\Big]\Delta(x-y,m^2) \\ &+\cdots\end{aligned} \tag{29.2.40}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

where

$$\int dX\ \delta^4(p-p_X)\Big\langle \mathrm{VAC}\Big|S^\mu_\alpha(0)\Big|X\Big\rangle\Big\langle X\Big|\bar{S}^\nu_\beta(0)\Big|\mathrm{VAC}\Big\rangle = H^{(1)}(-p^2)\gamma^\mu p^\nu + H^{(2)}(-p^2)\gamma^\nu p^\mu - H^{(3)}(-p^2)\ \not{p}p^\mu p^\nu + H^{(4)}(-p^2)\ \not{p}\eta^{\mu\nu} + H^{(5)}(-p^2)\epsilon^{\mu\nu\lambda\rho}p_\lambda\gamma_\rho + \cdots\ , \tag{29.2.41}$$

and the dots in Eqs. (29.2.40) and (29.2.41) denote a linear combination of the other independent Dirac covariant matrices 1, γ_5, $\gamma_5\gamma_\sigma$, and $[\gamma_\sigma,\gamma_\tau]$, which will not concern us here. The Majorana nature of the currents together with Eq. (26.A.20) tell us that the spectral functions $H^{(i)}$ are all real, and the conservation of the supersymmetry currents dictate that

$$H^{(1)}(m^2) = H^{(2)}(m^2) = -m^2\,H^{(3)}(m^2) - H^{(4)}(m^2)\,, \tag{29.2.42}$$

and

$$m^2 H^{(1)}(m^2) = 0\,. \tag{29.2.43}$$

Setting $\mu = \nu = 0$ and $x^0 = y^0$ in Eq. (29.2.40), integrating over $\mathbf{x}$, and using Eq. (29.2.42) then gives

$$\Big\langle \mathrm{VAC}\Big|\Big\{Q\,,\bar{S}^0(0)\Big\}\Big|\mathrm{VAC}\Big\rangle = (2\pi)^3\beta\int dm^2\left[H^{(1)}(m^2) + H^{(2)}(m^2) + m^2\,H^{(3)}(m^2) + H^{(4)}(m^2)\right] + \cdots = (2\pi)^3\beta\int dm^2\,H^{(1)}(m^2) + \cdots\,. \tag{29.2.44}$$

To evaluate this anticommutator, we first write it as

$$\Big\{Q\,,\bar{S}^\nu(x)\Big\} = -2i\gamma_\mu T^{\mu\nu}(x) + \cdots\ , \tag{29.2.45}$$

where $T^{\mu\nu}(x)$ is some tensor operator and the dots again denote some linear combination of the other independent Dirac covariants. The Majorana character of Q and S^μ tells us that $T^{\mu\nu}(x)$ is Hermitian; the conservation of the supersymmetry current tells us that it is conserved, in the sense that

$$\partial_\nu T^{\mu\nu} = 0\,, \tag{29.2.46}$$

and the anticommutation relation (25.2.36) tells us that

$$\int d^3x\ T^{\mu 0}(x) = P^\mu\,. \tag{29.2.47}$$

These properties allow us to identify $T^{\mu\nu}$ as the energy-momentum tensor. It is not in general equal to the *symmetric* energy-momentum tensor $\Theta^{\mu\nu}$ discussed in Section 7.4, but Eq. (29.2.47) shows that the energy density T^{00} can differ from Θ^{00} only by spatial derivative terms that cannot contribute in states of zero three-momentum, so the vacuum energy density is given by

$$\rho_{\mathrm{VAC}} = \Big\langle \mathrm{VAC}\Big|T^{00}\Big|\mathrm{VAC}\Big\rangle\,. \tag{29.2.48}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Thus Eqs. (29.2.44) and (29.2.45) yield

$$2\rho_{\rm VAC} = (2\pi)^3 \int dm^2\, H^{(1)}(m^2)\,. \tag{29.2.49}$$

But Eq. (29.2.43) tells us that $H^{(1)}(m^2)$ vanishes except perhaps at $m^2 = 0$, so

$$H^{(1)}(m^2) = 2(2\pi)^{-3}\delta(m^2)\rho_{\rm VAC}\,. \tag{29.2.50}$$

Thus we see again that a non-vanishing vacuum energy density entails the existence of a massless fermion, the goldstino. Using Eq. (29.2.7), a straightforward calculation of the one-goldstino contribution to the spectral functions yields

$$H^{(1)}(m^2) = (2\pi)^{-3}\delta(m^2)\,|F|^2\,. \tag{29.2.51}$$

Comparing this with Eq. (29.2.50) gives our previous result (29.2.8) for the vacuum energy density.

29.3 Non-Perturbative Corrections to the Superpotential

We saw in Section 27.6 that the superpotential in general supersymmetric theories of gauge and chiral superfields is not renormalized to any finite order of perturbation theory, so that if supersymmetry is not broken in the tree approximation then it can only be broken by non-perturbative corrections to the Wilsonian effective Lagrangian. We now take up a general analysis of these corrections. These were thoroughly studied in a series of papers by Affleck, Davis, Dine, and Seiberg in the early 1980s,[3] with special attention to the case of supersymmetric versions of quantum chromodynamics with arbitrary numbers of colors and flavors. Here we will present a somewhat simplified analysis of general supersymmetric gauge theories, based on the more recent holomorphy arguments of Seiberg[4] already used in Section 27.7.

To study non-perturbative effects, we shall here again consider a general renormalizable supersymmetric theory, but now including a possible θ-term in the Lagrangian density

$$\mathscr{L} = \left[\Phi^\dagger e^{-V}\Phi\right]_D + 2\,{\rm Re}\left[f(\Phi)\right]_{\mathscr{F}} + {\rm Re}\left[\frac{\tau}{8\pi i}\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{A\alpha L}\,W_{A\beta L}\right]_{\mathscr{F}}\,, \tag{29.3.1}$$

where the superpotential $f(\Phi)$ is a gauge-invariant cubic polynomial in the left-chiral superfields and τ is the parameter (27.3.23):

$$\tau = \frac{4\pi i}{g^2} + \frac{\theta}{2\pi}\,. \tag{29.3.2}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

As in Section 27.6, we introduce a pair of gauge-invariant left-chiral external superfields, now called Y and T, and replace the Lagrangian density with

$$\mathscr{L}^{\sharp} = \left[\Phi^{\dagger} e^{-V} \Phi\right]_D + 2\,\mathrm{Re}\left[Y f(\Phi)\right]_{\mathscr{F}} + \mathrm{Re}\left[\frac{T}{8\pi i}\sum_{A\alpha\beta} \epsilon_{\alpha\beta} W_{A\alpha L}\, W_{A\beta L}\right]_{\mathscr{F}} . \tag{29.3.3}$$

This becomes the same as (29.3.1) when we set the spinor and auxiliary components of Y and T equal to zero, and take their scalar components as $y = 1$ and $t = \tau$, respectively. Non-perturbative effects will in general invalidate both of the two symmetries on which the analysis of Section 27.6 was based. The translation operation, which in our present notation is $T \to T + \xi$ with real ξ, is not a symmetry because $\sum_A \epsilon_{\mu\nu\rho\sigma} f_A^{\mu\nu} f_A^{\rho\sigma}$ can have a non-vanishing integral over spacetime. The original R invariance (with T and Y having R values 0 and +2) is not a symmetry because the anomaly discussed in Chapter 22 gives the R-current a non-vanishing divergence. With θ_L and θ_R having $R = +1$ and $R = -1$, respectively, and V_A and Φ_n R-neutral, the fermion fields λ_{AL} and ψ_{nL} have $R = +1$ and $R = -1$, respectively, so Eq. (22.2.26) here gives

$$\partial_\mu J_R^\mu = -\frac{1}{32\pi^2}\left(C_1 - C_2\right)\sum_A \epsilon_{\mu\nu\rho\sigma} f_A^{\mu\nu} f_A^{\rho\sigma} , \tag{29.3.4}$$

where C_1 and C_2 are the constants defined in Eqs. (17.5.33) and (17.5.34):

$$\sum_{CD} C_{ACD} C_{BCD} = C_1\,\delta_{AB} , \qquad \mathrm{Tr}\{t_A t_B\} = C_2\,\delta_{AB} , \tag{29.3.5}$$

with the trace taken over all species of left-chiral superfield.* For instance, in the generalized supersymmetric version of quantum chromodynamics studied in Reference 3, with gauge group $SU(N_c)$ and N_f pairs of left-chiral quark superfields Q_n and $\bar{Q}_n$ in the defining representation and its complex conjugate, these constants have values given by Eq. (17.5.35) (with $n_f = 2N_f$) as

$$C_1 = N_c , \qquad C_2 = N_f .$$

Although T translation and R invariance are invalidated by non-perturbative effects, there is a remaining symmetry which is almost as

* A factor 32 instead of 16 appears in the denominator in Eq. (29.3.4) because gauginos do not have distinct antiparticles, and we are now counting antiparticles separately from particles in taking the trace in Eq. (29.3.5). Also, we are now adopting the convention, described at the end of Section 27.3, of including a gauge coupling factor in the gauge fields and not in structure constants and the matrix generators t_A. The gauge generators are thus normalized so that for t_A, t_B, and t_C in the standard $SU(2)$ subalgebra of the gauge algebra, the structure constant is $C_{ABC} = \epsilon_{ABC}$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

powerful. Consider a general R transformation

$$\theta_L \to e^{i\varphi}\theta_L\,, \quad \Phi \to \Phi\,, \quad V_A \to V_A\,, \quad Y \to e^{2i\varphi}Y\,, \tag{29.3.6}$$

with arbitrary real φ. This leaves the T-independent terms of the Lagrangian density (29.3.3) invariant, but according to Eq. (29.3.4), quantum effects violate this symmetry, just as if there were a term $\Delta\mathscr{L}$ in the Lagrangian density with a transformation

$$\Delta\mathscr{L} \to \Delta\mathscr{L} - \frac{1}{32\pi^2}\left(C_1 - C_2\right)\sum_A \epsilon_{\mu\nu\rho\sigma} f_A^{\mu\nu} f_A^{\rho\sigma}\varphi\,.$$

Recalling Eq. (27.3.18), this is cancelled if we give T a transformation

$$T \to T + (C_1 - C_2)\varphi/\pi\,. \tag{29.3.7}$$

Because $W_{A\alpha L}$ has $R = 2$, the whole theory including non-perturbative effects is invariant under the *combined* transformations (29.3.6) and (29.3.7). In particular, the superfield $\exp(2i\pi T)$, which for $T = \tau$ is periodic in θ, has $R = 2(C_1 - C_2)$.

We again introduce an ultraviolet cut-off, and consider the effective 'Wilsonian' Lagrangian

$$\mathscr{L}_\lambda^\sharp = \left[\mathscr{A}_\lambda(\Phi, \Phi^\dagger, V, T, T^\dagger, Y, Y^\dagger, \mathscr{D}\cdots)\right]_D + 2\,\mathrm{Re}\left[\frac{T}{8\pi i}\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{AL\alpha}W_{AL\beta} + \mathscr{B}_\lambda(\Phi, W_L, T, Y)\right]_{\mathscr{F}}\,, \tag{29.3.8}$$

with $\mathscr{A}_\lambda$ and $\mathscr{B}_\lambda$ both gauge-invariant functions of the displayed arguments. The term proportional to T has been separated from the function $\mathscr{B}_\lambda$ in order that the translation (29.3.7) of T should continue to cancel the anomaly in the R transformation (29.3.6). Invariance under the combined transformation (29.3.6), (29.3.7) then tells us that terms in the function $\mathscr{B}_\lambda$ must be proportional to powers of $\exp(2i\pi T)$, which have definite R values.

Furthermore, it is only *positive* powers of $\exp(2i\pi T)$ that may appear in $\mathscr{B}_\lambda$. According to Eq. (27.3.24), it is only instantons with positive winding number $\nu \geq 0$ that can make contributions to the effective Lagrangian that are holomorphic in T rather than T^*, and these give rise to factors $\exp(2i\pi\nu T)$. More generally, for $T = \tau$ any power $\exp(2ia\pi T)$ will depend on the gauge coupling through a factor $\exp(-8\pi^2 a/g^2)$, so that a must be positive in order for non-perturbative effects to be suppressed for small g. In consequence, non-perturbative effects enter in $\mathscr{L}_\lambda^\sharp$ through operators $\exp(2i\pi aT)$ that have positive-definite, zero, or negative-definite values of R, depending on whether $C_1 > C_2$, $C_1 = C_2$, or $C_1 < C_2$. (In the generalized supersymmetric version of quantum chromodynamics

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

described above, this corresponds to $N_c > N_f$, $N_c = N_f$, and $N_c < N_f$, respectively.) We shall now consider each of these cases.

$$C_1 > C_2$$

Here the powers $\exp(2ia\pi T)$ with $a > 0$ have positive-definite values of R, given by Eq. (29.3.7) as $R = 2(C_1 - C_2)a$. Lorentz invariance tells us that if any term in $\mathscr{B}_\lambda$ contains a factor $W_{A\alpha L}$, then it must contain at least two of them, so the only ways to construct terms in $\mathscr{B}_\lambda$ with $R = 2$ is to have two Ws and no dependence on Y or T, or one Y and no dependence on W or T, or one factor of $\exp(2i\pi T/(C_1 - C_2))$ and no dependence on W or Y:

$$\mathscr{B}_\lambda = Y f_\lambda(\Phi) + \sum_{\alpha\beta AB} \epsilon_{\alpha\beta} W_{A\alpha L} W_{B\beta B} \ell_{\lambda AB}(\Phi) + \exp\left(\frac{2i\pi T}{C_1 - C_2}\right) v_\lambda(\Phi) \,. \tag{29.3.9}$$

Because $f_\lambda(\Phi)$ does not depend on Y or T, it can only be the tree-approximation superpotential

$$f_\lambda(\Phi) = f(\Phi) \,, \tag{29.3.10}$$

just as in perturbation theory. Likewise, because $\ell_{\lambda AB}(\Phi)$ does not depend on Y or T it must have equal numbers of Φs and $\Phi^\dagger$s, so since it does not depend on $\Phi^\dagger$ it cannot depend on Φ either. Gauge invariance then requires (for a simple gauge group) that $\ell_{\lambda AB}(\Phi)$ is proportional to δ_{AB}, and since it does not depend on T or Y, the power-counting argument of Section 27.6 shows that the coefficient of δ_{AB} can only be the one-loop contribution to the running inverse-square Wilsonian gauge coupling.

To be more explicit about the running gauge coupling, recall that in non-supersymmetric gauge theories with fermions, Eq. (18.7.2) gives the one-loop renormalization group equation as

$$\lambda \frac{dg_\lambda}{d\lambda} = b g_\lambda^3 \,, \tag{29.3.11}$$

with

$$b = -\frac{1}{4\pi^2}\left(\frac{11}{12}C_1 - \frac{1}{6}C_2\right) \,, \tag{29.3.12}$$

with the coefficient of C_2 taken as $-1/6$ rather than $-1/3$ because we are now counting the left-chiral states of antifermions separately from the left-chiral states of particles. As we have seen in Section 28.2, the effect of gauginos is to multiply the C_1-term by a factor $9/11$, while the effect of the scalar components of the left-chiral superfields (such as squarks and sleptons) is to multiply the C_2-term by a factor $3/2$, so in supersymmetric

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

theories Eq. (29.3.12) becomes instead

$$b = -\frac{1}{16\pi^2}(3C_1 - C_2) \ . \tag{29.3.13}$$

The solution of Eq. (29.3.11) for the running gauge coupling is then

$$g_\lambda^{-2} = g^{-2} + \frac{3C_1 - C_2}{8\pi^2} \ln\left(\frac{\lambda}{K}\right) \ , \tag{29.3.14}$$

where K is an ultraviolet cut-off, introduced to give meaning to the otherwise ultraviolet-divergent bare gauge coupling g.

To summarize the results so far, setting $T = \tau$ and $Y = 1$, the Wilsonian effective Lagrangian for $C_1 > C_2$ takes the form

$$\mathscr{L}_\lambda^\sharp = \left[\mathscr{A}_\lambda(\Phi, \Phi^\dagger, V, \tau, \tau^*, \mathscr{D}\cdots)\right]_D + 2\,\mathrm{Re}\left[\frac{\tau_\lambda}{8\pi i}\sum_{A\alpha\beta}\epsilon_{\alpha\beta}W_{A\alpha L}W_{A\beta L}\right]_{\mathscr{F}}$$
$$+2\,\mathrm{Re}\,[f(\Phi)]_{\mathscr{F}} + \exp\left(\frac{2i\pi\tau_\lambda}{C_1 - C_2}\right)[v_\lambda(\Phi)]_{\mathscr{F}} \ , \tag{29.3.15}$$

where

$$\tau_\lambda = \frac{4\pi i}{g_\lambda^2} + \frac{\theta}{2\pi} \ . \tag{29.3.16}$$

We have been able to replace τ with τ_λ in the exponential in the last term in Eq. (29.3.15), because the difference is a constant times $\ln\lambda$, which yields a power of λ that can be absorbed into the definition of v_λ.

Non-perturbative effects have now been isolated in the last term in Eq. (29.3.15). This term can be generated by instantons of winding number $\nu > 0$ if $C_1 - C_2 = 1/\nu$. (In general $C_1 - C_2$ is a rational number. For the generalized supersymmetric version of quantum chromodynamics, $C_1 - C_2 = N_c - N_f$ is an integer, so the condition $C_1 - C_2 = 1/\nu$ requires that $N_c = N_f - 1$, and then only $\nu = 1$ instantons contribute. Detailed calculations[5] in this model show that instantons actually do make such contributions.) Whether or not it is instantons that generate the non-perturbative contribution $v_\lambda(\Phi)$, we can determine its form by considering the non-anomalous symmetries of the theory. Since this function is independent of Y, it can be evaluated as if $Y = 0$, so it shares all the non-anomalous symmetries of the first term in Eq. (29.3.1). These include the gauge symmetry itself and a global symmetry under $\prod_d SU(n(d))$, where d labels the different irreducible representations of the gauge group furnished by the left-chiral superfields, and $n(d)$ is the number of times representation d occurs. (For instance, in generalized supersymmetric quantum chromodynamics d takes two values, labelling the N_c and $\bar{N}_c$ representations of $SU(N_c)$, and $n(N_c) = n(\bar{N}_c) = N_f$.) Let us label the Φs as $\Phi_{ai}^{(d)}$, where a is a gauge index, and i is a 'flavor' index, labelling

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

the $n(d)$ different Φs that transform under the gauge group according to the representation d. The only way to construct a function of the Φs that is invariant under the global symmetry group $\prod_d SU(n(d))$ is as a product of Φs, with the $n(d)$ flavor indices contracted for each d with the antisymmetric $SU(n(d))$ tensor $\epsilon_{i_1 \ldots i_{n(d)}}$, and with the gauge indices contracted with constant tensors of the gauge group. (For instance, for generalized supersymmetric quantum chromodynamics, v_λ must be a function of the sole invariant

$$D \equiv \mathrm{Det}_{ij} \sum_a Q_{ai} \overline{Q}_{aj} \,,$$

which is non-zero only for $N_c \geq N_f$.)

In addition to the anomaly-free $SU(n(d))$ flavor symmetries, there is also a $U_d(1)$ symmetry for each of the irreducible representations d furnished by the Φs, with all $\Phi_{ai}^{(d)}$ for a given d undergoing the transformation

$$\Phi_{ai}^{(d)} \to e^{i\varphi_d} \Phi_{ai}^{(d)} \,. \tag{29.3.17}$$

This symmetry is anomalous, with the effects of the anomaly the same as if the Lagrangian underwent the change

$$\mathscr{L} \to \mathscr{L} - \sum_d \frac{n(d) C_{2d}}{32\pi^2} \sum_A \epsilon_{\mu\nu\rho\sigma} f_A^{\mu\nu} f_A^{\rho\sigma} \varphi_d \,, \tag{29.3.18}$$

where C_{2d} is the contribution to C_2 of any one left-chiral scalar superfield belonging to the irreducible representation d of the gauge group. The symmetry is restored if we give T the transformation property

$$T \to T + n(d) C_{2d} \varphi_d / \pi \,. \tag{29.3.19}$$

Since $v_\lambda(\Phi)$ is accompanied in Eq. (29.3.9) by a factor $\exp(2i\pi T/(C_1 - C_2))$, which undergoes the transformation

$$\exp\left(\frac{2i\pi T}{C_1 - C_2}\right) \to \prod_d \exp\left(\frac{+2in(d) C_{2d} \varphi_d}{C_1 - C_2}\right) \exp\left(\frac{2i\pi T}{C_1 - C_2}\right) \,, \tag{29.3.20}$$

we conclude that *for each representation d of the gauge group furnished by the left-chiral scalars, $v_\lambda(\Phi)$ must be a homogeneous function of the $\Phi_{ai}^{(d)}$ of negative order $-2n(d)C_{2d}/(C_1 - C_2)$.* (For instance, in generalized supersymmetric quantum chromodynamics we have two irreducible representations of $SU(N_c)$, the defining and antidefining representations, each with $n(d) = N_f$ and $C_{2d} = 1/2$, so v_λ is a homogeneous function of order $-N_f/(N_c - N_f)$ in the Q belonging to the defining representation and of the same order in the $\bar{Q}$ belonging to the antidefining representation. Thus it must be proportional to $D^{-1/(N_c - N_f)}$, where D is the determinant introduced earlier.) In general $C_2 = \sum_d n(d) C_{2d}$, so $v_\lambda(\Phi)$ is a homogeneous function of all the Φs, of order $-2C_2/(C_1 - C_2)$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

This result satisfies an important consistency check. Recall from Section 27.4 that any superpotential has dimensionality +3 (counting powers of mass, with $\hbar = c = 1$), while the scalar superfields Φ like ordinary scalar fields have dimensionality +1, so the Φ-dependent part of v_λ must appear with a coefficient of dimensionality

$$3 + \frac{2C_2}{C_1 - C_2} = \frac{3C_1 - C_2}{C_1 - C_2} .$$

This coefficient does not depend on the gauge coupling or on any of the couplings or masses in the superpotential, and because we have replaced the bare coupling g with g_λ in the second term of Eq. (29.3.11) it cannot depend on the ultraviolet cut-off K used to define g either, so it can only depend on λ. Therefore

$$v_\lambda(\Phi) = \lambda^{(3C_1 - C_2)/(C_1 - C_2)} H(\Phi) , \tag{29.3.21}$$

where $H(\Phi)$ is a homogeneous function of order $-2n(d)C_{2d}/(C_1 - C_2)$ in the Φs that belong to each representation d of the gauge group and is independent of any parameters of the theory. We can rewrite Eq. (29.3.14) in the form

$$\tau_\lambda = i\frac{3C_1 - C_2}{2\pi} \ln\left(\frac{\lambda}{\Lambda}\right) + \frac{\theta}{2\pi} , \tag{29.3.22}$$

where Λ is an energy parameter that characterizes the running gauge coupling, like the $\Lambda \approx 200$ MeV of quantum chromodynamics. Thus the last term in Eq. (29.3.15) is

$$\exp\left(\frac{2i\pi\tau_\lambda}{C_1 - C_2}\right) v_\lambda(\Phi) = \exp\left(\frac{i\theta}{C_1 - C_2}\right) \Lambda^{(3C_1 - C_2)/(C_1 - C_2)} H(\Phi) . \tag{29.3.23}$$

The whole effective superpotential, including the non-perturbative contribution (29.3.23), is therefore independent of the floating cut-off λ.

The function $H(\Phi)$ is homogeneous and of negative order in Φ, so in the absence of a bare superpotential the potential is positive-definite at finite values of the scalar fields and vanishes only at infinite field values. In such a theory, there is no stable vacuum state, and the question of supersymmetry breaking is moot. The vacuum may be stabilized by adding a suitable bare superpotential. For instance, in generalized supersymmetric quantum chromodynamics with $N_f < N_c$ the only renormalizable bare superpotential is a sum of mass terms

$$f(Q, \overline{Q}) = \sum_{ija} m_{ij} \overline{Q}_{ai} Q_{aj} . \tag{29.3.24}$$

To seek a supersymmetric vacuum state, we need first of all to find what

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

scalar components q_{ai} and $\overline{q}_{ai}$ satisfy Eq. (27.4.11), which here reads

$$\sum_{abi} q^*_{ai}(t_A)_{ab}q_{bi} - \sum_{abi} \overline{q}^*_{ai}(t_A)_{ab}\overline{q}_{bi} = 0\,, \tag{29.3.25}$$

for all generators t_A of $SU(N_c)$. The gauge interactions (but not the superpotential) are invariant under simultaneous $SU(N_c)$ transformations on the color indices a of both q_{ai} and $\overline{q}_{ai}$; under independent $SU(N_f)$ and $\overline{SU(N_f)}$ transformations on the flavor indices i of q_{ai} and $\overline{q}_{ai}$, respectively, and under a $U(1)$ transformation of both q_{ai} and $\overline{q}_{ai}$ by opposite phases. Using these symmetries, it is possible to put the general solution of these conditions in the form

$$q_{ai} = \overline{q}_{ai} = \begin{cases} u_i\delta_{ai} & a \le N_f \\ 0 & a > N_f \end{cases}\,, \tag{29.3.26}$$

where the u_i are complex numbers with the same phase. (Here is the proof. The $SU(N_c)$ generators t_A span the space of all traceless Hermitian matrices, so Eq. (29.3.25) is equivalent to the requirement that

$$\sum_{i} q^*_{ai}q_{bi} - \sum_{i} \overline{q}^*_{ai}\overline{q}_{bi} = k\delta_{ab}\,, \tag{29.3.27}$$

for some constant k. By a combined color and flavor transformation $q \to UqV$ with U and V unitary and unimodular, we can put the matrix q in the diagonal form (29.3.26), and by a unimodular change of phase of the diagonal elements we can arrange that they all have the same phase. Then Eq. (29.3.27) becomes

$$\sum_{i} \overline{q}^*_{ai}\overline{q}_{bi} = \begin{cases} \left(u_a^2 - k\right)\delta_{ab} & a \le N_f \\ -k\delta_{ab} & a > N_f \end{cases}\,.$$

The conditions for $a > N_f$ show that $k \le 0$. If k were non-zero, then the $\overline{q}_{ai}$ would furnish N_c non-zero orthogonal vectors with N_f components, which is impossible for $N_f < N_c$, so $k = 0$. We can then put the $\overline{q}_{ai}$ in the diagonal form (29.3.26) by a series of unitary flavor transformations: first rotate $\overline{q}_{1i}$ into the 1-direction; then keeping the 1-direction fixed, rotate in the space perpendicular to this direction to put $\overline{q}_{2i}$ in the 2-direction; and so on; and then perform a unimodular phase transformation so that all diagonal elements have the same phase. Eq. (29.3.27) then shows that the absolute values of the diagonal elements of q_{ai} and $\overline{q}_{ai}$ are equal, and by a non-anomalous opposite phase change of q_{ai} and $\overline{q}_{ai}$ we can arrange that their common phases are equal, as was to be proved.)

The function H in Eq. (29.3.23) is here

$$H(q,\overline{q}) = \mathscr{I}\left[\mathrm{Det}_{ij}\sum_a q_{ai}\overline{q}_{aj}\right]^{-1/(N_c-N_f)} = \mathscr{I}\left[\prod_i u_i\right]^{-2/(N_c-N_f)}\,, \tag{29.3.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

where $\mathscr{I}$ is a purely numerical constant. (Detailed calculations show that $\mathscr{I} = 2$ for $N_f = 2$ and $N_c = 3$.) Adding the terms (29.3.23) and (29.3.24), the complete effective superpotential is now

$$f_{\text{total}}(q, \bar{q}) = \mathscr{K}\left[\prod_i u_i\right]^{-2/(N_c - N_f)} + \sum_i m_i u_i^2 , \tag{29.3.29}$$

where

$$\mathscr{K} \equiv \mathscr{I} \exp\left(\frac{i\theta}{N_c - N_f}\right) \Lambda^{(3N_c - N_f)/(N_c - N_f)} \tag{29.3.30}$$

and the m_i are the diagonal elements of the mass matrix that results when the original mass matrix is subjected to the $SU(N_f) \times \overline{SU(N_f)}$ transformation used to put the scalars in the form (29.3.26). The condition (27.4.10), that $f_{\text{total}}(q, \bar{q})$ should be stationary, has the solution

$$u_i^2 = \frac{1}{m_i}\left(\frac{\mathscr{K}}{N_c - N_f}\right)^{1 - N_f/N_c}\left(\prod_j m_j\right)^{-1/N_c} . \tag{29.3.31}$$

Because we have put the scalar fields in a basis in which the u_i have a common phase, the m_i must also have a common phase in this basis. But the common phase of the u_i^2 is not unique — the $1/N_c$ powers in Eq. (29.3.31) tell us that the solution is undetermined by a factor $\exp(2i\pi n/N_c)$, with n an integer ranging from 0 to $N_c - 1$. (The two signs of u_i for a given u_i^2 are physically equivalent, because the whole theory is invariant under a non-anomalous symmetry with $q_{ai} \to e^{i\pi} q_{ai}$ and $\bar{q}_{ai} \to e^{-i\pi}\bar{q}_{ai}$.) The fact that there are N_c physically inequivalent solutions will turn up again in our discussion of the Witten index in the next section.

$C_1 = C_2$

This case is of some interest because, as we saw in Eq. (27.9.3), the simplest $N = 2$ supersymmetric Yang–Mills theory, when written in terms of $N = 1$ superfields, contains a single left-chiral superfield in the adjoint representation, for which** of course $C_2 = C_1$.

For $C_2 = C_1$ the function $\exp(2i\pi T)$ has $R = 0$, so that its appearance in $\mathscr{L}_\lambda^\sharp$ is not restricted by R invariance. The general form of the F-term

** Note that here C_2 refers to the representation of the gauge group furnished by the chiral superfields, so it is the same as the quantity C_2^b in Section 27.9, which refers to the representation furnished by complex scalars, but half of C_2^f, which refers to the representation furnished by all spinor fields, including the gauginos.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

in $\mathscr{L}_\lambda^\sharp$ is given here by Eq. (29.3.9), but with the last term absent:

$$\mathscr{B}_\lambda = Y f_\lambda(\Phi, \exp(2i\pi T)) + \sum_{\alpha\beta AB} \epsilon_{\alpha\beta} W_{A\alpha L} W_{B\beta B} \ell_{\lambda AB}(\Phi, \exp(2i\pi T)) \,. \tag{29.3.32}$$

Because f_λ may depend on T, we cannot now conclude that it is equal to the bare superpotential, but only that it depends linearly on whatever coupling coefficients and masses appear linearly in the bare superpotential. In particular, *if there is no superpotential to begin with, then none is generated by non-perturbative effects.*

To go further, we need to make use of the anomalous chiral symmetry under a $U(1)$ transformation of all the Φ_n. In order for the whole theory to be invariant under this symmetry, it is necessary to introduce a separate external left-chiral superfield Y_r for the terms in the bare superpotential of order r in the Φ_n. Then the theory is invariant under the combined transformations

$$\Phi_n \to e^{i\varphi}\Phi_n \,, \qquad T \to T + C_2\varphi/\pi \,, \qquad Y_r \to e^{-ir\varphi} Y_r \,. \tag{29.3.33}$$

This symmetry tells us that a term in $\mathscr{B}_\lambda$ that is of order $\mathscr{N}_r$ in the coefficients of the term in the superpotential of order r in Φ and proportional to $\exp(2ai\pi T)$ must be of an order $\mathscr{N}_\Phi$ in Φ, given by

$$\mathscr{N}_\Phi = \sum_r r\mathscr{N}_r - 2C_2 a \,. \tag{29.3.34}$$

The coefficients $\ell_{\lambda AB}$ of the terms in $\mathscr{B}_\lambda$ that are quadratic in W are shown in Eq. (29.3.32) to be independent of the parameters in the superpotential for $C_1 = C_2$, so in this case Eq. (29.3.34) becomes

$$\mathscr{N}_\Phi = -2C_2 a \,. \tag{29.3.35}$$

Thus there can be no terms in $\ell_{\lambda AB}$ of positive order in the Φ_r, and any term in $\ell_{\lambda AB}$ that is independent of the Φ_r must be independent of T. These Φ-independent terms in $\ell_{\lambda AB}$ are therefore again just the one-loop contribution to the running coupling parameter τ_λ.

The effective superpotential is shown by Eq. (29.3.32) to be linear in the parameters in the superpotential, so all of its terms have just one $\mathscr{N}_r = 1$ and the others zero. For such a term Eq. (29.3.34) gives

$$\mathscr{N}_\Phi = r - 2C_2 a \,. \tag{29.3.36}$$

A term in the effective superpotential with $\mathscr{N}_\Phi$ powers of Φ can therefore only arise from terms in the bare superpotential with $r \geq \mathscr{N}_\Phi$ powers of Φ. The terms with $r = \mathscr{N}_\Phi$ have $a = 0$, so they are given by the tree approximation as just the bare superpotential. The only other terms are non-perturbative corrections with $r > \mathscr{N}_\Phi$. Such a non-perturbative

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

term, of a given order in the Φs, can only arise from terms in the bare superpotential of *higher* order in the Φs.

$$C_1 < C_2$$

Here the R value $2(C_1 - C_2)$ of $\exp(2i\pi T)$ is *negative*, so positive powers of $\exp(2i\pi T)$ can compensate for the positive R values of Y and W_α, and $\mathscr{B}_\lambda$ may therefore contain terms of arbitrary order in Y and W_α. Using the chiral symmetry condition (29.3.34) and the R invariance condition

$$2 = \mathscr{N}_W + 2\sum_r \mathscr{N}_r - 2a(C_2 - C_1)\,, \tag{29.3.37}$$

we can, however, set limits on the structure of terms of a given order $\mathscr{N}_\Phi$ in the Φs. Eqs. (29.3.34) and (29.3.37) have a trivial solution with $\mathscr{N}_r = 1$ for $r = \mathscr{N}_\Phi$, $\mathscr{N}_r = 0$ for other values of r, and $a = \mathscr{N}_W = 0$; this solution just represents the presence in $\mathscr{B}_\lambda$ of the original bare superpotential, with no radiative corrections. If there is no superpotential to begin with, then $\mathscr{N}_r = 0$ for all r, so Eq. (29.3.35) does not allow any terms in the Wilsonian Lagrangian with $\mathscr{N}_W = 0$, and so *no superpotential can be generated.* (In the supersymmetric version of quantum chromodynamics this conclusion is usually derived by noting that there is no possible term in the superpotential that would be consistent with all symmetries, but as we have now seen, the conclusion is much more general.) For renormalizable asymptotically free theories there is a useful limit on the structure of Φ-independent terms in $\mathscr{B}_\lambda$. The condition of renormalizability tells us that $\mathscr{N}_r = 0$ for $r > 3$, so by subtracting 2/3 of Eq. (29.3.34) from Eq. (29.3.37) we find

$$2 \geq \frac{2}{3}\mathscr{N}_\Phi + \mathscr{N}_W + 2a\left(C_1 - \frac{1}{3}C_2\right)\,. \tag{29.3.38}$$

Asymptotic freedom requires that $3C_1 > C_2$, so for $\mathscr{N}_\Phi = 0$ (or $\mathscr{N}_\Phi > 0$) each term on the right-hand side is positive. It follows that we can have no Φ-independent terms of higher than second order in W, and these terms are also independent of T, so they again represent the one-loop contribution to the running coupling parameter τ_λ. But for $C_1 < C_2$ there is no general prohibition against terms of second or higher order in W and negative order in Φ.

29.4 Supersymmetry Breaking in Gauge Theories

We now turn to a question of great physical interest: in what gauge theories is supersymmetry spontaneously broken?

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Let us start with an Abelian gauge theory, the supersymmetric version of quantum electrodynamics described in Section 27.5. This is a $U(1)$ gauge theory with two chiral superfields $\Phi_\pm$ carrying $U(1)$ quantum numbers $\pm e$, and superpotential $f(\Phi) = m\Phi_+\Phi_-$. We saw in Section 27.5 that supersymmetry is broken in the tree approximation if we include a Fayet–Iliopoulos term $\xi[V]_D$ in the Lagrangian density, so the Witten index is zero for $\xi \neq 0$ and small e, and hence also for all values of e and ξ, including $\xi = 0$. Is supersymmetry unbroken for $\xi = 0$? It is unbroken in the tree approximation, but how can we tell if higher-order corrections or non-perturbative effects give the vacuum a finite energy in this case?

To answer this we shall use a symmetry principle of the theory for $\xi = 0$, in the way that was described in general terms in Section 29.1. The symmetry here is charge conjugation: the whole Lagrangian density is invariant under the charge conjugation transformation of the chiral and gauge superfields:

$$\mathsf{C}\Phi_\pm\mathsf{C}^{-1} = \Phi_\mp\,, \qquad \mathsf{C}V\mathsf{C}^{-1} = -V\,. \tag{29.4.1}$$

There is a massless fermion in the tree approximation, the photino, but if we take the vacuum to be even under C then the one-photino state is odd under C, so these states are not related to each other by multiplication with the supersymmetry generator. The chiral fermion here has mass m in the tree approximation, so for small e it is not paired with the vacuum either. With no fermionic state available to pair with, the vacuum state must have strictly zero energy, at least for e small enough so that perturbation theory gives a good qualitative picture of the spectrum. As we saw in Section 29.1, the zero energy of the vacuum implies that supersymmetry is not broken. Likewise, the photino must be strictly massless, since it has no bosonic state with which to pair.

Now, what about values of e that are so large that perturbation theory cannot be trusted at all? The Witten index itself is no help here, because it vanishes. Instead, let us consider the weighted Witten index, $\mathrm{Tr}\,\mathsf{C}(-1)^F$. We have seen that for small e and $\xi = 0$ the vacuum has zero energy, and there are also two zero-energy states containing a zero-momentum photino of spin up or down, and a bosonic zero-energy state containing two zero-momentum photinos of opposite spin. The vacuum makes a contribution $+1$ to $\mathrm{Tr}\,\mathsf{C}(-1)^F$; the two one-photino states make a contribution $+2$ (because both C and $(-1)^F$ are -1); and the two-photino state makes a contribution $+1$, giving a weighted Witten index $\mathrm{Tr}\,\mathsf{C}(-1)^F = 4$. This is independent of the value of e, so even for strong couplings the weighted Witten index is 4, and so supersymmetry is not broken.

There is a complication here.[2] In counting zero-momentum states in the tree approximation, we have not considered the zero-momentum components of the gauge field $V_\mu(x)$. A constant term in $V_0(x)$ is no problem,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

because it can be removed by a gauge transformation

$$V_\mu(x) \to V_\mu(x) + \partial_\mu \Lambda(x) \,, \tag{29.4.2}$$

with gauge parameter $\Lambda(x)$ proportional to x^0. On the other hand, we cannot simply remove a constant term in $V_i(x)$, because this would require a gauge transformation with $\Lambda(x)$ proportional to x^i, which would conflict with the assumed periodicity of the fields under translations by the box dimension $L \equiv V^{1/3}$. In the particular model under consideration here all fields have charges $\pm e$ or zero, so this periodicity is preserved if we limit ourselves to a lattice of gauge transformations with

$$\Lambda(x) = \frac{2\pi}{eL} \sum_i \ell_i x^i \,, \tag{29.4.3}$$

where ℓ_i are three positive or negative integers. Therefore although we cannot remove the zero-momentum components of $V_i(x)$ by a gauge transformation, we can freely shift them by amounts $2\pi\ell_i/eL$. The tree-approximation Lagrangian (*not* Lagrangian density) for the x^i-independent part of V_μ in a gauge with $V_0 = 0$ is simply $-\frac{1}{2}L^3 \sum_i (\partial_0 V_i)^2$, so the Hamiltonian is $+\frac{1}{2}L^{-3} \sum_i (\pi_i)^2$, where π_i is the canonical conjugate to V_i: $\pi_i = L^3 \partial_0 V_i$. The wave function $\Psi(\mathbf{V})$ for this field is then just like that of a free particle of unit mass in three dimensions, in a box of linear dimensions $2\pi/eL$ with periodic boundary conditions. The wave functions of definite energy are proportional to $\exp(i\mathbf{k} \cdot \mathbf{V})$, with energy $\mathbf{k}^2/2L^3$, and $k_i = eL\ell_i$, with ℓ_i integers. There is a unique zero-energy state of this field, with $k_i = 0$ and with a normalized wave function equal to the constant $(eL/2\pi)^{3/2}$. Because this state is unique, our counting of zero-energy states is unaffected by the gauge degree of freedom, and the weighted Witten index is indeed 4.

Now let us consider the theory of a simple non-Abelian gauge superfield, without chiral superfields. Witten's 1982 paper[2] on the Witten index presented an argument that for such theories this index is $r+1$ (or possibly $-r-1$), where r is the rank of the gauge group, the maximum number of commuting generators. In 1997 he found a correction to this calculation,[6] with the result that for the classical unitary, orthogonal, and symplectic groups the index is a Casimir invariant C_1, which for the unitary and symplectic groups is indeed equal to $r+1$, but for the orthogonal groups $O(N)$ with $N > 7$ and the exceptional groups takes a different value. In general, this Casimir invariant is defined by Eq. (17.5.33):

$$\sum_{CD} C_{ACD} C_{BCD} = g^2 C_1 \delta_{AB} \,, \tag{29.4.4}$$

where g is a coupling constant, whose definition can be made unambiguous by specifying that when the generators t_A, t_B, and t_C are restricted to the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

three generators of the 'standard' $SU(2)$ subalgebra that was used in the calculation of instanton effects in Section 23.5, the structure constant is simply $g\epsilon_{ABC}$. For the classical groups, we have

$$C_1 = \begin{cases} N & SU(N) \\ N-2 & SO(N) \text{ for } N > 3 \\ N+1 & USp(2N) \end{cases} . \tag{29.4.5}$$

For $SU(N)$ the rank is $r = N - 1$, while for $USp(2N)$ it is $r = N$, so in both cases $r + 1 = C_1$. But $SO(N)$ with $N > 6$ has rank $r = (N-1)/2 = (C_1+1)/2$ for N odd and $r = N/2 = (C_1+2)/2$ for N even, and the index is C_1, not $r + 1$. Of course, this does not affect the main conclusion that, since the Witten index is not zero, supersymmetry is not spontaneously broken. Kac and Smilga showed in 1999 that the index is also equal to C_1 for exceptional groups.[7] We shall calculate the Witten index here only for $SU(N)$ and $USp(2N)$ supersymmetric gauge theories with no chiral superfields, but in the course of this calculation we will also see why the orthogonal and exceptional groups present special difficulties.

The general strategy of this calculation is the same as for the Abelian theories considered previously. We first examine the states of zero energy, to see if there are any that cannot be paired by action of the supersymmetry generator. If there are, then as long as the coupling is weak enough so that the tree approximation gives a good qualitative picture of the spectrum, we know that these unpaired states really have precisely zero energy. We can then find some non-zero weighted Witten index, which will be constant even for stronger couplings, and conclude that supersymmetry is not broken for any coupling strength.

There is another complication here, that did not enter in the earlier example of an Abelian theory with charged chiral superfields. As usually formulated, general renormalizable theories of gauge bosons and gauginos with no chiral superfields do not contain any dimensionless parameters at all, so that there is no coupling parameter that can be adjusted to make the couplings weak. Instead we have a running coupling constant, depending on the ratio of the energy to a characteristic energy scale Λ, like the scale $\Lambda \approx 200$ MeV for quantum chromodynamics discussed in Section 18.7. For non-Abelian theories the gauge coupling constant becomes strong at energies below Λ, however weak it may be at higher energies. But here we are working in a box of volume L^3, which provides an infrared cut-off at an energy $\approx 1/L$, that is normally lacking in theories with unbroken gauge symmetries. When we speak of a weakly coupled gauge theory, it should be understood that we mean one with a coupling that is small down to energies of order $1/L$. In this case, it is essential to be able to argue that supersymmetry is broken for strong as well as weak couplings,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

in order to be able to draw any conclusions at all about the realistic limit of infinite volume.

We will work in temporal gauge, where $V_A^0 = 0$. The Lagrangian density (27.3.1) is then

$$\mathscr{L} = -\tfrac{1}{4}\sum_{Aij} f_{Aij}^2 - \tfrac{1}{2}\sum_A (\partial_0 V_{Ai})^2 - \tfrac{1}{2}\sum_A \left(\bar{\lambda}_A (\not{D}\lambda)_A\right) + \tfrac{1}{2}\sum_A D_A^2 , \quad (29.4.6)$$

where, in temporal gauge,

$$f_{Aij} = \partial_i V_{Aj} - \partial_j V_{Ai} + \sum_{BC} C_{ABC} V_{Bi} V_{Cj} , \quad (29.4.7)$$

$$(D_i\lambda)_A = \partial_i \lambda_A + \sum_{BC} C_{ABC} V_{Bi} \lambda_C , \quad (29.4.8)$$

$$(D_0\lambda)_A = \partial_0 \lambda_A . \quad (29.4.9)$$

(As usual, the gauge coupling constant or constants are included as factors in the structure constants C_{ABC}.) Without chiral superfields, there is no other dependence on the auxiliary field D_A; since it enters quadratically, it may be put equal to the value at which the Lagrangian is stationary, $D_A = 0$, and ignored from now on. Retaining only the **x**-independent modes, the effective Lagrangian becomes

$$\int d^3x\, \mathscr{L} = L^3\left[-\frac{1}{4}\sum_{Aij}\left(\sum_{BC} C_{ABC} V_{Bi} V_{Cj}\right)^2 - \frac{1}{2}\sum_{Ai}(\partial_0 V_{Ai})^2 \right.$$
$$\left. -\frac{1}{2}\sum_{ABCi} C_{ABC}\left(\bar{\lambda}_A \gamma_i V_{Bi} \lambda_C\right) - \tfrac{1}{2}\sum_A \bar{\lambda}_A \gamma^0 \partial_0 \lambda_A \right] . \quad (29.4.10)$$

The Hamiltonian is then

$$H = \frac{1}{2L^3}\sum_{Ai}\pi_{Ai}^2 + \frac{L^3}{4}\sum_{Aij}\left(\sum_{BC} C_{ABC} V_{Bi} V_{Cj}\right)^2 + \frac{L^3}{2}\sum_{ABCi} C_{ABC}\left(\bar{\lambda}_A \gamma_i V_{Bi} \lambda_C\right) , \quad (29.4.11)$$

where $\pi_{Ai} = L^3 \partial_0 V_{Ai}$ is the canonical conjugate to V_{Ai}.

The gauge field configurations with zero energy in the tree approximation are those for which $\sum_{BC} C_{ABC} V_{Bi} V_{Cj} = 0$ for all A, i, and j. This condition is always satisfied if V_{Bi} vanishes for all i except where t_B is in a Cartan subalgebra of the gauge Lie algebra.* For the unitary and

* A *Cartan subalgebra* is any subalgebra spanned by r-independent generators $t_{\mathscr{A}}$ that commute with one another, that is, for which $C_{\mathscr{A}\mathscr{B}A}$ vanishes for all t_A when $t_{\mathscr{A}}$ and $t_{\mathscr{B}}$ are in the Cartan subalgebra, where r is the rank, the maximum number of such generators. For instance, for the $SU(3)$ symmetry of strong interactions discussed in Section 19.7, the rank is $r = 2$, and the Cartan subalgebra can be taken to consist of the third component of isospin t_3 and the hypercharge t_8, which act on the light quarks with the commuting matrices denoted λ_3 and λ_8 in Eq. (19.7.2).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

symplectic groups and their direct products this is the only way that it can be satisfied. The same is true for the orthogonal gauge algebras $O(N)$ with $N \leq 6$, which are all equivalent to symplectic and/or unitary Lie algebras (see the Appendix to Chapter 15), but not for orthogonal gauge algebras $O(N)$ with $N \geq 7$, which is why Witten's original calculation needed correction for this case.

In the rest of this calculation we will only consider gauge algebras like the symplectic and unitary algebras for which the condition of zero energy in the tree approximation does require the V_{Bi} to vanish for all i except where t_B is in a Cartan subalgebra of the gauge Lie algebra. With all $V_{Ai} = 0$ except for the $V_{\mathscr{A}i}$ for which $t_{\mathscr{A}}$ is in the Cartan subalgebra, the zero-energy modes of the fermion field are those also with $\lambda_A = 0$, except for the $\lambda_{\mathscr{A}}$ for which $t_{\mathscr{A}}$ is in the Cartan subalgebra.

Now we must count these states. The eigenvalues of the $t_{\mathscr{A}}$ in any representation of a semi-simple Lie algebra are quantized, so by gauge transformations all values of the non-zero gauge fields $V_{\mathscr{A}i}$ are equivalent to values in a finite box with periodic boundary conditions. The quantization of these modes is just like that carried out for the gauge field in the $U(1)$ model considered above, so the zero-energy state of these fields is again unique, with a constant wave function in the box.

The multiplicity of states with zero energy in the tree approximation comes entirely from the fermion degrees of freedom. It is convenient to use a two-component notation, in which instead of a four-component Majorana field $\lambda_{\mathscr{A}\alpha}$ for each generator of the Cartan subalgebra, we have two left-handed fields $\lambda_{\mathscr{A}La}$ with $a = \pm 1/2$, and their right-handed Hermitian adjoints $\lambda^*_{\mathscr{A}La}$. These gaugino fields satisfy the canonical anticommutation relations

$$\left\{\lambda_{\mathscr{A}La}, \lambda^*_{\mathscr{B}Lb}\right\} = \delta_{\mathscr{A}\mathscr{B}}\delta_{ab}$$

and

$$\left\{\lambda_{\mathscr{A}La}, \lambda_{\mathscr{B}Lb}\right\} = \left\{\lambda^*_{\mathscr{A}La}, \lambda^*_{\mathscr{B}Lb}\right\} = 0\,.$$

By operating on an arbitrary state vector with as many factors of $\lambda_{\mathscr{A}La}$ as necessary, we can construct a state vector $|0\rangle$ which is annihilated by all $\lambda_{\mathscr{A}La}$. The general zero-energy state vector is then a linear combination of products of the $\lambda^*_{\mathscr{A}La}$ acting on $|0\rangle$.

To see which of these states may be paired with each other by the action of the supersymmetry generator, we must take account of a symmetry of the theory. The zero-energy condition, that the gauge and gaugino fields lie only in directions corresponding to the Cartan subalgebra, is invariant under the subgroup of the original gauge group consisting of elements h that leaves this subalgebra invariant, that is, for which $h^{-1}t_{\mathscr{A}}h$ is a linear combination of the $t_{\mathscr{B}}$. These form a finite group, known as the *Weyl*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

group. For instance, in the defining representation of $SU(N)$ the Weyl group consists of permutations of the N coordinate axes, together with multiplication by a phase needed to make the transformations unimodular. These can be represented by products of the finite gauge transformations $W(i,j) = \exp(i\pi\sigma(ij)/2) = i\sigma(ij)$ with $i \neq j$ that permute the ith and jth coordinate axes, where $\sigma(ij)$ is the $U(N)$ generator with $[\sigma(ij)]^i{}_j = [\sigma(ij)]^j{}_i = 1$, and with all other elements zero. These transformations induce orthogonal transformations in the space spanned by the diagonal traceless Hermitian matrices $t_{\mathscr{A}}$ that generate the Cartan subalgebra in the adjoint representation. For instance, for the group $SU(2)$ we can take the Cartan subalgebra to consist of just t_3, and the Weyl group then consists of the unit element and a single non-trivial gauge transformation $W(1,2) = i\sigma(1,2) = i\sigma_1$, for which $W^{-1}t_3W = -t_3$. For $SU(3)$ the Cartan subalgebra has the two generators λ_3 and λ_8, and the Weyl group consists of the six gauge transformations 1, $W(1,2)$, $W(2,3)$, $W(1,3)$, $W(1,2)W(2,3)$, and $W(2,3)W(1,2)$, which generate rotations by multiples of 60° in the space spanned by t_3 and t_8.

Assuming the vacuum to be invariant under the Weyl group, it can be paired by action of the supersymmetry generators only with other states that are invariant under the Weyl group.** This may or may not include the previously constructed zero-energy state $|0\rangle$. The condition that the state $|0\rangle$ be annihilated by all $\lambda_{\mathscr{A}La}$ is obviously invariant under the Weyl group, so if it is unique this state must furnish a one-dimensional representation of the Weyl group. The Weyl group always acts on the generators of the Cartan subalgebra by orthogonal transformations, so there are two such representations: the invariant representation, in which each Weyl transformation is represented by unity, and the pseudoinvariant representation, in which each Weyl transformation is represented by the determinant of its action on the generators of the Cartan subalgebra.

Let us first consider the case where $|0\rangle$ is invariant under the Weyl group. Obviously no linear combination of the one-fermion states $\lambda^*_{\mathscr{A}La}|0\rangle$ can be invariant under the Weyl group. There is just one Weyl-invariant linear combination of the two-fermion states; it is of the form $U|0\rangle$, where

$$U \equiv \sum_{ab\mathscr{A}} e_{ab}\,\lambda^*_{\mathscr{A}La}\,\lambda^*_{\mathscr{A}Lb}\,. \tag{29.4.12}$$

(The spin indices a and b are contracted with the antisymmetric tensor e_{ab} defined by Eq. (25.2.9), because the anticommutation relations make the

** Witten remarked that physical states are necessarily Weyl-invariant, but we will not need to go into this here, because the Weyl invariance of the vacuum and of the supersymmetry generators means that only Weyl-invariant states are relevant to spontaneous supersymmetry breaking.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

product $\lambda^*_{\mathscr{A}La}\lambda^*_{\mathscr{A}Lb}$ antisymmetric in a and b.) There are various Weyl-invariant linear combinations of products of three or more generators of the Cartan subalgebra, but the anticommutation of the $\lambda^*_{\mathscr{A}La}$ makes all of them vanish† except for powers of U. Also, in the product of more than r of the Us some of the $\lambda^*_{\mathscr{A}La}$ would have to appear twice, so $U^{r+1}=0$. We conclude then that the Weyl-invariant states are limited to the $r+1$ states

$$|0\rangle\,,\ U|0\rangle\,, U^2|0\rangle\,,\ \ldots\ , U^r|0\rangle\,. \tag{29.4.13}$$

These are all bosonic states, and there are no fermionic states with zero energy in the tree approximation with which they could be paired, so for sufficiently weak coupling these states must have precisely zero energy, and supersymmetry is unbroken. Also, the Weyl-invariant Witten index here is $r+1$, and this is independent of the coupling strength, so *supersymmetry is not spontaneously broken whatever the strength of the gauge coupling.*

In the case where the state $|0\rangle$ is pseudoinvariant, the only Weyl-invariant states are of the form

$$\sum_{\mathscr{A},\mathscr{B},\cdots}\epsilon_{\mathscr{A},\mathscr{B},\cdots}\lambda^*_{\mathscr{A}La}\lambda^*_{\mathscr{B}Lb}\cdots|0\rangle\,, \tag{29.4.14}$$

† To see this for $SU(N)$, note that each generator $t_{\mathscr{A}}$ of its Cartan subalgebra may be written as a linear combination $t_{\mathscr{A}}=\sum_i c_{\mathscr{A}i}T_i$ of generators T_i of the Cartan subalgebra of $U(N)$, with the only non-zero element of each T_i being $(T_i)^i{}_i=1$ (with indices not summed). In order for the $t_{\mathscr{A}}$ to be traceless, we must have $\sum_i c_{\mathscr{A}i}=0$. Since the fields $\lambda^*_{\mathscr{A}La}$ transform under the Weyl group like the $t_{\mathscr{A}}$, it follows that for a function $\sum_{\mathscr{A}\mathscr{B}\cdots}d_{\mathscr{A}\mathscr{B}\cdots}\lambda^*_{\mathscr{A}La}\lambda^*_{\mathscr{B}Lb}\cdots$ to be Weyl-invariant, the coefficients $d_{\mathscr{A}\mathscr{B}\cdots}$ must take the form

$$d_{\mathscr{A}\mathscr{B}\cdots}=\sum_{ij\cdots}c_{\mathscr{A}i}\,c_{\mathscr{B}j}\cdots D_{ij\cdots}\,,$$

where the $D_{ij\cdots}$ are invariant tensors, in the sense that for any vectors u_i, v_j, etc., the function $D(u,v,\ldots)\equiv\sum_{ij\cdots}D_{ij\cdots}u_iv_j\ldots$ is invariant under permutations of the coordinate axes. The most general such function is a linear combination of products of the function

$$S(x,\,y,\,z,\,\ldots)=\sum_i x_iy_iz_1\cdots\,,$$

with arguments x, y, z, ... taken as various subsets of the u, v, etc. But because $\sum_i c_{\mathscr{A}i}=0$, in our case the sum of the components of each vector vanishes, so $D(u)=S(u)=0$; $D(u,v)$ is proportional to $S(u,v)$; $D(u,v,w)$ is proportional to $S(u,v,w)$; $D(u,v,w,x)$ is a linear combination of $S(u,v,w,x)$, $S(u,v)S(w,x)$, $S(u,w)S(v,x)$, and $S(u,x)S(v,w)$; and so on. The important point is that, even though the function $D(u,v,\ldots)$ may not be symmetric in its arguments (because different products of Ss may appear with different coefficients), the functions $S(u,v,\ldots)$ *are* symmetric. In our case, the vectors are the anticommuting quantities $u(a)_i=\sum_{\mathscr{A}}\lambda^*_{\mathscr{A}L,a}c_{\mathscr{A}i}$, for which the only non-vanishing S function is $S(u(1/2),u(-1/2))$. With a suitable normalization of the generators, this is the operator U.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

where $\epsilon_{\mathscr{A},\mathscr{B},\ldots}$ is the totally antisymmetric tensor of rank r. This is completely symmetric among the $a, b, \ldots$, so each state is characterized by the number of these indices that are $+1/2$ rather than $-1/2$, a number that can take any value from 0 to r, so the number of independent states is $r+1$. Depending on whether r is even or odd, these states are either all bosonic or all fermionic, so the Witten index here is $\pm(r+1)$, and again supersymmetry is not spontaneously broken.

There is an interesting relation between the value of the Witten index obtained here and ideas of how certain global symmetries become broken. The Lagrangian for a supersymmetric gauge theory is invariant under the transformations of a global $U(1)$ 'R-symmetry,' which change the left- and right-handed parts of the gaugino fields by opposite phases:

$$\lambda_{AL} \to e^{i\varphi}\lambda_{AL}\,, \qquad \lambda_{AR} \to e^{-i\varphi}\lambda_{AR}\,, \tag{29.4.15}$$

with φ an arbitrary real constant phase. The conservation of the current J_5^μ associated with this symmetry is violated by an anomaly

$$\partial_\mu J_5^\mu = -\frac{1}{32\pi^2}\sum_{ABCD} C_{ACD}\,C_{BCD}\,f_A^{\mu\nu}\,f_B^{\rho\sigma}\,\epsilon_{\mu\nu\rho\sigma}\,, \tag{29.4.16}$$

where as usual $\epsilon_{\mu\nu\rho\sigma}$ is the totally antisymmetric quantity with $\epsilon^{0123} \equiv 1$. (This is obtained from Eq. (22.2.24) by taking the gauge group generators t_A as $(t_A)_{BC} = -iC_{ABC}$, because the gauginos are in the adjoint representation of the gauge group, and by multiplying the anomaly with a factor $1/2$, because the gauginos do not have distinct antiparticles.) With the definition of the gauge coupling specified above, Eq. (23.5.20) gives the integral of the product of field strengths in the anomaly as

$$\epsilon_{\mu\nu\rho\sigma}\int d^4x \sum_A f_A^{\mu\nu} f_A^{\rho\sigma} = 64\pi^2 \nu/g^2\,, \tag{29.4.17}$$

where the 'winding number' ν is an integer characterizing the topological class to which the gauge field belongs. Putting together Eqs. (29.4.4), (29.4.16), and (29.4.17), we see that an instanton of winding number ν induces a change in $R \equiv \int d^3x\, J_5^0$ given by

$$\Delta R = \int d^4x\, \partial_\mu J_5^\mu = -2\nu C_1\,. \tag{29.4.18}$$

That is, the effective action contains terms $(\sum_{Aab}\lambda_{AL\,a}\lambda_{AL\,b}e_{ab})^{C_1}$ and its integer powers that, instead of being invariant under the R-symmetry transformations (29.4.15), are transformed by integer powers of the phase $\exp(2i\varphi C_1)$. Thus instantons invalidate invariance under the general $U(1)$ R-transformation (29.4.15), reducing it to the group Z_{2C_1} of transformations (29.4.15) with φ an integer multiple of π/C_1. We might expect that the growth of the gauge coupling at low energy would lead (as in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

quantum chromodynamics) to the appearance of vacuum expectation values for gaugino bilinears, which would mean that the discrete symmetry group Z_{2C_1} is spontaneously broken to its Z_2 subgroup, generated by a simple sign change of the gaugino multiplet. Then there would be C_1 zero-energy states $|n\rangle$, given by acting on any one vacuum state $|X\rangle$ with the elements $\exp(in\pi R/C_1)$ of Z_{2C_1}, with n running only over the values $0, 1, \ldots, C_1 - 1$ because we treat as equivalent any pair of states that differ only by action of the generator $\exp(i\pi R)$ of Z_2. Since $|X\rangle$ can only be a linear combination of states with even values of R, we can form C_1 states with all values $\mathscr{R} = 0, 2, \ldots, 2C_1 - 2$ of R by taking the linear combinations

$$\sum_n \exp\left(\frac{in\pi\mathscr{R}}{C_1}\right)|n\rangle \ .$$

In particular, for $SU(N)$ and $USp(2N)$ these states are the same as the Weyl-invariant states encountered in the calculation of the Witten index in the case where the state $|0\rangle$ is R-invariant, and Weyl-invariant rather than pseudoinvariant. The operator U defined by Eq. (29.4.12) has $R = 2$, so there are $r + 1 = C_1$ states $U^n|0\rangle$ with $R = 2n$ running from zero to $2r = 2C_1 - 2$. More generally, the presence of C_1 zero-energy states with the same statistics helps to explain why the Witten index is equal to $\pm C_1$ for the exceptional and orthogonal as well as the unitary and symplectic groups.

The fact that the Witten index is non-vanishing for all pure gauge supersymmetric theories means that to find examples of spontaneous supersymmetry breaking we must add chiral superfields to the theory. It does not help to add *massive* chiral superfields to the theory, since for weak coupling the introduction of massive fields does not change the menu of zero-energy states. We have already seen an example of this: the work of Affleck, Dine, and Seiberg reviewed in Section 29.3 showed that an $SU(N_c)$ gauge theory with $N_f < N_c$ left-chiral superfields Q_{ai} in the defining representation of $SU(N_c)$ and an equal number of left-chiral superfields $\overline{Q}_{ai}$ in the complex conjugate representation, with a mass term $\sum_{aij} m_{ij}\overline{Q}_{ai}Q_{aj}$, has N_c zero-energy bosonic states and no zero-energy fermionic states. (This is one case where the Witten index is not left unchanged when a mass term vanishes, because this mass term is the term in the superpotential of highest order in the superfields, so that a vanishing mass changes the behavior of the superpotential for large superfields. In fact, Eq. (29.3.31) shows that as the masses go to zero the scalar field values in the state of zero energy go to infinity.)

On the other hand, there is no difficulty in finding theories with left-chiral superfields subject to a symmetry that keeps them massless, in which supersymmetry *is* dynamically broken. For instance, consider an

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

$SU(N_c)$ gauge theory with N_f left-chiral superfields Q_{ai} and N_f left-chiral superfields $\overline{Q}_{ai}$ again in the N_c and $\overline{N}_c$ representations of the gauge group, but now also with N_f left-chiral superfields L_i that are neutral under the gauge group $SU(N_c)$. Assume a global (or weakly coupled local) $SU(N_f)$ symmetry, acting on the 'flavor' i indices of the Qs and the Ls but not the $\overline{Q}$s, which among other things forbids a mass term linking the Qs and $\overline{Q}$s. We take Q_{ai} and L_i in the representations N_f and $\overline{N}_f$ of $SU(N_f)$, respectively, while the $\overline{Q}$s are taken to be $SU(N_f)$ singlets. The only renormalizable superpotential is then of the form

$$f(Q,\overline{Q},L) = \sum_{ija} \mathscr{G}_j \overline{Q}_{aj} Q_{ai} L_i \,, \tag{29.4.19}$$

where the $\mathscr{G}_j$ are a set of coupling constants, which by an $SU(N_f)$ rotation may be chosen to have only one non-vanishing component, say the one with $j = N_f$, which can also be chosen to be positive. (This is a generalization of a model treated by Affleck, Dine, and Seiberg,[8] in which they took $N_c = 3$ and $N_f = 2$.) The gauge-neutral superfields L_i have no effect on the non-perturbative terms in the effective superpotential, so we can use the result of Section 29.3, that for $N_c > N_f$ the gauge interactions yield a total effective superpotential

$$f_{\text{total}}(Q,\overline{Q},L) = \sum_{ija} \mathscr{G}_j \overline{Q}_{aj} Q_{ai} L_i + \mathscr{K} \left[\text{Det}_{ij} \sum_a \overline{Q}_{aj} Q_{ai} \right]^{-1/(N_c - N_f)} , \tag{29.4.20}$$

with $\mathscr{K}$ a constant. In order for supersymmetry to be unbroken it is necessary (though not sufficient) that the scalar components q_{aj}, $\overline{q}_{aj}$, and ℓ_i of the chiral superfields satisfy the condition $\partial f_{\text{total}}(q,\overline{q},\ell)/\partial \ell_i = 0$, so that, for all i,

$$\sum_{ja} \mathscr{G}_j \, \overline{q}_{aj} \, q_{ai} = 0 \,. \tag{29.4.21}$$

But this tells us that the matrix $\sum_a \overline{q}_{aj} q_{ai}$ has a zero eigenvalue, and therefore has zero determinant, so this is a singular point of the superpotential (29.4.20), at which it is impossible for $\partial f_{\text{total}}/\partial q_{ai}$ or $\partial f_{\text{total}}/\partial \overline{q}_{aj}$ to vanish. Supersymmetry is therefore necessarily broken in this class of models.

For instance, Affleck, Dine, and Seiberg[8] found the non-vanishing scalar components q_{aj}, $\overline{q}_{aj}$, and ℓ_i of the superfields Φ_{ai}, $\overline{\Phi}_{ai}$, L_i at the minimum of the potential in their model with $N_c = 3$ and $N_f = 2$ to be at the values

$$\begin{aligned} q_{11} &= \overline{q}_{11} = 1.286\,(\mathscr{K}/2\mathscr{G})^{1/7} \,, \\ q_{22} &= \overline{q}_{22} = 1.249\,(\mathscr{K}/2\mathscr{G})^{1/7} \,, \\ \ell_1 &= \sqrt{q_{11}^2 - q_{22}^2} \,, \end{aligned} \tag{29.4.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

where the vacuum energy density is

$$\rho_{\rm VAC} = 3.593\,(16\mathscr{G}^{10}/\mathscr{K}^4)^{1/7}\,. \tag{29.4.23}$$

Also, $\mathscr{K} = 2\Lambda^7$ for $\theta = 0$.

The fact that supersymmetry is spontaneously broken in this case encourages us to think that it will be broken by strong gauge forces in a wide range of asymptotically free gauge theories, and thus lends legitimacy to the speculations about supersymmetry breaking in Section 28.3.

29.5 The Seiberg–Witten Solution*

It often happens that the tree-approximation potential in a supersymmetric theory will take the value zero for a continuous range of scalar field values. (For one example, see Eq. (29.1.9).) In this case the theory has a number of scalar excitations with zero mass in the tree approximation, which since supersymmetry is unbroken must be accompanied with suitable fermionic superpartners. At low energies the theory will then be described by a family of supersymmetric effective Lagrangians, whose members are parameterized by one or more *moduli*, the scalar expectation values in the underlying theory. Quantum effects in the underlying theory can modify the dependence of the effective Lagrangian on these moduli, and even alter the topology of the space of moduli.[9]

In one of the most striking accomplishments of the 1990s in supersymmetry theory, Seiberg and Witten[10] were able to calculate the exact dependence of the low-energy effective Lagrangian on a modular parameter in gauge theories with $N = 2$ supersymmetry. The ideas behind this calculation can be made apparent by running through only the simplest special case, that of an $SU(2)$ gauge theory with $N = 2$ supersymmetry and no additional matter hypermultiplets.

We saw in Section 27.9 that the Lagrangian density for this theory is given after elimination of the auxiliary fields by

$$\mathscr{L} = \frac{1}{e^2}\left[-\sum_A (D_\mu\phi)^*_A\,(D^\mu\phi)_A - \frac{1}{2}\sum_A \left(\overline{\psi_A}\,(\not{D}\psi)_A\right)\right.$$
$$\left. -2\sqrt{2}\,{\rm Re}\sum_{ABC}\epsilon_{ABC}\left(\lambda^{\rm T}_{AL}\,\epsilon\,\psi_{CL}\right)\phi^*_B - \frac{1}{4}\sum_A f_{A\mu\nu}f^{\mu\nu}_A\right.$$

* This section lies somewhat out of the book's main line of development and may be omitted in a first reading.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

$$-\frac{1}{2}\sum_A \left(\overline{\lambda_A}(\not{D}\lambda)_A\right)\Bigg] + \frac{\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_A f_A^{\mu\nu} f_A^{\rho\sigma} - V(\phi,\phi^*)\,, \tag{29.5.1}$$

where A, B, and C now run over the values 1, 2, and 3. We have now rescaled all the fields by multiplying each of them by a factor e, so that e does not appear in the covariant derivatives:

$$(D_\mu\psi)_A = \partial_\mu\psi_A + \sum_{BC}\epsilon_{ABC}V_{B\mu}\psi_C\,, \tag{29.5.2}$$

$$(D_\mu\lambda)_A = \partial_\mu\lambda_A + \sum_{BC}\epsilon_{ABC}V_{B\mu}\lambda_C\,, \tag{29.5.3}$$

$$(D_\mu\phi)_A = \partial_\mu\phi_A + \sum_{BC}\epsilon_{ABC}V_{B\mu}\phi_C\,, \tag{29.5.4}$$

$$f_{A\mu\nu} = \partial_\mu V_{A\nu} - \partial_\nu V_{A\mu} + \sum_{BC}\epsilon_{ABC}V_{B\mu}V_{C\nu}\,, \tag{29.5.5}$$

and the potential is

$$V(\phi,\phi^*) = 2\sum_A\left[\sum_{BC}\epsilon_{ABC}\,\mathrm{Re}\,\phi_B\,\mathrm{Im}\,\phi_C\right]^2. \tag{29.5.6}$$

This potential takes the value zero for a family of scalar field expectation values, which (up to a gauge transformation) may be parameterized as

$$\phi_1 = \phi_2 = 0\,, \qquad \phi_3 = a\,, \tag{29.5.7}$$

with a a complex parameter, known as the vacuum modulus. This vacuum expectation value gives masses $2|a|$ to the vector fields $V_{1\mu}$ and $V_{2\mu}$, the gauginos λ_1 and λ_2, the chiral fermions ψ_1 and ψ_2, and the scalars ϕ_1 and ϕ_2, leaving $V_{3\mu}$, λ_3, ψ_3, and $\phi_3 = a$ all massless.

Taking account only of these zero-mass modes (and dropping the subscript 3), the tree approximation gives the effective low-energy theory as simply the free-field theory with Lagrangian density

$$\mathscr{L}_{\text{eff}} = \frac{1}{e^2}\Bigg[-(\partial_\mu a)^*(\partial^\mu a) - \frac{1}{2}\left(\overline{\psi}(\not{\partial}\psi)\right) - \frac{1}{4}f_{\mu\nu}f^{\mu\nu} - \frac{1}{2}\left(\overline{\lambda}(\not{D}\lambda)\right)\Bigg] + \frac{\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}f^{\mu\nu}f^{\rho\sigma}\,, \tag{29.5.8}$$

with $f_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$. Indeed, any renormalizable theory of a single gauge boson and its gauge-neutral $N = 2$ superpartners must be a free-field theory, because $N = 2$ supersymmetry does not allow a superpotential for this theory.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

But this is not the whole story. In integrating out the massive degrees of freedom in this theory, quantum corrections produce non-renormalizable interaction terms in the low-energy effective field theory. We can classify the interactions and Feynman diagrams containing them that predominate at low energy by the same sort of counting of powers of energy that we used in dealing with low-momentum pions and nucleons in Section 19.5. When we use the effective Lagrangian perturbatively to calculate low-energy scattering amplitudes, the number v of powers of energy contributed by a connected graph with L loops, I_f external fermion lines, I_b internal boson (a or V_μ) lines, I_a internal auxiliary field lines, and V_i vertices of each type i, is

$$v = 4L + \sum_i V_i d_i - 2I_b - I_f \,, \tag{29.5.9}$$

where d_i is the number of derivatives in the interaction of type i. (Internal auxiliary field lines do not contribute in Eq. (29.5.9) because their propagators are independent of momentum.) These quantities are subject to the topological relations

$$L = I_b + I_f + I_a - \sum_i V_i + 1 \,, \tag{29.5.10}$$

and

$$2I_b + E_b = \sum_i V_i b_i \,, \quad 2I_f + E_f = \sum_i V_i f_i \,, \quad 2I_a + E_a = \sum_i V_i a_i \,, \tag{29.5.11}$$

where E_b, E_f, and E_a are the numbers of external boson, fermion, and auxiliary field lines, and b_i, f_i, and a_i are the numbers of boson, fermion, and auxiliary fields in interactions of type i. We can therefore write the number of powers of energy as

$$v = \sum_i V_i \left(d_i + \tfrac{1}{2} f_i + a_i - 2\right) + 2L - E_f - 2E_a + 2 \,. \tag{29.5.12}$$

According to Eqs. (26.8.4) and (27.4.42), both the D-term of a function of left-chiral scalar $N = 1$ superfields and their adjoints and the $\mathcal{F}$-term of a pair of $N = 1$ gauge superfields W_α times an arbitrary function of $N = 1$ left-chiral scalar superfields have $d_i + \frac{1}{2} f_i + a_i = 2$, while adding any additional W_α factors or superderivatives $\mathcal{D}_\alpha$ would give $d_i + \frac{1}{2} f_i + a_i > 2$, so in our case supersymmetry rules out any interactions with $d_i + \frac{1}{2} f_i + a_i$ less than 2. The dominant contribution to low-energy scattering amplitudes is thus given by the tree approximation ($L = 0$), calculated with an effective Lagrangian containing only terms with $d_i + \frac{1}{2} f_i + a_i = 2$, which takes the general form discussed in Section 27.4:

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}\left[K(\Phi, \Phi^\dagger)\right]_D - \frac{1}{2}\text{Re}\left[T(\Phi)\left(W_L^{\mathrm{T}} \epsilon W_L\right)\right]_{\mathcal{F}} \,. \tag{29.5.13}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Eq. (27.4.42) then gives the Lagrangian in terms of component fields as

$$\begin{aligned}\mathscr{L} = {} & \frac{\partial^2 K(a,a^*)}{\partial a\,\partial a^*}\left[-\frac{1}{2}\left(\overline{\psi}\,\not{\partial}\psi\right) + |\mathscr{F}|^2 - \partial_\mu a\,\partial^\mu a^*\right] \\ & - \mathrm{Re}\left\{\frac{\partial^3 K(a,a^*)}{\partial^2 a\,\partial a^*}\left(\overline{\psi}\psi_L\right)\mathscr{F}^*\right\} - \frac{1}{2}\mathrm{Re}\left\{\frac{\partial^3 K(a,a^*)}{\partial^2 a\,\partial a^*}\left(\overline{\psi}\gamma^\mu\gamma_5\psi\right)\partial_\mu a\right\} \\ & + \frac{1}{4}\frac{\partial^4 K(a,a^*)}{\partial^2 a\,\partial^2 a^*}\left(\overline{\psi}\psi_L\right)\left(\overline{\psi}\psi_R\right) \\ & + \frac{1}{4}\mathrm{Re}\left\{\left(\overline{\lambda}\lambda_L\right)\left(\overline{\psi}\psi_L\right)\frac{d^2T(a)}{da^2}\right\} - \frac{1}{2}\mathrm{Re}\left\{\left(\overline{\lambda}\lambda_L\right)\mathscr{F}\frac{dT(a)}{da}\right\} \\ & + \mathrm{Re}\left\{T(a)\Big[-\frac{1}{2}\left(\overline{\lambda}\,\not{\partial}(1-\gamma_5)\lambda\right) - \frac{1}{4}f_{\mu\nu}f^{\mu\nu} \right. \\ & \qquad \left. + \frac{1}{8}i\,\epsilon_{\mu\nu\rho\sigma}f^{\mu\nu}f^{\rho\sigma} + \frac{1}{2}D^2\Big]\right\} \\ & + \frac{\sqrt{2}}{4}\mathrm{Re}\left\{\frac{dT(a)}{da}\left[-\left(\overline{\lambda}\gamma^\mu\gamma^\nu\psi_L\right)f_{\mu\nu} + 2i\left(\overline{\lambda}\psi_L\right)D\right]\right\}. \end{aligned} \tag{29.5.14}$$

In order to implement $N = 2$ supersymmetry, we now want to impose invariance under the discrete R-symmetry transformation (27.9.2):

$$\psi \to \lambda\,, \qquad \lambda \to -\psi\,, \tag{29.5.15}$$

with a and V_μ unchanged. The condition that the coefficients of $(\overline{\psi}\,\not{\partial}\psi)$ and $(\overline{\lambda}\,\not{\partial}\lambda)$ should be equal is

$$\frac{\partial^2 K(a,a^*)}{\partial a\,\partial a^*} = \mathrm{Re}\,T(a)\,. \tag{29.5.16}$$

The right-hand side is the sum of a function of a and a function of a^*, so $\partial^4 K/\partial^2 a\,\partial^2 a^* = 0$, and therefore the term quartic in ψ, which would have had no counterpart for λ, is absent. By an integration by parts, the term $\frac{1}{2}\mathrm{Re}\{T(a)(\overline{\lambda}\,\not{\partial}\gamma_5\lambda)\}$ may be replaced with $-\frac{1}{4}\mathrm{Re}\{(\overline{\lambda}\gamma^\mu\gamma_5\lambda)\partial_\mu T(a)\}$. The condition that the coefficients of $(\overline{\psi}\gamma^\mu\gamma_5\psi)$ and $(\overline{\lambda}\gamma^\mu\gamma_5\lambda)$ should be equal is then satisfied if

$$\frac{1}{4}\partial_\mu T = \frac{1}{2}\frac{\partial^3 K}{\partial^2 a\,\partial a^*}\partial_\mu a\,,$$

which is also an automatic consequence of Eq. (29.5.16). According to Eq. (26.A.7), the terms proportional to $f_{\mu\nu}(\overline{\lambda}[\gamma^\mu,\gamma^\nu]\psi)$ and $f_{\mu\nu}(\overline{\lambda}[\gamma^\mu,\gamma^\nu]\gamma_5\psi)$ are automatically invariant under the transformation (29.5.15), as are also the terms proportional to $(\overline{\lambda}\lambda_L)(\overline{\psi}\psi_L)$ and its adjoint. On the other hand, the invariance of the terms proportional to $(\overline{\lambda}\psi)D$ and $(\overline{\lambda}\gamma_5\psi)D$ requires

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

that we extend the transformation (29.5.15) so that

$$D \to -D\,, \tag{29.5.17}$$

which also leaves the term $\frac{1}{2}D^2$ invariant. Finally, Eq. (29.5.16) tells us that the coefficients of $(\overline{\psi}\psi_L)\mathscr{F}^*$ and $(\overline{\lambda}\lambda_L)\mathscr{F}$ are equal, so the transformation (29.5.15) must also be extended so that

$$\mathscr{F} \to \mathscr{F}^*\,, \tag{29.5.18}$$

which also leaves the term $|\mathscr{F}|^2$ invariant. We conclude then that the condition (29.5.16) ensures that the whole action derived from Eq. (29.5.14) is invariant under the combined transformation (29.5.15), (29.5.17), and (29.5.18). Since the Lagrangian is invariant under $N = 1$ supersymmetry with a left-chiral scalar supermultiplet $(a, \psi, \mathscr{F})$ and a gauge supermultiplet (V_μ, λ, D), it is also invariant under a second supersymmetry with a left-chiral scalar supermultiplet $(a, \lambda, \mathscr{F}^*)$ and a gauge supermultiplet $(V_\mu, -\psi, -D)$. Eq. (29.5.16) is therefore enough to ensure that the action obtained from Eq. (29.5.13) or (29.5.14) is invariant (without imposing the field equations) under $N = 2$ supersymmetry.

The general solution of Eq. (29.5.16) can be expressed as**

$$T(a) = \frac{1}{4\pi i}\frac{dh(a)}{da}\,, \qquad K(a, a^*) = \mathrm{Im}\left\{\frac{a^* h(a)}{4\pi}\right\}\,, \tag{29.5.19}$$

with h a function of a alone. In terms of h, the Lagrangian density (29.5.14) now reads

$$\begin{aligned}
\mathscr{L} = {} & \frac{1}{4\pi}\mathrm{Im}\left\{\left[\frac{dh}{da}\right]\left[-\frac{1}{2}\left(\overline{\psi}\,\partial\!\!\!/(1-\gamma_5)\psi\right) - \frac{1}{2}\left(\overline{\lambda}\,\partial\!\!\!/(1-\gamma_5)\lambda\right)\right.\right. \\
& \left.\left. -\partial_\mu a\,\partial^\mu a^* + |\mathscr{F}|^2 + \frac{1}{2}D^2 - \frac{1}{4}f_{\mu\nu}f^{\mu\nu} + \frac{1}{8}i\,\epsilon_{\mu\nu\rho\sigma}f^{\mu\nu}f^{\rho\sigma}\right]\right\} \\
& + \frac{1}{4\pi}\mathrm{Im}\left\{\left[\frac{d^2h}{da^2}\right]\left[-\frac{1}{2}\left(\overline{\psi}\psi_L\right)\mathscr{F}^* - \frac{1}{2}\left(\overline{\lambda}\lambda_L\right)\mathscr{F} + \frac{\sqrt{2}}{2}\,i\left(\overline{\lambda}\psi_L\right)D\right.\right. \\
& \qquad \left.\left. -\frac{\sqrt{2}}{4}\left(\overline{\lambda}\gamma^\mu\gamma^\nu u\psi_L\right)f_{\mu\nu}\right]\right\} \\
& + \frac{1}{16\pi}\mathrm{Im}\left\{\frac{d^3h(a)}{da^3}\left(\overline{\lambda}\lambda_L\right)\left(\overline{\psi}\psi_L\right)\right\}\,.
\end{aligned} \tag{29.5.20}$$

This is invariant under an $SU(2)$ R-symmetry, under which (ψ, λ) transforms as a doublet and $(\mathrm{Im}\,\mathscr{F}, \mathrm{Re}\,\mathscr{F}, D/\sqrt{2})$ as a triplet. The discrete

** The factor $1/4\pi i$ is included in order to simplify the duality transformation introduced below. Seiberg and Witten introduced a function $\mathscr{F}(a)$ (not related to the auxiliary field $\mathscr{F}$), known as the *prepotential*, for which $h(a) = d\mathscr{F}(a)/da$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

transformation (29.5.15), (29.5.17), and (29.5.18) is a finite element of this $SU(2)$ group: a rotation of π radians around the 2-axis.

Comparing Eqs. (29.5.8) and (29.5.20), we see that in the tree approximation

$$h(a)_{\text{tree}} = \left[\frac{4\pi i}{e^2} + \frac{\theta}{2\pi}\right] a \,. \tag{29.5.21}$$

The achievement of Seiberg and Witten was to calculate $h(a)$ exactly.

The first step in this calculation is to recognize that there are various linear transformations on a and $dh(a)/da$ that give physically equivalent theories. This is because of a remarkable property of the low-energy effective theory, related to the duality property discussed briefly in Section 27.9. To demonstrate this property, let us return to the Lagrangian density (29.5.13) expressed in terms of $N = 1$ superfields, now using the relations (29.5.19) required by $N = 2$ supersymmetry

$$\mathscr{L}_{\text{eff}} = \frac{1}{8\pi}\text{Im}\left[\Phi^* h(\Phi)\right]_D - \frac{1}{8\pi}\text{Im}\left[h'(\Phi)\left(W_L^{\text{T}}\epsilon W_L\right)\right]_{\mathscr{F}} \,. \tag{29.5.22}$$

In path integrals the spinor field-strength superfield W_L is constrained by the supersymmetric extension (27.2.20) of the homogeneous Maxwell equations:

$$\text{Re}\left(\mathscr{D}_L^{\text{T}}\epsilon W_L\right) = 0 \,. \tag{29.5.23}$$

This is usually imposed by requiring that W_L take the form (27.2.15)

$$W_L = \frac{i}{4}\left(\mathscr{D}_R^{\text{T}}\epsilon\mathscr{D}_R\right)\mathscr{D}_L V \,, \tag{29.5.24}$$

where V is an unrestricted real superfield. Instead, we can implement the condition (29.5.23) by introducing a Lagrange multiplier term in the action

$$\Delta I_{\text{eff}} = \frac{1}{8\pi}\text{Re}\int d^4x\,\left[\tilde{V}\left(\mathscr{D}_L^{\text{T}}\epsilon W_L\right)\right]_D \,, \tag{29.5.25}$$

where $\tilde{V}$ is an unrestricted real superfield. (The numerical factor $1/8\pi$ serves to fix the normalization of $\tilde{V}$ in a way we shall find convenient later.) Then in path integrals we can integrate over both $\tilde{V}$ and W_L, with no restrictions on either except that $\tilde{V}$ is a real superfield and W_L is a left-handed spinor superfield satisfying the left-chirality condition $\mathscr{D}_{R\alpha}W_{L\beta} = 0$. Integrating by parts in superspace allows us to write the new term in the action as

$$\Delta I_{\text{eff}} = -\frac{1}{8\pi}\text{Re}\int d^4x\,\left[\left((\mathscr{D}_L\tilde{V})^{\text{T}}\epsilon W_L\right)\right]_D \,, \tag{29.5.26}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

or, using Eq. (26.3.31) and the fact that W_L is left-chiral,

$$\Delta I_{\text{eff}} = \text{Re}\left[\frac{i}{4\pi}\int d^4x\,\left[\left(\tilde{W}_L^{\text{T}}\epsilon W_L\right)\right]_{\mathscr{F}}\right], \tag{29.5.27}$$

where $\tilde{W}_L$ is defined in terms of $\tilde{V}$ by the same relation that previously gave W_L in terms of V:

$$\tilde{W}_L = \frac{i}{4}\left(\mathscr{D}_R^{\text{T}}\epsilon\mathscr{D}_R\right)\mathscr{D}_L\tilde{V}\,. \tag{29.5.28}$$

But now we are integrating over W_L with no such constraint, with an action that is quadratic in W_L

$$I_{\text{eff}} + \Delta I_{\text{eff}} = \text{Im}\int d^4x\,\left[-\frac{1}{8\pi}h'(\Phi)\left(W_L^{\text{T}}\epsilon W_L\right) - \frac{1}{4\pi}\left(\tilde{W}_L^{\text{T}}\epsilon W_L\right)\right]_{\mathscr{F}} + \frac{1}{8\pi}\text{Im}\int d^4x\,\left[\Phi^* h(\Phi)\right]_D\,. \tag{29.5.29}$$

This integration is done by setting W_L equal to the value where the action is stationary in W_L:

$$W_L = -\frac{\tilde{W}_L}{h'(\Phi)} \tag{29.5.30}$$

and the whole effective action becomes

$$\tilde{I}_{\text{eff}} = +\frac{1}{8\pi}\text{Im}\int d^4x\,\left[\frac{1}{h'(\Phi)}\left(\tilde{W}_L^{\text{T}}\epsilon\tilde{W}_L\right)\right]_{\mathscr{F}} + \frac{1}{8\pi}\text{Im}\int d^4x\,\left[\Phi^* h(\Phi)\right]_D\,. \tag{29.5.31}$$

Now, if we define a new left-chiral scalar superfield and a new h function

$$\tilde{\Phi} \equiv h(\Phi)\,, \qquad \tilde{h}(\tilde{\Phi}) \equiv -\Phi\,, \tag{29.5.32}$$

then

$$\frac{d\tilde{h}}{d\tilde{\Phi}}\frac{dh}{d\Phi} = -\frac{d\Phi}{d\tilde{\Phi}}\frac{d\tilde{\Phi}}{d\Phi} = -1\,, \tag{29.5.33}$$

so the action (29.5.31) may be written as

$$\tilde{I}_{\text{eff}} = -\frac{1}{8\pi}\text{Im}\int d^4x\,\left[\tilde{h}'(\tilde{\Phi})\left(\tilde{W}_L^{\text{T}}\epsilon\tilde{W}_L\right)\right]_{\mathscr{F}} + \frac{1}{8\pi}\text{Im}\int d^4x\,\left[\tilde{\Phi}^*\tilde{h}(\tilde{\Phi})\right]_D\,. \tag{29.5.34}$$

From the method by which we have derived it, we know that the theory based on this effective action is equivalent to the original effective field theory, so *the $N = 2$ effective field theory with scalar field value a and h function $h(a)$ is physically equivalent to one with scalar field value*[†] $a_D \equiv$

[†] The subscript D stands for 'dual,' and of course has nothing to do with the D-term of a superfield.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

h(a) and h function $\tilde{h}(a_D) \equiv -a$. This is the version of duality that applies in the present context.

There is another transformation of the function $h(\Phi)$ that also yields an equivalent Lagrangian (now with no change in Φ or W) and can be combined with the transformation (29.5.32) to give a larger group of duality transformations. Suppose we shift $h(\Phi)$ by a linear term with a real coefficient:

$$h(\Phi) \to h(\Phi) + b\,\Phi\,, \tag{29.5.35}$$

where b is a real constant. Then the first term in the effective Lagrangian density (29.5.22) is shifted by $b\,\mathrm{Im}\{[\Phi^*\Phi]_D\}/8\pi$, which vanishes since the D-term of $\Phi^*\Phi$ is real. The change in the effective Lagrangian density (29.5.22) is therefore given by the shift in the second term

$$\mathscr{L}_{\text{eff}} \to \mathscr{L}_{\text{eff}} - \frac{b}{8\pi}\mathrm{Im}\left[\left(W_L^{\mathrm{T}}\epsilon W_L\right)\right]_{\mathscr{F}}\,, \tag{29.5.36}$$

or, according to Eq. (27.2.13),

$$\mathscr{L}_{\text{eff}} \to \mathscr{L}_{\text{eff}} - \frac{b}{8\pi}\left[i\left(\bar{\lambda}\,\not{\partial}\gamma_5\lambda\right) + \frac{1}{4}\epsilon^{\mu\nu\rho\sigma}f_{\mu\nu}f_{\rho\sigma}\right]\,. \tag{29.5.37}$$

The first term in the brackets on the right-hand side is a spacetime derivative, and therefore does not affect the effective action. The second term in the brackets can also be written as a spacetime derivative $(1/2)\partial_\mu(\epsilon^{\mu\nu\rho\sigma}A_\nu f_{\rho\sigma})$ wherever $f_{\mu\nu}$ can be written as $\partial_\mu A_\nu - \partial_\nu A_\mu$. However, as discussed in Section 23.3, the gauge transformation that we used to put ϕ_A in the 3-direction must be singular somewhere, so we cannot write $f_{\mu\nu}$ everywhere in terms of a single A_μ. In consequence, physical quantities *can* be affected by a term in the action of the form

$$-\frac{\theta}{64\pi^2}\int d^4x\;\epsilon^{\mu\nu\rho\sigma}f_{\mu\nu}f_{\rho\sigma}\,. \tag{29.5.38}$$

In particular, Witten[11] has shown that, in the theory of magnetic monopoles described in Section 23.3, the electric charge of a magnetic monopole with minimum monopole moment in the presence of such a term is $e\theta/2\pi$. But, as mentioned in Section 27.9, in this theory there are also dyons, particles with both magnetic monopole moments and charges given by any integer multiple of e, so the whole pattern of monopole and dyon charges is periodic in θ with period 2π. In fact, all physical quantities have this periodicity, because the θ of the low-energy effective theory is inherited from the θ appearing in the Lagrangian density (29.5.1) of the underlying theory (note Eq. (29.5.21)), and we have seen in Section 23.5 that all physical quantities are periodic in this θ. According to Eq. (29.5.37), the transformation (29.5.35) changes θ by $2\pi b$, and therefore with b equal to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

an arbitrary positive or negative integer it yields an equivalent effective action.

This sort of transformation is closely related to an exact invariance of the underlying theory. The Lagrangian (29.5.1) is invariant under a continuous R-symmetry, under which

$$\theta_L \to \exp(i\alpha)\theta_L\,, \quad W_{AL} \to \exp(i\alpha)W_{AL}\,, \quad \Phi_A \to \exp(2i\alpha)\Phi_A\,. \tag{29.5.39}$$

This symmetry is broken by an anomaly: both λ_{AL} and ψ_{AL} have R quantum numbers $+1$, and of course λ_{AR} and ψ_{AR} have R quantum numbers -1, so according to Eqs. (23.5.21) and (23.5.23), the measure in the integral over fermion fields changes under the transformation (29.5.39) by a factor $\exp(2i\alpha N\nu)$, where ν is an integer, the winding number of the vector field configuration, and N is defined by

$$\mathrm{Tr}\,(t_A t_B) = \frac{1}{2}N\delta_{AB}\,.$$

(The factor $\exp(2i\alpha N\nu)$ is the same as given in Section 23.5, even though here we have two fields ψ_A and λ_A in the same representation of the gauge group, because these are Majorana fields, in contrast with the Dirac fields used in Section 23.5.) The generators here are $(t_A)_{BC} = -i\epsilon_{ABC}$, so $N = 4$, and therefore the measure remains invariant when $\exp(8i\alpha) = 1$. In other words, the continuous R-symmetry is broken by instantons to a $\mathbb{Z}_8$ subgroup, generated by the transformation

$$\psi_{AL} \to \sqrt{i}\,\psi_{AL}\,, \quad \lambda_{AL} \to \sqrt{i}\,\lambda_{AL}\,, \quad \phi_A \to i\,\phi_A\,. \tag{29.5.40}$$

This symmetry then must carry over to the effective low-energy theory. However, it does *not* tell us that the effective Lagrangian (29.5.20) is invariant under the discrete transformation (29.5.40), which would require that $-ih(ia) = h(a)$ for all a. This condition is satisfied by the tree approximation (29.5.21), but (as we will see) it is violated even in one-loop order. The $\mathbb{Z}_8$ R-symmetry is realized in the effective theory by the condition that the effective theory with h function $-ih(ia)$ is *equivalent* to one with h function $h(a)$. That is, $-ih(ia)$ must be related to $h(a)$ by some combination of the transformations (29.5.32) and (29.5.35), with b an integer.

We have seen that the physical significance of the theory is left unchanged by two sorts of transformation: the transformation $\Phi \to \tilde{\Phi} = h(\Phi)$ and $h(\Phi) \to \tilde{h}(\tilde{\Phi}) = -\Phi$, which can be written as

$$\begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix} \to \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix}\,, \tag{29.5.41}$$

and the transformations $h(\Phi) \to h(\Phi) + b\Phi$, $\Phi \to \Phi$, which can be written

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

as

$$\begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix} , \tag{29.5.42}$$

where b is an arbitrary integer. By combining these transformations, we can make generalized duality transformations

$$\begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix} \to \begin{pmatrix} n & m \\ k & l \end{pmatrix} \begin{pmatrix} \Phi \\ h(\Phi) \end{pmatrix} , \tag{29.5.43}$$

where n, m, k, and l are any integers satisfying the condition that the matrix in Eq. (29.5.43) should, like those in Eqs. (29.5.41) and (29.5.42), have unit determinant:

$$nl - mk = 1 . \tag{29.5.44}$$

These transformations therefore form the group $SL(2, \mathbb{Z})$.

The physical significance of the duality transformation can be brought forward by considering the central charge of the $N = 2$ supersymmetry algebra. As shown in Section 25.5, the eigenvalue Z_{12} of this central charge in any one-particle state sets a lower bound $M \geq |Z_{12}|/2$ on the mass of the state, a bound that is reached for particles that belong to 'short' supermultiplets. We saw in Section 27.9 that Z_{12} is given in the underlying $SU(2)$ gauge theory with $N = 2$ supersymmetry by Eq. (27.9.22):

$$Z_{12} = 2\sqrt{2}\, v\, [iq - \mathscr{M}] ,$$

where q and $\mathscr{M}$ are the charge and magnetic moment of the particle, and v is the vacuum expectation value of the conventionally normalized neutral scalar field. In the notation used in the present section, we have absorbed a factor e into the normalization of the field a, so here we have

$$Z_{12} = 2\sqrt{2}\, a\, [iq - \mathscr{M}]/e . \tag{29.5.45}$$

This theory contains particles like the massive elementary scalars, spinors, and vector bosons, which have charge e and zero magnetic monopole moment, so such particles are eigenstates of Z_{12} with eigenvalue

$$Z_{12}^{\text{charged elementary}} = 2\sqrt{2}\, i\, a . \tag{29.5.46}$$

As we have seen, the theory with scalar field a and h function $h(a)$ is equivalent to one with scalar field $n\,a + m\,h(a)$ and h function $k\,a + l\,h(a)$. Therefore for all integers n and m the theory with scalar field a and h function $h(a)$ must *also* contain a particle that looks like a massive particle of charge e and zero magnetic moment in the version of the theory with scalar field $na + mh(a)$ and h function $ka + lh(a)$, and that therefore has a central charge

$$Z_{12} = 2\sqrt{2}\, i\Big[n\, a + m\, h(a)\Big] . \tag{29.5.47}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Comparing this with Eq. (29.5.45), we see that, in the version of this theory with scalar field a and h function $h(a)$, this particle appears as one with a charge q and a magnetic monopole moment $\mathscr{M}$ given by

$$q/e = n + m\,\mathrm{Re}\,[h(a)/a]\,, \qquad \mathscr{M}/e = m\,\mathrm{Im}\,[h(a)/a]\,. \tag{29.5.48}$$

This is a dyon, with both charge and magnetic monopole moment. Note that the formula (29.5.48) for the charge of this particle bears out Witten's earlier result about the charge of magnetic monopoles: adding a term $b\,a$ to the function $h(a)$ changes the charge of a monopole with $m = 1$ by an amount $be = e\Delta\theta/2\pi$.

Using the tree-approximation result (29.5.21) for $h(a)$ in Eq. (29.5.48) shows that in this approximation the magnetic monopole moments in this theory are multiples of a value $4\pi/e$. This is the same as the magnetic monopole moment derived in Section 23.3 (recall that the magnetic monopole moment g used there is $\mathscr{M}/4\pi$), but this is a semi-classical result, subject to quantum corrections. Note that the duality transformation (29.5.41) with $n = 0$ and $m = 1$ takes an elementary particle with charge e and magnetic monopole moment zero into a non-elementary particle with magnetic monopole moment $\mathrm{Im}\,[h(a)/a]$, which in the tree approximation is $4\pi/e$.

The beta function for the electric charge in the underlying $N = 2$ supersymmetric $SU(2)$ gauge theory is given perturbatively by the one-loop result Eq. (27.9.50), with the first Casimir invariant taken now as $C_1 = 2e$ and the number of hypermultiplets set at $H = 0$:

$$\beta_{\text{perturbative}}(e) = -\frac{e^3}{4\pi^2}\,. \tag{29.5.49}$$

We here take a as the renormalization scale, so the running charge $e(a)$ satisfies

$$a\frac{d}{da}e(a) = \beta\Big(e(a)\Big)\,. \tag{29.5.50}$$

Using the perturbation theory formula (29.5.49), this gives

$$[e^{-2}(a)]_{\text{perturbative}} = \frac{1}{2\pi^2}\ln\left(\frac{a}{\Lambda}\right)\,, \tag{29.5.51}$$

where Λ is an integration constant. In the formalism we are adopting here, with a factor e absorbed into the definition of the gauge field, the quantity $e^{-2}(a)$ appears as the coefficient $h'(a)/4\pi i$ of the term $-\frac{1}{4}f_{\mu\nu}f^{\mu\nu}$ in the low-energy effective Lagrangian (29.5.20), so the function $h(a)$ is

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

given in perturbation theory by[††]

$$[h(a)]_{\text{perturbative}} = \frac{2i}{\pi}\left[a\ln\left(\frac{a}{\Lambda}\right) - a\right] = 4\pi i a\left[[e^{-2}(a)]_{\text{perturbative}} - \frac{1}{2\pi^2}\right] . \tag{29.5.52}$$

This is a good approximation[‡] for sufficiently large values of $|a|$, where Eq. (29.5.51) gives a small value of $e(a)$. In this case, Eq. (29.5.48) gives the magnetic monopole moment of monopoles and dyons as

$$\mathscr{M}_{\text{perturbative}}/e = 4\pi m a\left[[e^{-2}(a)]_{\text{perturbative}} - \frac{1}{2\pi^2}\right] , \tag{29.5.53}$$

where m is an arbitrary integer. Note that this is *not* the same as would be found by simply replacing e with the running charge $e(a)$ on the right-hand side of the semi-classical formula $\mathscr{M}/e = 4\pi m/e^2$.

For sufficiently large values of $|a|$, Eq. (29.5.52) also satisfies a necessary consistency condition: the coefficient $\text{Im}\,h'(a)/4\pi$ of the kinematic terms $-\frac{1}{2}(\bar{\psi}\,\partial\!\!\!/\psi)$, $-\frac{1}{2}(\bar{\lambda}\,\partial\!\!\!/\lambda)$, $-\partial_\mu a^* \partial^\mu a$ and $-\frac{1}{4}f_{\mu\nu}f^{\mu\nu}$ must be *positive*. The same condition tells us also that Eq. (29.5.52) cannot be a good approximation for all values of a, because it gives a negative value for $\text{Im}\,h'(a)$ for $|a|$ sufficiently small.

Eq. (29.5.52) shows that if a is carried counterclockwise around a circle with a large value of $|a|$, where perturbation theory is valid, then $h(a)$ is shifted by $-4a$. This tells us that $h(a)$ must have one or more singularities at finite values of a.

We can easily rule out the possibility that $h(a)$ has just one singularity, because this would make it impossible to satisfy the condition that $h'(a)$ should have a positive imaginary part for all non-singular values of a, which as we have seen is necessary for the positivity of the coefficients of the kinematic terms in the effective Lagrangian. A single singularity would have to be at $a = 0$, because the $\mathbb{Z}_8$-symmetry tells us that if $h(a)$ is singular at a then it is also singular at ia. The function $h(a) - h_{\text{perturbative}}(a)$ would then be analytic except at infinity and perhaps at $a = 0$, but this function vanishes as $|a| \to \infty$, and in particular does not change when carried around a circle at large a, so it could at most be a polynomial in $1/a$, with no constant term. If the polynomial does not vanish, then $h'(a)$ goes as some negative power of a as $a \to 0$, which would not

[††] The additive constant needed in integrating $h'(a)$ has no effect on the low-energy effective action derived from Eq. (29.5.22), because $[\Phi^*]_D$ is a derivative. This constant can be fixed by reference to the $\mathbb{Z}_8$-symmetry discussed earlier. Eq. (29.5.52) satisfies $-ih(ia) = h(a) - 1$, which is a special case with $b = -1$ of the transformations (29.5.35) that leave the physical significance of the theory unchanged. This would not be the case if we had added a constant to Eq. (29.5.52).

[‡] There are non-perturbative contributions[12] to $\beta(e)$, arising from instantons, but these vanish rapidly for $e \to 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

have a positive imaginary part, while if the polynomial did vanish then $h'(a)$ would be equal to $h'_{\text{perturbative}}(a)$, which also does not have positive imaginary part for $a \to 0$.

In studying the singularity structure of $h(a)$, it is helpful to keep track of duality by treating a and $a_D = h(a)$ in the same way, expressing them both as functions of some complex variable u. Seiberg and Witten took u to be the expectation value of the gauge-invariant quantity $\frac{1}{2}\sum_A \phi_A \phi_A$. This changes sign under the $\mathbb{Z}_8$ transformations $\phi_A \to \pm i\phi_A$, and of course under this transformation $a \equiv \phi_3 \to \pm ia$, so

$$a(-u) = \pm i\, a(u)\,. \tag{29.5.54}$$

The sign here has no physical significance because a and $-a$ are related by a finite $SU(2)$ gauge transformation. (Where necessary it can be made definite by adopting the convention that the upper and lower signs apply if $\mathrm{Re}\, u > 0$ or $\mathrm{Re}\, u < 0$, respectively.) We can trust perturbation theory for large values of $|a|$, so for $|u| \to \infty$

$$a \to \sqrt{2u}\,, \qquad a_D \to \frac{i}{\pi}\left[\sqrt{2u}\,\ln\left(\frac{2u}{\Lambda^2}\right) - 2\sqrt{2u}\right]\,. \tag{29.5.55}$$

Note that when u is taken counterclockwise around a circle at a fixed large value of $|u|$, the logarithm $\ln(2u/\Lambda^2)$ is shifted by $2i\pi$, while $\sqrt{2u}$ changes sign, so the changes in a and a_D are given by a *monodromy matrix*:

$$\begin{pmatrix} a \\ a_D \end{pmatrix} \to \begin{pmatrix} -1 & 0 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a \\ a_D \end{pmatrix}\,. \tag{29.5.56}$$

The functions $a(u)$ and $a_D(u)$ must therefore have two or more singularities at finite values u_n of u, such that the combined effect of going counterclockwise around each singularity is the same as (29.5.56).

Let us consider the possibility that there are just two singularities. (This was shown to be the case in the second Seiberg–Witten paper of Reference 10.) Under the $\mathbb{Z}_8$-symmetry that takes $a_A \to ia_A$ we have $u \to -u$, so the singularities must be at a pair of u values u_0 and $-u_0$. Going from a non-singular base point P in the u plane counterclockwise around the singularity at $\pm u_0$ all the way back to P should yield an equivalent theory. Therefore it must take the form of a duality transformation, in general depending on P, in which the vector (a, a_D) is multiplied with an $SL(2,\mathbb{Z})$ *monodromy matrix* $M_\pm$, as in Eq. (29.5.43). The counterclockwise contour around the circle with fixed large u can be deformed into a contour that starts at P and goes counterclockwise back to P around $-u_0$ and then counterclockwise back to P around $+u_0$. (See Figure 29.1.) Since this deformation cannot change the integral, the product of the monodromy matrices *in this order* (reading from right to left) must be equal to the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

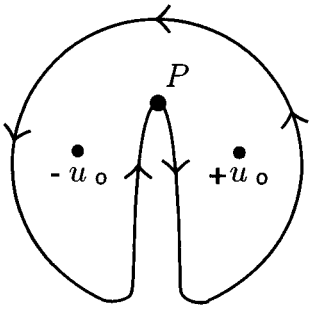

Figure 29.1. Deformation of a contour that circles the complex u plane counterclockwise at large $|u|$ into a contour that starts at a base point P, circles the singularity at $u = -u_0$ counterclockwise back to P, and then circles the singularity at $u = +u_0$ counterclockwise back to P.

matrix in Eq. (29.5.56):

$$M_+ M_- = M_\infty \equiv \begin{pmatrix} -1 & 0 \\ 2 & -1 \end{pmatrix} . \tag{29.5.57}$$

Singularities occur when a and a_D take values at which some particle has zero mass. For instance, the singularity at $a = 0$ in the perturbative formula (29.5.52) for $h(a)$ occurs because this is the value of a at which elementary charged particles become massless in perturbation theory. We have already ruled out the possibility of a single singularity at $a = 0$, so the singularities at $\pm u_0$ must arise from the vanishing of the mass of some other particles.

The most striking part of the Seiberg–Witten calculation was their realization that these particles are among the non-elementary magnetic monopoles or dyons found in the underlying $SU(2)$ supersymmetric theory. Semi-classical calculations of the sort done in Section 23.3 show that the stable monopoles and dyons have magnetic quantum number $m = \pm 1$ and any integer electric quantum number n, and belong to hypermultiplets, each consisting of a pair of Majorana spinors and a pair of complex scalars. These are 'short' multiplets, with a mass given by the BPS value, which according to Eqs. (27.9.24) and (29.5.47) is

$$M = |Z_{12}|/2 = \sqrt{2} \left| Na + h(a) \right| , \tag{29.5.58}$$

where $N \equiv \pm n$. The easiest way to calculate what happens when this mass goes to zero is to consider the more familiar problem of what happens when the mass of a hypermultiplet of ordinary charged particles goes to zero, and then use duality to switch to the case of a light monopole. The

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

beta function for the $U(1)$ gauge coupling is given here by Eq. (27.9.45) with $C_1 = 0$, $C_2^f = C_2^s = 2$, so

$$\beta(e) = +\frac{e^3}{8\pi^2} .$$

The solution of the renormalization group equation is then

$$e^{-2}(a) = -\frac{1}{4\pi^2} \ln\left(\frac{a}{\text{constant}}\right) .$$

As we saw in deriving Eq. (29.5.52), this gives

$$h'(a) = 4\pi i e^{-2}(a) = -\frac{i}{\pi} \ln\left(\frac{a}{\text{constant}}\right) .$$

Because this gives a value of $e^2(a)$ that is positive and small for $a \to 0$, this *would* be a reliable result if the theory really contained a hypermultiplet of ordinary charged particles whose mass goes to zero in this limit. Instead, we are assuming that there is a hypermultiplet of monopoles or dyons whose mass goes to zero; to deal with this case, we can apply the duality transformation (29.5.43) that takes a into $Na + h(a)$

$$\begin{pmatrix} a \\ h(a) \end{pmatrix} \to \begin{pmatrix} \hat{a} \\ \hat{h}(\hat{a}) \end{pmatrix} = \begin{pmatrix} N & +1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a \\ h(a) \end{pmatrix} . \tag{29.5.59}$$

We conclude that when u approaches a point u_0 where $\hat{a} = Na + h(a) \to 0$, we have

$$\frac{d\hat{h}(\hat{a})}{d\hat{a}} \to -\frac{i}{\pi} \ln\left(\frac{\hat{a}}{\text{constant}}\right) , \tag{29.5.60}$$

or, in other words,

$$\frac{da}{d\,(h(a) + Na)} \to +\frac{i}{\pi} \ln\left(\frac{h(a) + Na}{\text{constant}}\right) . \tag{29.5.61}$$

The solution is

$$a(u) = a_0 + \frac{i}{\pi}\Big(h(u) + Na(u)\Big) \ln\left(\frac{h(u) + Na(u)}{\Lambda_0}\right) , \tag{29.5.62}$$

where a_0 and Λ_0 are integration constants. We are also assuming that $h + Na \to 0$ for $u \to u_0$, so we can write the leading term as

$$h(u) + Na(u) \to c_0(u - u_0) . \tag{29.5.63}$$

Thus Eq. (29.5.62) has the leading term

$$a(u) \to a_0 + \frac{i\,c_0}{\pi}(u - u_0) \ln\left(\frac{c_0(u - u_0)}{\Lambda_0}\right) . \tag{29.5.64}$$

When we take u counterclockwise all the way around a circle surrounding u_0 there is no change in $h(a) + Na$, but a is shifted by $-2(h(a) + Na)$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

so the monodromy matrix for this singularity is

$$M_+ = \begin{pmatrix} 1-2N & -2 \\ 2N^2 & 1+2N \end{pmatrix} . \tag{29.5.65}$$

Likewise, if the singularity at $-u_0$ is associated with the vanishing mass of a monopole or dyon with magnetic quantum number $\pm'1$ (the prime distinguishing this sign from that for u_0) and electric quantum number n', then at this singularity $h(u) + N'a(u) \to 0$, and

$$a \to a'_0 + \frac{i}{\pi}\Big(h(u) + N'a(u)\Big) \ln\left(\frac{h(u) + N'a(u)}{\Lambda'_0}\right) , \tag{29.5.66}$$

where $N' \equiv \pm' n'$ and a'_0 and Λ'_0 are new integration constants. The leading terms are

$$h(u) + N'a(u) \to c'_0(u + u_0) , \tag{29.5.67}$$

$$a \to a'_0 + \frac{i\,c'_0}{\pi}(u + u_0) \ln\left(\frac{c'_0(u + u_0)}{\Lambda'_0}\right) . \tag{29.5.68}$$

The monodromy matrix for this singularity is

$$M_- = \begin{pmatrix} 1-2N' & -2 \\ 2N'^2 & 1+2N' \end{pmatrix} . \tag{29.5.69}$$

It is straightforward then to see that the condition (29.5.57) on these matrices is satisfied if and only if

$$N' = N - 1 . \tag{29.5.70}$$

It makes no difference what value we take for N, because we can shift it by an even integer $2M$ by going around the circle at infinity M times, and we can shift it by unity by reflecting $u \leftrightarrow -u$. Seiberg and Witten chose to take $N = 0$, so that $N' = -1$. Then for $u \to u_0$

$$h(u) \to c_0(u - u_0) , \tag{29.5.71}$$

$$a(u) \to a_0 + \frac{i\,c_0}{\pi}(u - u_0) \ln\left(\frac{u - u_0}{\Lambda_0}\right) , \tag{29.5.72}$$

and for $u \to -u_0$

$$h(u) - a(u) \to c'_0(u + u_0) , \tag{29.5.73}$$

$$a(u) \to a'_0 + \frac{i\,c'_0}{\pi}(u + u_0) \ln\left(\frac{u + u_0}{\Lambda'_0}\right) . \tag{29.5.74}$$

We now impose the unbroken $\mathbb{Z}_8$-symmetry condition (29.5.54). We can calculate $a(u)$ for $u \to -u_0$ by writing it as $-ia(-u)$ and using Eq. (29.5.72).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Then for $u \to -u_0$

$$a(u) \to -ia_0 + \frac{c_0}{\pi} \ln\left(\frac{-c_0(u+u_0)}{\Lambda_0}\right) . \tag{29.5.75}$$

Comparing this with Eq. (29.5.74), we see that $c_0' = i c_0$, so Eq. (29.5.73) becomes, for $u \to -u_0$,

$$h(u) - a(u) \to i c_0 (u + u_0) . \tag{29.5.76}$$

The field a is defined to include a factor of the gauge coupling e, which can be given any value by an appropriate choice of the renormalization point at which e is evaluated. Seiberg and Witten chose to define the scale of a and u (keeping $u = a^2/2$ at infinity) so that $u_0 = 1$; that is, the singularities are at $u = \pm 1$. With this convention, they obtained the solutions (defined in the complex plane cut from -1 to $+1$)

$$a_{\rm SW}(u) = \frac{\sqrt{2}}{\pi} \int_{-1}^{1} dx \sqrt{\frac{u-x}{1-x^2}} , \tag{29.5.77}$$

$$h_{\rm SW}(u) = \frac{i\sqrt{2}}{\pi} \int_{1}^{u} dx \sqrt{\frac{u-x}{x^2-1}} . \tag{29.5.78}$$

The mathematical methods originally used to obtain these results are beyond the scope of this book, but fortunately it is not difficult after the fact to check that they are correct.

First, we can check that $a_{\rm SW}(u)$ and $h_{\rm SW}(u)$ have a singular behavior near $u = \pm 1$ of the same form (29.5.71), (29.5.72), (29.5.75), and (29.5.76) as the true solution. For $u \to 1$, Eq. (29.5.77) gives

$$\begin{aligned} a_{\rm SW}(u) &\to \frac{\sqrt{2}}{\pi} \int_{-1}^{1} \frac{dx}{\sqrt{x+1}} + \frac{u-1}{2\pi} \int_{-1}^{1} \frac{dx}{\sqrt{(1-x)(u-x)}} \\ &= \frac{4}{\sqrt{\pi}} + \frac{u-1}{2\pi} \ln\left(\frac{u-1}{3+u-2\sqrt{2}\sqrt{1+u}}\right) \\ &\to \frac{4}{\sqrt{\pi}} - \frac{u-1}{2\pi} \ln\big(4(u-1)\big) . \end{aligned} \tag{29.5.79}$$

Also, Eq. (29.5.78) gives

$$h_{\rm SW}(u) \to \frac{i}{\pi} \int_{1}^{u} dx \sqrt{\frac{u-x}{x-1}} = \frac{i(u-1)}{2} . \tag{29.5.80}$$

Eqs. (29.5.79) and (29.5.80) agree with our previous results (29.5.71) and (29.5.72), with $c_0 = i$ and $a_0 = 4/\sqrt{\pi}$. Eq. (29.5.77) satisfies the $\mathbb{Z}_8$ reflection property (29.5.54), so $a_{\rm SW}(u)$ automatically has the behavior (29.5.75) for $u \to -1$, and (with due regard to the signs of square roots dictated by the cut from $u = -1$ to $u = +1$) it follows directly from

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

Eqs. (29.5.77) and (29.5.78) that $h_{SW}(u) - a_{SW}(u)$ is proportional to $u+1$ for u near -1.

Now since $a_{SW}(u)$ and $h_{SW}(u)$ have the same singularity structure at $u \to \pm 1$ as $a(u)$ and $h(u)$, they have the same monodromy (Eqs. (29.5.59) and (29.5.69) with $N = 0$ and $N' = -1$): as u circles the point $+1$, we have

$$\begin{pmatrix} a_{SW} \\ h_{SW} \end{pmatrix} \to \begin{pmatrix} 0 & +1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a_{SW} \\ h_{SW} \end{pmatrix}, \tag{29.5.81}$$

while as u circles the point -1,

$$\begin{pmatrix} a_{SW} \\ h_{SW} \end{pmatrix} \to \begin{pmatrix} 3 & -2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a_{SW} \\ h_{SW} \end{pmatrix}. \tag{29.5.82}$$

Let us now consider the quantity

$$f(u) \equiv a(u)h_{SW}(u) - a_{SW}(u)h(u). \tag{29.5.83}$$

This is an $SL(2,\mathbb{Z})$ invariant, so it has trivial monodromy: it returns to the same value when u circles either $+1$ or -1. The only finite singularities in $a(u)$, $h(u)$, $a_{SW}(u)$, or $h_{SW}(u)$ are the logarithmic singularities at ± 1, but since $f(u)$ has trivial monodromy it does not have these singularities, and it is therefore analytic at all finite points.

To evaluate the entire function $f(u)$, let's first check that the leading terms in the asymptotic behavior of the functions $a_{SW}(u)$ and $h_{SW}(u)$ are the same as those of $a(u)$ and $h(u)$, respectively. For $u \to \infty$, Eq. (29.5.77) gives

$$a_{SW}(u) \to \frac{\sqrt{2u}}{\pi} \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} = \sqrt{2u}, \tag{29.5.84}$$

while Eq. (29.5.78) gives

$$h_{SW}(u) \to \frac{i\sqrt{2u}}{\pi} \int_{1}^{u} \frac{dx}{\sqrt{x^2-1}} \to \frac{i\sqrt{2}}{\pi} \sqrt{u} \ln u. \tag{29.5.85}$$

This is the same as the leading behavior (29.5.55) of $a(u)$ and $h(u)$.

This in itself only shows that $f(u)/u \ln u \to 0$ for $u \to \infty$. But note that the reflection symmetry (29.5.54) for $a(u)$ and its counterpart for $a_{SW}(u)$ tell us that the next-to-leading terms in both functions are of order $\sqrt{u}/u^2$. Also, the next-to-leading terms in $h(u)$ and $h_{SW}(u)$ are of order $\sqrt{u}$. It follows that we can calculate the leading term in the asymptotic behavior of $f(u)$ by setting $a(u) = a_{SW}(u) = \sqrt{2u}$, so that $f(u) = O(u)$ for $u \to \infty$. Since $f(u)$ is an entire function, this means that $f(u)$ is linear in u. But $f(u)$ vanishes at $u = +1$, where $h(u)$ and $h_{SW}(u)$ are $O(u-1)$, and $f(u)$ also vanishes at $u = -1$, where $a(u) - h(u)$ and $a_{SW}(u) - h_{SW}(u)$ are $O(u-1)$,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

so it must vanish everywhere. We conclude then that

$$\frac{a_{SW}(u)}{a(u)} = \frac{h_{SW}(u)}{h(u)} \equiv g(u) \,. \tag{29.5.86}$$

Now we must consider the properties of the function $g(u)$. Because $h_{SW}(u)$ and $h(u)$ are analytic at all finite points except $u = -1$, and $a_{SW}(u) - h_{SW}(u)$ and $a(u) - h(u)$ are analytic at all finite points except $u = +1$, it follows that $g(u)$ (which can also be written as $[a_{SW}(u) - h_{SW}(u)]/[a(u) - h(u)]$) is analytic everywhere. (There are no zeroes of $a(u)$ or $h(u)$ at $u \neq \pm 1$, let alone zeroes of both, because a zero of $a(u)$ or $h(u)$ would be associated with a point where a charged particle or monopole mass vanishes, which has been assumed to be not the case for any $u \neq \pm 1$.) Also, the fact that the leading terms in the asymptotic behavior for $u \to \infty$ of the functions $a_{SW}(u)$ and $h_{SW}(u)$ are the same as those of $a(u)$ and $h(u)$, respectively, means that $g(u) \to 1$ for $u \to \infty$. It follows then that the entire function $g(u)$ must equal unity for all u, and therefore

$$a(u) = a_{SW}(u) \,, \qquad h(u) = h_{SW}(u) \,, \tag{29.5.87}$$

as was to be shown.

Problems

1. What is the Witten index of the model in Problem 4 of Chapter 26 for $a \neq 0$? For $a = 0$? Can supersymmetry be broken by higher-order effects in this model for $a = 0$? Explain.

2. Consider the renormalizable supersymmetric theory with $SO(N_c)$ gauge symmetry and N_f left-chiral scalar superfields Φ_n in the N-vector representation. What can you say about the structure of the non-perturbative Wilsonian Lagrangian density when the bare superpotential vanishes? What if there is a bare superpotential $\sum_n \Phi_n \Phi_n$?

3. In the theory with Lagrangian density (29.5.22), what is the relation between the component fields of the spinor field strength superfield W_α and its dual $\tilde{W}_\alpha$?

References

1. E. Witten, *Nucl. Phys.* **B185**, 513 (1981). This article is reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

1*a*. J. Hughes and J. Polchinski, *Nucl. Phys.* **B278**, 147 (1986); J. Hughes, J. Liu, and J. Polchinski, *Phys. Lett.* **B 180**, 370 (1986); J. Bagger and A. Galperin, *Phys. Lett.* **B336**, 25 (1994); *Phys. Rev.* **D55**, 1091 (1997); *Phys. Lett.* **B412**, 296 (1997); L. Antoniadis, H. Partouche, and T. R. Taylor, *Phys. Lett.* **B372**, 83 (1996); S. Ferrara, L. Girardello, and M. Porrati, *Phys. Lett.* **B376**, 275 (1996).

2. E. Witten, *Nucl. Phys.* **B202**, 253 (1982). This article is reprinted in *Supersymmetry*, Ref. 1.

3. A. C. Davis, M. Dine, and N. Seiberg, *Phys. Lett.* **125B**, 487 (1983); I. Affleck, M. Dine, and N. Seiberg, *Phys. Rev. Lett.* **51**, 1026 (1983); *Phys. Lett.* **137B**, 187 (1984), reprinted in *Supersymmetry*, Ref. 1; *Nucl. Phys.* **B241**, 493 (1984); *Nucl. Phys.* **B52**, 1677 (1984); *Phys. Lett.* **140B**, 59 (1984).

4. N. Seiberg, *Phys. Lett.* **B318**, 469 (1993).

5. V. A. Novikov, M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Nucl. Phys.* **B229**, 381 (1983); *Nucl. Phys.* **B260**, 157 (1985); G. C. Rossi and G. Veneziano, *Phys. Lett.* **138B**, 195 (1984), reprinted in *Supersymmetry*, Ref. 1; S. F. Cordes, *Nucl. Phys.* **B273**, 629 (1986). For a review, see M. A. Shifman and A. I. Vainshtein, hep-th/9902018, to be published.

6. E. Witten, *J. High Energy Phys.* **9802**, 006 (1998).

7. V. G. Kac and A. V. Smilga, hep-th/9902029, to be published.

8. I. Affleck, M. Dine, and N. Seiberg, *Nucl. Phys.* **B256**, 557 (1985).

9. N. Seiberg, *Phys. Rev.* **D49**, 6857 (1994).

10. N. Seiberg and E. Witten, *Nucl. Phys.* **B426**, 19 (1994); *erratum*, *Nucl. Phys.* **B430**, 485 (1994). The extension to an $N = 2$ theory with matter hypermultiplets is given by N. Seiberg and E. Witten, *Nucl. Phys.* **B431**, 484 (1994). For a review, see: K. Intriligator and N. Seiberg, *Nucl. Phys. Proc. Suppl.* **45BC**, 1 (1996); also published in *Summer School in High Energy Physics and Cosmology, Trieste, 1995*, E. Gava, ed. (World Scientific, Singapore, 1997), and in *QCD and Beyond: Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder Colorado, 1995*, D. E. Soper, ed. (World Scientific, Singapore, 1996).

11. E. Witten, *Phys. Lett.* **86B**, 283 (1979).

12. N. Seiberg, *Phys. Lett.* **B206**, 75 (1988).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.008

30

Supergraphs

The introduction of Feynman graphs in the late 1940s provided the enormous advantage of preserving manifest Lorentz invariance at every stage of perturbative calculations. For this purpose it was necessary that the exchange of all spin states of any virtual particle should be described by a single propagator. Fortunately, it has become possible to go a step further, and develop a supergraph formalism, in which supersymmetry as well as Lorentz invariance is kept manifest at every stage.[1] To do so, it is necessary to describe the exchange of all the particles described by a given superfield by a single superpropagator.

There is a problem here in that the left-chiral superfields Φ over which we integrate are subject to a differential constraint, $\mathscr{D}_R\Phi = 0$. This is analogous to the problem in electrodynamics that the field-strength tensor is subject to a differential constraint, the homogeneous Maxwell equation. In electrodynamics this problem is dealt with by expressing the field-strength tensor in terms of a vector potential, and doing path integrals over the vector potential instead of the field strength tensor. In much the same way, here we will impose the left-chiral constraint by expressing left-chiral superfields in terms of potential superfields, over which we integrate in path integrals. In this formalism we encounter problems like those produced by gauge invariance in electrodynamics, and deal with them in much the same way.

The most important result to come out of the supergraph formalism was the non-renormalization theorem for the superpotential.[2] We have already proved this theorem in Section 27.6, using a much easier indirect technique developed by Seiberg, which can be extended also to describe non-perturbative effects. Nevertheless, it is interesting to see how these surprising cancellations of renormalization effects occur in actual perturbative calculations.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

30.1 Potential Superfields

Consider a theory of left-chiral superfields $\Phi_n(x,\theta)$ and their complex conjugates, but for simplicity without gauge superfields. All of the vacuum expectation values of time-ordered products of component fields may be calculated from vacuum expectation values of the time-ordered products of these superfields. We might try to calculate these from the path-integral formula

$$\left\langle T\left\{\Phi_{n_1}(x_1,\theta_1),\,\Phi_{n_2}(x_2,\theta_2),\,\cdots\right\}\right\rangle = \int\left[\prod_{n,x,\theta} d\Phi_n(x,\theta)\right]\exp\left(iI[\Phi]\right) \times \Phi_{n_1}(x_1,\theta_1)\,\Phi_{n_2}(x_2,\theta_2)\cdots\,, \tag{30.1.1}$$

where $I[\Phi]$ is the action

$$I[\Phi] = \frac{1}{2}\int d^4x\Big[\sum_n \Phi_n^*(x,\theta)\Phi_n(x,\theta)\Big]_D + 2\,\mathrm{Re}\int d^4x\Big[f(\Phi)\Big]_{\mathscr{F}}\,. \tag{30.1.2}$$

(The factor $1/2$ is introduced in the first term, as in Eq. (26.4.3), to give the component fields of Φ a conventional normalization.) But we *cannot* simply read off the supergraph Feynman rules from Eq. (30.1.1), because the functional integral over the superfields Φ_n must be constrained to satisfy the left-chiral condition $\mathscr{D}_R\Phi_n = 0$.

This is analogous to a similar problem in electrodynamics. As discussed in Section 12.3, at energies below the electron mass the interactions of soft photons with each other are described by an effective action of the form:

$$I[f] = -\frac{1}{4}\int d^4x\left[f_{\mu\nu}f^{\mu\nu} + c_1\left(f_{\mu\nu}f^{\mu\nu}\right)^2 + c_2\left(\epsilon_{\mu\nu\rho\sigma}f^{\mu\nu}f^{\rho\sigma}\right)^2\right].$$

But we cannot read off the Feynman rules from this formula without taking into account the fact that the path integral is constrained by the homogeneous Maxwell equations

$$\partial_\mu f_{\nu\rho} + \partial_\nu f_{\rho\mu} + \partial_\rho f_{\mu\nu} = 0\,.$$

As everyone knows, we deal with this constraint by introducing a four-vector potential A_μ, with $f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, so that the constraint is automatically satisfied, and integrate over $A_\mu(x)$, not $f_{\mu\nu}(x)$.

In the same way, we may adopt a trick already used in Section 26.6 to derive the field equations for superfields, and introduce a non-chiral potential superfield $S_n(x,\theta)$, with

$$\Phi_n = \mathscr{D}_R^2 S_n\,, \tag{30.1.3}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

where

$$\mathscr{D}_R^2 \equiv \sum_{\alpha\beta} \epsilon_{\alpha\beta} \mathscr{D}_{R\alpha} \mathscr{D}_{R\beta} \,, \tag{30.1.4}$$

so that Φ_n automatically satisfies the left-chiral constraint, $\mathscr{D}_{R\alpha}\Phi_n = 0$. In place of Eq. (30.1.1), we then have the path-integral formula

$$\left\langle T\left\{\Phi_{n_1}(x_1,\theta_1),\, \Phi_{n_2}(x_2,\theta_2), \cdots\right\}\right\rangle = \int \left[\prod_{n,x,\theta} dS_n(x,\theta)\right] \exp\left(i I[\mathscr{D}_R^2 S]\right) \times \mathscr{D}_R^2 S_{n_1}(x_1,\theta_1)\, \mathscr{D}_R^2 S_{n_2}(x_2,\theta_2) \cdots \,. \tag{30.1.5}$$

In expressing the action (30.1.2) in terms of S_n, we may recall that the D-term of a superderivative does not contribute to the action, so we may shift the operator $(\mathscr{D}_R^2)^* = \mathscr{D}_L^2$ acting on S_n^* in the first term of Eq. (30.1.2) to act instead on S_n, which gives

$$I[\mathscr{D}_R^2 S] = \frac{1}{2}\int d^4x \left[\sum_n S_n^* \mathscr{D}_L^2 \mathscr{D}_R^2 S_n\right]_D + 2\,\mathrm{Re}\int d^4x \left[f(\mathscr{D}_R^2 S)\right]_{\mathscr{F}} \,. \tag{30.1.6}$$

The $\mathscr{D}_R^2$ operator in any one of the factors $\mathscr{D}_R^2 S$ in any term in $f(\mathscr{D}_R^2 S)$ may be taken to act on the whole term, since $\mathscr{D}_R$ gives zero when acting on any other factor of $\mathscr{D}_R^2 S$ in that term. In this way, we may write

$$f(\mathscr{D}_R^2 S) = \mathscr{D}_R^2 \tilde{f}(S) \,, \tag{30.1.7}$$

where $\tilde{f}(S)$ is obtained from $f(\mathscr{D}_R^2 S)$ by omitting any *one* operator $\mathscr{D}_R^2$ in each term. For instance, for a single type of superfield, if $f(\Phi) = \sum_r c_r \Phi^r$, then

$$\tilde{f}(S) = \sum_r c_r S\,(\mathscr{D}_R^2 S)^{r-1} \,.$$

Using Eqs. (30.1.6), (30.1.7), and (26.3.31), we may put the whole action in the form of a D-term

$$I[\mathscr{D}_R^2 S] = \frac{1}{2}\int d^4x \sum_n \left[S_n^* \mathscr{D}_L^2 \mathscr{D}_R^2 S_n\right]_D + 2\,\mathrm{Re}\int d^4x \left[\tilde{f}(S)\right]_D \,. \tag{30.1.8}$$

Eq. (26.6.5) shows that this may also be written as a superspace integral:

$$I[\mathscr{D}_R^2 S] = -\frac{1}{4}\int d^4x \int d^4\theta \sum_n S_n^* \mathscr{D}_L^2 \mathscr{D}_R^2 S_n - \mathrm{Re}\int d^4x \int d^4\theta\, \tilde{f}(S) \,. \tag{30.1.9}$$

This is the action we will use in the path-integral derivation of the supergraph formalism.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

30.2 Superpropagators

Usually the propagator can be obtained directly from the part of the action of second order in the fields. If we write this quadratic part for a complex field ϕ_i (where i is a compound index including spacetime coordinates as well as spin and species labels) in the form

$$I_{\text{quad}}[\phi] = -\sum_{ij} D_{ij}\phi_i^*\phi_j \,, \tag{30.2.1}$$

with D_{ij} Hermitian, then as explained in Section 9.4 the propagator is simply $\Delta = D^{-1}$. The problem with this arises when the quadratic part of the action is invariant under a (linearized) gauge transformation

$$\phi_i \to \phi_i + \xi_i \,, \tag{30.2.2}$$

for some class of 'vectors' ξ. In this case we have

$$\sum_i D_{ij}\xi_j = 0 \,, \tag{30.2.3}$$

and we clearly cannot invert the 'matrix' D_{ij}. In electrodynamics the problem arises because the Lagrangian density is invariant under a gauge transformation $A_\mu \to A_\mu + \partial_\mu\Lambda$. We have this problem here too: since the action is actually a functional of $\mathscr{D}_R^2 S_n$ and not of S_n itself, it is invariant under the transformation

$$S_n \to S_n + \mathscr{D}_R X_n \,, \tag{30.2.4}$$

for any superfields X_n.

In the electrodynamics of charged particles the problems introduced by gauge invariance are typically handled by choosing a gauge, for instance by the Faddeev–Popov–de Witt method described in Section 15.5. But the problems raised here by invariance under the transformation (30.2.2) are much more like the problems of the effective field theory of photons at energies below the threshold for producing charged particles, where the theory is gauge-invariant simply because the action only involves gauge-invariant fields. In such theories there is a simpler option. In addition to the eigenvector ξ_i of D_{ij} with eigenvalue zero (taken for simplicity to be in a unique direction) we can find a set of orthonormal eigenvectors u_{vi} with eigenvalues $d_v \neq 0$:

$$\sum_j D_{ij}\,u_{vj} = d_v\,u_{vi} \,, \qquad \sum_i u_{vi}^*\,u_{v'i} = \delta_{vv'} \,, \qquad \sum_i u_{vi}^*\,\xi_i = 0 \,. \tag{30.2.5}$$

We can introduce a new set of integration variables ϕ' and ϕ'_v:

$$\phi_i = \phi'\,\xi_i + \sum_v \phi'_v\,u_{vi} \,. \tag{30.2.6}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

The sort of integral encountered in Feynman diagram calculations of quantum expectation values may then be written

$$\int \left[\prod_i d\phi_i\, d\phi_i^*\right] \exp\{iI_{\text{quad}}[\phi]\}\phi_a \cdots \phi_b^* \cdots = \mathscr{J} \int d\phi'\, d\phi'^* \times \int \left[\prod_v d\phi_v'\, d\phi_v'^*\right] \exp\left\{-i\sum_v d_v|\phi_v'|^2\right\} \left[\phi'\xi_a + \sum_v \phi_v' u_{va}\right] \cdots \times \left[\phi'\xi_b + \sum_v \phi_v' u_{vb}\right]^* \cdots , \tag{30.2.7}$$

where $\mathscr{J}$ is the Jacobian of the transformation (30.2.6). The integral over ϕ' and ϕ'^* is of course ill defined, because they do not appear in the argument of the exponential. But this doesn't matter if the action only involves gauge-invariant fields, because then ϕ_a, ϕ_b, etc. will be contracted with 'currents' J_a, J_b, etc. for which

$$\sum_a \xi_a J_a = 0 . \tag{30.2.8}$$

We may thus write Eq. (30.2.7) as

$$\int \left[\prod_i d\phi_i\, d\phi_i^*\right] \exp\{iI_{\text{quad}}[\phi]\}\phi_a \cdots \phi_b^* \cdots = \mathscr{C} \int \left[\prod_v d\phi_v' d\phi_v'^*\right] \times \exp\left\{-i\sum_v d_v|\phi_v'|^2\right\} \left[\sum_v \phi_v' u_{va}\right] \cdots \left[\sum_v \phi_v' u_{vb}\right]^* \cdots + \xi\text{-terms} , \tag{30.2.9}$$

where 'ξ-terms' denotes terms proportional to one or more factors of ξ_a, ξ_b, etc., which vanish when contracted with Js satisfying Eq. (30.2.8), and $\mathscr{C}$ is the infinite constant $\mathscr{J} \int d\phi'$. The integral over the ϕ_v' then gives

$$\int \left[\prod_i d\phi_i\, d\phi_i^*\right] \exp\{iI_{\text{quad}}[\phi]\}\phi_a \cdots \phi_b^* \cdots \propto \sum_{\text{pairings}} \left[-i\Delta_{ab}\right] \cdots + \xi\text{-terms} , \tag{30.2.10}$$

where the sum over pairings is over all ways of pairing indices on the ϕs with indices on the ϕ^*s, and Δ_{ab} is the propagator

$$\Delta_{ab} = \sum_v \frac{u_{va}\, u_{vb}^*}{d_v} . \tag{30.2.11}$$

Instead of actually evaluating the sum (30.2.11), we can use its defining property that

$$\sum_c D_{ac}\Delta_{cb} = \sum_v u_{va}\, u_{vb}^* \equiv \Pi_{ab} , \tag{30.2.12}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

with Π the projection operator on the space orthogonal to ξ:

$$\Pi^2 = \Pi\,, \qquad \Pi\,\xi = 0\,. \tag{30.2.13}$$

The solution of Eq. (30.2.12) is unique only up to ξ-terms, but these do not matter if the fields ϕ_i appear in the action only in gauge-invariant combinations.

For instance, in electrodynamics, we can write the kinematic part of the action as

$$I_{\text{quad}}[A] = -\frac{1}{4}\int d^4x\, f_{\mu\nu}f^{\mu\nu} = +\frac{1}{2}\int d^4x\, A^\mu\Big(\Box\delta^\nu_\mu - \partial_\mu\partial^\nu\Big)A_\nu\,.$$

The differential operator $-\Box\delta^\nu_\mu + \partial^\nu\partial_\mu$ is not invertible, because it has a zero eigenvalue with eigenvectors of the form $\xi^\mu = \partial^\mu\Lambda$. The projection matrix onto the space orthogonal to these vectors is

$$\Pi_\mu{}^\nu(x,y) = [\delta_\mu{}^\nu - \partial_\mu\partial^\nu\Box^{-1}]\delta^4(x-y)\,,$$

where $\Box^{-1}\delta^4(x-y)$ is any solution of the equation $\Box[\Box^{-1}\delta^4(x-y)] = \delta^4(x-y)$. The defining equation for the propagator is

$$\Big(-\Box\delta^\lambda_\mu + \partial_\mu\partial^\lambda\Big)\Delta_\lambda{}^\nu(x,y) = \Pi_\mu{}^\nu(x,y)\,,$$

with the solution

$$\Delta_\lambda{}^\nu(x,y) = \delta^\nu_\mu\,\Delta_F(x-y) + \partial_\mu\partial^\nu\text{-terms}\,,$$

where $\Delta_F(x-y)$ is the usual Feynman propagator (6.2.16), satisfying $\Box\Delta_F(x-y) = -\delta^4(x-y)$. (The factor $1/2$ in the action does not appear in the defining equation for the propagator here because A_μ is a real field. The origin of the $-i\epsilon$ in the denominator of the Fourier integral in Eq. (6.2.16) is explained in the path-integral formalism in Section 9.2.)

Inspection of the first term in Eq. (30.1.9) shows that the defining equation for the superpropagator of the potential superfields is

$$\frac{1}{4}\mathscr{D}_L^2\,\mathscr{D}_R^2\,\Delta^S_{nm}(x,\theta;x',\theta') = \mathscr{P}\delta^4(x-x')\delta^4(\theta-\theta')\delta_{nm}\,, \tag{30.2.14}$$

where $\mathscr{P}$ is a superspace differential operator satisfying the conditions for a projection operator

$$\mathscr{P}^2 = \mathscr{P}\,, \qquad \mathscr{P}\mathscr{D}_R = 0\,, \tag{30.2.15}$$

and $\delta^4(\theta-\theta')$ is the fermionic delta function introduced in Eq. (26.6.8). The solution is

$$\mathscr{P} = \frac{-1}{16\Box}\mathscr{D}_L^2\,\mathscr{D}_R^2\,. \tag{30.2.16}$$

(It is obvious that $\mathscr{P}\mathscr{D}_R = 0$. To check that $\mathscr{P}^2 = \mathscr{P}$, we need to use (26.6.12), which shows that $\mathscr{D}_R^2\mathscr{D}_L^2\mathscr{D}_R^2 = -16\Box\mathscr{D}_R^2$.) The solution of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

Eq. (30.2.14) is

$$\Delta^S_{nm}(x,\theta;x',\theta') = -\frac{1}{4\Box}\delta^4(x-x')\delta^4(\theta-\theta')\delta_{nm}$$
$$= \frac{1}{4}\Delta_F(x-x')\delta^4(\theta-\theta')\delta_{nm} + \mathscr{D}_R\text{-terms}\,. \tag{30.2.17}$$

This is the superpropagator for a line created by a potential superfield $S^*_m(x',\theta')$ and destroyed by a potential superfield $S_n(x,\theta)$, which we must use in evaluating supergraphs for the action (30.1.9). To make contact with ordinary propagators, it is also of some interest to consider the superpropagator of the line created by a left-chiral superfield $\Phi^*_m(x',\theta')$ and destroyed by a left-chiral superfield $\Phi_n(x,\theta)$. These chiral superfields are obtained by operating on $S_n(x,\theta)$ with $\mathscr{D}^2_R$ and on $S^*_m(x',\theta')$ with $\mathscr{D}'^2_R{}^* = \mathscr{D}'^2_L$, so the propagator for the left-chiral superfields is

$$\Delta^\Phi_{nm}(x,\theta;x',\theta') = \frac{1}{4}\mathscr{D}^2_R\mathscr{D}'^2_L\Delta_F(x-x')\,\delta^4(\theta-\theta')\,\delta_{nm}\,. \tag{30.2.18}$$

For instance, the term in the superpropagator of zeroth order in θ and θ' is

$$\left[\Delta^\Phi_{nm}(x,\theta;x',\theta')\right]_{\theta=\theta'=0} = \frac{1}{4}\left(\frac{\partial}{\partial\theta_R}\right)^2\left(\frac{\partial}{\partial\theta'_L}\right)^2\Delta_F(x-x')\,\delta^4(\theta-\theta')\,\delta_{nm}\,. \tag{30.2.19}$$

To evaluate this, we recall Eq. (26.6.8) for the fermionic delta function:

$$\delta^4(\theta-\theta') = \frac{1}{4}\Big((\theta_L-\theta'_L)^{\mathrm{T}}\epsilon(\theta_L-\theta'_L)\Big)\Big((\theta_R-\theta'_R)^{\mathrm{T}}\epsilon(\theta_R-\theta'_R)\Big)\,,$$

from which we find $(\partial/\partial\theta_R)^2(\partial/\partial\theta'_L)^2\delta^4(\theta-\theta') = 4$. Eq. (30.2.19) therefore yields

$$\left[\Delta^\Phi_{nm}(x,\theta;x',\theta')\right]_{\theta=\theta'=0} = \Delta_F(x-x')\,\delta_{nm}\,, \tag{30.2.20}$$

which is just the usual propagator for the scalar components of the superfield.

30.3 Calculations with Supergraphs

We now consider how to use the results of the previous sections to calculate the quantum effective action $\Gamma[S,S^*]$ for a set of classical potential superfields $S_n(x,\theta)$ and their adjoints. Following the prescription discussed in Section 16.1, this may be defined in terms of a sum over all connected one-particle-irreducible supergraphs, consisting of vertices to which are attached directed internal and external lines. For each external line of type n

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

ending or beginning in a vertex labelled x, θ, we include a c-number factor $S_n(x, \theta)$ or $S_n^*(x, \theta)$, respectively (but no propagator). A vertex labelled x, θ with N incoming or N outgoing lines labelled $n_1, n_2, \ldots, n_N$ yields a factor equal to i times the coefficient of the $S_1 S_2 \cdots S_N$ term in the superpotential $\tilde{f}(S)$ or times the complex conjugate of this coefficient, respectively. An internal line of any type coming out of a vertex labelled x, θ and going into a vertex labelled x', θ' yields a propagator factor, given by Eqs. (30.2.10) and (30.2.17) as

$$-\frac{i}{4}\delta^4(\theta - \theta')\,\Delta_F(x - x')\,. \tag{30.3.1}$$

In addition, as shown by Eq. (30.1.7), a superderivative $\mathscr{D}_R^2$ acts on the propagators or external line S-factors of all but one of the internal or external lines coming into any vertex, and a superderivative $\mathscr{D}_L^2$ acts on the propagators or external-line S^*-factors of all but one of the lines coming out of any vertex. The product of these factors must be integrated over all xs and θs; the quantum effective action is the sum of these integrals for all one-particle-irreducible graphs.

By integrating by parts in superspace, the $\mathscr{D}_L^2$ and/or $\mathscr{D}_R^2$ operators accompanying any one propagator (say, one that connects vertices labelled x, θ and x', θ') may be moved to other propagators or external line factors. This leaves the factor contributed by this internal line proportional to a factor $\delta^4(\theta - \theta')$. Integrating over θ' then eliminates this delta function, and lets us replace θ' with θ everywhere else. (In cases where several internal lines connect the same pair of vertices, we need to use the property of fermionic delta functions that $[\delta^4(\theta - \theta')]^2 = 0$.) Continuing in this way, we wind up with a single four-dimensional θ integral, with all $\mathscr{D}$s acting on the external line factors S_n and S_m^*. That is, though not usually local in spatial coordinates, $\Gamma[S, S^*]$ *is local in the fermionic coordinates.*

The structure of this functional is governed by its invariance under the 'gauge' transformations (30.2.4). This tells us that every one of the external line factors S_n or S_n^* must be acted on by an operator $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$, respectively, with two possible exceptions. The exceptions are that a term with only incoming or only outgoing external lines, in which all but one of the external line factors S_n or S_n^* are acted on by $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$, respectively, and by no other superderivatives, is still invariant under the transformation (30.2.4), even though it cannot be expressed as a four-dimensional θ-integral of a functional only of the Φ_n and/or Φ_n^* alone. This is because the change in such an amplitude under this transformation could only come from the change in the external line factor S_n or S_n^* that is *not* acted on by an operator $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$, and this change is eliminated if we use integration by parts to move one of the other $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$ operators to act on it. As we saw in Section 30.1, such a term in $\Gamma[S, S^*]$ is the $\mathscr{F}$-term

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

of a functional of the Φ_n or of the Φ_n^*, and thus makes a correction to the superpotential or to its complex conjugate. Aside from these exceptions, every term in $\Gamma[S,S^*]$ may be written as a four-dimensional θ-integral of a functional only of the Φ_n and/or Φ_n^*, and therefore makes a correction to the D-term part of the effective action.

Also note that a term in the quantum effective action in which all but one of the external line factors S_n or S_n^* is acted on by $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$, respectively, *and* also by additional $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$ operators (such as in combinations like $\mathscr{D}_R^2\mathscr{D}_L^2\mathscr{D}_R^2 S_n$ or $\mathscr{D}_L^2\mathscr{D}_R^2\mathscr{D}_L^2 S_n^*$) may also be written by integration by parts as a term in which one of the extra $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$ acts on the previously undifferentiated external line S_n or S_n^* factor. Such a term may therefore be expressed as a four-dimensional θ-integral of a functional only of the Φ_n and/or Φ_n^*, and therefore just makes another correction to the D-term part of the action. The only terms in $\Gamma[S,S^*]$ that cannot be written in this way are those that have E incoming external lines and *only* $E-1$ operators $\mathscr{D}_R^2$ operating on the S_n external line factors, or E^* incoming external lines and *only* E^*-1 operators $\mathscr{D}_L^2$ operating on the S_n^* external line factors. We can therefore tell whether a supergraph can make a contribution to the superpotential or its adjoint by simply counting the number of $\mathscr{D}_R^2$ or $\mathscr{D}_L^2$ operators contributed by the supergraph to the corresponding term in $\Gamma[S,S^*]$.

Let us count these superderivatives. Consider a connected graph with: V_n vertices with n lines coming in; V_n^* vertices with n lines going out; I internal lines; E external incoming lines; and E^* external outgoing lines. These numbers are related by

$$I+E=\sum_n nV_n\,, \qquad I+E^*=\sum_n nV_n^*\,. \tag{30.3.2}$$

The total numbers of $\mathscr{D}_R^2$ and $\mathscr{D}_L^2$ operators are then

$$N_R=\sum_n V_n(n-1)=I+E-V=L+V^*+E-1\,, \tag{30.3.3}$$

and

$$N_L=\sum_n V_n^*(n-1)=I+E^*-V^*=L+V+E^*-1\,, \tag{30.3.4}$$

where $V=\sum_n V_n$ is the total number of vertices with incoming lines; $V^*=\sum_n V_n^*$ is the total number of vertices with outgoing lines; and $L=I-V-V^*+1$ is the number of loops. We see that a graph with any loops will have $N_R\geq E$ and $N_L\geq E^*$, so that there are at least enough $\mathscr{D}_R^2$ operators to convert all S_n into $\Phi_n=\mathscr{D}_R^2 S_n$ and its derivatives, and enough $\mathscr{D}_L^2$ operators to convert all S_n^* into $\Phi_n^*=\mathscr{D}_L^2 S_n^*$ and its derivatives. *Any graph with loops thus yields only a contribution to the four-dimensional*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

θ-integral — in other words the D-term — of a functional of the left-chiral superfields Φ_n and their adjoints.

The only way to obtain a contribution to an $\mathscr{F}$-term or its adjoint is to have just $N_R = E - 1$ operators $\mathscr{D}_R^2$ or just $N_L = E^* - 1$ operators $\mathscr{D}_L^2$, respectively. According to Eqs. (30.3.3) and (30.3.4), such a graph would have $L = 0$ and $V^* = 0$ or $V = 0$, respectively. In other words, since we are considering only one-particle-irreducible graphs, we can only get an $\mathscr{F}$-term from a graph with a single vertex with incoming lines, and we can only get its adjoint from a graph with a single vertex with outgoing lines. This contribution is nothing but the integrated $\mathscr{F}$-term of the original superpotential, or its adjoint. Thus we see again that *there is no finite or infinite renormalization of $\mathscr{F}$-terms to any order of perturbation theory.*

Problems

1. Use Eq. (30.2.18) to calculate the propagators of the spinor and auxiliary components of a chiral superfield.

2. Consider a supersymmetric theory of a single left-chiral superfield Φ, with Lagrangian density

$$\mathscr{L} = \frac{1}{2}\left[\Phi^*\Phi\right]_D + 2\mathrm{Re}\left(g[\Phi^3]_{\mathscr{F}}\right),$$

with g an arbitrary complex constant. Use the supergraph formalism to calculate the one-loop contribution to the quantum effective action. Express the answer as an integral over coordinates and over a single Grassman coordinate θ.

3. What is the superpropagator for a gauge superfield $V(x,\theta)$ in a supersymmetric Abelian gauge theory with kinematic term (27.3.17)?

References

1. A. Salam and J. Strathdee, *Phys. Rev.* **D11**, 1521 (1975); *Nucl. Phys.* **B86**, 142 (1975); D. M. Capper, *Nuovo Cimento* **25A**, 259 (1975); R. Delbourgo, *Nuovo Cimento* **25A**, 646 (1975); D. M. Capper and G. Leibrandt, *Nucl. Phys.* **B85**, 492 (1975); F. Krause, M. Scheunert, J. Honerkamp, and M. Schlindwein, *Phys. Lett.* **53B**, 60 (1974); K. Fujikawa and W. Lang, *Nucl. Phys.* **B88**, 61 (1975); J. Honerkamp, M. Schlindwein, F. Krause, and M. Scheunert, *Nucl. Phys.* **B95**, 397 (1975); S. Ferrara and O. Piguet, *Nucl. Phys.* **B93**, 261 (1975); R.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

Delbourgo, *J. Phys.* **G1**, 800 (1975). This formalism was extended to supergravity by W. Siegel, *Phys. Lett.* **84B**, 197 (1979).

2. M. T. Grisaru, W. Siegel, and M. Roček, *Nucl. Phys.* **B159**, 429 (1979).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.009

31

Supergravity

Gravity exists, so if there is any truth to supersymmetry then any realistic supersymmetry theory must eventually be enlarged to a supersymmetric theory of matter and gravitation, known as *supergravity*. Supersymmetry without supergravity is not an option, though it may be a good approximation at energies far below the Planck scale.

There are two leading approaches to the construction of the theory of supergravity. First, supergravity can be presented as a theory of curved superspace.[1] This approach is analogous to the development of supersymmetric gauge theories in Sections 27.1–27.3; the gravitational field appears as a component of a superfield with unphysical as well as physical components, like the unphysical C, M, N, and ω components of the gauge superfield V. The task of deriving the full non-linear supergravity theory in this way is forbiddingly complicated, and so far has not been freed of steps that are apparently arbitrary. At one point or another in the derivation, it has been necessary simply to state that some set of constraints on the graviton superfield are the proper ones to adopt.

Here we will follow a second approach that is less elegant but more transparent.[2] In our discussion here, we begin in Sections 31.1–31.5 with the case where the gravitational field is weak,[3] analyzing supergravity by the same flat-space superfield methods that we used in Chapters 26 and 27 to study ordinary supersymmetry theories. In this way we can identify the physical components of the gravitational superfield (including auxiliary fields analogous to the D-component of the gauge superfield V.) The weak-field approximation will allow us to obtain some of the most important consequences of supergravity theory, including the general formula for the gravitino mass in Section 31.3 and the results for gaugino masses and the A and B parameters given by anomaly-mediated supersymmetry breaking in Section 31.4.

In Section 31.6 we add terms of higher order in G to the supersymmetry transformation rules for the physical fields and to the Lagrangian that describes their interactions, subject to the condition that the supersymmetry transformations and general coordinate transformations form

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

a closed algebra, and that the Lagrangian is invariant under these transformations. This approach is in some respects analogous to the treatment of supersymmetric gauge theories in Section 27.8; we work only with the physical components of the gravitational superfield, and obtain transformation rules that involve covariant rather than ordinary derivatives. This is the approach that has been used in deriving the most important application of supergravity theory beyond the weak-field approximation: the derivation of the low-energy effective Lagrangian in theories of gravitationally mediated supersymmetry breaking. This application is the subject of Section 31.7.

31.1 The Metric Superfield

Supergravity necessarily involves spinor as well as tensor fields, so we will have to describe gravitational fields in terms of a vierbein (or tetrad) $e^a{}_\mu(x)$ rather than a metric, which is related to the vierbein by

$$g_{\mu\nu}(x) = \eta_{ab}\, e^a{}_\mu(x)\, e^b{}_\nu(x)\,. \tag{31.1.1}$$

The indices μ, ν, etc. label general coordinates, while indices a, b, etc. label coordinates in a locally inertial coordinate system, with η_{ab} the usual diagonal matrix with elements $+1, +1, +1, -1$. In the vierbein formalism, the action must be supposed to be invariant under two different sorts of symmetry transformations: general coordinate transformations $x^\mu \to x'^\mu(x)$, under which the vierbein $e^a{}_\mu(x)$ is transformed into $e'^a{}_\mu(x)$, where

$$e'^a{}_\mu(x') = \frac{\partial x^\nu}{\partial x'^\mu} e^a{}_\nu(x)\,, \tag{31.1.2}$$

and local Lorentz transformations, under which

$$e^a{}_\mu(x) \to \Lambda^a{}_b(x)\, e^b{}_\mu(x)\,, \tag{31.1.3}$$

with $\Lambda^a{}_b(x)$ an arbitrary real matrix subject to the constraints

$$\eta_{ab}\Lambda^a{}_c(x)\Lambda^b{}_d(x) = \eta_{cd}\,. \tag{31.1.4}$$

An elementary review of this formalism is given in the appendix to this chapter.

A weak gravitational field is one for which the vierbein is close to the unit matrix. In such a field, the vierbein may conveniently be written as

$$e^a{}_\mu(x) = \delta^a_\mu + 2\kappa\, \phi^a{}_\mu(x)\,, \tag{31.1.5}$$

where $\phi^a{}_\mu(x)$ is small. As we shall see in Section 31.2, if $\phi^a{}_\mu$ is to be a conventionally normalized field, then the constant κ should be expressed in

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

terms of Newton's constant G by $\kappa = \sqrt{8\pi G}$. The closeness of the vierbein to the unit matrix is preserved by a small coordinate transformation

$$x^\mu \to x^\mu + \xi^\mu(x)\,, \tag{31.1.6}$$

and also by small local Lorentz transformations

$$\Lambda^a{}_b(x) = \delta^a_b + \omega^a{}_b(x)\,, \tag{31.1.7}$$

where $\xi^\mu(x)$ and $\omega^a{}_b(x)$ are of the same order as $\phi^a{}_\mu(x)$, and $\omega_{ab}(x) \equiv \eta_{ac}\omega^c{}_b(x)$ is constrained by Eq. (31.1.4) to satisfy

$$\omega_{ab}(x) = -\omega_{ba}(x)\,. \tag{31.1.8}$$

The combined transformations (31.1.2) and (31.1.3) then become

$$\phi_{\mu\nu}(x) \to \phi_{\mu\nu}(x) + \frac{1}{2\kappa}\left[-\frac{\partial\xi_\mu(x)}{\partial x_\nu} + \omega_{\mu\nu}(x)\right]\,. \tag{31.1.9}$$

We are now dropping the distinction between general coordinate indices μ, ν, etc. and local Lorentz coordinate indices a, b, etc., and raising and lowering all indices with $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$. In terms of the metric (31.1.1), the weak field assumption (31.1.5) becomes

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + 2\kappa\, h_{\mu\nu}(x)\,, \tag{31.1.10}$$

with

$$h_{\mu\nu}(x) \equiv \phi_{\mu\nu}(x) + \phi_{\nu\mu}(x)\,, \tag{31.1.11}$$

while the transformation law (31.1.9) reads

$$h_{\mu\nu}(x) \to h_{\mu\nu}(x) - \frac{1}{2\kappa}\left[\frac{\partial\xi_\mu(x)}{\partial x^\nu} + \frac{\partial\xi_\nu(x)}{\partial x^\mu}\right]\,. \tag{31.1.12}$$

By use of the supersymmetry algebra we showed in Section 25.4 that the graviton has a fermionic superpartner, the gravitino, with helicities $\pm 3/2$. We saw in Section 5.9 that a self-charge-conjugate massless particle with helicities ± 1 can only have low-energy interactions if described by a real field $A_\mu(x)$ whose interactions are invariant under a gauge transformation $A_\mu(x) \to A_\mu(x) + \partial_\mu\Lambda(x)$. In the same way, a self-charge-conjugate massless particle with helicities $\pm 3/2$ can only have low-energy interactions if it is described by a Majorana field $\psi_\mu(x)$ with an extra vector index μ, with invariance under a gauge transformation

$$\psi_\mu(x) \to \psi_\mu(x) + \partial_\mu\psi(x)\,, \tag{31.1.13}$$

where $\psi(x)$ is an arbitrary Majorana field.[3a] We need now to consider how fields $\phi_{\mu\nu}(x)$ and $\psi_\mu(x)$ with these transformation properties can be put into a superfield.

We found in Section 27.1 that a gauge field $V_\nu(x)$ can be regarded as the V_ν-component, defined in Eq. (26.2.10) as the coefficient of $i(\bar\theta\gamma_5\gamma_\nu\theta)/2$, of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

a real scalar superfield $V(x,\theta)$. Similarly, we would like to incorporate the vierbein field $\phi_{\mu\nu}(x)$ and the gravitino field $\psi_\mu(x)$ into a *vector* superfield $H_\mu(x,\theta)$, known as the *metric superfield*. The question is: how are $\phi_{\mu\nu}(x)$ and $\psi_\mu(x)$ related to the components of this superfield?

To address this question, note that supersymmetry requires the 'gauge' transformations (31.1.9) and (31.1.13) to be special cases of a transformation of the whole metric superfield

$$H_\mu(x,\theta) \to H_\mu(x,\theta) + \Delta_\mu(x,\theta)\,. \tag{31.1.14}$$

Furthermore, a *weak* gravitational field $h_{\mu\nu}$ interacts with an energy-momentum tensor $T^{\mu\nu}$ which as we found in Section 26.7 is a linear combination of components of a real vector superfield Θ^μ, so we expect the interaction of the whole superfield H_μ with matter to take the form

$$I_{\rm int} = 2\kappa \int d^4x \left[H_\mu\,\Theta^\mu\right]_D\,. \tag{31.1.15}$$

(Later in this section we will confirm the correctness of the coefficient 2κ, with Θ^μ normalized as in Section 26.7.) The supercurrent Θ^μ was shown in Section 26.7 to satisfy the conservation conditions

$$\gamma^\mu \mathscr{D}\Theta_\mu = \mathscr{D}\,X\,, \tag{31.1.16}$$

where X is a real chiral scalar superfield (the sum of a left-chiral scalar superfield and its complex conjugate) and $\mathscr{D}$ is the four-component superderivative (26.2.26). It follows that this interaction is invariant under a transformation of the form (31.1.14), with Δ_μ of the form

$$\Delta_\mu = \left(\bar{\mathscr{D}}\gamma_\mu \Xi\right)\,, \tag{31.1.17}$$

where $\Xi(x,\theta)$ is a superfield subject to the manifestly supersymmetric condition

$$\left(\bar{\mathscr{D}}\mathscr{D}\right)\left(\bar{\mathscr{D}}\Xi\right) = 0\,. \tag{31.1.18}$$

This can be seen by recalling from Section 26.7 that the chirality condition on X allows us to write it in the form

$$X = \left(\bar{\mathscr{D}}\mathscr{D}\right)\Omega\,, \tag{31.1.19}$$

with Ω in general a non-local superfield. From Eq. (31.1.16) we then find

$$\begin{aligned}\int \left[\Theta^\mu\left(\bar{\mathscr{D}}\gamma_\mu\Xi\right)\right]_D &= -\int \left[\left((\bar{\mathscr{D}}\Theta^\mu)\gamma_\mu\Xi\right)\right]_D = -\int \left[\left((\bar{\mathscr{D}}X)\Xi\right)\right]_D \\ &= +\int \left[\Omega\left(\bar{\mathscr{D}}\mathscr{D}\right)\left(\bar{\mathscr{D}}\Xi\right)\right]_D = 0\,,\end{aligned}$$

so that the interaction (31.1.15) is invariant under the transformation $H_\mu \to H_\mu + (\bar{\mathscr{D}}\gamma_\mu\Xi)$. (We can also obtain the same result without using the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

representation (31.1.19) for X, by noting from Eq. (26.2.25) that the D-term of the product of a chiral superfield like X which satisfies Eqs. (26.3.1) and (26.3.2) and a linear superfield like $(\bar{\mathscr{D}}\Xi)$ which satisfies Eq. (26.3.45) is a spacetime derivative.)

At the end of this section we shall show that Eqs. (31.1.17) and (31.1.18) yield conditions on the components of the superfield Δ_μ:

$$V^\Delta_{\mu\nu}(x) + V^\Delta_{\nu\mu}(x) = \frac{\partial v_\mu(x)}{\partial x_\nu} + \frac{\partial v_\nu(x)}{\partial x_\mu} - 2\,\eta_{\mu\nu}\frac{\partial v^\lambda(x)}{\partial x^\lambda}\,, \tag{31.1.20}$$

$$\lambda^\Delta_\mu(x) - \tfrac{1}{3}\gamma_\mu\gamma^\rho\lambda^\Delta_\rho(x) - \tfrac{1}{3}\gamma_\mu\partial^\rho\omega^\Delta_\rho(x) = \partial_\mu\chi(x)\,, \tag{31.1.21}$$

$$-\tfrac{1}{2}\epsilon^{\nu\mu\kappa\sigma}\partial_\kappa V^\Delta_{\nu\mu}(x) = D^{\Delta\,\sigma}(x) + \partial^\sigma\partial^\rho C^\Delta_\rho(x)\,, \tag{31.1.22}$$

$$\partial^\mu M^\Delta_\mu(x) = \partial^\mu N^\Delta_\mu(x) = 0\,, \tag{31.1.23}$$

with $v_\mu(x)$ a real vector field, and $\chi(x)$ a Majorana spinor field. (Here we introduce a notation that will be used throughout this chapter; following Eq. (26.3.9), the components C^S, ω^S, M^S, N^S, V^S_ν, λ^S, and D^S of an arbitrary superfield $S(x,\theta)$ are defined by the expansion

$$\begin{aligned} S(x,\theta) = {} & C^S(x) - i\big(\bar\theta\,\gamma_5\,\omega^S(x)\big) - \frac{i}{2}\big(\bar\theta\,\gamma_5\,\theta\big)M^S(x) - \frac{1}{2}\big(\bar\theta\,\theta\big)N^S(x) \\ & + \frac{i}{2}\big(\bar\theta\,\gamma_5\,\gamma^\nu\,\theta\big)V^S_\nu(x) - i\big(\bar\theta\,\gamma_5\,\theta\big)\Big(\bar\theta\Big[\lambda^S(x) + \frac{1}{2}\,\partial\!\!\!/\,\omega^S(x)\Big]\Big) \\ & - \frac{1}{4}\big(\bar\theta\,\gamma_5\,\theta\big)^2\Big(D^S(x) + \frac{1}{2}\Box C^S(x)\Big)\,. \end{aligned} \tag{31.1.24}$$

Also, $V^\Delta_{\mu\nu}(x)$ is the V_ν-component of Δ_μ.) This leads us to define the fields

$$\phi_{\mu\nu}(x) \equiv V^H_{\mu\nu}(x) - \tfrac{1}{3}\eta_{\mu\nu}\,V^{H\,\lambda}{}_\lambda(x)\,, \tag{31.1.25}$$

$$\tfrac{1}{2}\psi_\mu(x) \equiv \lambda^H_\mu(x) - \tfrac{1}{3}\gamma_\mu\gamma^\rho\lambda^H_\rho(x) - \tfrac{1}{3}\gamma_\mu\partial^\rho\omega^H_\rho(x)\,, \tag{31.1.26}$$

$$b^\sigma(x) \equiv D^{H\,\sigma}(x) + \tfrac{1}{2}\epsilon^{\nu\mu\kappa\sigma}\partial_\kappa V^H_{\nu\mu}(x) + \partial^\sigma\partial^\rho C^H_\rho(x)\,. \tag{31.1.27}$$

(As discussed in Section 31.3, the factor 1/2 introduced on the left-hand side of Eq. (31.1.26) will give the field ψ_μ a conventional normalization.) From Eqs. (31.1.20) and (31.1.21) it follows that the transformation (31.1.14) induces on $\phi_{\mu\nu}(x)$ and $\psi_\mu(x)$ the gauge transformations (31.1.9) and (31.1.13), with

$$\xi_\mu = -2\kappa\,v_\mu\,, \quad \omega_{\mu\nu} = \kappa\Big[-\frac{\partial v_\mu}{\partial x^\nu} + \frac{\partial v_\nu}{\partial x^\mu} + V^\Delta_{\mu\nu} - V^\Delta_{\nu\mu}\Big]\,, \quad \psi = 2\chi\,, \tag{31.1.28}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

while Eq. (31.1.22) shows that $b_\mu(x)$ is invariant. Also, Eq. (31.1.23) shows that the transformation (31.1.14) induces shifts of $M^H_\mu(x)$ and $N^H_\mu(x)$ that leave invariant the fields

$$s \equiv \partial^\mu M^H_\mu(x)\,, \qquad p \equiv \partial^\mu N^H_\mu(x)\,. \tag{31.1.29}$$

Finally, since $C^\Delta_\mu(x)$, $V^\Delta_{\mu\nu}(x) - V^\Delta_{\nu\mu}(x)$, and $\omega^\Delta_\mu(x)$ are unconstrained by supersymmetry, the transformation (31.1.14) allows us to make the components $C^H_\mu(x)$, $V^H_{\mu\nu}(x) - V^H_{\nu\mu}(x) = \phi_{\mu\nu} - \phi_{\nu\mu}$, and $\omega^H_\mu(x)$ anything we like. In particular, in analogy with the Wess–Zumino gauge for gauge superfields discussed in Section 27.1, we can take $C^H_\mu(x)$, $V^H_{\mu\nu}(x) - V^H_{\nu\mu}(x) = \phi_{\mu\nu} - \phi_{\nu\mu}$, and $\omega^H_\mu(x)$ all to vanish. The fields $h_{\mu\nu}(x)$ and $\psi_\mu(x)$ are identified by their transformation properties as the fields of the graviton and gravitino, respectively, while $b_\mu(x)$, $s(x)$, and $p(x)$ are auxiliary fields[4] that are important in understanding the coupling of the graviton superfield to matter.

Incidentally, note that the number of independent components of the symmetric tensor $h_{\mu\nu}$ modulo the gauge transformations (31.1.12) is $10 - 4 = 6$, which with the auxiliary fields s, p, and b_μ gives a total of $6+6 = 12$ independent physical bosonic fields, while the number of independent physical components of the Majorana spinor field ψ_μ modulo the gauge transformations (31.1.13) is $16 - 4 = 12$. This satisfies the condition, discussed at the end of Section 26.2, that in any supermultiplet of fields that furnishes a representation of the supersymmetry algebra, there must be equal numbers of independent bosonic and fermionic field components.

Let us now return to the interaction of matter and gravitation. In general, the integrated D-term of the product of two superfields Θ^μ and H_μ is given by Eq. (26.2.25) as

$$\begin{aligned}\int d^4x \left[\Theta^\mu H_\mu\right]_D = \int d^4x \Big[&- \partial_\mu C^{H\,\sigma}\, \partial^\mu C^\Theta_\sigma + C^{H\,\sigma}\, D^\Theta_\sigma + D^{H\,\sigma}\, C^\Theta_\sigma \\ &- \left(\overline{\omega}^{H\,\sigma}\, [\lambda^\Theta_\sigma + \tfrac{1}{2}\, \not{\partial} \omega^\Theta_\sigma]\right) - \left([\overline{\lambda}^{H\,\sigma} + \tfrac{1}{2}\overline{\omega}^{H\,\sigma}\, \not{\partial}]\omega^\Theta_\sigma\right) \\ &+ M^{H\,\sigma}\, M^\Theta_\sigma + N^{H\,\sigma}\, N^\Theta_\sigma - V^{H\,\kappa\sigma}\, V^\Theta_{\kappa\sigma}\Big]\,.\end{aligned} \tag{31.1.30}$$

Using Eqs. (31.1.25)–(31.1.27), we may express $V^H_{\mu\nu}$, λ^H_μ, and D^H_μ respectively in terms of $\phi_{\mu\nu}$, ψ_μ, and b_μ, and find

$$\begin{aligned} I_{\rm int} &= 2\kappa \int d^4x \left[\Theta^\mu H_\mu\right]_D \\ &= 2\kappa \int d^4x \left[C^{H\,\sigma}\left[\Box C^\Theta_\sigma - \partial_\sigma \partial^\rho C^\Theta_\rho + D^\Theta_\sigma\right] + b^\sigma\, C^\Theta_\sigma\right. \end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

$$+\left(\overline{\omega}^{H\sigma}\left[-\lambda_\sigma^\Theta - \not{\partial}\omega_\sigma^\Theta + \partial_\sigma\gamma^\rho\omega_\rho^\Theta\right]\right) - \left(\bar{\psi}^\sigma\,\omega_\sigma^\Theta\right) + \left(\bar{\psi}^\sigma\,\gamma_\sigma\gamma^\rho\omega_\rho^\Theta\right)$$
$$+ M^{H\sigma}M_\sigma^\Theta + N^{H\sigma}N_\sigma^\Theta$$
$$+\tfrac{1}{2}\epsilon^{\nu\mu\kappa\sigma}\phi_{\nu\mu}\partial_\kappa C_\sigma^\Theta - \phi^{\lambda\sigma}\left[V_{\lambda\sigma}^\Theta - \eta_{\lambda\sigma}V^{\Theta\,\rho}{}_\rho\right]\Big] . \tag{31.1.31}$$

We saw in Section 26.7 that the conservation condition (31.1.16) yields conditions (26.7.44), (26.7.39), and (26.7.35):

$$D_\mu^\Theta = -\Box C_\mu^\Theta + \partial_\mu\partial^\nu C_\nu^\Theta ,$$
$$\lambda_\nu^\Theta = -\not{\partial}\omega_\nu^\Theta + \partial_\nu\gamma^\mu\omega_\mu^\Theta ,$$
$$0 = V_{\mu\nu}^\Theta - V_{\nu\mu}^\Theta + \epsilon_{\mu\nu\rho\sigma}\partial^\sigma C^{\Theta\rho} ,$$

which respectively tell us that the coefficients in Eq. (31.1.31) of $C^{H\sigma}$, of $\overline{\omega}^{H\sigma}$, and of the antisymmetric part of $\phi^{\mu\nu}$ all vanish. Also, we may write the remaining terms of Eq. (31.1.31) in terms of the supersymmetry current (26.7.20), the energy-momentum tensor (26.7.42), the $\mathscr{R}$-current (26.7.51):

$$S^\mu = -2\omega^{\Theta\mu} + 2\gamma^\mu\gamma^\nu\omega_\nu^\Theta ,$$
$$T_{\mu\nu} = -\tfrac{1}{2}V_{\mu\nu}^\Theta - \tfrac{1}{2}V_{\nu\mu}^\Theta + \eta_{\mu\nu}V^{\Theta\lambda}{}_\lambda ,$$
$$\mathscr{R}^\mu = 2\,C^{\Theta\mu} ,$$

and densities $\mathscr{M}$ and $\mathscr{N}$ defined by

$$M_\mu^\Theta = \partial_\mu\mathscr{M} , \qquad N_\mu^\Theta = \partial_\mu\mathscr{N} , \tag{31.1.32}$$

and given by Eqs. (26.7.33) and (26.7.34) as:

$$\mathscr{N} = -A^X , \qquad \mathscr{M} = B^X , \tag{31.1.33}$$

where X is the real chiral superfield appearing on the right-hand side of the conservation equation (31.1.16). (A label 'new' on the supersymmetry current will be understood in this chapter.) The first-order interaction between matter and the components of the metric superfield is then

$$2\kappa\int d^4x\left[\Theta^\mu H_\mu\right]_D = \kappa\int d^4x\left[\mathscr{R}_\sigma b^\sigma + \tfrac{1}{2}\bar{S}^\sigma\psi_\sigma - 2\mathscr{M}s - 2\mathscr{N}p + T^{\kappa\sigma}h_{\sigma\kappa}\right] . \tag{31.1.34}$$

We note that the gravitino field interacts with the supersymmetry current, in much the same way as the gravitational field interacts with the energy-momentum tensor.

We can now check the constant factor appearing in this interaction. The usual definition of $T^{\mu\nu}$ is that the variation of the matter action under a change $\delta g_{\mu\nu}$ in the metric is[5]

$$\delta I_M = \frac{1}{2}\int d^4x\,\sqrt{\mathrm{Det}\,g}\,T^{\mu\nu}\,\delta g_{\mu\nu} .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The interaction between matter and a weak gravitational field given by Eq. (31.1.10) is then

$$\kappa \int d^4x\; T^{\mu\nu}(x)\, h_{\mu\nu}(x)\,,$$

which agrees with the $h_{\mu\nu}$-dependent part of Eq. (31.1.34), thus confirming the normalization of the interaction (31.1.15).

* * *

We shall now check that Eqs. (31.1.17) and (31.1.18) yield the conditions (31.1.20)–(31.1.23). The superfield $\Upsilon \equiv \left(\bar{\mathscr{D}}\Xi\right)$ has the components

$$\begin{aligned}
C^\Upsilon &= -i\mathrm{Tr}\,(\epsilon\omega^\Xi)\,,\\
\omega^\Upsilon &= -i\gamma_5\;\not\partial C^\Xi + M^\Xi - i\gamma_5 N^\Xi + \not V^\Xi\,,\\
M^\Upsilon &= -\mathrm{Tr}\,(\epsilon\gamma_5\lambda^\Xi)\,,\\
N^\Upsilon &= i\mathrm{Tr}\,(\epsilon\lambda^\Xi)\,,\\
V_\nu^\Upsilon &= -\mathrm{Tr}\,(\epsilon\gamma_5\lambda_\nu\lambda^\Xi) - \tfrac{1}{2}\mathrm{Tr}\,(\epsilon\gamma_5\,[\gamma_\nu\,,\not\partial]\lambda^\Xi)\,,\\
\lambda^\Upsilon &= -\not\partial M^\Xi - i\gamma_5\;\not\partial N^\Xi - i\gamma_5(D^\Xi + \Box C^\Xi) - \partial_\nu V^{\Xi\nu}\,,\\
D^\Upsilon &= i\mathrm{Tr}\,(\epsilon\;\not\partial\lambda^\Xi) + i\mathrm{Tr}\,(\epsilon\Box\omega^\Xi)\,,
\end{aligned}$$

and the condition (31.1.18) yields

$$M^\Upsilon = N^\Upsilon = D^\Upsilon + \Box C^\Upsilon = \partial^\lambda V_\lambda^\Upsilon = \lambda^\Upsilon + \not\partial\omega^\Upsilon = 0\,.$$

The vanishing of M^Υ and N^Υ tells us that λ^Ξ is a linear combination of the form

$$\lambda^\Xi\epsilon = f_\mu\gamma^\mu + g_\mu\gamma_5\gamma^\mu + k_{\mu\nu}\,[\gamma^\mu\,,\gamma^\nu]\,. \tag{31.1.35}$$

The vanishing of $\partial^\lambda V_\lambda^\Upsilon$ and $D^\Upsilon + \Box C^\Upsilon$ then implies that

$$\partial_\mu f^\mu = \partial_\mu g^\mu = 0\,. \tag{31.1.36}$$

Also, the vanishing of $\lambda^\Upsilon + \not\partial\omega^\Upsilon$ yields

$$D^\Xi = \frac{i}{2}\gamma_5\,[\gamma^\nu\,,\not\partial]V_\nu^\Xi\,. \tag{31.1.37}$$

The components of $\Delta_\mu(x,\theta)$ are

$$C_\mu^\Delta = i\,\mathrm{Tr}\,(\epsilon\gamma_5\gamma_\mu\omega^\Xi)\,, \tag{31.1.38}$$

$$\omega_\mu^\Delta = i\;\not\partial\gamma_\mu C^\Xi + \gamma_5\gamma_\mu M^\Xi - i\gamma_\mu N^\Xi + \gamma_5\gamma^\rho\gamma_\mu V_\rho^\Xi\,, \tag{31.1.39}$$

$$M_\mu^\Delta = -\,\mathrm{Tr}\,(\epsilon\gamma_\mu\lambda^\Xi)\,, \tag{31.1.40}$$

$$N_\mu^\Delta = i\,\mathrm{Tr}\,(\epsilon\gamma_5\gamma_\mu\lambda^\Xi)\,, \tag{31.1.41}$$

$$V_{\mu\nu}^\Delta = \mathrm{Tr}\,(\epsilon\gamma_\nu\gamma_\mu\lambda^\Xi) - \tfrac{1}{2}\mathrm{Tr}\,(\epsilon\,[\gamma_\nu\,,\not\partial]\,\gamma_\mu\,\omega^\Xi)\,, \tag{31.1.42}$$

$$\lambda_\mu^\Delta = \gamma_5\;\not\partial\gamma_\mu M^\Xi + i\;\not\partial\gamma_\mu N^\Xi - i\gamma_\mu(D^\Xi + \Box C^\Xi) + \gamma_5\gamma_\mu\partial^\nu V_\nu^\Xi\,, \tag{31.1.43}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

$$D_\mu^\Delta = i\,\mathrm{Tr}\,(\epsilon\ \partial\!\!\!/\gamma_5\gamma_\mu\lambda^\Xi) + i\,\Box\,\mathrm{Tr}\,(\epsilon\gamma_5\gamma_\mu\omega^\Xi)\,. \tag{31.1.44}$$

The symmetric part of Eq. (31.1.42) gives condition (31.1.20), with

$$v_\mu = -\,\mathrm{Tr}\,(\epsilon\gamma_\mu\omega^\Xi) + \text{constant}\,. \tag{31.1.45}$$

A linear combination of Eqs. (31.1.39) and (31.1.43) yields condition (31.1.21), with

$$\chi = 2\gamma_5 M^\Xi + 2i\,N^\Xi + \text{constant}\,. \tag{31.1.46}$$

We then use the antisymmetric part of Eq. (31.1.42), together with Eqs. (31.1.37), (31.1.38), and the identities

$$[\gamma_\nu\,,\,\gamma_\rho]\gamma_\mu - [\gamma_\mu\,,\,\gamma_\rho]\gamma_\nu = 2\eta_{\mu\rho}\gamma_\nu - 2\eta_{\nu\rho}\gamma_\mu + 2i\epsilon_{\nu\rho\mu\lambda}\,\gamma_5\gamma^\lambda\,,$$

$$\epsilon^{\nu\mu\kappa\sigma}[\gamma_\nu\,,\,\gamma_\mu] = 2i\gamma_5[\gamma^\kappa\,,\,\gamma^\sigma]\,,$$

and find the condition (31.1.22). Finally, Eqs. (31.1.40) and (31.1.41) along with Eqs. (31.1.35) and (31.1.36) yield the condition (31.1.23).

31.2 The Gravitational Action

To find a suitable gravitational action, we must construct a superfield that is invariant under the generalized gauge transformation $H_\mu \to H_\mu + \Delta_\mu$. As a starting point, we recall that the field b_μ defined by Eq. (31.1.27) is invariant under this transformation. By making successive supersymmetry transformations we can see that b_μ is the C-component of an 'Einstein' superfield E_μ, whose components are

$$C_\mu^E = b_\mu\,, \tag{31.2.1}$$

$$\omega_\mu^E = \frac{3}{2}L_\mu - \frac{1}{2}\gamma_\mu\gamma^\nu L_\nu\,, \tag{31.2.2}$$

$$M_\mu^E = \partial_\mu s\,, \qquad N_\mu^E = \partial_\mu p\,, \tag{31.2.3}$$

$$V_{\mu\nu}^E = -\frac{3}{2}E_{\mu\nu} + \frac{1}{2}\eta_{\mu\nu}E^\rho{}_\rho + \frac{1}{2}\epsilon_{\nu\mu\sigma\rho}\partial^\sigma b^\rho\,, \tag{31.2.4}$$

$$\lambda_\mu^E = \partial_\mu\gamma^\nu\omega_\nu^E - \partial\!\!\!/\omega_\mu^E\,, \tag{31.2.5}$$

$$D_\mu^E = \partial_\mu\partial^\nu b_\nu - \Box b_\mu\,, \tag{31.2.6}$$

where $E_{\mu\nu}$ is the linearized Einstein tensor

$$E_{\mu\nu} \equiv \frac{1}{2}\Big(\partial_\mu\partial_\nu h^\lambda{}_\lambda + \Box h_{\mu\nu} - \partial_\mu\partial^\lambda h_{\lambda\nu} - \partial_\nu\partial^\lambda h_{\lambda\mu} - \eta_{\mu\nu}\Box h^\lambda{}_\lambda + \eta_{\mu\nu}\partial^\lambda\partial^\rho h_{\lambda\rho}\Big) \tag{31.2.7}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

and

$$L^{\nu} \equiv i\,\epsilon^{\nu\mu\kappa\rho}\,\gamma_5\,\gamma_\mu\,\partial_\kappa\psi_\rho\,, \tag{31.2.8}$$

which will be shown in Section 31.3 to be the left-hand side of the wave equation for a free massless field of spin 3/2. For instance, by applying the supersymmetry transformation rules (26.2.11), (26.2.15), and (26.2.17) for C^H, V_μ^H, and D^H to Eqs. (31.2.1) and (31.1.27) we find

$$\delta C^{E\,\sigma} = i\left(\bar{\alpha}\gamma_5\left[\not{\partial}\lambda^{H\,\sigma} - \tfrac{1}{2}i\,\epsilon^{\nu\mu\kappa\sigma}\,\gamma_5\,\gamma_\mu\,\partial_\kappa\lambda_\nu^H + \partial^\sigma\partial^\kappa\omega_\kappa^H\right]\right)\,.$$

Comparison of this with the transformation rule (26.2.11) for C^E shows that

$$\omega^{E\,\sigma} = \not{\partial}\lambda^{H\,\sigma} - \tfrac{1}{2}i\,\epsilon^{\nu\mu\kappa\sigma}\,\gamma_5\,\gamma_\mu\,\partial_\kappa\lambda_\nu^H + \partial^\sigma\partial^\kappa\omega_\kappa^H\,.$$

We can use Eq. (31.1.26) to express λ_μ^H in terms of ψ_μ and ω_μ^H:

$$\lambda_\mu^H = \psi_\mu - \gamma_\mu\gamma^\rho\psi_\rho - \gamma_\mu\partial^\rho\omega_\rho^H\,,$$

and find that ω_σ^E may be written in terms of ψ_σ alone

$$\omega_\sigma^E = \not{\partial}\,\psi_\sigma - \partial_\sigma\gamma^\rho\psi_\rho - \tfrac{1}{2}i\,\epsilon_{\nu\mu\kappa\sigma}\gamma_5\,\gamma^\mu\,\partial^\kappa\psi^\nu\,.$$

Using the identity

$$\eta_{\mu\nu}\gamma_\lambda - \eta_{\mu\lambda}\gamma_\nu = i\,\epsilon_{\mu\lambda\nu\rho}\gamma_5\gamma^\rho - \tfrac{1}{2}i\,\gamma_\mu\gamma^\sigma\,\epsilon_{\sigma\lambda\nu\rho}\gamma_5\gamma^\rho\,,$$

we find Eq. (31.2.2) for ω_μ^E. Continuing in the same way gives the other formulas (31.2.3)–(31.2.6) for the components of E_μ and confirms that these components make up a real superfield.

We can now form a quadratic action for the metric superfield that is invariant under both supersymmetry and the extended gauge transformation $H_\mu \to H_\mu + \Delta_\mu$, by taking the Lagrangian density as

$$\mathscr{L}_E = \tfrac{4}{3}\left(E_\mu H^\mu\right)_D = E_{\mu\nu}h^{\mu\nu} - \tfrac{1}{2}\bar{\psi}_\mu L^\mu - \tfrac{4}{3}(s^2 + p^2 - b_\mu b^\mu)\,. \tag{31.2.9}$$

The factor 4/3 is chosen to give the kinematic Lagrangian for the graviton field a conventional sign and normalization: apart from terms involving $\partial^\mu h_{\mu\nu}$ or $h^\lambda{}_\lambda$, it is a sum of Klein–Gordon Lagrangians for the components of $h_{\mu\nu}$. The normalization of the gravitino field is discussed in the next section.

To see that the first two terms in the final expression in Eq. (31.2.9) are invariant under the gauge transformations (31.1.12) and (31.1.13), we note that $E_{\mu\nu}$ and L_μ are manifestly invariant under these transformations, and that the action is symmetric between the two factors of $h_{\mu\nu}$ or ψ_μ appearing in these terms. The absence in Eq. (31.2.9) of derivatives of s, p, and b_μ shows that these are auxiliary fields. The field equations make

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

them vanish for pure gravitation, but not when gravitation is coupled to matter.

Before taking up the coupling of matter and gravitation, let us pause to consider what value we should give the normalization constant κ in Eqs. (31.1.5) and (31.1.10). The Einstein–Hilbert action for a pure gravitational field is

$$I_{GR} = -\frac{1}{16\pi G}\int d^4x\,\sqrt{g}\,R\,, \tag{31.2.10}$$

where G is the Newton constant of gravitation, $g(x)$ is the determinant of the metric tensor $g_{\mu\nu}(x)$, and $R(x)$ is the curvature scalar calculated from $g_{\mu\nu}(x)$. To calculate I_{GR} for a weak gravitational field with $g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa h_{\mu\nu}$, we may recall that for arbitrary variations $\delta g_{\mu\nu}(x)$ in the gravitational field[6]

$$\delta I_{GR} = \frac{1}{16\pi G}\int d^4x\,\sqrt{g}\left[R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right]\delta g_{\mu\nu}\,, \tag{31.2.11}$$

where $R^{\mu\nu}(x)$ is the Ricci tensor calculated from $g_{\mu\nu}(x)$. For a weak field with $g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa h_{\mu\nu}$, the Ricci tensor is[7]

$$R^{\mu\nu} = \kappa\Big(\Box h^{\mu\nu} - \partial_\lambda\partial^\mu h^{\lambda\nu} - \partial_\lambda\partial^\nu h^{\lambda\mu} + \partial^\mu\partial^\nu h^\lambda{}_\lambda\Big)\,, \tag{31.2.12}$$

so for weak fields

$$R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R = 2\kappa E^{\mu\nu}\,, \tag{31.2.13}$$

and therefore Eq. (31.2.10) gives

$$\delta I_{GR} = \frac{\kappa^2}{4\pi G}\int d^4x\,E^{\mu\nu}\delta h_{\mu\nu}\,.$$

On the other hand, taking account of the symmetry of $\int d^4x\,E_{\mu\nu}\,h^{\mu\nu}$ between the two factors of h it contains, we have

$$\delta\int d^4x\,E^{\mu\nu}\,h_{\mu\nu} = 2\int d^4x\,E^{\mu\nu}\,\delta h_{\mu\nu}\,.$$

In order for the term $E_{\mu\nu}\,h^{\mu\nu}$ in Eq. (31.2.9) to give the usual gravitational Lagrangian density, it is therefore necessary to take

$$\kappa = \sqrt{8\pi G}\,. \tag{31.2.14}$$

Let's now combine the Einstein Lagrangian density (31.2.9) with the interaction (31.1.34) between gravitation and matter and with the matter Lagrangian $\mathscr{L}_M$, to form a total Lagrangian density:

$$\begin{aligned}\mathscr{L} = \mathscr{L}_M + E_{\mu\nu}h^{\mu\nu} - \tfrac{1}{2}\bar\psi_\mu L^\mu - \tfrac{4}{3}(s^2 + p^2 - b_\mu b^\mu)\\ +2\kappa\Big[\tfrac{1}{2}\mathscr{R}_\sigma b^\sigma + \tfrac{1}{4}\bar S^\sigma\psi_\sigma - \mathscr{M}\,s - \mathscr{N}\,p + \tfrac{1}{2}T^{\kappa\sigma}h_{\sigma\kappa}\Big]\,.\end{aligned} \tag{31.2.15}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The field equations for the auxiliary fields give

$$s = -6\kappa\mathscr{M}/8\,, \qquad p = -6\kappa\mathscr{N}/8\,, \qquad b_\mu = -6\kappa\mathscr{R}_\mu/16\,. \tag{31.2.16}$$

Using this to eliminate the auxiliary fields, the Lagrangian density (31.2.15) now gives

$$\begin{aligned}\mathscr{L} = \mathscr{L}_M + E_{\mu\nu}h^{\mu\nu} - \tfrac{1}{2}\bar{\psi}_\mu L^\mu + \tfrac{3}{4}\kappa^2(\mathscr{M}^2 + \mathscr{N}^2 - \tfrac{1}{4}\mathscr{R}_\mu\mathscr{R}^\mu)\\ + \tfrac{1}{2}\kappa\bar{S}^\sigma\psi_\sigma + \kappa T^{\kappa\sigma}h_{\sigma\kappa}\,.\end{aligned} \tag{31.2.17}$$

The sources of the fields ψ_μ and $h_{\mu\nu}$ are of order κ, so we can regard these fields as being of this order, which makes all terms in Eq. (31.2.17) beyond $\mathscr{L}_M$ of order κ^2.

In the vacuum only the scalar fields s and p can have tree-approximation expectation values, giving the vacuum an energy density, to order G:

$$\rho_{VAC} = -\mathscr{L}_{VAC} = -\mathscr{L}_{M\,VAC} - \tfrac{3}{4}\kappa^2(\mathscr{M}^2 + \mathscr{N}^2)\,. \tag{31.2.18}$$

The negative sign of the term in the vacuum energy density of first order in G is a characteristic difference between theories of supergravity and ordinary supersymmetry.

For instance, in the theory of a single left-chiral superfield Φ with superpotential $f(\Phi)$, the zeroth-order vacuum energy is $-\mathscr{L}_{M\,VAC} = |df(\phi)/d\phi|^2$, while Eqs. (31.1.33) and (26.7.27) give

$$\mathscr{M} + i\mathscr{N} = -\frac{1}{3}\left[\phi\frac{df(\phi)}{d\phi} - 3f(\phi)\right]\,,$$

so to first order in G, the total vacuum energy is

$$\rho_{VAC} = \left|\frac{df(\phi)}{d\phi}\right|^2 - \frac{8\pi G}{3}\left|\phi\frac{df(\phi)}{d\phi} - 3f(\phi)\right|^2\,, \tag{31.2.19}$$

where ϕ is the scalar component of Φ. This is for the definition of the metric for which the energy-momentum tensor is given by Eq. (26.7.42), so that in particular $T^\lambda{}_\lambda$ vanishes for $f = 0$. For other definitions the vacuum energy would be changed by terms of order $8\pi G|\phi|^2|df(\phi)/d\phi|^2$. However, this ambiguity is unimportant in calculating the minimum value of the vacuum energy. Inspection of Eq. (31.2.19) shows that if $f(\phi)$ is stationary at some point ϕ_0, then ρ_{VAC} has a local minimum for $|\phi - \phi_0|$ and $|df/d\phi|$ of first order in G, and at any such point the vacuum energy to first order in G is

$$\rho_{VAC} = -24\pi G|f(\phi_0)|^2\,. \tag{31.2.20}$$

There is an algebraic reason why vacuum states with unbroken supersymmetry in supergravity theories cannot have positive energy density.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The solutions of the Einstein field equations for a uniform non-zero vacuum energy density ρ_V take the form of a de Sitter space for $\rho_V > 0$ and anti-de Sitter space for $\rho_V < 0$. These spaces may be described as the surfaces

$$x_5^2 \pm \eta_{\mu\nu} x^\mu x^\nu = R^2 \,, \tag{31.2.21}$$

in a quasi-Euclidean five-dimensional space with line element

$$ds^2 = \eta_{\mu\nu} x^\mu x^\nu \pm dx_5^2 \,, \tag{31.2.22}$$

with the upper sign for de Sitter space and the lower sign for anti-de Sitter space. The spacetime symmetry of these spaces is no longer the Poincaré group consisting of translations and Lorentz transformations, but instead $O(4,1)$ for de Sitter space and $O(3,2)$ for anti-de Sitter space, where $O(n,m)$ is the group of linear transformations that leave invariant a diagonal metric with n positive and m negative elements on the main diagonal. The supersymmetry that is unbroken in theories with a de Sitter or anti-de Sitter vacuum state therefore has the simple group $O(4,1)$ or $O(3,2)$, respectively, as its spacetime symmetry. Nahm[7a] has cataloged all supersymmetries with simple spacetime symmetries. There are simple $O(3,2)$ supersymmetries, as well as N-extended $O(3,2)$ supersymmetries, but for $O(4,1)$ there is only an $N = 2$ supersymmetry. We have been considering $N = 1$ supergravity theories here, so they can have vacuum states with unbroken supersymmetry and $\rho_V < 0$, which gives $O(3,2)$ spacetime symmetry, but not with unbroken supersymmetry and $\rho_V > 0$, which would give $O(4,1)$ spacetime symmetry.

The possibility of vacuum field configurations with negative energy density may seem at first sight to threaten the stability of our universe. It is common for $f(\phi)$ to have several stationary points, at which it takes different values. Even if we fine-tune the parameters in $f(\phi)$ so that $f(\phi)$ vanishes at one of these stationary points, accounting for the observed nearly flat space of our universe, any other stationary point with a non-zero value of $f(\phi)$ will yield a *lower* vacuum energy, raising the possibility of a collapse to a state of negative energy density, with a metric that would have to be of the 'anti-de Sitter' form rather than flat.

Fortunately, the value (31.2.20) is just barely insufficient to compensate for the positive energy that has to go into surface tension in making a bubble of anti-de Sitter space, so that the transition from ordinary flat space is not actually favored energetically. Coleman and de Luccia[8] have applied the equations of general relativity to a bubble of negative internal energy density $-\epsilon$ and positive surface tension S in a flat space of energy density zero, and have shown that any such bubble that does not entail gravitational singularities will have positive energy if

$$\epsilon \leq 6\pi G S^2 \,. \tag{31.2.23}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The surface tension is the energy per area in the bubble surface, given to zeroth order in G by the integral of the energy density through the bubble wall:

$$S_1 = \int_{r_-}^{r_+} dr \left[\left| \frac{d\phi}{dr} \right|^2 + \left| \frac{df(\phi)}{d\phi} \right|^2 \right] , \tag{31.2.24}$$

with r_- and r_+ taken just inside and just outside the bubble wall. This can be rewritten as

$$S_1 = \int_{r_-}^{r_+} dr \left[\left| \frac{d\phi}{dr} + \xi \left(\frac{df(\phi)}{d\phi} \right)^* \right|^2 - 2\mathrm{Re} \left(\xi^* \frac{d\phi}{dr} \frac{df(\phi)}{d\phi} \right) \right] ,$$

where ξ is an arbitrary phase factor with $|\xi| = 1$. The integral in the second term is trivial: since we assume that $\phi(r_+)$ is at a value where $f(\phi)$ is stationary and vanishes, while $\phi(r_-)$ is at some value ϕ_0 where $f(\phi)$ is stationary but non-zero, the integral is

$$\int_{r_-}^{r_+} dr \, \frac{d\phi}{dr} \frac{df(\phi)}{d\phi} = -f(\phi_0) .$$

To maximize this term, we take $\xi = f(\phi_0)/|f(\phi_0)|$, and obtain the inequality[9]

$$S_1 \geq 2\,|f(\phi_0)| , \tag{31.2.25}$$

with equality attained only if (as is usual) there is a solution of the differential equation $d\phi/dr = -\xi(df/d\phi)^*$ with the appropriate boundary conditions. The inequality (31.2.23) is thus satisfied if the internal energy density is not less than $-24\pi G|f(\phi_0)|^2$, which is precisely the value (31.2.20). This calculation leaves open the possibility of an instability of flat space due to radiative corrections, but the reader need not worry: it has been proved that in supergravity theories in which there is a vacuum field configuration with zero vacuum energy, the energy of any disturbance in the fields which is limited to a finite region is positive.[10]

* * *

Eq. (26.7.48) shows that the energy-momentum tensor for a set of left-chiral scalar superfields Φ_n contains a term

$$\Delta T^{\mu\nu} = \frac{1}{3} \left(\eta^{\mu\nu} \Box - \partial^\mu \partial^\nu \right) \sum_n |\phi_n|^2 . \tag{31.2.26}$$

Integrating by parts, the corresponding interaction $\kappa h_{\mu\nu} \Delta T^{\mu\nu}$ makes a contribution to the action of the form

$$\frac{\kappa}{3} \int \delta^4 x \sum_n |\phi_n|^2 \left(\eta^{\mu\nu} \Box - \partial^\mu \partial^\nu \right) h_{\mu\nu} = \frac{1}{6} R^{(1)} \sum_n |\phi_n|^2 , \tag{31.2.27}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

where $R^{(1)}$ is the curvature scalar in linear approximation. This is added to the usual Einstein–Hilbert action $-\sqrt{g}R/2\kappa^2$ (which appears in Eq. (31.2.17) as the term $E_{\mu\nu}h^{\mu\nu}$) and has the effect of replacing the coefficient of this term with

$$-\frac{1}{2\kappa^2}+\frac{1}{6}\sum_n|\phi_n|^2=-\frac{1}{2\kappa^2}\left(1-\frac{\kappa^2}{3}\sum_n|\phi_n|^2\right)\,. \tag{31.2.28}$$

In order to restore the usual constant of gravitation, we can subject the metric to a *Weyl transformation*, replacing the vierbein $e^a{}_\mu$ with

$$\tilde{e}^a{}_\mu=e^a{}_\mu\sqrt{1-\frac{\kappa^2}{3}\sum_n|\phi_n|^2}\,. \tag{31.2.29}$$

That is, we replace the metric $g_{\mu\nu}$ with

$$\tilde{g}_{\mu\nu}=\left(1-\frac{\kappa^2}{3}\sum_n|\phi_n|^2\right)g_{\mu\nu}\,, \tag{31.2.30}$$

or for weak fields

$$\tilde{h}_{\mu\nu}=h_{\mu\nu}-\frac{\kappa}{6}\sum_n|\phi_n|^2\,\eta_{\mu\nu}\,. \tag{31.2.31}$$

The weak field Einstein tensor (31.2.7) for the new metric is

$$\begin{aligned}\tilde{E}_{\mu\nu}&\equiv\frac{1}{2}\left(\partial_\mu\partial_\nu\tilde{h}^\lambda{}_\lambda+\Box\tilde{h}_{\mu\nu}-\partial_\mu\partial^\lambda\tilde{h}_{\lambda\nu}-\partial_\nu\partial^\lambda\tilde{h}_{\lambda\mu}-\eta_{\mu\nu}\Box\tilde{h}^\lambda{}_\lambda+\eta_{\mu\nu}\partial^\lambda\partial^\rho\tilde{h}_{\lambda\rho}\right)\\&=E_{\mu\nu}-\frac{\kappa}{6}\left(\partial_\mu\partial_\nu-\eta_{\mu\nu}\Box\right)\sum_n|\phi_n|^2\,.\end{aligned} \tag{31.2.32}$$

The sum of the Einstein term in the original action and the term (31.2.27) is then

$$\begin{aligned}\int d^4x\left[h_{\mu\nu}E^{\mu\nu}+\frac{\kappa}{3}\sum_n|\phi_n|^2\left(\eta^{\mu\nu}\Box-\partial^\mu\partial^\nu\right)h_{\mu\nu}\right]&=\int d^4x\left[\tilde{h}_{\mu\nu}\tilde{E}^{\mu\nu}\right.\\&\left.+\frac{\kappa^2}{12}\left(\partial_\mu\sum_n|\phi_n|^2\right)\left(\partial^\mu\sum_n|\phi_n|^2\right)\right]\,,\end{aligned} \tag{31.2.33}$$

so the effective gravitational constant is now actually constant. This redefinition of the metric also produces a change in the potential. The original Lagrangian density contains a term $-e\sum_n|\partial f(\phi)/\partial\phi_n|^2$, which in terms of the new vierbein reads

$$\begin{aligned}-e\sum_n|\partial f(\phi)/\partial\phi_n|^2&=-\tilde{e}\sum_n|\partial f(\phi)/\partial\phi_n|^2\\&\quad-\frac{2\kappa^2}{3}\left(\sum_n|\phi_n|^2\right)\sum_n|\partial f(\phi)/\partial\phi_n|^2\,.\end{aligned}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

With the new definition of the metric, the potential (31.2.19) is therefore replaced with

$$\rho_{VAC} = \sum_n \left| \frac{\partial f(\phi)}{\partial \phi_n} \right|^2 - \frac{\kappa^2}{3} \left| \sum_n \phi_n \frac{\partial f(\phi)}{\partial \phi_n} - 3 f(\phi) \right|^2 + \frac{2\kappa^2}{3} \left(\sum_n |\phi_n|^2 \right) \sum_n \left| \frac{\partial f(\phi)}{\partial \phi_n} \right|^2 . \tag{31.2.34}$$

The new term does not change the value of ρ_{VAC} at its stationary point to order κ^2, so no change is needed in our previous discussion of vacuum stability.

31.3 The Gravitino

In this section we will use the weak field formalism developed in Sections 31.1–31.3 to derive some properties of the gravitino. In particular, by using an argument of continuity as $G \to 0$ we shall obtain a formula for the gravitino mass when supersymmetry is spontaneously broken that is valid to first order in G but to all orders in all other interactions. (The original derivation of this formula will be given in Section 31.6.)

First we must verify that the term $-\frac{1}{2}\bar{\psi}_\mu L^\mu$ in Eq. (31.2.9) is the correct free-field Lagrangian for a massless self-charge-conjugate particle of spin 3/2. The time-honored approach to the construction of suitable free-field Lagrangians for particles with spin is to guess at the Lagrangian and then check that it gives a physically satisfactory field equation and propagator. This has led to some uncertainty for particles of spin 3/2 — for instance, what is usually called the Rarita–Schwinger Lagrangian in papers on supersymmetry is not the Lagrangian originally proposed by Rarita and Schwinger.[11] Here we shall follow an approach in the spirit of Section 6.2: we will first derive the propagator from the requirements of Lorentz invariance for a massive particle of spin 3/2 and then invert it to find the Lagrangian. We work here with massive gravitinos for the sake of simplicity and because in the real world we must take account of the breakdown of supersymmetry, but we will be able to apply the results obtained in this way to the case of massless gravitinos by noting that supersymmetry current conservation makes the zero-mass singularities in the propagator inconsequential.

A spinor field ψ^μ with an extra vector index belongs to the $[(\frac{1}{2},0) + (0,\frac{1}{2})] \times (\frac{1}{2},\frac{1}{2})$ representation of the homogeneous Lorentz group. To isolate the $(1,\frac{1}{2}) + (\frac{1}{2},1)$ parts of the free field, we impose the irreducibility

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

condition

$$\gamma_\mu \psi^\mu = 0\,. \tag{31.3.1}$$

Rotational invariance and Eq. (31.3.1) tell us that the matrix elements of this field between the vacuum and a massive spin 3/2 particle of momentum $\mathbf{q} = 0$ and spin z-component s will satisfy the conditions

$$\langle 0|\psi^0(0)|s\rangle = 0 \tag{31.3.2}$$

and

$$\sum_{s=-3/2}^{3/2} \langle 0|\psi^i(0)|s\rangle\, \langle 0|\psi^j(0)|s\rangle^* \propto \delta_{ij} - \tfrac{1}{3}\gamma_i\gamma_j\,, \tag{31.3.3}$$

with a coefficient that may depend on the rotationally invariant matrix $\beta \equiv i\gamma^0$. With the usual Dirac convention that $\beta\langle 0|\psi^i(0)|s\rangle = \langle 0|\psi^i(0)|s\rangle$ (chosen to simplify the space inversion property of the field) and a conventional choice of normalization of the $\beta = +1$ component analogous to Eq. (5.5.23), Eq. (31.3.3) may be written

$$\sum_{s=-3/2}^{3/2} \langle 0|\psi^i(0)|s\rangle\langle 0|\psi^j(0)|s\rangle^* = (2\pi)^{-3}\left(\frac{1+\beta}{2}\right)\left[\delta_{ij} - \tfrac{1}{3}\gamma_i\gamma_j\right]. \tag{31.3.4}$$

It follows that the momentum-space propagator for a spin 3/2 particle of mass m_g takes the form

$$\Delta^{\mu\nu}(q) = \frac{P^{\mu\nu}(q)}{q^2 + m_g^2 - i\epsilon}\,, \tag{31.3.5}$$

where $P^{\mu\nu}(q)$ is a Lorentz-covariant polynomial in the four-vector q, subject to the conditions that for $\mathbf{q} = 0$ and $q^0 = m_g$, we have

$$P^{ij} = \left(\frac{1+\beta}{2}\right)\left[\delta_{ij} - \tfrac{1}{3}\gamma_i\gamma_j\right], \qquad P^{i0} = P^{0i} = P^{00} = 0\,. \tag{31.3.6}$$

Apart from possible terms that vanish on the mass shell (and whose effect therefore would be the same as direct current–current interactions) the unique covariant function with this limit is

$$P^{\mu\nu}(q) = \left(\eta^{\mu\nu} + \frac{q^\mu q^\nu}{m_g^2}\right)\left(-i\not{q} + m_g\right) - \frac{1}{3}\left(\gamma^\mu - i\frac{q^\mu}{m_g}\right)\left(i\not{q} + m_g\right)\left(\gamma^\nu - i\frac{q^\nu}{m_g}\right). \tag{31.3.7}$$

(The difference between Eq. (31.3.7) and any other covariant function with the limit (31.3.6) would be a covariant function whose components all vanish for $\mathbf{q} = 0$ and $q^0 = m_g$ and hence vanish everywhere on the mass shell.) The free-field Lagrangian density is then of the form

$$\mathscr{L}_0 = -\tfrac{1}{2}\left(\bar{\psi}^\mu D_{\mu\nu}(-i\partial)\,\psi^\nu\right), \tag{31.3.8}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

where

$$\Delta^{\mu\nu}(q)\, D_{\nu\lambda}(q) = \delta^{\mu}_{\lambda}\,. \tag{31.3.9}$$

A tedious but straightforward calculation gives

$$D_{\nu\lambda}(q) = -\epsilon_{\nu\mu\kappa\lambda}\,\gamma_5\,\gamma^{\mu}\,q_{\kappa} - \tfrac{1}{2}m_g\,[\gamma_{\nu}\,,\,\gamma_{\lambda}]\,, \tag{31.3.10}$$

so that the Lagrangian density (31.3.8) is

$$\mathscr{L}_0 = -\tfrac{1}{2}i\,\epsilon^{\nu\mu\kappa\lambda}\left(\bar{\psi}_{\nu}\,\gamma_5\,\gamma_{\mu}\,\partial_{\kappa}\,\psi_{\lambda}\right) + \tfrac{1}{4}m_g\left(\bar{\psi}_{\nu}\,[\gamma^{\nu}\,,\,\gamma^{\lambda}]\,\psi_{\lambda}\right). \tag{31.3.11}$$

For $m_g = 0$ this result verifies that the term $-\frac{1}{2}\bar{\psi}_{\mu}L^{\mu}$ in Eq. (31.2.9) is the correct conventionally normalized free-field Lagrangian for a massless self-charge-conjugate particle of spin 3/2. In the limit $m_g \to 0$ the propagator given by Eqs. (31.3.5) and (31.3.7) is singular (which just reflects the impossibility shown in Section 5.9 of forming a $(1, \frac{1}{2}) + (\frac{1}{2}, 1)$ field from the creation and annihilation operators for a massless particle of helicity $\pm 3/2$), but the terms in Eq. (31.3.7) for $P^{\mu\nu}(q)$ that blow up for $m_g \to 0$ are all proportional to q^{μ} and/or q^{ν}, and hence give no contribution when the current with which ψ_{μ} interacts is conserved.

As a further check on the validity of the Lagrangian density (31.3.11), including the peculiar-looking mass term, we note that it yields a field equation

$$-i\,\epsilon^{\nu\mu\kappa\lambda}\,\gamma_5\,\gamma_{\mu}\,\partial_{\kappa}\,\psi_{\lambda} + \tfrac{1}{2}m_g\,[\gamma^{\nu}\,,\,\gamma^{\lambda}]\,\psi_{\lambda} = 0\,. \tag{31.3.12}$$

Taking the divergence of this equation yields the result that

$$[\not{\partial}\,,\,\gamma^{\lambda}]\,\psi_{\lambda} = 0\,.$$

Also, contracting Eq. (31.3.12) with γ_{ν} shows that

$$\gamma_{\nu}\psi^{\nu} \propto \epsilon^{\nu\mu\kappa\lambda}\,\gamma_5\,\gamma_{\nu}\gamma_{\mu}\,\partial_{\kappa}\,\psi_{\lambda} \propto [\not{\partial}\,,\,\gamma^{\lambda}]\,\psi_{\lambda} = 0\,,$$

so that the irreducibility condition (31.3.1) is satisfied for free fields (though not necessarily when interactions are taken into account). From these two results there follows another irreducibility condition

$$\partial_{\lambda}\psi^{\lambda} = \tfrac{1}{2}\{\not{\partial}\,,\,\gamma_{\lambda}\}\,\psi^{\lambda} = \tfrac{1}{2}[\not{\partial}\,,\,\gamma_{\lambda}]\,\psi^{\lambda} = 0\,.$$

Using these irreducibility results allows the field equation (31.3.12) to be put in the form of a Dirac equation

$$\left(\not{\partial} + m_g\right)\psi^{\lambda} = 0\,, \tag{31.3.13}$$

which among other things shows that this is the free field of a particle of mass m_g.

We will now consider the effects in supergravity theories of a spontaneous breakdown of supersymmetry. Broken global supersymmetry entails

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

the existence of a massless particle of spin 1/2, the goldstino, but in supergravity theories the goldstino field χ can be eliminated by a gauge transformation $\psi_\mu \to \psi_\mu - \partial_\mu \chi$. Since the gauge is then fixed by the condition that the goldstino field is eliminated in this way, gauge invariance no longer keeps the gravitino massless, and it acquires a mass (to be denoted m_g from now on), in much the same way as we saw in Section 21.3 that the vector bosons $W^\pm$ and Z^0 get masses from the spontaneous breakdown of the $SU(2) \times U(1)$ gauge symmetry of the electroweak interactions.

As discussed at the beginning of Chapter 28, if supersymmetry is relevant at all to accessible phenomena then the characteristic energy scale at which it is broken must be much less than the Planck mass. In this case we may use a continuity argument to give a universal formula for the gravitino mass m_g. According to Eq. (31.1.34), the gravitino field ψ_μ couples to the supersymmetry current S^μ with a coupling constant $\frac{1}{2}\kappa = \frac{1}{2}\sqrt{8\pi G}$, so the exchange of a virtual gravitino of four-momentum q in a transition $A + B \to C + D$ contributes to the invariant amplitude a term

$$M(A + B \to C + D) = \tfrac{1}{4}(8\pi G) \langle C|\bar{S}_\mu|A\rangle_N \Delta^{\mu\nu}(q) \langle D|S_\nu|B\rangle_N , \quad (31.3.14)$$

where the subscript N indicates that the one-goldstino pole at $q^2 = 0$ has been removed from the matrix element of the supersymmetry current. For a supersymmetry-breaking scale sufficiently small compared with the Planck mass, there will be a range of momentum transfers that are much larger than the gravitino mass but much less than the Planck mass. For such momenta, the matrix element is dominated by the $1/m_g^2$ terms in the propagator numerator (31.3.7)

$$M(A + B \to C + D) \to \tfrac{1}{4}(8\pi G) \langle C|\bar{S}_\mu|A\rangle_N \left(\frac{-2i \not{q} q^\mu q^\nu}{3m_g^2 q^2}\right) \langle D|S_\nu|B\rangle_N . \quad (31.3.15)$$

But for sufficiently large Planck mass the coupling of the gravitino becomes negligible, and the matrix element must be the same as would have been produced by goldstino exchange in a theory without gravitinos. According to Eq. (29.2.10), this is

$$M(A+B \to C+D) \to \langle C|\bar{S}_\mu|A\rangle_N \left(\frac{-i \not{q}}{q^2}\right) \left(\frac{q^\mu q^\nu}{F^2}\right) \langle D|S_\nu|B\rangle_N , \quad (31.3.16)$$

where F is the parameter characterizing the strength of supersymmetry breaking (here taken real), defined so that the vacuum energy density is $F^2/2$. In order for Eqs. (31.3.15) and (31.3.16) to agree, the gravitino mass must have the value

$$m_g = \sqrt{\frac{4\pi G F^2}{3}} . \quad (31.3.17)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

This formula is valid only to lowest order in GF^2, but to all orders (and even non-perturbatively) in the non-gravitational interactions responsible for the spontaneous breakdown of supersymmetry.

It is convenient for some purposes to express m_g in terms of the expectation values $\langle s \rangle$ and $\langle p \rangle$ of the spinless auxiliary gravitational fields. Note that for the vacuum state to have zero spacetime curvature, the vacuum energy density $F^2/2$ of matter fields must be balanced by the negative vacuum energy of gravitation and its interaction with the hidden sector fields, which is given in terms of $\langle s \rangle$ and $\langle p \rangle$ by Eqs. (31.2.18) and (31.2.16) as $-(4/3)(\langle s \rangle^2 + \langle p \rangle^2)$, so

$$F^2/2 = (4/3)(\langle s \rangle^2 + \langle p \rangle^2) \,. \tag{31.3.18}$$

We can therefore write Eq. (31.3.17) as

$$m_g = \frac{2\kappa}{3}\sqrt{\langle s \rangle^2 + \langle p \rangle^2} \,. \tag{31.3.19}$$

It is sometimes convenient to introduce a complex gravitino mass, defined as

$$\tilde{m}_g \equiv \frac{2\kappa}{3}\Big(\langle s \rangle + i\langle p \rangle\Big) \,, \tag{31.3.20}$$

whose absolute magnitude is the physical gravitino mass (31.3.19).

31.4 Anomaly-Mediated Supersymmetry Breaking

In Section 28.3 the possibility was raised that supersymmetry may be broken in some sort of hidden sector of superfields that do not carry the $SU(3) \times SU(2) \times U(1)$ quantum numbers of the standard model, and communicated to observable particles gravitationally. In this section we will deal with one class of supersymmetry-breaking effects in the minimum supersymmetric standard model, those of first order in $\kappa \equiv \sqrt{8\pi G}$. This includes the gaugino masses and the parameters A_{ij} and B in the Lagrangian density (28.4.1). Other supersymmetry-breaking effects such as squark and slepton squared masses are of second order in κ, and will be taken up in Section 31.7, when we consider gravity-mediated supersymmetry breaking using the general supergravity formalism described in Section 31.6.

We can find the effects of gravity-mediated supersymmetry breaking to first order in κ by simply replacing the component fields of the gravitational supermultiplet in the interaction (31.1.34) with their expectation values. The only ones of these component fields that can acquire non-vanishing vacuum expectation values from the spontaneous breakdown

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

of supersymmetry in a hidden sector of matter superfields are the spinless auxiliary fields s and p, so with Eq. (31.1.33) this gives a first-order supersymmetry-breaking interaction

$$\mathscr{L}^{(1)} = 2\kappa\left[-A^X\langle p\rangle + B^X\langle s\rangle\right] = 3\,\mathrm{Im}\left[\tilde{m}_g^*\,(A^X + iB^X)\right], \tag{31.4.1}$$

where $\tilde{m}_g$ is the complex gravitino mass (31.3.20), and A^X and B^X are the A- and B-components of the real chiral scale-non-invariance superfield X discussed in Section 26.7.

We showed in Section 26.7 that, for a renormalizable theory of left-chiral superfields Φ_n with superpotential $f(\Phi)$, the X superfield is given by

$$X = \frac{2}{3}\,\mathrm{Im}\left[\sum_n \Phi_n \frac{\partial f(\Phi)}{\partial \Phi_n} - 3f(\Phi)\right]. \tag{31.4.2}$$

This can be put in a form that allows an immediate extension to more general theories by writing the coupling parameters of the superpotential in terms of dimensionless parameters and a parameter $\mathscr{M}$ with the dimensions of mass. Since the superpotential has the dimensions of (mass)^3, we have

$$\mathscr{M}\frac{\partial f(\Phi)}{\partial \mathscr{M}} + \sum_n \Phi_n \frac{\partial f(\Phi)}{\partial \Phi_n} = 3f(\Phi)\,. \tag{31.4.3}$$

Therefore Eq. (31.4.2) may be written

$$X = \frac{2}{3}\,\mathrm{Im}\left[\mathscr{M}\frac{\partial f(\Phi)}{\partial \mathscr{M}}\right]. \tag{31.4.4}$$

This formula can be generalized to give the contribution to X of the scale dependence of any sort of $\mathscr{F}$-term in the Lagrangian. A term $2\mathrm{Re}\,[f(\Phi, W)]_{\mathscr{F}}$ in the Lagrangian makes a contribution to X given by the obvious generalization of Eq. (31.4.4):

$$X = \frac{2}{3}\mathrm{Im}\left[\mathscr{M}\frac{\partial f(\Phi, W)}{\partial \mathscr{M}}\right]. \tag{31.4.5}$$

(There is also a contribution to X from any mass scale dependence in the D-terms in the Lagrangian.) Comparing Eq. (31.4.5) with Eqs. (26.3.10) and (26.3.13), we see that

$$A^X + iB^X = \frac{2\mathscr{M}}{3i}\frac{\partial}{\partial \mathscr{M}}\left[f(\Phi, W)\right]_{\theta=0} = \frac{2\mathscr{M}}{3i}\frac{\partial f(\phi, \lambda_L)}{\partial \mathscr{M}}\,. \tag{31.4.6}$$

There will then be a supersymmetry-breaking term in the effective Lagrangian of first order in κ, given by Eqs. (31.4.1) and (31.4.6) as

$$\mathscr{L}_f^{(1)} = -2\,\mathrm{Re}\left[\tilde{m}_g^*\,\frac{\mathscr{M}\,\partial f(\phi, \lambda_L)}{\partial \mathscr{M}}\right]. \tag{31.4.7}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Eq. (31.4.7) applies not only for terms in the Lagrangian with explicit scale dependence, but also for the scale dependence of coupling constants described by the renormalization group.[11a] This scale dependence arises from the quantum mechanical anomaly that gives a non-zero value to the trace of the energy-momentum tensor and also to the divergence of the R-current, so the observable supersymmetry-breaking effects that arise in this way are said to be *anomaly-mediated.*

Consider in particular the kinematic term $\mathscr{L}_{\text{gauge}}$ for renormalizable supersymmetric gauge theories, given by Eqs. (27.3.22) and (27.3.23) as

$$\mathscr{L}_{\text{gauge}} = -\frac{1}{2g^2}\text{Re}\sum_{A\alpha\beta}\left[\epsilon_{\alpha\beta}W_{A\alpha}W_{A\beta}\right]_{\mathscr{F}} . \tag{31.4.8}$$

This does not depend explicitly on any mass scale, but the coupling constant g has the familiar dependence on a renormalization scale $\mathscr{M}$, given by the renormalization group equation

$$\mathscr{M}\frac{dg(\mathscr{M})}{d\mathscr{M}} = \beta\Big(g(\mathscr{M})\Big) . \tag{31.4.9}$$

Then Eq. (31.4.7) shows that the gauge Lagrangian (31.4.8) yields a supersymmetry-breaking term in the Lagrangian

$$\mathscr{L}^{(1)}_{\text{gauge}} = -\frac{\beta(g)}{g^3}\text{Re}\left[\tilde{m}_g^*\sum_{A\alpha\beta}\epsilon_{\alpha\beta}\lambda_{AL\alpha}\lambda_{AL\beta}\right] , \tag{31.4.10}$$

where $\lambda_{AL\alpha}$ is the left-handed part of the gaugino field, normalized like $W_{A\alpha}$ by multiplication with the gauge coupling g, so that g does not appear in the structure constants or the interaction of the gauge superfield with quark superfields. Taking account of this normalization convention, the gaugino mass equals g^2 times the absolute value of the coefficient of $\frac{1}{2}\sum_{A\alpha\beta}\epsilon_{\alpha\beta}\lambda_{AL\alpha}\lambda_{AL\beta}$, or

$$m_{\text{gaugino}} = m_g\left|\frac{\beta(g)}{g}\right| . \tag{31.4.11}$$

In this formula m_{gaugino} and g are cut-off-dependent bare parameters in the Lagrangian, governed by Wilsonian renormalization group equations. In Section 27.6 we saw that $\beta(g)$ arises purely from one-loop diagrams, so that $\beta(g) = bg^3$, with b a constant, and therefore

$$m_{\text{gaugino}} = m_g\,|b|\,g^2 . \tag{31.4.12}$$

The physical mass of the gaugino differs from Eq. (31.4.12) by corrections of higher order in g but, since we know that gauginos must be much heavier than the characteristic scale of quantum chromodynamics, these corrections are small for the gluino as well as for the wino and bino.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Supersymmetry does not allow any explicitly scale-dependent terms in the Lagrangian for the gluon and quark superfields, so, with electroweak interactions neglected, Eq. (31.4.12) gives the only contribution of order κ to the gluino mass. For three quark generations, Eq. (28.2.10) gives $b = -3g_s^3/16\pi^2$. Taking $g_s^2/4\pi = 0.118$, Eq. (31.4.12) gives the gluino mass as

$$m_{\text{gluino}} = \frac{3g_s^2 m_g}{16\pi^2} = 2.8 \times 10^{-2}\, m_g \,. \tag{31.4.13}$$

On the other hand, there is a scale-dependent interaction in the Lagrangian for the Higgs superfields, produced by the μ-term $-\mu(H_2^{\mathrm{T}} e H_1)$ in Eq. (28.1.7), so there is a term in the Lagrangian given by Eq. (31.4.7) as $2\mathrm{Re}\,[\tilde{m}_g^* \mu(\mathscr{H}_2^{\mathrm{T}} e \mathscr{H}_1)]$. Comparing with the $B\mu$-term in Eq. (28.4.1), we see that this gives

$$B = -\tilde{m}_g^* \,. \tag{31.4.14}$$

With three generations of quarks and lepton superfields and one pair of Higgs doublet superfields H_1 and H_2, Eqs. (28.2.8) and (28.2.9) give $b = 11/16\pi^2$ for the $U(1)$ gauge coupling g' and $b = 1/16\pi^2$ for the $SU(2)$ gauge coupling g. Taking $g'^2/4\pi = 0.0102$ and $g^2/4\pi = 0.0338$, Eq. (31.4.11) would give masses $8.9 \times 10^{-3} m_g$ and $2.7 \times 10^{-3} m_g$ for the bino and wino, respectively. However, the bino and wino masses also receive contributions of order $g'^2 m_g/16\pi^2$ and $g^2 m_g/16\pi^2$, respectively, from diagrams in which the bino or wino is attached to a higgs–higgsino loop, with supersymmetry breaking introduced by the term $2\mathrm{Re}\,[\tilde{m}_g^* \mu(\mathscr{H}_2^{\mathrm{T}} e \mathscr{H}_1)]$ in the Lagrangian density. This gives the bino and wino masses[11b]

$$m_{\text{bino}} = \frac{g'^2 m_g}{16\pi^2} \left| 11 - f\left(\frac{\mu^2}{m_A^2}\right) \right| , \tag{31.4.15}$$

$$m_{\text{wino}} = \frac{g^2 m_g}{16\pi^2} \left| 1 - f\left(\frac{\mu^2}{m_A^2}\right) \right| , \tag{31.4.16}$$

where m_A is the pseudoscalar particle mass defined by Eq. (28.5.21), and

$$f(x) \equiv \frac{2x \ln x}{x - 1} \,. \tag{31.4.17}$$

The implications of these results will be considered further in Section 31.7.

Finally, there is a scale-dependent field-renormalization factor Z_r multiplying the kinematic Lagrangian density $[\Phi_r^* e^{-V} \Phi_r]_D$ for any left-chiral superfield Φ_r. This can be moved from the kinematic D-terms to the superpotential $\mathscr{F}$-terms by absorbing a factor $Z_r^{1/2}$ in Φ_r. The Yukawa couplings h_{rst} (such as the h_{ij}^E, h_{ij}^D and h_{ij}^U in Eq. (28.1.7)) in the trilinear superpotential terms $\sum_{rst} h_{rst} \Phi_r \Phi_s \Phi_t$ are then multiplied with a factor $Z_r^{-1/2} Z_s^{-1/2} Z_t^{-1/2}$ that depends on the cut-off $\mathscr{M}$. According to

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Eq. (31.4.7), the interaction $\mathscr{L}^{(1)}$ will then make a contribution to the Lagrangian density:

$$\mathscr{L}^{(1)}_{\text{Yukawa}} = -2\sum_{rst}\gamma_{rst}\text{Re}\left[\tilde{m}_g^* h_{rst}\phi_r\phi_s\phi_t\right], \tag{31.4.18}$$

where

$$\begin{aligned}\gamma_{rst} &\equiv \mathscr{M}\frac{\partial \ln h_{rst}(\mathscr{M})}{\partial \mathscr{M}} \\ &= -\frac{1}{2}\mathscr{M}\frac{\partial \ln Z_r(\mathscr{M})}{\partial \mathscr{M}} - \frac{1}{2}\mathscr{M}\frac{\partial \ln Z_s(\mathscr{M})}{\partial \mathscr{M}} - \frac{1}{2}\mathscr{M}\frac{\partial \ln Z_t(\mathscr{M})}{\partial \mathscr{M}}.\end{aligned} \tag{31.4.19}$$

We see that the coefficients A^E_{ij}, A^D_{ij}, and A^U_{ij} in Eq. (28.1.7) are given by[11b]

$$A^N_{ij} = \tilde{m}_g^*\gamma^N_{ij} = \tilde{m}_g^*\mathscr{M}\frac{\partial \ln h^N_{ij}}{\partial \mathscr{M}}, \tag{31.4.20}$$

where $N = E$, D, or U. This gives A^D_{ij} and A^U_{ij} of order $g_s^2 m_g/8\pi^2$, while A^E_{ij} is of order $g^2 m_g/8\pi^2$ or $g'^2 m_g/8\pi^2$.

31.5 Local Supersymmetry Transformations

As a last step before considering the effects of higher order in G on supersymmetry transformation rules, we will now complete our discussion of the transformation rules of the physical components of the graviton superfield H_μ and other superfields to lowest order in G.

Let us first note the form that these transformation rules take when expressed in terms of the physical fields $h_{\mu\nu}$, ψ_μ, b_μ, s, and p in a 'Wess–Zumino' gauge in which

$$C^H_\mu = \omega^H_\mu = V^H_{\mu\nu} - V^H_{\nu\mu} = 0. \tag{31.5.1}$$

Using the identification (31.1.25)–(31.1.27), (31.1.29), and (31.1.11) of the physical fields with components of the superfield H_μ, together with the general transformation rules (26.2.11)–(26.2.17) and the 'gauge' conditions (31.5.1), we find the transformation rules

$$\delta h_{\mu\nu} = \tfrac{1}{2}\left(\bar{\alpha}\left[\gamma_\mu\psi_\nu + \gamma_\nu\psi_\mu\right]\right), \tag{31.5.2}$$

$$\begin{aligned}\delta\psi_\mu = \Big[&\tfrac{1}{2}[\gamma^\nu, \gamma^\lambda]\,\partial_\lambda h_{\mu\nu} + \partial_\mu h^\lambda{}_\lambda + 2i\gamma_5 b_\mu - \tfrac{2}{3}i\gamma_\mu\gamma_\rho\gamma_5 b^\rho \\ &+ \tfrac{2}{3}\gamma_\mu s - \tfrac{2}{3}i\gamma_\mu\gamma_5 p\Big]\alpha,\end{aligned} \tag{31.5.3}$$

$$\delta s = \tfrac{1}{4}\left(\bar{\alpha}\gamma_\lambda L^\lambda\right), \tag{31.5.4}$$

$$\delta p = -\tfrac{1}{4}i\left(\bar{\alpha}\gamma_5\gamma_\lambda L^\lambda\right), \tag{31.5.5}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

$$\delta b_\mu = \tfrac{3}{4}i\left(\bar{\alpha}\gamma_5 L_\mu\right) - \tfrac{1}{4}i\left(\bar{\alpha}\gamma_5\gamma_\mu\gamma^\nu L_\nu\right), \tag{31.5.6}$$

where L_μ is given by Eq. (31.2.8):

$$L^\mu \equiv i\,\epsilon^{\mu\nu\kappa\rho}\,\gamma_5\,\gamma_\nu\,\partial_\kappa\psi_\rho\,.$$

This transformation shifts the components C_μ^H, ω_μ^H, and $V_{\mu\nu}^H - V_{\nu\mu}^H$ away from zero, by the amounts

$$\delta C_\mu^H = 0\,, \quad \delta\omega_\mu^H = V_{\mu\nu}^H\gamma^\nu\alpha\,, \tag{31.5.7}$$

$$\delta[V_{\nu\mu}^H - V_{\mu\nu}^H] = \left(\bar{\alpha}\,[\gamma_\mu\lambda_\nu^H - \gamma_\nu\lambda_\mu^H]\right). \tag{31.5.8}$$

We can then return to the gauge satisfying Eq. (31.5.1) by making a suitable gauge transformation $H_\mu \to H_\mu + \Delta_\mu$, where Δ_μ is a superfield of the form (31.1.17), (31.1.18), with components

$$C_\mu^\Delta = 0\,, \quad \omega_\mu^\Delta = -V_{\mu\nu}^H\gamma^\nu\alpha\,, \tag{31.5.9}$$

$$V_{\nu\mu}^\Delta - V_{\mu\nu}^\Delta = \left(\bar{\alpha}\,[\gamma_\nu\lambda_\mu^H - \gamma_\mu\lambda_\nu^H]\right). \tag{31.5.10}$$

So far, this has been for global supersymmetry transformations, with α an infinitesimal constant Majorana spinor. At least in lowest order, this symmetry can be easily extended to *local* supersymmetry transformations, with $\alpha(x)$ given an arbitrary dependence on x^μ. According to Eq. (26.7.11), under such a transformation the matter action undergoes the change

$$\delta\int d^4x\,\mathscr{L}_M = -\int d^4x\,\left(\bar{S}^\mu(x)\,\partial_\mu\alpha(x)\right). \tag{31.5.11}$$

Inspection of Eq. (31.2.17) shows that this change in the action is cancelled if we add an inhomogeneous term $(2/\kappa)\partial_\mu\alpha(x)$ to the right-hand side of Eq. (31.5.3), so that the change in the gravitino field is now

$$\delta\psi_\mu(x) = (2/\kappa)\partial_\mu\alpha(x) + \Big[\tfrac{1}{2}[\gamma^\nu\,,\,\gamma^\lambda]\,\partial_\lambda h_{\mu\nu}(x) + \partial_\mu h^\lambda{}_\lambda(x) + 2\,i\gamma_5\,b_\mu(x)$$
$$-\tfrac{2}{3}i\,\gamma_\mu\gamma_\rho\gamma_5\,b^\rho(x) + \tfrac{2}{3}\gamma_\mu\,s(x) - \tfrac{2}{3}\,i\,\gamma_\mu\gamma_5\,p(x)\Big]\alpha(x)\,. \tag{31.5.12}$$

It is useful to rewrite Eq. (31.5.12) in a form that makes the generalization to general coordinates more transparent. First, we note that the term $\partial_\mu h^\lambda{}_\lambda\alpha$ on the right-hand side of Eq. (31.5.12) may be eliminated by replacing the parameter $\alpha(x)$ with

$$\tilde{\alpha} \equiv (\mathrm{Det}\,g)^{1/4}\alpha \simeq \alpha + \tfrac{1}{2}\kappa h^\lambda{}_\lambda\alpha\,. \tag{31.5.13}$$

Dropping the tilde, to zeroth order in κ Eq. (31.5.12) then becomes

$$\delta\psi_\mu(x) = (2/\kappa)\partial_\mu\alpha(x) + \Big[\tfrac{1}{2}[\gamma^\nu\,,\,\gamma^\lambda]\,\partial_\lambda h_{\mu\nu}(x) + 2\,i\gamma_5\,b_\mu(x)$$
$$-\tfrac{2}{3}i\,\gamma_\mu\gamma_\rho\gamma_5\,b^\rho(x) + \tfrac{2}{3}\gamma_\mu\,s(x) - \tfrac{2}{3}\,i\,\gamma_\mu\gamma_5\,p(x)\Big]\alpha(x)\,. \tag{31.5.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

We can express this in terms of the covariant derivative of $\alpha(x)$, which in general coordinates takes the form

$$D_\mu\alpha(x) = \partial_\mu\alpha(x) + \tfrac{1}{2}i\mathcal{J}_{bc}\,\omega_\mu^{bc}(x)\,\alpha(x)\,, \tag{31.5.15}$$

where $\mathcal{J}_{bc}$ is the matrix (5.4.6) representing the generator of Lorentz transformations in the Dirac representation

$$\mathcal{J}^{bc} \equiv -\frac{i}{4}\left[\gamma^b\,,\,\gamma^c\right], \tag{31.5.16}$$

and $\omega_\mu^{bc}(x)$ is the spin connection,

$$\omega_\mu^{bc} = e^b{}_\lambda\, e^c{}_{\nu;\mu}\, g^{\lambda\nu} = e^b{}_\lambda\,\frac{\partial e^c{}_\nu}{\partial x^\mu}\, g^{\lambda\nu} - \Gamma^\rho_{\nu\mu}\, e^b{}_\lambda\, e^c{}_\rho\, g^{\lambda\nu}\,. \tag{31.5.17}$$

Using the weak field approximations (31.1.5), (31.1.10), and (31.1.11), together with the gauge condition $\phi_{\mu\nu} = \phi_{\nu\mu}$, and in this approximation again ignoring the difference between local Lorentz indices a, b, etc. and spacetime indices μ, ν, etc., this gives

$$D_\mu\alpha(x) \simeq \partial_\mu\alpha(x) + \tfrac{1}{4}\kappa\,[\gamma^\nu\,,\,\gamma^\lambda]\,\partial_\lambda h_{\mu\nu}(x)\,\alpha(x)\,. \tag{31.5.18}$$

Thus the local supersymmetry transformation rule (31.5.12) may be written

$$\begin{aligned}\delta\psi_\mu(x) = (2/\kappa)D_\mu\alpha(x) + \Big[&2i\gamma_5\, b_\mu(x) - \tfrac{2}{3}i\gamma_\mu\gamma_\rho\gamma_5\, b^\rho(x)\\ &+ \tfrac{2}{3}\gamma_\mu\, s(x) - \tfrac{2}{3}\,i\gamma_\mu\gamma_5\, p(x)\Big]\alpha(x)\,.\end{aligned} \tag{31.5.19}$$

We see that in Wess–Zumino gauge the derivatives in the supersymmetry transformation become covariant derivatives. In this sense, the approach outlined here is similar to the de Wit–Freedman approach to supersymmetric gauge theories, described in Section 27.8.

The transformation $\psi_\mu(x) \to \psi_\mu(x) + (2/\kappa)\partial_\mu\alpha(x)$ is a gauge transformation of the same type as Eq. (31.1.13), so it leaves the zeroth-order gravitino action $-\frac{1}{2}\int d^4x(\bar\psi_\mu\, L^\mu)$ invariant. The whole action obtained from the Lagrangian density (31.2.15) is therefore invariant to order zero in κ under the local supersymmetry transformations (31.5.2), (31.5.4)–(31.5.6), (31.5.12) (or (31.5.19)) and matter superfield transformations like (26.7.15). We conclude that *the combination of gravitation and supersymmetry automatically leads to local supersymmetry.*

31.6 Supergravity to All Orders

Although the action derived from the Lagrangian density (31.2.15) is invariant to zeroth order in $\kappa \equiv \sqrt{8\pi G}$ under the local supersymmetry transformations constructed in Sections 26.7 and 31.5, it is *not* invariant to first order in κ, because the interaction of matter with the gravitational

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

supermultiplet introduces terms of order κ in $\partial_\mu S^\mu$. To make possible the invariance of the full action, we will have to add terms of higher order in κ to the Lagrangian and to the supersymmetry transformation rules for the components of both the matter and the gravitational supermultiplets. This can be done by first adding terms of higher order in κ to the transformation rules, chosen so that these transformations form a closed algebra along with local Lorentz transformations and general coordinate transformations, and then adding terms to the action to make it invariant under all these transformations.

This is a long and tedious process. Here we shall give only the results,[12] and then turn in the next section to what has been their most important application. The local supersymmetry transformation of the vierbein, gravitino field, and auxiliary fields takes the form:

$$\delta e^a{}_\mu = \kappa \left(\bar{\alpha}\gamma^a \psi_\mu\right), \tag{31.6.1}$$

$$\delta\psi_\mu = (2/\kappa) D_\mu \alpha + 2i\gamma_5 (b_\mu - \tfrac{1}{3}\gamma_\mu \gamma_\rho b^\rho)\alpha + \tfrac{2}{3}\gamma_\mu (s - i\gamma_5 p)\alpha, \tag{31.6.2}$$

$$\delta s = \frac{1}{4e}\left(\bar{\alpha}\gamma_\mu L^\mu\right) + \frac{\kappa}{2}\left(\bar{\alpha}[i\gamma_5 b^\nu - s\gamma^\nu - ip\gamma_5\gamma^\nu]\psi_\nu\right), \tag{31.6.3}$$

$$\delta p = -\frac{i}{4e}\left(\bar{\alpha}\gamma_5\gamma_\mu L^\mu\right) + \frac{\kappa}{2}\left(\bar{\alpha}[b^\nu + is\gamma_5\gamma^\nu - p\gamma^\nu]\psi_\nu\right), \tag{31.6.4}$$

$$\begin{aligned}\delta b_\mu = {} & \frac{3i}{4e}\left(\bar{\alpha}\gamma_5 (L_\mu - \tfrac{1}{3}\gamma_\mu\gamma_\rho L^\rho)\right) + \frac{\kappa}{2} b_\nu \left(\bar{\alpha}\gamma^\nu \psi_\mu\right) \\ & + i\frac{\kappa}{2}\left(\bar{\psi}_\mu \gamma_5 (s - i\gamma_5 p)\alpha\right) - i\frac{\kappa}{4}\epsilon_{\mu\nu\kappa\sigma} b^\nu \left(\bar{\alpha}\gamma_5\gamma^\kappa \psi^\sigma\right).\end{aligned} \tag{31.6.5}$$

Here D_μ is again the covariant derivative given by (31.5.15) and (31.5.16):

$$D_\mu \equiv \partial_\mu + \tfrac{1}{8}[\gamma_a, \gamma_b]\,\omega_\mu^{ab}, \tag{31.6.6}$$

but now with the spin connection including terms bilinear in the gravitino field

$$\begin{aligned}\omega_\mu^{ab} = {} & e^a{}_\lambda e^b{}_{\nu;\mu} g^{\lambda\nu} \\ & + \frac{\kappa^2}{4}\left[e^b{}_\nu \left(\bar{\psi}_\mu \gamma^a \psi^\nu\right) + e^a{}_\nu e^b{}_\rho \left(\bar{\psi}^\nu \gamma_\mu \psi^\rho\right) - e^a{}_\nu \left(\bar{\psi}_\mu \gamma^b \psi^\nu\right)\right].\end{aligned} \tag{31.6.7}$$

Also, L^μ is the covariant version of the Rarita–Schwinger operator (31.2.8)

$$L^\mu = i\gamma_5\gamma_\nu D_\rho \psi_\sigma \epsilon^{\mu\nu\rho\sigma}, \tag{31.6.8}$$

γ_μ is defined in terms of the usual Dirac matrices γ_a as

$$\gamma_\mu = e^a{}_\mu \gamma_a, \tag{31.6.9}$$

and e is here the determinant of the vierbein

$$e \equiv \sqrt{\text{Det}\, g}. \tag{31.6.10}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

It is easy to see that these transformation rules reduce in the weak field limit to Eqs. (31.5.2), (31.5.19), and (31.5.4)–(31.5.6).

The action for pure supergravity that is invariant under these transformations, and that reduces to the form (31.2.9) in the weak field limit, is

$$I_{\text{SUGRA}} = \int d^4x \left[-\frac{e}{2\kappa^2} R - \frac{1}{2}\left(\bar{\psi}_\mu L^\mu\right) - \frac{4e}{3}\left(s^2 + p^2 - b_\mu b^\mu\right)\right] , \tag{31.6.11}$$

where R is the curvature scalar, calculated using the spin connection (31.6.7). In the absence of matter, the action would be stationary for $s = p = b_\mu = 0$, leaving us with the simpler action

$$I_{\text{SUGRA}} = \int d^4x \left[-\frac{e}{2\kappa^2} R - \frac{1}{2}\left(\bar{\psi}_\mu L^\mu\right)\right] .$$

The transformation of matter fields is also now more complicated. For the components of a general scalar supermultiplet, these are

$$\delta C = i\left(\bar{\alpha}\gamma_5 \omega\right) , \tag{31.6.12}$$

$$\delta\omega = [-i\gamma_5 \not{\mathscr{D}} C - M + i\gamma_5 N + \not{V}]\,\alpha , \tag{31.6.13}$$

$$\delta M = -\left(\bar{\alpha}\,[\lambda + \not{\mathscr{D}}\omega]\right) + \frac{2\kappa}{3}\left(\bar{\alpha}\,[s - i\gamma_5 p + i\gamma_5 \not{b}]\,\omega\right) , \tag{31.6.14}$$

$$\delta N = i\left(\bar{\alpha}\gamma_5\,[\lambda + \not{\mathscr{D}}\omega]\right) + \frac{2i\kappa}{3}\left(\bar{\alpha}\,[s - i\gamma_5 p + i\gamma_5 \not{b}]\,\gamma_5\,\omega\right) , \tag{31.6.15}$$

$$\delta V_a = \left(\bar{\alpha}\gamma_a\lambda\right) + \left(\bar{\alpha}\mathscr{D}_a\omega\right) + \frac{\kappa}{3}\left(\bar{\alpha}\,[s - i\gamma_5 p + i\gamma_5 \not{b}]\,\gamma_a\,\omega\right) , \tag{31.6.16}$$

$$\delta\lambda = -\frac{1}{4}[\gamma^a, \gamma^b]\,\alpha\, F_{ab} + i D\,\gamma_5\,\alpha , \tag{31.6.17}$$

$$\delta D = i\left(\bar{\alpha}\gamma_5 \not{\mathscr{D}}\lambda\right) , \tag{31.6.18}$$

where the covariant derivatives here are

$$\mathscr{D}_a C = e_a{}^\mu\left[\partial_\mu C - \frac{i\kappa}{2}\left(\bar{\psi}_\mu\gamma_5\,\omega\right)\right] , \tag{31.6.19}$$

$$\begin{aligned}\mathscr{D}_a\omega = e_a{}^\mu\Big[&\partial_\mu\omega + \frac{1}{8}\omega_\mu^{cb}\,[\gamma_c, \gamma_b]\,\omega - i\kappa\, b_\mu\gamma_5\,\omega \\ &-\frac{\kappa}{2}\left(\not{V} - i\gamma_5 \not{\partial} C - M + i\gamma_5 N\right)\psi_\mu\Big] ,\end{aligned} \tag{31.6.20}$$

$$\begin{aligned}\mathscr{D}_a\lambda = e_a{}^\mu\Big[&\partial_\mu\lambda + \frac{1}{8}\omega_\mu^{cb}[\gamma_c, \gamma_b]\,\lambda \\ &+i\kappa\, b_\mu\gamma_5\,\lambda + \frac{\kappa}{8}[\gamma_b, \gamma_c]\,\psi_\mu F_{bc} - \frac{i\kappa}{2}\gamma_5 D\,\psi_\mu\Big] ,\end{aligned} \tag{31.6.21}$$

$$F_{ab} = e_a{}^\mu e_b{}^\nu\left[\partial_\mu V_\nu + \frac{\kappa}{2}\partial_\mu\left(\bar{\psi}_\mu\,\omega\right) - \frac{\kappa}{2}\left(\bar{\psi}_\mu\gamma_\nu\,\lambda\right)\right] - a \leftrightarrow b , \tag{31.6.22}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

with $V_\mu \equiv e^a{}_\mu V_a$. The rules for multiplying general scalar multiplets are the same as Eqs. (26.2.19)–(26.2.25), except that ∂_μ is everywhere replaced with $\mathscr{D}_a$, so that the components of a supermultiplet $S = S_1 S_2$ are

$$C = C_1 C_2 \,, \tag{31.6.23}$$

$$\omega = C_1 \omega_2 + C_2 \omega_1 \,, \tag{31.6.24}$$

$$M = C_1 M_2 + C_2 M_1 + \tfrac{1}{2} i \left(\overline{\omega_1} \gamma_5 \omega_2 \right) , \tag{31.6.25}$$

$$N = C_1 N_2 + C_2 N_1 - \tfrac{1}{2} \left(\overline{\omega_1} \omega_2 \right) , \tag{31.6.26}$$

$$V^a = C_1 V_2^a + C_2 V_1^a - \tfrac{1}{2} i \left(\overline{\omega_1} \gamma_5 \gamma^a \omega_2 \right) , \tag{31.6.27}$$

$$\begin{aligned} \lambda = {} & C_1 \lambda_2 + C_2 \lambda_1 - \tfrac{1}{2} \gamma^a \omega_1 \mathscr{D}_a C_2 - \tfrac{1}{2} \gamma^a \omega_2 \mathscr{D}_a C_1 + \tfrac{1}{2} i \not{V}_1 \gamma_5 \omega_2 \\ & + \tfrac{1}{2} i \not{V}_2 \gamma_5 \omega_1 + \tfrac{1}{2} (N_1 - i \gamma_5 M_1) \omega_2 + \tfrac{1}{2} (N_2 - i \gamma_5 M_2) \omega_1 \,, \end{aligned} \tag{31.6.28}$$

$$\begin{aligned} D = {} & - \mathscr{D}_a C_1 \mathscr{D}^a C_2 + C_1 D_2 + C_2 D_1 + M_1 M_2 + N_1 N_2 \\ & - \left(\overline{\omega_1} \left[\lambda_2 + \tfrac{1}{2} \not{\mathscr{D}} \omega_2 \right] \right) - \left(\overline{\omega_2} \left[\lambda_1 + \tfrac{1}{2} \not{\mathscr{D}} \omega_1 \right] \right) - V_{1a} V_2^a \,. \end{aligned} \tag{31.6.29}$$

Just as in flat space, there are supersymmetric constraints that can be imposed on these general supermultiplets. One such constraint is reality. Inspection of Eqs. (31.6.12)–(31.6.22) together with Eqs. (26.A.20)–(26.A.21) shows that if C, M, N, V_a, D, ω, and λ form a supermultiplet S, then C^*, M^*, N^*, V_a^*, D^*, $\beta \epsilon \gamma_5 \omega^*$, and $\beta \epsilon \gamma_5 \lambda^*$ are also the components of a supermultiplet, called S^*. In particular, a *real* supermultiplet is one with $S = S^*$, so that C, M, N, V_a, and D are real and ω and λ are Majorana spinors.

We can also impose a supersymmetric condition of chirality. Suppose we set

$$\lambda = 0 \,, \quad D = 0 \,, \quad V_\nu + \tfrac{1}{2} \kappa \left(\bar{\psi}_\nu \omega \right) = \partial_\nu Z \,, \tag{31.6.30}$$

for some field Z. Then Eq. (31.6.22) gives $F_{ab} = 0$, so Eq. (31.6.17) shows that $\delta \lambda = 0$, while Eq. (31.6.21) gives $\mathscr{D}_a \lambda = 0$, so Eq. (31.6.18) shows that $\delta D = 0$. The conditions $\lambda = D = 0$ are therefore preserved by local supersymmetry transformations. With a little more work one can show that

$$\delta \left[V_\nu + \tfrac{1}{2} \kappa \left(\bar{\psi}_\nu \omega \right) \right] = \partial_\nu \left(\bar{\alpha} \omega \right) , \tag{31.6.31}$$

so the remaining condition, that $V_\nu + \tfrac{1}{2} \kappa \left(\bar{\psi}_\nu \omega \right)$ is a spacetime gradient, is also supersymmetric. A supermultiplet of component fields satisfying Eq. (31.6.30) is said to be *chiral*. Just as in the case of global supersymmetry, the components of a chiral supermultiplet are conventionally renamed as A, B, ψ, F, and G, defined by

$$C = A \,, \quad \omega = -i \gamma_5 \psi \,, \quad M = G \,, \quad N = -F \,, \quad Z = B \,. \tag{31.6.32}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

A chiral supermultiplet is real if A, B, F, and G are real and ψ is a Majorana spinor. Such a real chiral supermultiplet can be written as the sum of a left-chiral supermultiplet Φ, with components conventionally defined by

$$\phi \equiv \frac{A+iB}{\sqrt{2}}\,, \qquad \psi_L \equiv \left(\frac{1+\gamma_5}{2}\right)\psi\,, \qquad \mathscr{F} \equiv \frac{F-iG}{\sqrt{2}}\,, \tag{31.6.33}$$

and its complex conjugate, a right-chiral supermultiplet. From these definitions and Eqs. (31.6.12)–(31.6.15) and (31.6.31) we see that the components of a left-chiral supermultiplet have the transformation properties

$$\delta\phi = \sqrt{2}\left(\bar{\alpha}\,\psi_L\right), \tag{31.6.34}$$

$$\delta\psi_L = \sqrt{2}(\not{\partial}\phi)\alpha_R - \kappa\gamma^\mu\left(\bar{\psi}_\mu\psi_L\right)\alpha_R + \sqrt{2}\mathscr{F}\,\alpha_L\,, \tag{31.6.35}$$

$$\delta\mathscr{F} = \sqrt{2}\left(\bar{\alpha}\not{\mathscr{D}}\psi_L\right) - \frac{2\kappa}{3}\left(\bar{\alpha}\,[s - ip - i\not{b}]\psi_L\right), \tag{31.6.36}$$

with $\mathscr{D}_a\psi$ given by Eqs. (31.6.20) and (31.6.32). The rules for multiplying left-chiral supermultiplets are the same as the corresponding rules (26.3.27)–(26.3.29) in the case of global supersymmetry: The product of left-chiral supermultiplets Φ_1 and Φ_2 is a left-chiral supermultiplet denoted $\Phi_1\Phi_2$, with components

$$\phi = \phi_1\phi_2\,, \tag{31.6.37}$$

$$\psi_L = \phi_1\psi_{2L} + \phi_2\psi_{1L}\,, \tag{31.6.38}$$

$$\mathscr{F} = \phi_1\mathscr{F}_2 + \phi_2\mathscr{F}_1 - \left(\psi_{1L}^{\mathrm{T}}\,\epsilon\,\psi_{2L}\right). \tag{31.6.39}$$

Now we must consider how to construct actions that are invariant under local supersymmetry transformations as well as general coordinate transformations and local Lorentz transformations. Eqs. (31.6.18) and (31.6.21) show that the change under a supersymmetry transformation of the D-component of a general supermultiplet S is no longer simply a spacetime derivative, so the integral of such a D-component is not suitable as a term in the action. Instead, from an arbitrary superfield S we can form a density whose integral *is* supersymmetric

$$\begin{aligned}[S]_D = e\Bigg[\, & D^S - \frac{i\kappa}{2}\left(\bar{\psi}^\mu\gamma_\mu\gamma_5\lambda^S\right) + \frac{4\kappa}{3}[-sN^S + pM^S - b^\mu V_\mu^S] \\ & - \frac{i\kappa}{3}\left(\overline{\omega^S}\gamma_5\not{L}\right) - \frac{\kappa^2}{4}\epsilon^{\mu\rho\sigma\tau}V_\sigma^S\left(\bar{\psi}_\rho\gamma_\tau\psi_\mu\right) \\ & - \frac{\kappa^2}{8}\epsilon^{\mu\rho\sigma\tau}\left(\overline{\omega^S}\psi_\sigma\right)\left(\bar{\psi}_\rho\gamma_\tau\psi_\mu\right)\Bigg] - \frac{2\kappa^2}{3}C^S\mathscr{L}_{\mathrm{SUGRA}}\,,\end{aligned} \tag{31.6.40}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

where $\mathscr{L}_{\rm SUGRA}$ is the supergravity Lagrangian density in Eq. (31.6.11):

$$\mathscr{L}_{\rm SUGRA} = -\frac{e}{2\kappa^2}R - \frac{1}{2}\left(\bar{\psi}_\mu L^\mu\right) - \frac{4e}{3}\left(s^2 + p^2 - b_\mu b^\mu\right) . \qquad (31.6.41)$$

Likewise, Eqs. (31.6.36) and (31.6.20) show that the change under a supersymmetry transformation of the $\mathscr{F}$-component of a left-chiral superfield X is not a spacetime derivative, so terms must be added to this $\mathscr{F}$-term to form a density whose integral is supersymmetric:

$$[X]_{\mathscr{F}} = e\left[\mathscr{F}^X + \frac{\kappa}{\sqrt{2}}\left(\bar{\psi}_{\mu R}\gamma^\mu\psi_L^X\right) + \frac{\kappa^2}{4}\left(\bar{\psi}_{\mu R}[\gamma^\mu, \gamma^\nu]\psi_{\nu R}\right)\phi^X \right.$$
$$\left. +2\kappa(s - ip)\phi^X\right] . \qquad (31.6.42)$$

There are some supermultiplets of special physical interest. One is the real non-chiral supermultiplet **1** whose only non-zero component is $C = 1$. According to Eq. (31.6.40), this supermultiplet has

$$[\mathbf{1}]_D = -\frac{2\kappa^2}{3}\mathscr{L}_{\rm SUGRA} , \qquad (31.6.43)$$

so this gives nothing new.

A more interesting example is provided by a *left-chiral* supermultiplet **I** whose only non-zero component is $\phi = 1$. According to Eq. (31.6.42), this supermultiplet has

$$\mathrm{Re}\,[\mathbf{I}]_{\mathscr{F}} = e\left[\frac{\kappa^2}{4}\left(\bar{\psi}_\mu[\gamma^\mu, \gamma^\nu]\psi_\nu\right) + 4\kappa s\right] . \qquad (31.6.44)$$

If a term $c\int d^4x\,\mathrm{Re}\,[\mathbf{I}]_{\mathscr{F}}$ appears in the Lagrangian density with a real coefficient c, then the sum of this term and the supergravity action (31.6.11) is stationary with respect to variation of auxiliary fields at

$$s = 3c\kappa/2 , \qquad p = b^\mu = 0 , \qquad (31.6.45)$$

so that after eliminating the auxiliary fields, the action will contain a cosmological constant term

$$3\kappa^2 c^2 \int d^4x\, e , \qquad (31.6.46)$$

corresponding to a vacuum energy density $-3\kappa^2 c^2$. Such a term is needed to keep the vacuum state Lorentz-invariant when supersymmetry is spontaneously broken; to cancel the positive vacuum energy density $F^2/2$ associated with supersymmetry breaking, we must take

$$3\kappa^2 c^2 = F^2/2 . \qquad (31.6.47)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Looking back at Eq. (31.6.44) and comparing with Eq. (31.3.11), we see that this gives a gravitino mass[13]

$$m_g = c\,\kappa^2 = \frac{F\kappa}{\sqrt{6}} = \sqrt{\frac{4\pi G F^2}{3}}\,, \tag{31.6.48}$$

in agreement with our previous result (31.3.17).

As an illustration of the use of these formulas that will also provide results needed in the next section, let us calculate the bosonic part of the Lagrangian density*

$$\mathscr{L} = \mathscr{L}_{\text{SUGRA}} + \tfrac{1}{2}[K(\Phi,\Phi^*)]_D + 2\text{Re}\,[f(\Phi)]_{\mathscr{F}}\,, \tag{31.6.49}$$

where $\mathscr{L}_{\text{SUGRA}}$ is here the bosonic part of the supergravity Lagrangian density (31.6.41), $K(\Phi,\Phi^*)$ is a real function of a set of left-chiral supermultiplets Φ_n and their complex conjugates, and $f(\Phi)$ is a function of the Φ_n alone. The purely bosonic terms in the multiplication rules (31.6.23), (31.6.25)–(31.6.27), and (31.6.29) are the same as in global supersymmetry, so we can use either these rules or the superspace formalism of Chapter 26 to calculate that the supermultiplet $K(\Phi,\Phi^*)$ has the bosonic components

$$C^K = K(\phi,\phi^*)\,, \tag{31.6.50}$$

$$M^K = -2\,\text{Im}\sum_n\left(\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\mathscr{F}_n\right) + \cdots\,, \tag{31.6.51}$$

$$N^K = -2\,\text{Re}\sum_n\left(\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\mathscr{F}_n\right) + \cdots\,, \tag{31.6.52}$$

$$V_\mu^K = 2\,\text{Im}\sum_n\left(\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\partial_\mu\phi_n\right) + \cdots\,, \tag{31.6.53}$$

$$D^K = 2\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\partial\phi_m^*}\left(-g^{\mu\nu}\,\partial_\mu\phi_n\,\partial_\nu\phi_m^* + \mathscr{F}_n\mathscr{F}_m^*\right) + \cdots\,, \tag{31.6.54}$$

where the dots indicate terms involving fermion fields. Also, the rules (31.6.37)–(31.6.39) for multiplying left-chiral supermultiplets are the same as those in global supersymmetry, so we can use either these rules or the superfield formalism of Chapter 26 to calculate that the left-chiral supermultiplet $f(\Phi)$ has the bosonic components

$$\phi^f = f(\phi)\,, \tag{31.6.55}$$

$$\mathscr{F}^f = \sum_n\frac{\partial f(\phi)}{\partial\phi_n}\mathscr{F}_n + \cdots\,, \tag{31.6.56}$$

* Often the term $\mathscr{L}_{\text{SUGRA}}$ is omitted, with the supergravity Lagrangian introduced instead by including a constant term $-3/\kappa^2$ in $K(\phi,\phi^*)$. We will not follow this practice; here, with a conventional normalization of the scalar fields, the term in $K(\phi,\phi^*)$ of leading order in κ is $\sum_n|\phi_n|^2$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

with dots again indicating terms involving fermions. Inserting these results into Eqs. (31.6.40) and (31.6.42) then gives the bosonic terms of the Lagrangian density (31.6.49) as

$$\begin{aligned}\mathscr{L}_{\text{bosonic}} = &\left[-\frac{e}{2\kappa^2}R - \frac{4e}{3}\left(s^2+p^2-b_\mu b^\mu\right)\right]\left[1-\frac{\kappa^2}{3}K(\phi,\phi^*)\right]\\ &-e\sum_{nm}\frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m^*}\left(g^{\mu\nu}\partial_\mu\phi_n\partial_\nu\phi_m^* - \mathscr{F}_n\mathscr{F}_m^*\right)\\ &+\frac{4\kappa e}{3}\operatorname{Re}\sum_n\frac{\partial K(\phi,\phi^*)}{\partial\phi_n}\left(\mathscr{F}_n(s+ip)+ib^\mu\partial_\mu\phi_n\right)\\ &+2e\operatorname{Re}\left(\sum_n\frac{\partial f(\phi)}{\partial\phi_n}\mathscr{F}_n+2\kappa(s-ip)f(\phi)\right)\ . \qquad (31.6.57)\end{aligned}$$

Now we must eliminate the auxiliary fields by setting them at values where the Lagrangian density is stationary.** This gives the auxiliary fields

$$\begin{aligned}\mathscr{F}_n = &\frac{\kappa^2}{3N}\sum_m\left(\mathscr{G}^{-1}\right)_{mn}\frac{\partial K}{\partial\phi_m^*}\left(-\sum_{k\ell}\left(\mathscr{G}^{-1}\right)_{k\ell}\frac{\partial K}{\partial\phi_\ell}\left(\frac{\partial f}{\partial\phi_k}\right)^* + 3f^*\right)\\ &-\sum_m\left(\mathscr{G}^{-1}\right)_{mn}\left(\frac{\partial f}{\partial\phi_m}\right)^*, \qquad (31.6.58)\end{aligned}$$

$$s-ip = \frac{\kappa}{2N}\left(-\sum_{\ell k}\left(\mathscr{G}^{-1}\right)_{k\ell}\frac{\partial K}{\partial\phi_\ell}\left(\frac{\partial f}{\partial\phi_k}\right)^* + 3f^*\right), \qquad (31.6.59)$$

$$b_\mu = \frac{\kappa}{2(1-\kappa^2K/3)}\operatorname{Im}\left(\sum_n\frac{\partial K}{\partial\phi_n}\partial_\mu\phi_n\right), \qquad (31.6.60)$$

where

$$N \equiv 1-\frac{\kappa^2}{3}K+\frac{\kappa^2}{3}\sum_{k\ell}\left(\mathscr{G}^{-1}\right)_{k\ell}\frac{\partial K}{\partial\phi_\ell}\frac{\partial K}{\partial\phi_k^*}$$

and $\mathscr{G}(\phi,\phi^*)$ is the Kahler metric

$$\mathscr{G}_{nm}(\phi,\phi^*) \equiv \frac{\partial^2 K(\phi,\phi^*)}{\partial\phi_n\,\partial\phi_m^*}\ . \qquad (31.6.61)$$

** Though this procedure is generally followed, it is not strictly correct, for even though the Lagrangian density is quadratic in the auxiliary fields, the coefficients of the terms of second order in the auxiliary fields are not field-independent. In consequence, in doing the path integration over auxiliary fields we encounter determinants of the coefficients of the quadratic terms, which are equivalent to adding terms to the Lagrangian proportional to $\delta^4(0) = (2\pi)^{-4}\int d^4k\,1$. Such terms can be eliminated by using dimensional regularization, for which $\int d^4k\,1 = 0$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Using this in Eq. (31.6.57) gives the bosonic Lagrangian

$$\begin{aligned}\mathscr{L}_{\text{bosonic}} = &-\frac{e}{2\kappa^2}R\left[1-\frac{\kappa^2}{3}K\right]-e\sum_{nm}\mathscr{G}_{nm}\,g^{\mu\nu}\partial_\mu\phi_n\partial_\nu\phi_m^* \\ &+\frac{e\kappa^2}{3N}\left|\sum_{mn}\left(\mathscr{G}^{-1}\right)_{nm}\frac{\partial f}{\partial\phi_m}\frac{\partial K}{\partial\phi_n^*}-3f\right|^2 \\ &-\frac{e\kappa^2}{3\left(1-\kappa^2K/3\right)}\,\mathrm{Im}\left[\sum_n\frac{\partial K}{\partial\phi_n}\partial_\mu\phi_n\right]\mathrm{Im}\left[\sum_n\frac{\partial K}{\partial\phi_n}\partial_\nu\phi_n\right]g^{\mu\nu} \\ &-e\sum_{mn}\left(\mathscr{G}^{-1}\right)_{nm}\frac{\partial f}{\partial\phi_m}\left(\frac{\partial f}{\partial\phi_n}\right)^*. \end{aligned} \tag{31.6.62}$$

As already noted in the weak field case in Section 31.2, the Lagrangian density (31.6.62) has the uncomfortable feature that the Einstein–Hilbert term $-eR/2\kappa^2$ is multiplied with a factor $(1-\kappa^2K(\phi,\phi^*)/3)$, so that the effective gravitational constant varies from point to point in spacetime. To remedy this, we perform a Weyl transformation, defining a new metric by

$$\tilde{g}_{\mu\nu} = \left(1-\kappa^3K/3\right)\,g_{\mu\nu}\,. \tag{31.6.63}$$

The Einstein Lagrangian density is given in terms of the new metric by

$$eg^{\mu\nu}R_{\mu\nu} = \left(1-\frac{\kappa^2}{3}K\right)^{-1}\tilde{e}\,\tilde{g}^{\mu\nu}\left(\tilde{R}_{\mu\nu}+\frac{3}{2}\partial_\mu\ln\left(1-\frac{\kappa^2}{3}\right)\partial_\nu\ln\left(1-\frac{\kappa^2}{3}\right)\right),$$

where $\tilde{R}_{\mu\nu}$ is the curvature tensor calculated using a metric $\tilde{g}_{\mu\nu}$ in place of $g_{\mu\nu}$, and $\tilde{e}\equiv\sqrt{\mathrm{Det}\,\tilde{g}}$. A straightforward calculation then gives the bosonic Lagrangian (31.6.62) as

$$\begin{aligned}\mathscr{L}_{\text{bosonic}} = &-\frac{\tilde{e}}{2\kappa^2}\tilde{g}^{\mu\nu}\,\tilde{R}_{\mu\nu} \\ &-\tilde{e}\,\tilde{g}^{\mu\nu}\sum_{nm}\partial_\mu\phi_n\partial_\nu\phi_m^*\left[\left(1-\frac{\kappa^2}{3}K\right)^{-1}\frac{\partial^2K}{\partial\phi_n\phi_m^*}\right. \\ &\qquad\left.+\frac{\kappa^2}{3}\left(1-\frac{\kappa^2}{3}K\right)^{-2}\frac{\partial K}{\partial\phi_n}\frac{\partial K}{\partial\phi_m^*}\right] \\ &+\frac{\tilde{e}\,\kappa^2}{3N}\left(1-\frac{\kappa^2}{3}K\right)^{-2}\left|\sum_{mn}\left(\mathscr{G}^{-1}\right)_{nm}\frac{\partial f}{\partial\phi_m}\frac{\partial K}{\partial\phi_n^*}-3f\right|^2 \\ &-\tilde{e}\left(1-\frac{\kappa^2}{3}K\right)^{-2}\sum_{mn}\left(\mathscr{G}^{-1}\right)_{nm}\frac{\partial f}{\partial\phi_m}\left(\frac{\partial f}{\partial\phi_n}\right)^*. \end{aligned} \tag{31.6.64}$$

The Weyl transformation has not only removed the factor $(1-\kappa^2K/3)$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

from the Einstein–Hilbert term; it has also eliminated terms proportional to $\partial_\mu \phi_n \, \partial_\nu \phi_m$ and $\partial_\mu \phi_n^* \, \partial_\nu \phi_m^*$.

This result may be further simplified by introducing a *modified Kahler potential* $d(\phi, \phi^*)$ in place of $K(\phi, \phi^*)$, which we define by

$$1 - \frac{\kappa^2}{3} K \equiv \exp\left(-\frac{\kappa^2 d}{3}\right) . \tag{31.6.65}$$

We also introduce a new metric on the scalar field space

$$g_{nm} \equiv \frac{\partial^2 d}{\partial \phi_n \, \partial \phi_m^*} . \tag{31.6.66}$$

The reciprocals of the new and old metrics are related by

$$\mathscr{G}_{\ell k}^{-1} = \exp(\kappa^2 d/3) \left[g_{\ell k}^{-1} + \frac{\kappa^2}{3} \frac{\sum_{mn} g_{\ell n}^{-1} g_{mk}^{-1} (\partial d/\partial \phi_n)(\partial d/\partial \phi_m^*)}{1 - (\kappa^2/3) \sum_{mn} g_{mn}^{-1} (\partial d/\partial \phi_n)(\partial d/\partial \phi_m^*)} \right] .$$

The bosonic Lagrangian density (31.6.64) now takes the simpler form

$$\mathscr{L}_{\text{bosonic}} = -\frac{\tilde{e}}{2\kappa^2} \tilde{g}^{\mu\nu} \tilde{R}_{\mu\nu} - \tilde{e} \, \tilde{g}^{\mu\nu} \sum_{nm} g_{nm} \partial_\mu \phi_n \partial_\nu \phi_m^* - \tilde{e} V , \tag{31.6.67}$$

where $V(\phi, \phi^*)$ is the potential

$$V = \exp(\kappa^2 d) \left[\sum_{nm} g_{nm}^{-1} L_m \, L_n^* - 3 \kappa^2 |f|^2 \right] \tag{31.6.68}$$

and

$$L_m \equiv \frac{\partial f}{\partial \phi_m} + \kappa^2 f \frac{\partial d}{\partial \phi_m} . \tag{31.6.69}$$

The potential (31.6.68) has an obvious stationary point at field strengths satisfying the condition $L_m = 0$. However, as we found in the weak field case, at this point the vacuum energy in general takes the negative value $-3\kappa^2 |f|^2$. To have the stationary point $L_m = 0$ give a solution with flat space, it is necessary that both $f(\phi)$ and $\partial f(\phi)/\partial \phi_n$ should vanish at these field values. Inspection of Eqs. (31.6.58) and (31.6.59) shows that the scalar auxiliary fields $\mathscr{F}_n$, s, and p vanish for such field values, so that the vacuum expectation values of the variations (31.6.2) and (31.6.35) of the gravitino and chiral spinor fields under global supersymmetry transformations with constant α vanish. Thus a vacuum field value for which $f(\phi)$ and $\partial f(\phi)/\partial \phi_n$ all vanish is one for which global supersymmetry is unbroken in the classical limit. In the next section we shall consider vacuum configurations in which supersymmetry *is* broken.

We will not show it here, but the additional terms in the bosonic Lagrangian required by the inclusion of gauge superfields are unaffected by gravitation, aside from an over-all determinantal factor $\tilde{e}$ and metric

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

factors needed to raise and lower indices. After elimination of auxiliary fields and a Weyl transformation, the complete bosonic Lagrangian for a theory with gauge as well as chiral and gravitational superfields is

$$\mathscr{L}_{\text{bosonic}}/\tilde{e} = -\frac{1}{2\kappa^2}\tilde{R}^{\mu}{}_{\mu} - \sum_{nm} g_{nm} D_\mu \phi_n D^\mu \phi_m^* - \tfrac{1}{4}\sum_{AB} \operatorname{Re} f_{AB}\, F^A_{\mu\nu} F^{B\mu\nu} - \tfrac{1}{8}\sum_{AB} \operatorname{Im} f_{AB}\, F^A_{\mu\nu} F^B_{\rho\sigma} \epsilon^{\mu\nu\rho\sigma} - V \,. \tag{31.6.70}$$

Here D_μ, $F^A_{\mu\nu}$, and t_A denote the gauge-covariant derivatives, the field-strength tensors, and the representatives of the gauge generators on the chiral scalar superfields, using the notation described in Section 15.1; f_{AB} is an independent holomorphic function of the ϕ_n; all spacetime indices are raised and lowered with $\tilde{g}_{\mu\nu}$; and the potential V now takes the form

$$V = \exp(\kappa^2 d)\left[\sum_{nm} g^{-1}_{nm} L_m L_n^* - 3\kappa^2 |f|^2\right] + \frac{1}{2}\operatorname{Re}\sum_{AB} f^{-1}_{AB}\left(\sum_{nm}\frac{\partial d}{\partial \phi_n}(t_A)_{nm}\phi_m\right)\left(\sum_{kl}\frac{\partial d}{\partial \phi_k}(t_B)_{kl}\phi_l\right)^* \,. \tag{31.6.71}$$

The form (31.6.71) of the bosonic potential is simple enough to make it apparent that terms in d that depend only on ϕ_n or only on ϕ_n^* may be traded for corrections to the superpotential. Specifically, if we write

$$d(\phi,\phi^*) = \tilde{d}(\phi,\phi^*) + a(\phi) + a(\phi)^* \,, \qquad f(\phi) = \tilde{f}(\phi)\exp\left(-\kappa^2 a(\phi)\right) \,, \tag{31.6.72}$$

with $a(\phi)$ an arbitrary holomorphic function satisfying the gauge invariance condition

$$\sum_{nm}\frac{\partial a(\phi)}{\partial \phi_n}(t_A)_{nm}\phi_m = 0 \,,$$

then the potential (31.6.71) takes the same form in terms of $\tilde{d}$ and $\tilde{f}$ as it did in terms of d and f. With a suitable redefinition of the superpotential, we can then eliminate any terms in the power series expansion of $d(\phi,\phi^*)$ that depend only on ϕ_n or only on ϕ_n^*. With this understanding, the leading term in the power series expansion of $d(\phi,\phi^*)$ (now dropping the tilde) is of the form $\sum_{nm} d_{nm}\phi_n\phi_m^*$. By a suitable linear transformation of the superfields, we can then make the matrix d_{nm} equal to δ_{nm}, so that the power series expansion of $d(\phi,\phi^*)$ begins

$$d(\phi,\phi^*) = \sum_n |\phi_n|^2 + \cdots \,, \tag{31.6.73}$$

and the power series expansion of the metric (31.6.66) begins

$$g_{nm} = \delta_{nm} + \cdots \,. \tag{31.6.74}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Inspection of the second term on the right-hand side of Eq. (31.6.67) shows that the scalar fields defined in this way are canonically normalized.

The fermion terms are much more complicated. Here we will quote only the terms quadratic in the gaugino fields

$$\mathscr{L}^{(2)}_{\text{gaugino}}/\tilde{e} = -\frac{1}{2}\,\mathrm{Re}\sum_{AB} f_{AB}\left(\bar{\lambda}_A \not{D}\lambda_B\right) + \frac{1}{2}\exp(\kappa^2 d/2)\,\mathrm{Re}\sum_{mn}\sum_{AB} g^{-1}_{nm} L_m \left(\frac{\partial f_{AB}}{\partial \phi_n}\right)^* \left(\bar{\lambda}_A \lambda_B\right), \quad (31.6.75)$$

with L_m given by Eq. (31.6.69). We see that if the gauge fields are canonically normalized, then the constant term in the expansion of the function f_{AB} in powers of the scalar fields is δ_{AB}, and then the gaugino fields λ_A are also canonically normalized.

* * *

Instead of moving all the holomorphic terms in $d(\phi, \phi^*)$ and their complex conjugates into the superpotential, we can use the transformation (31.6.72) to make the new superpotential $\tilde{f}(\phi)$ equal to a constant, which can be chosen to be equal to unity, by taking $a(\phi) = -\kappa^{-2}\ln f(\phi)$. The potential then depends only on the function

$$\mathscr{D}(\phi, \phi^*) \equiv d(\phi, \phi^*) + 2\,\kappa^{-2}\mathrm{Re}\,\ln f(\phi)\,, \quad (31.6.76)$$

and takes the form

$$V = \exp(\kappa^2 \mathscr{D})\left[\kappa^4 \sum_{nm} g^{-1}_{nm}\left(\frac{\partial \mathscr{D}}{\partial \phi_m}\right)\left(\frac{\partial \mathscr{D}}{\partial \phi_n}\right)^* - 3\,\kappa^2\right] + \frac{1}{2}\mathrm{Re}\sum_{AB} f^{-1}_{AB}\left(\sum_{nm}\frac{\partial \mathscr{D}}{\partial \phi_n}(t_A)_{nm}\phi_m\right)\left(\sum_{kl}\frac{\partial \mathscr{D}}{\partial \phi_k}(t_B)_{kl}\phi_l\right)^*. \quad (31.6.77)$$

Also, the metric (31.6.66) for the scalar fields may be written

$$g_{nm} = \frac{\partial^2 \mathscr{D}}{\partial \phi_n\, \partial \phi^*_m}\,. \quad (31.6.78)$$

Although we have not shown it here, the symmetry that allows us to replace the Kahler potential and superpotential with the single function $\mathscr{D}(\phi, \phi^*)$ also allows us to make this replacement in the whole Lagrangian, including all terms involving fermions and gauge fields.

There is an interesting class of 'no-scale' theories,[13a] in which the potential V vanishes for all values of ϕ_m. For instance, this is the case for a single gauge-neutral chiral scalar superfield, with

$$\mathscr{D} = -3\kappa^{-2}\ln\left(h(\phi) + h(\phi)^*\right), \quad (31.6.79)$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

where $h(\phi)$ is an arbitrary function of ϕ. But there is no known principle that would require $\mathscr{D}$ to take this form.

31.7 Gravity-Mediated Supersymmetry Breaking

We now take up again the problem of supersymmetry breaking. As discussed at the beginning of Chapter 28, if supersymmetry is to be of use in solving the hierarchy problem — that is, in understanding the large ratio of the Planck mass $m_{\rm Pl} \equiv 1/\sqrt{8\pi G}$ to the mass scale of observed particles — then supersymmetry must be unbroken at the Planck scale, and broken spontaneously only at some much lower mass scale. The only plausible mechanism known that would naturally produce a very large ratio of mass scales is the non-perturbative effect of asymptotically free gauge interactions. If these interactions are moderately weak at the Planck scale then their slow growth with decreasing energy will make them become strong at a much lower scale $\Lambda \ll m_{\rm Pl}$. The known elementary particles do not feel such strong forces, so whatever supersymmetry breakdown is produced directly or indirectly by these strong gauge interactions must be communicated to the observed particles by some interaction in which they do participate.

In Section 28.3 we noted two possible mechanisms for communicating the breakdown of supersymmetry to observable particles. One mechanism, gauge-mediated supersymmetry breaking, was discussed in detail in Section 28.6. We are now ready to consider the other mechanism, mediation of supersymmetry breaking by effects of gravitational strength.

In the early 1980s, when gravitation was first considered as the mediator of supersymmetry breaking,[14] it was generally assumed that the superpotential consisted of two terms: a function $f(\Phi)$ of various left-chiral superfields Φ_r of an *observable sector*, including all the superfields of observable particles, plus a function $\tilde{f}(Z)$ of various left-chiral superfields Z_k of a *hidden sector*,[15] all of which are neutral under the $SU(3)\times SU(2)\times U(1)$ gauge group of the standard model. Further, the superpotential of the hidden sector was assumed to take the form

$$\tilde{f}(Z) = \epsilon^3 F(\kappa Z)\,, \tag{31.7.1}$$

where ϵ is some mass that is much less than the Planck mass, and $F(\kappa Z)$ is a power series in κZ with coefficients of order unity. The assumption that the total superpotential should be a sum $f(\Phi) + \tilde{f}(Z)$ is somewhat arbitrary but, as we shall see, it is not difficult to think of reasons why this should be at least approximately true. A more serious criticism of this approach is that it did not offer any hope of solving the hierarchy

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

problem; the energy ϵ was simply assumed to be much less than the Planck mass.

After the development of these first models of gravity-mediated supersymmetry breaking there appeared other models in which the hierarchy of energy scales is explained naturally, in terms of the slow growth with decreasing energy of a gauge coupling which becomes strong at an energy $\Lambda \ll m_{\rm Pl}$. These models come in two versions that are distinguished by different assumptions regarding the source of supersymmetry breaking. As we shall see, in both versions the squarks and sleptons get supersymmetry-breaking masses of the order of the gravitino mass m_g, but they differ in the formula for m_g; in the first version $m_g \approx \kappa\Lambda^2$, while in the second version $m_g \approx \kappa^2\Lambda^3$, giving $\Lambda \approx 10^{11}$ GeV and $\Lambda \approx 10^{13}$ GeV, respectively. The two versions will turn out also to differ in the formulas for other soft supersymmetry-breaking parameters, including $B\mu$, the A-parameters, and the gaugino masses.

First Version[16]

In this version of gravity-mediated supersymmetry breaking it is assumed that the superfields of the theory fall into two sectors:

Observable Sector: These are the superfields of the minimum supersymmetric standard model: the $SU(3) \times SU(2) \times U(1)$ gauge superfields together with the quark, antiquark, lepton, antilepton, and Higgs left-chiral superfields which we will generically call Φ_r.

Hidden Sector: These are gauge superfields of an asymptotically free gauge interaction that becomes strong at an intermediate energy scale Λ with $m_W \ll \Lambda \ll m_{\rm Pl}$, together with left-chiral superfields Z_k that feel this gauge interaction.

The Z_k must be assumed to be neutral under the $SU(3) \times SU(2) \times U(1)$ gauge group, since otherwise we would be back in the case of gauge-mediated supersymmetry breaking. Also, we know enough about the observable sector to be sure that its chiral superfields do not feel the gauge interactions of the hidden sector.

In order naturally to have the renormalizable part of the total superpotential of the form $f(\Phi) + \tilde{f}(Z)$, we can assume that the symmetries that survive below the Planck scale include a group G_H (which may be part of the gauge group of the hidden sector), under which all of the fields of the observable sector and none of the fields of the hidden sector are invariant, and a group G_O (which may be part of the $SU(3) \times SU(2) \times U(1)$ gauge group of the observable sector), under which all of the fields of the hidden

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

sector and none of the fields of the observable sector are invariant. In this case if any observable sector fields appear in a term of the superpotential then there must be at least two of them, and also if any hidden sector fields appear in a term of the superpotential then there must be at least two of them, so there can be no term in a cubic polynomial superpotential that involves both hidden and observable sector fields. This argument leaves open the possibility of non-renormalizable terms in the superpotential that involve two or more factors of both hidden and observable sector superfields, a possibility to which we will return later. Of course, we are assuming that the strong interactions of the hidden sector produce additional non-perturbative terms in the total hidden sector superpotential $\tilde{f}(Z)$, but these too depend only on the hidden sector superfields.

Assuming then that the superpotential takes the form $f(\Phi) + \tilde{f}(Z)$, the potential of the scalar components of these superfields is given by Eq. (31.6.71) as

$$\begin{aligned} V = e^{\kappa^2 d}\Bigg[& \sum_{rs} g^{-1}_{rs} \left(\frac{\partial f}{\partial \phi_r} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial \phi_r}\right)\left(\frac{\partial f}{\partial \phi_s} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial \phi_s}\right)^* \\ & + 2\mathrm{Re}\sum_{rk} g^{-1}_{rk} \left(\frac{\partial f}{\partial \phi_r} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial \phi_r}\right)\left(\frac{\partial \tilde{f}}{\partial z_k} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial z_k}\right)^* \\ & + \sum_{kl} g^{-1}_{kl} \left(\frac{\partial \tilde{f}}{\partial z_k} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial z_k}\right)\left(\frac{\partial \tilde{f}}{\partial z_l} + \kappa^2 (f+\tilde{f})\frac{\partial d}{\partial z_l}\right)^* \\ & - 3\kappa^2 \left|f+\tilde{f}\right|^2 \Bigg] \\ & + \frac{1}{2}\mathrm{Re}\sum_{AB} f^{-1}_{AB} \left(\sum_{kl.} \frac{\partial d}{\partial z_k}(t_A)_{kl} z_l\right)\left(\sum_{mn} \frac{\partial d}{\partial z_m}(t_B)_{mn} z_n\right)^* \\ & + \frac{1}{2}\mathrm{Re}\sum_{AB} f^{-1}_{AB} \left(\sum_{rs} \frac{\partial d}{\partial \phi_r}(t_A)_{rs} \phi_s\right)\left(\sum_{tu} \frac{\partial d}{\partial \phi_t}(t_B)_{tu} \phi_u\right)^* . \end{aligned} \tag{31.7.2}$$

In writing the terms arising from gauge interactions, we are assuming here that there is no mixing of the gauge bosons of the hidden and observable sectors — that is, f^{-1}_{AB} vanishes for any pair of gauge generators t_A and t_B, for which t_A acts non-trivially on the ϕ_r and t_B acts non-trivially on the z_k, or vice-versa.

We are interested in exploring a region of field space where the scalar fields z_k of the hidden sector are of order Λ, the variable part of the superpotential $\tilde{f}(z)$ of the hidden sector is of order Λ^3, and, on dimensional grounds, $\partial \tilde{f}/\partial z_k$ is of order Λ^2. We will leave the magnitude of the constant part of $\tilde{f}$ open for the moment; as we shall see, we must include a constant

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

term in $\tilde{f}$ that is much larger than Λ^3 in order to cancel the cosmological constant.

We further define the region of field space to be explored as one in which the fields ϕ_r of the observable sector are of order $\kappa\Lambda^2$ because, as we shall see, this is the characteristic mass scale arising in the observable sector from gravitational effects of supersymmetry breaking in the hidden sector. The observable sector superpotential $f(\phi)$ for fields of order $\kappa\Lambda^2$ is assumed to be of order $\kappa^3\Lambda^6$, and its derivatives $\partial f(\phi)/\partial\phi_r$ are taken to be of order $\kappa^2\Lambda^4$.

In consequence of the definition of the superpotential and the scalar fields discussed at the end of the previous section, and the symmetry $G_H \times G_O$ assumed above, the modified Kahler potential takes the form*

$$\begin{aligned} d(\phi,\phi^*,z,z^*) = &\sum_r |\phi_r|^2 + \sum_k |z_k|^2 + O(\kappa^2 z^{*2} z^2) \\ &+O(\kappa^2 z^* z^3) + O(\kappa^2 z^{*3} z) + O(\kappa^2 \phi^{*2} z^2) + O(\kappa^2 \phi^{*2} z^* z) \\ &+O(\kappa^2 z^{*2}\phi^2) + O(\kappa^2 z^* z \phi^2) + O(\kappa^2 \phi^* \phi z^2) \\ &+O(\kappa^2 z^{*2}\phi^*\phi) + O(\kappa^2 \phi^* z^* \phi z) + O(\kappa^2 \phi^{*2}\phi^2) \\ &+O(\kappa^2 \phi^* \phi^3) + O(\kappa^2 \phi^{*3}\phi) + \cdots , \end{aligned} \tag{31.7.3}$$

with dots indicating terms of higher order. The metric (31.6.66) then has components

$$\begin{aligned} g_{rs} = &\delta_{rs} + O(\kappa^2 z^2) + O(\kappa^2 z^{*2}) + O(\kappa^2 z^* z) \\ &+O(\kappa^2\phi^2) + O(\kappa^2\phi^{*2}) + O(\kappa^2\phi^*\phi) + \cdots , \end{aligned} \tag{31.7.4}$$

$$\begin{aligned} g_{kl} = &\delta_{kl} + O(\kappa^2 z^2) + O(\kappa^2 z^{*2}) + O(\kappa^2 z^* z) \\ &+O(\kappa^2\phi^2) + O(\kappa^2\phi^{*2}) + O(\kappa^2\phi^*\phi) + \cdots , \end{aligned} \tag{31.7.5}$$

$$g_{rk} = g^*_{kr} = O(\kappa^2\phi z^*) + O(\kappa^2\phi z) + O(\kappa^2\phi^* z^*) + O(\kappa^2\phi^* z) + \cdots . \tag{31.7.6}$$

(The characteristic energy scale in d is assumed to be $1/\kappa$, because d is the modified Kahler potential produced by unknown dynamical effects at the Planck scale, in contrast with $\tilde{f}$, which gets its structure from dynamical effects of the strong gauge couplings at scale Λ.) The g_{rs} and g_{kl} are generically of order unity, while the mixed components g_{rk} and g_{kr} are of order of $\kappa^2(\kappa\Lambda^2)\Lambda = \kappa^3\Lambda^3 \ll 1$. It follows that the same is also true of the components of g^{-1}: $(g^{-1})_{rs}$ and $(g^{-1})_{kl}$ are generically of order unity, while the $(g^{-1})_{rk}$ and $(g^{-1})_{kr}$ are of order of $\kappa^3\Lambda^3 \ll 1$.

It follows from these estimates that unless cancellations intervene the dominant terms in the potential (31.7.2) will be at least of order Λ^4, and

* By $O(\kappa^2 z^{*2} z^2)$ is meant a term of the form $\kappa^2 \sum_{klmn} C_{klmn} z_k^* z_l^* z_m z_n$ with constant coefficients C_{klmn} of order unity, and likewise for the other terms in Eqs. (31.7.3)–(31.7.6).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

take the form

$$[V]_{\Lambda^4} = \sum_k \left| \frac{\partial \tilde{f}}{\partial z_k} \right|^2 - 3\kappa^2 \left| \tilde{f}^0 \right|^2 , \tag{31.7.7}$$

with $\tilde{f}^0$ the constant term in $\tilde{f}$, which will be needed to cancel the vacuum energy. We assume that supersymmetry is spontaneously broken in the hidden sector, which requires that there is a point z_k^0 at which $\sum_k |\partial \tilde{f}/\partial z_k|^2$ is at least a local minimum, but not zero. Then in order to cancel the vacuum energy to this order, we must take

$$3\kappa^2 \left| \tilde{f}^0 \right|^2 = \sum_k \left| \left(\frac{\partial \tilde{f}}{\partial z_k} \right)^0 \right|^2 , \tag{31.7.8}$$

the superscript 0 on the right-hand side indicating that the quantity is to be evaluated at $z = z^0$. Hence $\tilde{f}^0$ must be given an anomalously large value, of order Λ^2/κ. This is a more extreme fine tuning than will turn out to be necessary in the second version of gravitationally mediated supersymmetry breaking but, in the absence of a real understanding of the cosmological constant, some fine tuning will be necessary in any theory of supersymmetry breaking.

We can calculate the supersymmetry-breaking parameter F in our formula (31.3.17) for the gravitino mass by setting the vacuum energy density $F^2/2$ equal to the flat-space value $\sum_k |(\partial \tilde{f}/\partial z_k)^0|^2$. The gravitino mass is then given by Eqs. (31.3.17) and (31.7.8) as

$$m_g = \kappa \sqrt{\frac{1}{3} \sum_k \left| \left(\frac{\partial \tilde{f}(z)}{\partial z_k} \right)^0 \right|^2} = \kappa^2 \left| \tilde{f}^0 \right| . \tag{31.7.9}$$

This is of the same order $\approx \kappa\Lambda^2$ as the scalar fields of the observable sector.

Now let us turn to the terms in Eq. (31.7.2) that *do* depend on the observable sector scalars ϕ_r. We are considering field values for which the usual supersymmetric term $\sum_r |\partial f/\partial \phi_r|^2$ is of order $m_g^4 \approx \kappa^4\Lambda^8$, so we have to collect all ϕ-dependent terms in Eq. (31.7.2) of this order or greater. Let us look in turn at each of the six lines on the right-hand side of Eq. (31.7.2).

The leading term in $\kappa^2(f + \tilde{f})\partial d/\partial \phi_r$ is $\kappa^2 \tilde{f}^0 \phi_r^*$, which like $\partial f/\partial \phi_r$ is of order $\kappa^2(\Lambda^2/\kappa)(\kappa\Lambda^2) = \kappa^2\Lambda^4$, while other terms in $\kappa^2(f + \tilde{f})\partial d/\partial \phi_r$ are much smaller. To leading order we can approximate $\exp(\kappa^2 d)$ by unity and g_{rs}^{-1} by δ_{rs}, so in this order the first line of Eq. (31.7.2) gives

$$\sum_r \left| \frac{\partial f}{\partial \phi_r} + \kappa^2 \tilde{f}^0 \phi_r^* \right|^2 .$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

This is of the desired order $\kappa^4\Lambda^8$, so it is not necessary to consider higher-order corrections.

The leading term in g_{rk}^{-1} is of order $\kappa^3\Lambda^3$; the leading term in $\partial\tilde{f}/\partial z_k$ is of order Λ^2, while $\kappa^2(f+\tilde{f})\partial d/\partial z_k$ is smaller, of order $\kappa\Lambda^3$; and we have seen that the leading term in $\partial f/\partial\phi_r + \kappa^2(f+\tilde{f})\partial d/\partial\phi_r$ is of order $\kappa^2\Lambda^4$, so the term on the second line of Eq. (31.7.2) is of order $(\kappa^3\Lambda^3)(\Lambda^2)(\kappa^2\Lambda^4) = \kappa^5\Lambda^9$, which may be neglected in comparison with the terms in Eq. (31.7.2) of order $\kappa^4\Lambda^8$.

The leading terms in the third and fourth lines of Eq. (31.7.2) are of order Λ^4, but these are independent of ϕ_r. There is a ϕ dependence coming from the terms on the third line involving f; the leading terms of this sort are $2\kappa^2\mathrm{Re}\,[f\sum_k z_k^*(\partial\tilde{f}/\partial z_k)^*]$, which are of order $\kappa^5\Lambda^9 \ll \kappa^4\Lambda^8$, and hence may be neglected. Also, there is a ϕ dependence arising from the terms on the fourth line of Eq. (31.7.2) involving f. The leading terms of this sort are

$$-6\kappa^2\mathrm{Re}\,[f\tilde{f}^{0*}]$$

which are of order $\kappa^2(\kappa\Lambda^2)^3\Lambda^2/\kappa = \kappa^4\Lambda^8$. The factor $\exp(\kappa^2 d)$ contains ϕ-dependent terms of order $\kappa^2(\kappa\Lambda^2)^2$, but these multiply a potential whose leading terms (31.7.7) have been adjusted to cancel to order Λ^4, so the ϕ-dependent terms arising from this source are much less than of order $\kappa^4\Lambda^8$. There is one other type of ϕ-dependent term, arising from the ϕ-dependent terms in g_{kl}^{-1}. According to Eqs. (31.7.4)–(31.7.6), these terms may be written as $\kappa^2 u_{kl}(\phi,\phi^*)$, with u_{kl} a homogeneous quadratic polynomial in ϕ_r and ϕ_r^* with coefficients of order unity. These give rise to a ϕ-dependent term in the third line of Eq. (31.7.2), of the form

$$\kappa^2\sum_{kl} u_{kl}(\phi,\phi^*)\left(\frac{\partial\tilde{f}}{\partial z_k}\right)^0\left(\frac{\partial\tilde{f}}{\partial z_l}\right)^{0*},$$

which is of order $\kappa^2(\kappa\Lambda^2)^2\Lambda^4 = \kappa^4\Lambda^8$.

The leading ϕ-dependent term in the fifth line of Eq. (31.7.2) arises either from $\kappa^2\phi^2$ terms in f_{AB}^{-1}, with both ds given by their leading-order terms, of order z^*z, or from terms in one of the ds with two factors of κ, two factors of ϕ and/or ϕ^*, and two factors of z and/or z^*, with f_{AB} and the other d given by their leading-order terms, of order 1 and z^*z respectively. Both types of ϕ-dependent term make a contribution of order $\kappa^2(\kappa\Lambda^2)^2\Lambda^4 = \kappa^4\Lambda^8$, so higher-order terms may be neglected.

The leading term in the sixth line of Eq. (31.7.2) arises from the leading term in f_{AB}^{-1}, which is of order unity, and the leading ϕ-dependent terms in d, which are of order $\phi^*\phi$. This makes a contribution to the potential of order $(\kappa\Lambda^2)^4$, so here too higher-order terms may be neglected.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The potential (31.7.2) also contains ϕ-independent terms of order $\kappa^2\Lambda^6$, $\kappa^4\Lambda^8$, and so on. The terms of order $\kappa^2\Lambda^6$ may be cancelled and the terms of order $\kappa^4\Lambda^8$ can be given an arbitrary value $\mathscr{C}$ by making a small shift of the constant term in $\tilde{f}(z)$ away from the value given by Eq. (31.7.8).

Putting these results together, to order $\kappa^4\Lambda^8 \approx m_g^4$ the potential of the observable sector is now

$$V_O(\phi,\phi^*) = \sum_r \left|\frac{\partial f(\phi)}{\partial \phi_r} + \kappa^2 \tilde{f}^0 \phi_r^*\right|^2 - 6\kappa^2 \mathrm{Re}\left[f(\phi)\tilde{f}^{0*}\right] + \frac{1}{2}\sum_A \left|\sum_{rs} \phi_r^*(t_A)_{rs}\phi_s\right|^2 + Q(\phi,\phi^*) + \mathscr{C}\,, \tag{31.7.10}$$

where $Q(\phi,\phi^*)$ is a quadratic polynomial in ϕ and/or ϕ^* with coefficients of order $\kappa^2\Lambda^4 \approx m_g^2$, which arises from the ϕ-dependent terms in g_{kl}^{-1} on the third line of Eq. (31.7.2) and from the ϕ-dependent terms in f_{AB}^{-1} and d on the fifth line of Eq. (31.7.2). We have normalized the gauge superfields so that $f_{AB} = \delta_{AB}$ when all scalar fields vanish. (There is also a ϕ dependence arising from a ϕ-dependent shift in the equilibrium value of the hidden sector scalars z_k, but this shift is at most of order $(\kappa\Lambda^2)^4/\Lambda^3$, and since (31.7.7) is supposed to be stationary at $z = z^0$ this shift enters quadratically in the effective potential of the observable sector fields, and may therefore be neglected.) The constant $\mathscr{C}$ may be chosen so that the value of this potential at its minimum is zero.

Finally, let us return to the non-renormalizable terms in the superpotential. As already mentioned, the leading Φ-dependent terms of this sort may be expected to be of order $\kappa\Phi^2Z^2$. When the hidden sector superfields Z_k are set equal to their equilibrium values z_k^0 these terms become a second-order polynomial in the Φ_r, with coefficients of order $\kappa\Lambda^2$. Thus to leading order, we can take account of these non-renormalizable terms by simply including a quadratic polynomial function of the Φ_r in the superpotential, with coefficients of order $\kappa\Lambda^2 \approx m_g$.

In this way theories of gravity-mediated supersymmetry breaking avoid the problem of the μ-term, discussed in Sections 28.1 and 28.5. Recall that the $SU(3) \times SU(2) \times U(1)$ symmetry of the standard model allows a single superrenormalizable term in the superpotential of the minimal supersymmetric standard model, a term $\mu(H_1^{\mathrm{T}} e H_2)$. In order to explain in a natural way why the coefficient μ is not of the order of the Planck mass, it is necessary to impose some sort of symmetry, like the 'Peccei–Quinn' symmetry discussed in connection with strong CP violation in Section 23.6, under which the product $(H_1^{\mathrm{T}} e H_2)$ is not neutral. But a μ-term with μ of the same order m_g as other supersymmetry-breaking

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

masses was found phenomenologically necessary in Section 28.5. Such a term can arise naturally from the breaking of the Peccei–Quinn symmetry by the vacuum expectation value of the hidden sector fields,** if the superpotential contains a non-renormalizable term in which $(H_1^{\mathrm{T}} e H_2)$ appears multiplied with two powers of the hidden sector fields z_k, with a coefficient of order κ.

We assume then that the effective superpotential $f(\phi)$ consists of a homogeneous polynomial $f^{(3)}(\phi)$ of third order in the fields ϕ_r, with coefficients very roughly of order unity, plus a μ-term that takes the form of a homogeneous polynomial $f^{(2)}(\phi)$ of second order in the fields ϕ_r, with coefficients of order $m_g \approx \kappa\Lambda^2$. The potential (31.7.10) then becomes

$$\begin{aligned} V_O(\phi,\phi^*) = &\sum_r \left|\frac{\partial f(\phi)}{\partial \phi_r}\right|^2 + \frac{1}{2}\sum_{AB}\left|\sum_{rs}\phi_r^*(t_A)_{rs}\phi_s\right|^2 \\ &-2\kappa^2 \mathrm{Re}\left[f^{(2)}(\phi)\tilde{f}^{0*}\right] + \kappa^4|\tilde{f}^0|^2\sum_r|\phi_r|^2 \\ &+Q(\phi,\phi^*) + \mathscr{C}\,. \end{aligned} \tag{31.7.11}$$

The terms on the first line of the right-hand side give a supersymmetric potential, while the terms on the second and third lines represent a soft breaking of supersymmetry. With $\tilde{f}^0 \approx \Lambda^2/\kappa$ and $(\partial\tilde{f}/\partial z)^0 \approx \Lambda^2$, the dimensional constants in the soft supersymmetry-breaking terms in Eq. (31.7.10) are all powers of $\kappa\Lambda^2 \approx m_g$, so that this is where we expect to find the expectation values of the observable sector scalar fields, justifying our choice of this as the region of field space to explore.

Setting $\kappa\Lambda^2$ equal to a typical mass ≈ 1 TeV in the effective Lagrangian of the supersymmetric standard model, we now find $\Lambda \approx 10^{11}$ GeV. It is mildly encouraging that, as discussed in Section 23.6, the spontaneous breakdown of the Peccei–Quinn symmetry at a scale $\Lambda \approx 10^{11}$ GeV is just what is needed to resolve the strong CP problem with a symmetry-breaking scale in the window from 10^{10} GeV to 10^{12} GeV allowed by astronomical observations.

Comparing the potential with the scalar field terms in the Lagrangian density (28.4.1) of the minimum supersymmetric standard model, we see that this version of gravity-mediated supersymmetry breaking predicts that supersymmetry is broken only by soft scalar mass terms with coefficients

** This is known as the Giudice–Masiero mechanism[17]. It is often described in terms of non-renormalizable holomorphic and antiholomorphic terms in the modified Kahler potential d but, as discussed at the end of the previous section, any such terms can be traded for holomorphic factors in the superpotential. Here we have *defined* d to not contain holomorphic and antiholomorphic terms, and with this definition the μ-term can only arise from non-renormalizable terms in the superpotential.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

(including $B\mu$) of order m_g^2. To leading order in $\kappa\Lambda$, the coefficients A and C of trilinear supersymmetry-breaking terms vanish.

One serious problem with these results is that there is no reason why the quadratic polynomial $Q(\phi, \phi^*)$ in Eq. (31.7.11) should respect the degeneracy among squark masses and among slepton masses discussed in Section 28.4, that would avoid unobserved flavor-changing processes. However, the fourth term in Eq. (31.7.11) for the potential makes an additional contribution $\kappa^4|\tilde{f}^0|^2 \approx \kappa^2\Lambda^4$ to scalar squared masses that is the same for all scalars, so the constraints imposed by experimental upper bounds on flavor-changing processes might be satisfied if the coefficients in $Q(\phi, \phi^*)$ (which arise from terms on the third and fifth lines of Eq. (31.7.2)) happen to be small compared with $\kappa^2\Lambda^4$.

We would have an interesting relation among the parameters of the minimum supersymmetric standard model if it were really true that $Q(\phi, \phi^*)$ could be neglected. Taking the quadratic part $f^{(2)}$ of the superpotential as $\mu(\phi_1^{\mathrm{T}} e \phi_2)$, the coefficient $B\mu$ in Eq. (28.4.1) would be given by the second term of Eq. (31.7.11) as $B\mu = -\kappa^2\mu\tilde{f}^{0*}$, so

$$|B| = \kappa^2|\tilde{f}^0| = m_g \,,$$

in agreement with Eq. (31.4.13). Also, all the squark and slepton masses M_s would be given by the third term in Eq. (31.7.11) as $\kappa^2|\tilde{f}^0|$, and so we would have the new relation

$$|B| = M_s \,. \tag{31.7.12}$$

With all squark and slepton masses equal, there would be no inconsistency with limits on flavor-changing processes. Furthermore, with Q neglected there would be only one complex parameter $\tilde{f}^0$ in the supersymmetry-breaking part of the potential (31.7.11), which can be chosen to be real by a redefinition of the over-all phase of the superpotential, so now the supersymmetry-breaking part of the potential would introduce no new violation of CP invariance. But there is no known reason why Q should be small.

Another serious problem with this version of gravitationally-mediated supersymmetry breaking is that it does not yield sufficiently large gaugino masses.[18] According to Eq. (31.6.75), the mass matrix of the $SU(3) \times SU(2) \times U(1)$ gauginos is given in the tree approximation by

$$m_{AB} = \exp(\kappa^2 d/2) \sum_{NM} [g^{-1}]_{NM} L_N \left(\frac{\partial f_{AB}}{\partial \varphi_M} \right)^* \,, \tag{31.7.13}$$

where here φ_N runs over all the scalar fields ϕ_r and z_k on which f_{AB} may depend, with g_{NM} and L_M given by Eqs. (31.6.66) and (31.6.69). According to the estimates we have made here, $\kappa^2 d = O(\kappa^2\Lambda^2) \ll 1$; $L_k = O(\Lambda^2)$ while L_r is much smaller; and g_{kl}^{-1} is of order unity. Also, we assume that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

f_{AB} is a term of order unity plus a term of order κ^2 times a bilinear in scalar fields and their complex conjugates, so $\partial f_{AB}/\partial z_k$ is of order $\kappa^2\Lambda$. This gives gaugino masses of order $\Lambda^2 \times \kappa^2\Lambda$. This is much smaller than the gravitino mass $m_g \approx \kappa\Lambda^2$ (which sets the scale of the supersymmetry-breaking terms in the potential (31.7.11) of the observable sector scalar fields) by a factor of order $\kappa\Lambda \approx 10^{-7}$, so if the scalar masses produced by supersymmetry breaking are of order 1 TeV, then the gaugino masses will be of order 100 keV, which is far too low to be consistent with the fact that gauginos have not yet been observed.

There are several ways that this problem might be avoided. One is to include gauge-singlet scalar fields among the z_k of the hidden sector, which can appear linearly in f_{AB}.[19] In this case $\partial f_{AB}/\partial z_k$ will be of order κ rather than $\kappa^2\Lambda$, yielding gaugino masses of order $\kappa\Lambda^2$, which is comparable to the squark and slepton masses. One trouble with this approach is that the inclusion of scalars that are neutral with respect to all gauge groups would make it no longer natural for the renormalizable part of the superpotential to take the form $f(\phi) + \tilde{f}(z)$.

Even without gauge singlets, there are one-loop contributions to the gluino, wino, and bino masses (and to the A-parameters), calculated in Section 31.4. For instance, if we take m_g at the largest value ≈ 10 TeV allowed by the 'naturalness' bound discussed in Section 28.1, then with $g_s^2/4\pi = 0.118$, Eq. (31.4.13) would give a gluino mass $3g_s^2 m_g/16\pi^2 = 280$ GeV, which is certainly high enough to allow the gluino to have escaped detection. The bino and wino masses depend on the unknown ratio of the μ parameter and the pseudoscalar Higgs mass m_A. Taking this ratio equal to unity and $m_g < 10$ TeV, Eqs. (31.4.15) and (31.4.16) would give $m_{\text{bino}} = 9g'^2 m_g/16\pi^2 < 73$ GeV and $m_{\text{wino}} = g^2 m_g/16\pi^2 < 27$ GeV.[19a] This bound on the wino mass is in conflict with the fact that wino pairs have not been seen at LEP in e^+–e^- collisions at an energy sufficiently high to produce W boson pairs, so that $m_{\text{wino}} > m_W$. In order to avoid this contradiction it would be necessary[11b] either to take $m_g > 30$ TeV, which is awkward from the point of view of naturalness, or $\mu^2/m_A^2 > 8$. In any case, this model has the general consequence that the gauginos are much lighter than the squarks and sleptons.

If the polynomial $Q(\phi, \phi^*)$ can be neglected then, as shown in Section 31.4, the A-parameters are also given by one-loop corrections. These are of order $g_s^2 m_g/16\pi^2$ for squarks and $g^2 m_g/16\pi^2$ or $g'^2 m_g/16\pi^2$ for sleptons.

Second Version[20]

This version of gravity-mediated supersymmetry breaking has an observable sector with chiral superfields Φ_r and a hidden sector with chiral

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

superfields Z_k, just as in the first version. The difference is that supersymmetry is now assumed to be *not* spontaneously broken in the hidden sector. Instead, the gauge couplings of the hidden sector, which become strong at an energy $\Lambda \ll m_{\rm Pl}$, produce a non-perturbative superpotential for the scalar fields of a third sector of superfields, the *modular superfields.* In various theories such as modern superstring theories there are extra dimensions that are not observed because they have been 'rolled up' into a tiny compact manifold, with size roughly of the order of κ. Typically some of the parameters that are needed to describe this compact manifold are not fixed in any order of perturbation theory. The values of these parameters may vary from point to point in four-dimensional spacetime and appear at energies far below the Planck scale κ^{-1} as gauge-invariant scalar fields y_a, known as *modular fields.* (The indices a, b, etc. here of course have nothing to do with the local Lorentz frame indices used in Section 31.6.) Assuming that supersymmetry is not broken in the compactification of the extra dimensions, these fields must be accompanied by fermionic superpartners and auxiliary fields, which together form gauge-invariant left-chiral modular superfields Y_a and their adjoints.[†]

Just below the compactification scale we have a supersymmetric theory with a superpotential that can depend on all the superfields, but in which κ is the only dimensional parameter. The circumstance that the y_a are not fixed in perturbation theory typically arises because the compactification does not result in any superpotential for the modular superfields Y_a alone. As we saw in Section 28.1, the $SU(3) \times SU(2) \times U(1)$ gauge symmetries rule out any terms in the superpotential with just one or two factors of observable sector superfields, except for a possible term bilinear in the Higgs superfields H_1 and H_2. We will again assume that this bilinear term in the bare superpotential is either accidentally absent (in which case it does not appear in any order of perturbation theory), or is ruled out by some symmetry, such as the 'Peccei–Quinn' symmetry discussed in connection with strong CP violation in Section 23.6. The gauge symmetry of the hidden sector rules out any terms in the superpotential with just one factor of the Z_k, and we shall assume that either accident or some symmetry (perhaps the same Peccei–Quinn symmetry) also rules out terms with two factors of the Z_k.

The bare superpotential therefore takes the form

$$f_{\rm bare}(\Phi, Y, Z) = \sum_{rst} f_{rst}(\kappa Y)\,\Phi_r\Phi_s\Phi_t + \sum f_{klm}(\kappa Y) Z_k Z_l Z_m + \cdots , \tag{31.7.14}$$

[†] Of course, modular fields may exist even under the assumptions of the first version of gravity-mediated supersymmetry breaking, but because of the smaller value of Λ in that case the couplings of the modular fields are too weak to be of interest there.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

where f_{rst} and f_{klm} are power series in their arguments with coefficients roughly of order unity, and the dots indicate terms involving $n > 3$ factors of the Φs and Zs, as well as any number of factors of κY_a, suppressed by factors proportional to κ^{n-3}.

We assume that non-perturbative effects in the hidden sector such as 'gaugino condensation' (the appearance of expectation values of bilinear functions of gaugino fields) without themselves breaking supersymmetry produce a superpotential for the modular superfields. Since Λ is the only scale in the problem (aside from gravitational effects suppressed by factors of $\kappa\Lambda$), this superpotential would have to be of the form

$$\hat{f}(Y) = \Lambda^3 F(\kappa Y)\,. \tag{31.7.15}$$

Such terms can also be produced by expectation values of the scalar components of the Z superfields in the Z^3 term in Eq. (31.7.14). At the same time, the replacement of the Z_k by the expectation values of their scalar components in the non-renormalizable terms indicated by dots in Eq. (31.7.14) will produce additional Φ-dependent terms in the effective superpotential, about which more later. The superpotential (31.7.15) is of the form (31.7.1) originally assumed in theories of gravitationally mediated supersymmetry breaking, but with ϵ now identified as the intermediate scale Λ at which the gauge interactions of the hidden sector become strong.

It is plausible that the superpotential (31.7.15) can lead to supersymmetry breaking by the appearance of $\mathscr{F}$-terms for the modular chiral superfields. For the moment, let's ignore the other superfields, leaving the justification for later. The potential for the modular scalars is given by Eqs. (31.6.68) and (31.6.69) as

$$\hat{V}(y, y^*) = \exp\left(\kappa^2 \hat{d}(y, y^*)\right) \left[\sum_{ab} [\hat{g}^{-1}(y, y^*)]_{ab} \hat{L}_a(y) \hat{L}_b(y)^* - 3\kappa^2 |\hat{f}(y)|^2\right]\,, \tag{31.7.16}$$

where $\hat{d}(y, y^*)$ is the Kahler d function with the scalar components of the Z_k and Φ_r neglected, and

$$\hat{g}_{ab} = \frac{\partial^2 \hat{d}}{\partial y_a \partial y_b^*}\,, \tag{31.7.17}$$

$$\hat{L}_a = \frac{\partial \hat{f}}{\partial y_a} + \kappa^2 \hat{f} \frac{\partial \hat{d}}{\partial y_a}\,. \tag{31.7.18}$$

We are assuming here that $\hat{V}$ has a stationary point, labelled with a superscript 0, where $\hat{L}_a^0 \neq 0$, so that supersymmetry is broken, but that $\hat{V}^0$ is very small, so that it can be cancelled by terms arising from the observable sector, leaving us with a flat spacetime. Since $\hat{f}$ is of the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

form (31.7.15) and $\hat{d}$ equals κ^{-2} times a power series in κy_a and κy_a^* with coefficients of order unity, the whole potential is of the form $\kappa^2\Lambda^6$ times a power series in κy_a and κy_a^*, again with coefficients of order unity. The orders of magnitude of the various ingredients in the potential are therefore

$$\begin{aligned} &y_a^0 = O(\kappa^{-1}) , \qquad \hat{f}^0 = O(\Lambda^3) , \qquad \hat{d}^0 = O(\kappa^{-2}) , \\ &\hat{L}_a^0 = O(\kappa\Lambda^3) , \qquad \hat{g}_{ab}^0 = O(1) . \end{aligned} \tag{31.7.19}$$

With the modular and hidden sector fields fixed at their expectation values, the superpotential of the observable sector is now of the form

$$f(\Phi) = \sum_{rs} \mu_{rs}\Phi_r\Phi_s + \sum_{rst} g_{rst}\Phi_r\Phi_s\Phi_t + \cdots , \tag{31.7.20}$$

where g_{rst} is $f_{rst}(\kappa y_0)$, which is assumed to be roughly of order unity, plus terms suppressed by powers of $\kappa\Lambda$. The dots here denote terms with more than three factors of Φ that are suppressed by additional factors of $\kappa\Phi$. The coefficient μ_{rs} arises from the non-renormalizable terms denoted by dots in Eq. (31.7.14); if it comes from a term with $n > 1$ factors of Zs as well as two factors of Φs, then it has an order of magnitude

$$\mu_{rs} = O(\kappa^{n-1}\Lambda^n) . \tag{31.7.21}$$

We shall see that the desired order of magnitude of μ is $m_g = O(\kappa^2\Lambda^3)$, which would come from a term with $n = 3$ factors of Zs.

The supersymmetry breaking in the modular sector will be transmitted to the observable sector by effects of the gravitational field and its superpartners. Eq. (31.6.68) gives the potential of the observable sector as

$$\begin{aligned} V_O = e^{\kappa^2 d^0} \Bigg[&\sum_{rs} [g^{0\,-1}]_{rs} \left(\frac{\partial f}{\partial \phi_r} + \kappa^2 (f + \hat{f}^0) \frac{\partial d^0}{\partial \phi_r} \right) \left(\frac{\partial f}{\partial \phi_s} + \kappa^2 (f + \hat{f}^0) \frac{\partial d^0}{\partial \phi_s} \right)^* \\ &+ 2\,\mathrm{Re} \sum_{ra} [g^{0\,-1}]_{ra} \left(\frac{\partial f}{\partial \phi_r} + \kappa^2 (f + \hat{f}^0) \frac{\partial d^0}{\partial \phi_r} \right) L_a^{0*} \\ &+ \sum_{ab} [g^{0\,-1}]_{ab} \hat{L}_a^0 \hat{L}_b^{0*} - 3\kappa^2 \left| f + \hat{f}^0 \right|^2 \Bigg] \\ &+ \frac{1}{2} \mathrm{Re} \sum_{AB} [f_{AB}^{-1}]^0 \left(\sum_{rs} \frac{\partial d^0}{\partial \phi_r} (t_A)_{rs} \phi_s \right) \left(\sum_{tu} \frac{\partial d^0}{\partial \phi_t} (t_A)_{tu} \phi_u \right)^* , \end{aligned} \tag{31.7.22}$$

with the superscript zero again indicating that the modular and hidden sector scalar fields are fixed at their equilibrium values. (This will be reconsidered later.) Note that, although Eq. (31.7.22) involves terms like $\hat{L}_a^0$ that arise in the modular sector, there are no terms here that refer explicitly to the hidden sector. This is because in this version of gravity-mediated supersymmetry breaking, supersymmetry is assumed not to be

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

broken in this sector, so that $L_k^0 = 0$, and $D_A^0 = 0$ for any gauge field that interacts with the hidden sector fields.

We are interested in exploring a region of field space in which the observable sector fields are of order $\kappa^2\Lambda^3$, because, as we shall see, this is the characteristic mass arising in the observable sector from gravitational effects of supersymmetry breaking in the hidden sector. According to Eq. (31.6.48) and our estimate $\hat{f}^0 \approx \Lambda^3$, this is also the order of magnitude of the gravitino mass m_g:

$$m_g \approx \kappa^2\Lambda^3 \,.$$

In order to calculate the potential for fields of this order, we note that the Kahler d function for the observable and modular scalar fields takes the form

$$d(\phi,\phi^*,y,y^*) = \kappa^{-2}\hat{d}(\kappa y,\kappa y^*) + \sum_{rs}\phi_r\phi_s^* A_{rs}(\kappa y,\kappa y^*) + \sum_{rs}\phi_r\phi_s B_{rs}(\kappa y,\kappa y^*) + \sum_{rs}\phi_r^*\phi_s^* B_{rs}^*(\kappa y,\kappa y^*) + \cdots \,, \quad (31.7.23)$$

where $\hat{d}$, A_{rs}, and B_{rs} are power series in their arguments with coefficients of order unity, and the dots indicate terms with $n > 2$ factors of ϕ and/or ϕ^*, suppressed by factors κ^{n-2}. According to Eq. (31.6.72), we can remove any holomorphic term in d along with its complex conjugate by multiplying the total superpotential by a suitable holomorphic factor. In particular, by multiplying the total superpotential with a factor $\exp[\kappa^2\hat{d}^0 + \kappa^2\sum_{rs}B_{rs}^0\phi_r\phi_s]$, we can arrange that the transformed d function has

$$\hat{d}^0 = 0 \,, \qquad B_{rs}^0 = 0 \,. \quad (31.7.24)$$

We shall assume that this has been done. Note that, because the total superpotential contains a constant term $\hat{f}^0$ of order Λ^3, while B_{rs}^0 was of order unity, this transformation generates a term in the superpotential that is quadratic in ϕ_r, making a contribution to its coefficient μ_{rs} of order $\kappa^2\Lambda^3$, of the same order as ϕ and m_g.

There can be another contribution to this coefficient of the same order of magnitude, arising from terms in the superpotential with $n > 1$ factors of Ys as well as two factors of Φs. Eq. (31.7.21) tells us that, in order for the contribution to the μ_{rs} coefficients from this also to be of order $\kappa^2\Lambda^3$, we must have $n = 3$. We can arrange that this is allowed, while terms with two factors of Φs and $n = 2$ factors of Ys are forbidden, by giving H_1 and H_2 Peccei–Quinn quantum numbers $+1$ and the Ys quantum numbers $-2/3$. The breaking of this Peccei–Quinn symmetry by the vacuum expectation values of the Y_k then produces an axion.

With μ_{rs} of order $\kappa^2\Lambda^3$, both the bilinear and trilinear terms in the superpotential of the observable sector are of order $\kappa^6\Lambda^9$. The usual

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

supersymmetric potential term $\sum_r |\partial f/\partial\phi_r|^2$ is then of order $\kappa^8\Lambda^{12}$, so we have to collect all terms in Eq. (31.7.22) of this order or greater.

With ϕ_r of order $\kappa^2\Lambda^3$ and y_a fixed at its equilibrium value $y_a^0 \approx \kappa^{-1}$, Eq. (31.7.23) gives

$$g_{rs}^0 = A_{rs}^0 + O(\kappa^3\Lambda^3)\,, \tag{31.7.25}$$

$$g_{ab}^0 = \left(\frac{\partial^2\hat{d}}{\partial y_a \partial y_b^*}\right)^0 + \sum_{rs}\phi_r\phi_s^*\left(\frac{\partial^2 A_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \sum_{rs}\phi_r\phi_s\left(\frac{\partial^2 B_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \sum_{rs}\phi_r^*\phi_s^*\left(\frac{\partial^2 B_{rs}^*}{\partial y_a\partial y_b^*}\right)^0 + O(\kappa^6\Lambda^6)\,, \tag{31.7.26}$$

$$g_{ra}^0 = \sum_s \phi_s\left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0 + O(\kappa^6\Lambda^6) = g_{ar}^{0\,*}\,, \tag{31.7.27}$$

where the zero superscript again indicates that the y_a are fixed at their equilibrium values y_a^0. We will subject the observable-sector superfields Φ_r and the modular superfields Z_a to separate linear transformations, designed so that

$$A_{rs}^0 = \delta_{rs}\,, \qquad \left(\frac{\partial^2\hat{d}}{\partial y_a\partial y_b^*}\right)^0 = \delta_{ab}\,. \tag{31.7.28}$$

With the metric (31.7.25)–(31.7.27) given by the unit matrix plus terms that are much less than unity, it is easy to calculate the inverse:

$$g_{rs}^{0\,-1} = \delta_{rs} + O(\kappa^3\Lambda^3)\,, \tag{31.7.29}$$

$$g_{ab}^{0\,-1} = \delta_{ab} - \sum_{rs}\phi_r\phi_s^*\left(\frac{\partial^2 A_{rs}}{\partial y_a\partial y_b^*}\right)^0 - \sum_{rs}\phi_r\phi_s\left(\frac{\partial^2 B_{rs}}{\partial y_a\partial y_b^*}\right)^0 - \sum_{rs}\phi_r^*\phi_s^*\left(\frac{\partial^2 B_{rs}^*}{\partial y_a\partial y_b^*}\right)^0 + O(\kappa^6\Lambda^6)\,, \tag{31.7.30}$$

$$g_{ra}^{0\,-1} = -\sum_s \phi_s\left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0 + O(\kappa^6\Lambda^6) = g_{ar}^{0\,-1\,*}\,. \tag{31.7.31}$$

In particular, our assumption about the form of the functions A_{rs} and B_{rs} gives the order-of-magnitude estimates

$$\left(\frac{\partial^2 A_{rs}}{\partial y_a\partial y_b^*}\right)^0 = O(\kappa^2)\,, \qquad \left(\frac{\partial^2 B_{rs}}{\partial y_a\partial y_b^*}\right)^0 = O(\kappa^2)\,, \qquad \left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0 = O(\kappa)\,, \tag{31.7.32}$$

so that, for $\phi_r = O(\kappa^2\Lambda^3)$,

$$[g^{0\,-1}]_{rs} = O(1)\,, \quad [g^{0\,-1}]_{ab} = O(1)\,, \quad [g^{0\,-1}]_{ra} = O(\kappa^3\Lambda^3)\,. \tag{31.7.33}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Further,

$$f = O(\phi^3) = O(\kappa^6\Lambda^9)\,, \qquad \frac{\partial f}{\partial \phi_r} = O(\phi^2) = O(\kappa^4\Lambda^6)\,, \tag{31.7.34}$$

and

$$\frac{\partial d^0}{\partial \phi_r} = O(\phi) = O(\kappa^2\Lambda^3)\,. \tag{31.7.35}$$

With $\hat{f}^0$ of order Λ^3, the quantities $\kappa^2(f+\hat{f}^0)\partial d/\partial\phi_r$ in Eq. (31.7.22) are dominated by $\kappa^2\hat{f}^0\phi_r^*$, which is of order $\kappa^2\times\Lambda^3\times\kappa^2\Lambda^3 = \kappa^4\Lambda^6$. This is of the same order of magnitude as $\partial f/\partial\phi_r$, so in leading order we must keep both terms:

$$\frac{\partial f}{\partial \phi_r} + \kappa^2(f+\hat{f}^0)\frac{\partial d}{\partial \phi_r} \simeq \frac{\partial f}{\partial \phi_r} + \kappa^2\hat{f}^0\,\phi_r^* = O(\kappa^4\Lambda^6)\,. \tag{31.7.36}$$

In this approximation, and with $g^{0\,-1}_{rs}$ replaced by its dominant term δ_{rs}, the first term in the square brackets in Eq. (31.7.22) is already of the desired order $\kappa^8\Lambda^{12}$, so we may use these approximations to write

$$\sum_{rs}[g^{0\,-1}]_{rs}\left(\frac{\partial f}{\partial \phi_r} + \kappa^2(f+\hat{f}^0)\frac{\partial d^0}{\partial \phi_r}\right)\left(\frac{\partial f}{\partial \phi_s} + \kappa^2(f+\hat{f}^0)\frac{\partial d^0}{\partial \phi_s}\right)^* \simeq \sum_r\left|\frac{\partial f}{\partial \phi_r} + \kappa^2\hat{f}^0\,\phi_r^*\right|^2\,. \tag{31.7.37}$$

The second term in the square brackets in Eq. (31.7.22) is of order $\kappa^3\Lambda^3\times\kappa^4\Lambda^6\times\kappa\Lambda^3 = \kappa^8\Lambda^{12}$, so we may evaluate it using only leading terms and find

$$2\mathrm{Re}\sum_{ra}[g^{0\,-1}]_{ra}\left(\frac{\partial f}{\partial \phi_r} + \kappa^2(f+\hat{f}^0)\frac{\partial d^0}{\partial \phi_r}\right)\hat{L}_a^{0*} \simeq -2\mathrm{Re}\sum_{ras}\phi_s\left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0\left[\frac{\partial f}{\partial \phi_r} + \kappa^2\hat{f}^0\,\phi_r^*\right]\hat{L}_a^{0*}\,. \tag{31.7.38}$$

The third and fourth terms in the square brackets in Eq. (31.7.22) are individually of order $\kappa^2\Lambda^6$, but are assumed nearly to cancel, so non-leading terms must be included in evaluating their contribution to the quantity in square brackets. One contribution comes from the terms in Eq. (31.7.30) for $g^{0\,-1}_{ab}$ that are of second order in ϕ and/or ϕ^*; this contribution is of order $\kappa^2\phi^2\times(\kappa\Lambda^3)^2$, which is of order $\kappa^8\Lambda^{12}$. Another non-cancelling contribution to the potential from the last term in square brackets in Eq. (31.7.22) comes from the interference between f and $\hat{f}^0$, which makes a contribution of order $\kappa^2\phi^3\Lambda^3$, which is also of order $\kappa^8\Lambda^{12}$. There is also a constant contribution $\mathscr{C}$ from a possible failure of the last two terms in the square brackets in Eq. (31.7.22) to cancel when $\phi_r = 0$;

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

to avoid a large cosmological constant, we shall have to assume that $\mathscr{C}$ is also of order $\kappa^8\Lambda^{12}$. (This is an unnatural fine-tuning, which so far has been necessary in any theory that avoids a huge cosmological constant.) Putting these estimates together, to order $\kappa^8\Lambda^{12}$ the last two terms in the square brackets in Eq. (31.7.22) are

$$\sum_{ab}[g_{ab}^{-1}]^0\hat{L}_a^0\hat{L}_b^{0*} - 3\kappa^2\left|f+\hat{f}^0\right|^2 \simeq -\sum_{abrs}\left[\phi_r\phi_s^*\left(\frac{\partial^2 A_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \phi_r\phi_s\left(\frac{\partial^2 B_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \phi_r^*\phi_s^*\left(\frac{\partial^2 B_{rs}^*}{\partial y_a\partial y_b^*}\right)^0\right]\hat{L}_a^0\hat{L}_b^{0*} - 6\kappa^2\mathrm{Re}\,(f\,\hat{f}^{0*}) + \mathscr{C}\,. \tag{31.7.39}$$

All these terms are of the desired order $\kappa^8\Lambda^{12}$, while Eq. (31.7.24) gives $\kappa^2 d^0 = O(\kappa^2\phi^2) = O(\kappa^6\Lambda^6)$, so we can ignore the factor $\exp(\kappa^2 d^0)$ in Eq. (31.7.22). Finally, the gauge terms on the last line of Eq. (31.7.22) are of order $\phi^4 = O(\kappa^8\Lambda^{12})$, so these can be evaluated using the leading term ϕ_r^* in $\partial d^0/\partial\phi_r$.

Putting this all together, the complete scalar potential of the observable sector to order $\kappa^8\Lambda^{12}$ is

$$\begin{aligned} V_O \simeq & \sum_r\left|\frac{\partial f}{\partial\phi_r} + \kappa^2\hat{f}^0\,\phi_r^*\right|^2 \\ & -2\mathrm{Re}\sum_{ras}\phi_s\left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0\left[\frac{\partial f}{\partial\phi_r} + \kappa^2\hat{f}^0\,\phi_r^*\right]\hat{L}_a^{0*} \\ & -\sum_{abrs}\left[\phi_r\phi_s^*\left(\frac{\partial A_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \phi_r\phi_s\left(\frac{\partial B_{rs}}{\partial y_a\partial y_b^*}\right)^0 + \phi_r^*\phi_s^*\left(\frac{\partial B_{rs}^*}{\partial y_a\partial y_b^*}\right)^0\right]\hat{L}_a^0\hat{L}_b^{0*} - 6\kappa^2\mathrm{Re}\,(f\,\hat{f}^{0*}) \\ & +\frac{1}{2}\mathrm{Re}\sum_{AB}[f_{AB}^{-1}]^0\left(\sum_{rs}\phi_r^*(t_A)_{rs}\phi_s\right)\left(\sum_{tu}\phi_t^*(t_B)_{tu}\phi_u\right)^* + \mathscr{C}\,. \end{aligned} \tag{31.7.40}$$

The potential (31.7.40) of the observable sector scalar fields takes the form assumed in the minimal supersymmetric standard models discussed in Section 28.4: it is the sum of a supersymmetric term V_{susy} and a soft supersymmetry-breaking term V_{soft}

$$V_O = V_{\mathrm{susy}} + V_{\mathrm{soft}}\,. \tag{31.7.41}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

The supersymmetric term is, as usual,

$$V_{\text{susy}} = \sum_r \left|\frac{\partial f}{\partial \phi_r}\right|^2 + \frac{1}{2}\text{Re}\sum_{AB}[f_{AB}^{-1}]^0 \left(\sum_{rs}\phi_r^*(t_A)_{rs}\phi_s\right)\left(\sum_{tu}\phi_t^*(t_B)_{tu}\phi_u\right)^* \tag{31.7.42}$$

and the soft supersymmetry breaking term here is

$$\begin{aligned} V_{\text{soft}} \simeq\ & 2\kappa^2\text{Re}\sum_r\left(\phi_r\frac{\partial f}{\partial\phi_r}\hat{f}^{0*}\right) + \kappa^4\left|\hat{f}^0\right|^2\sum_r|\phi_r|^2 \\ & -2\text{Re}\sum_{ras}\phi_s\left(\frac{\partial A_{rs}}{\partial y_a^*}\right)^0\left[\frac{\partial f}{\partial\phi_r}+\kappa^2\hat{f}^0\,\phi_r^*\right]\hat{L}_a^{0*} \\ & -\sum_{abrs}\left[\phi_r\phi_s^*\left(\frac{\partial A_{rs}}{\partial y_a\partial y_b^*}\right)^0+\phi_r\phi_s\left(\frac{\partial B_{rs}}{\partial y_a\partial y_b^*}\right)^0\right. \\ & \left.+\phi_r^*\phi_s^*\left(\frac{\partial B_{rs}^*}{\partial y_a\partial y_b^*}\right)^0\right]\hat{L}_a^0\hat{L}_b^{0*} - 6\kappa^2\text{Re}\,(f\,\hat{f}^{0*}) + \mathscr{C}\,. \end{aligned} \tag{31.7.43}$$

With $f(\phi)$ given by Eq. (31.7.20), this takes the form

$$V_{\text{soft}} = \sum_{rs}M_{rs}^2\phi_r\phi_s^* + 2\text{Re}\sum_{rs}N_{rs}^2\phi_r\phi_s + 2\text{Re}\sum_{rst}A_{rst}\phi_r\phi_s\phi_t + \mathscr{C}\,, \tag{31.7.44}$$

where

$$M_{rs}^2 = \kappa^4|\hat{f}^0|^2\delta_{rs} - 2\kappa^2\text{Re}\left[\hat{f}^0\sum_a\left(\frac{\partial A_{sr}}{\partial y_a^*}\right)^0\hat{L}_a^{0*}\right] - \sum_{ab}\left(\frac{\partial^2 A_{rs}}{\partial y_a\partial y_b^*}\right)^0\hat{L}_a^0\hat{L}_b^{0*}\,, \tag{31.7.45}$$

$$\begin{aligned} N_{rs}^2 =\ & 2\kappa^2\mu_{rs}\hat{f}^{0*} - \frac{1}{2}\sum_{at}\mu_{tr}\left(\frac{\partial A_{ts}}{\partial y_a^*}\right)^0\hat{L}_a^{0*} - \frac{1}{2}\sum_{at}\mu_{ts}\left(\frac{\partial A_{tr}}{\partial y_a^*}\right)^0\hat{L}_a^{0*} \\ & -\sum_{ab}\left(\frac{\partial^2 B_{rs}}{\partial y_a\partial y_b^*}\right)^0\hat{L}_a^0\hat{L}_b^{0*}\,, \end{aligned} \tag{31.7.46}$$

and

$$A_{rst} = -\sum_{au}\left[\left(\frac{\partial A_{ur}}{\partial y_a^*}\right)^0 g_{ust} + \left(\frac{\partial A_{us}}{\partial y_a^*}\right)^0 g_{urt} + \left(\frac{\partial A_{ut}}{\partial y_a^*}\right)^0 g_{urs}\right]\hat{L}_a^{0*}\,. \tag{31.7.47}$$

Our previous order-of-magnitude estimates give

$$M_{rs}^2 = O(\kappa^4\Lambda^6)\,, \qquad N_{rs}^2 = O(\kappa^4\Lambda^6)\,, \qquad A_{rst} = O(\kappa^2\Lambda^3)\,. \tag{31.7.48}$$

Together with our estimates $g_{rst} = O(1)$ and $\mu_{rs} = O(\kappa^2\Lambda^3)$ for the constants in V_{susy}, this shows that if the potential has a stationary point at some $\phi_a^0 \neq 0$, then $\phi_a^0 = O(\kappa^2\Lambda^3)$, justifying our decision to explore fields of

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

this order. For an equilibrium value of the ϕ_r of this order, the equilibrium value of the various terms in the potential is $O(\phi^4) = O(\kappa^8\Lambda^{12})$, so this is the order of magnitude of the constant $\mathscr{C}$ that is needed to cancel the vacuum energy.

In order for the characteristic mass $\kappa^2\Lambda^3$ in the minimal supersymmetric standard model to be of order 1 TeV we need $\Lambda \approx 10^{13}$ GeV. If it is a Peccei–Quinn symmetry that forbids a term in the bare superpotential with two superfield factors each from the observable and hidden sectors, as suggested above, then vacuum expectation values of the hidden sector scalars will break this symmetry, with a symmetry-breaking scale (called M in Section 23.6) of order 10^{13} GeV, and an axion mass which is then given by Eq. (23.6.26) as of order 10^{-6} eV. This value for the symmetry-breaking scale is somewhat above the upper bound of 10^{12} GeV quoted in Section 23.6, but given the uncertainty of cosmological arguments, this contradiction is not decisive.

Before considering further physical implications of these results, let's pause to reconsider a short-cut that we took in deriving them. In calculating the potential for the observable sector scalars ϕ_r, we fixed the modular fields at the equilibrium values y_a^0 that they would have in the absence of the observable sector fields ϕ_r. Instead, we ought to set the modular fields at their equilibrium values $y_a(\phi)$ for the actual values of the ϕ_r, by finding the stationary point of the potential

$$V_{\text{total}}(\phi, \phi^*, y, y^*) = \hat{V}(y, y^*) + V_O(\phi, \phi^*, y, y^*) \,, \tag{31.7.49}$$

and only then seek an equilibrium value for the ϕ_r. Since V_O is much smaller than $\hat{V}$ for fields of interest, the equilibrium value of y_a can be written as

$$y_a(\phi, \phi^*) = y_a^0 + \delta y_a(\phi, \phi^*) \,, \tag{31.7.50}$$

where y_a^0 is at the minimum of $\hat{V}(y, y^*)$ and

$$\sum_b \frac{\partial^2 \hat{V}}{\partial y_a \partial y_b} \delta y_b + \sum_b \frac{\partial^2 \hat{V}}{\partial y_a \partial y_b^*} \delta y_b^* = -\frac{\partial V_O}{\partial y_a} \,. \tag{31.7.51}$$

The second derivatives of $\hat{V}$ are of order $\kappa^2 \times (\kappa\Lambda^3)^2 = \kappa^4\Lambda^6$, while the first derivatives of V_O are of order $\kappa \times \kappa^8\Lambda^{12} = \kappa^9\Lambda^{12}$, so the δy_a are of order $\kappa^5\Lambda^6$. The change in the potential due to this ϕ-dependent shift in the equilibrium values of the y_a is quadratic in the δy_a and δy_a^*, with coefficients given by second derivatives of $\hat{V}$ with respect to y_a and/or y_a^*, and is therefore of order

$$(\kappa^5\Lambda^6)^2 \times \kappa^2 \times (\kappa\Lambda^3)^2 = \kappa^{14}\Lambda^{18} \,,$$

which is less than the potential we have calculated by a factor $(\kappa\Lambda)^6 \ll 1$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Without further assumptions regarding the functions $A_{rs}(y, y^*)$ and $B_{rs}(y, y^*)$, we can learn nothing whatever from Eqs. (31.7.45)–(31.7.47) about the precise values of the coefficients M_{rs}^2, N_{rs}^2, and A_{rst} in the soft supersymmetry-breaking potential (31.7.44). The only definite prediction that emerges from these results is that the coefficients C_{ij} in the soft supersymmetry-breaking Lagrangian (28.4.1) are all negligible, as has generally been assumed anyway.

The greatest problem presented by the results (31.7.44)–(31.7.48) is that without further assumptions they do not insure the degeneracy of squark masses and of slepton masses that would avoid the quark and lepton flavor-changing processes discussed in Section 28.4. The $\phi_r\phi_s$-terms in the soft supersymmetry-breaking potential (31.7.44) and in the superpotential (31.7.20) are not allowed by $SU(3) \times SU(2) \times U(1)$ to depend on anything but the Higgs scalars, so they cannot lead to flavor-changing processes. The problem therefore arises only from the fact that the coefficients M_{rs}^2 and A_{rst} in Eq. (31.7.44) may not conserve flavor in the same basis as the Yukawa couplings g_{rst}. One way to avoid this problem is to assume that for some reason the functions $A_{rs}(y, y^*)$ happen to depend only weakly on the y_a and y_a^*, so that Eq. (31.7.45) gives $M_{rs}^2 \propto \delta_{rs}$ and Eq. (31.7.47) makes A_{rst} (and hence the A_{ij} coefficients in Eq. (28.4.1)) anomalously small. Another possibility is that, although not slowly varying, for some reason the whole function $A_{rs}(y, y^*)$ (or at least its first and second derivatives at $y_a = y_a^0$) is proportional to δ_{rs}. In this case, Eq. (31.7.45) again gives $M_{rs}^2 \propto \delta_{rs}$, and now Eq. (31.7.47) gives trilinear couplings $A_{rst} \propto g_{rst}$, so that the A_{ij} coefficients in Eq. (28.4.1) would be all equal.[21]

We must also check the gaugino masses produced by this version of gravitationally mediated supersymmetry breaking. According to Eq. (31.6.75), the mass matrix of the $SU(3) \times SU(2) \times U(1)$ gauginos is given in general by

$$m_{AB} = \exp(\kappa^2 d/2) \sum_{NM} [g^{-1}]_{NM} L_N \left(\frac{\partial f_{AB}}{\partial \varphi_M} \right)^* , \tag{31.7.52}$$

where φ_N here runs over all the fields on which f_{AB} may depend, with g_{NM} and L_M given by Eqs. (31.6.66) and (31.6.69). According to the estimates we have made here, $\kappa^2 d \ll 1$; $L_a = O(\kappa\Lambda^3)$ for the modular fields y_a and is much smaller for the other fields; and $g_{ab}^{-1} \simeq \delta_{ab}$. Also, we assume that f_{AB} is a power series in κy with coefficients of order unity, so $\partial f_{AB}/\partial y_a$ is of order κ. The gaugino masses (31.7.52) are therefore of order $\kappa^2\Lambda^3$, which is the same as the order of magnitude of the scalar masses and expectation values, and hence is likely to be large enough to avoid conflict with observation. The one-loop corrections considered in Section 31.4 are much smaller here, and do not need to be taken into account.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

In summary, the first version of gravity-mediated supersymmetry breaking has the advantage of giving an axion mass that is within cosmological bounds, while the second version has the advantage of giving the gauginos masses that are comparable to the masses of the squarks and sleptons. Both versions of gravity-mediated supersymmetry breaking have an advantage over theories of gauge-mediated supersymmetry breaking: they naturally give μ-terms of the experimentally necessary order of magnitude. On the other hand, theories of gauge-mediated supersymmetry breaking have the advantage of naturally yielding generation-independent squark and slepton masses.

Theories with either version of gravity-mediated supersymmetry breaking may naturally entail the existence of slowly decaying superheavy particles, which could have interesting astrophysical effects.[22] It is plausible that the gauge interactions of the hidden sector that become strong at energy Λ can bind composite particles with masses of order Λ. These superheavy particles may be long-lived if their decay is forbidden by accidental symmetries of the renormalizable part of the hidden sector Lagrangian and only occurs through non-renormalizable terms in the Lagrangian, which are suppressed by factors of $\kappa\Lambda$.

Appendix The Vierbein Formalism

The familiar formulation of gravity in terms of a metric tensor $g_{\mu\nu}$ is adequate for theories with matter fields restricted to scalars, vectors, and tensors, but not for supergravity, where spinors are an indispensable ingredient. Unlike vectors and tensors, spinors have a Lorentz transformation rule that has no natural generalization to arbitrary coordinate systems. Instead, to deal with spinors, we have to introduce systems of coordinates $\xi^a_X(x)$ with $a = 0, 1, 2, 3$ that are locally inertial at any given point X in an arbitrary coordinate system. The Principle of Equivalence tells us that gravitation has no effect in these locally inertial coordinates, so the action may then be expressed in terms of matter fields like spinors, vectors, etc. that are defined in these locally inertial frames, as well as the vierbein, which arises from the transformation between the locally inertial and general coordinates

$$e^a{}_\mu(X) \equiv \left.\frac{\partial \xi^a_X(x)}{\partial x^\mu}\right|_{x=X} . \tag{31.A.1}$$

The action will be invariant under general coordinate transformations $x^\mu \to x'^\mu$ *and* local Lorentz transformations $\xi^a \to \xi'^a = \Lambda^a{}_b(x)\xi^b$ with $\Lambda^a{}_c(x)\Lambda^b{}_d(x)\eta_{ab} = \eta_{cd}$. The definition (31.A.1) of the vierbein shows that,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

under general coordinate transformations $x \to x'$, it transforms as

$$e^a{}_\mu(x) \to e'^a{}_\mu(x') = \frac{\partial x^\nu}{\partial x'^\mu} e^a{}_\nu(x) , \tag{31.A.2}$$

while, under a local Lorentz transformation $\xi^a(x) \to \Lambda^a{}_b(x)\xi^b(x)$, it transforms as

$$e^a{}_\mu(x) \to \Lambda^a{}_b(x) e^b{}_\mu(x) . \tag{31.A.3}$$

For instance, theories of pure gravitation may be expressed in terms of a field that is invariant under local Lorentz transformations and that transforms as a tensor under general coordinate transformations, the metric

$$g_{\mu\nu} \equiv e^a{}_\mu e^b{}_\nu \eta_{ab} . \tag{31.A.4}$$

Vectors may be regarded either as quantities V^a that transform as vectors under local Lorentz transformations

$$V^a(x) \to \Lambda^a{}_b(x) V^b(x) \tag{31.A.5}$$

but as scalars under general coordinate transformations, or as quantities v^μ that transform as scalars under local Lorentz transformations but as vectors under general coordinate transformations, the two being related by

$$V^a = e^a{}_\mu v^\mu .$$

But the supergravity action also involves spinor fields, which necessarily transform like scalars under general coordinate transformations but as spinors under local Lorentz transformations

$$\psi_\alpha(x) \to D_{\alpha\beta}(\Lambda(x))\psi_\beta(x) , \tag{31.A.6}$$

where $D_{\alpha\beta}(\Lambda)$ is the spinor representation of the homogeneous Lorentz group.

Because the Lorentz transformations in Eqs. (31.A.5) and (31.A.6) depend on the coordinate x^μ, the spacetime derivative of a quantity like $V^a(x)$ or $\psi_\alpha(x)$ is not just another quantity that transforms in the same way under local Lorentz transformations and as a covariant vector under general coordinate transformations. For instance, the derivative of Eq. (31.A.6) gives the local Lorentz transformation rule

$$\partial_\mu \psi_\alpha \to D_{\alpha\beta}(\Lambda) \left\{ \partial_\mu \psi_\beta + \left[D^{-1}(\Lambda) \partial_\mu D(\Lambda) \right]_{\beta\gamma} \psi_\gamma \right\} .$$

To cancel the second term in the brackets on the right-hand side we introduce a connection matrix Ω_μ with the local Lorentz transformation property

$$\Omega_\mu \to D(\Lambda)\Omega_\mu D^{-1}(\Lambda) - \left(\partial_\mu D(\Lambda)\right) D^{-1}(\Lambda) \tag{31.A.7}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

and define a covariant derivative

$$\mathscr{D}_\mu \psi \equiv \partial_\mu \psi + \Omega_\mu \psi \,, \tag{31.A.8}$$

which transforms under local Lorentz transformations like ψ itself:

$$\mathscr{D}_\mu \psi \to D(\Lambda) \mathscr{D}_\mu \psi \,. \tag{31.A.9}$$

Also, Ω_μ must transform like a covariant vector under general coordinate transformations, so that $\mathscr{D}_\mu$ will give a covariant vector when acting on a coordinate scalar. In order that $\mathscr{D}_\mu$ should give a tensor with one extra lower index when acting on a tensor, it must be supplemented with the usual affine connection term. For instance, when acting on the gravitino field ψ_μ, the covariant derivative is defined as

$$\mathscr{D}_\mu \psi_\nu \equiv \psi_{\nu;\mu} + \Omega_\mu \psi_\nu \equiv \partial_\mu \psi_\nu - \Gamma^\lambda_{\mu\nu} \psi_\lambda + \Omega_\mu \psi_\nu \,. \tag{31.A.10}$$

Eqs. (31.A.8)–(31.A.10) apply not only for spinors, but for fields that transform under local Lorentz transformations according to arbitrary representations $D(\Lambda)$ of the Lorentz group. The matrix Ω_μ depends on this representation, but in any representation it can be written in the form

$$[\Omega_\mu]_{\alpha\beta}(x) = \frac{1}{2} i [\mathscr{J}_{ab}]_{\alpha\beta} \omega_\mu^{ab}(x) \,, \tag{31.A.11}$$

where $\mathscr{J}_{ab}$ are the matrices representing the generators of the homogeneous Lorentz group in the representation furnished by the fields in question:

$$i [\mathscr{J}_{ab}, \mathscr{J}_{cd}] = \eta_{bc} \mathscr{J}_{ad} - \eta_{ac} \mathscr{J}_{bd} + \eta_{bd} \mathscr{J}_{ca} - \eta_{ad} \mathscr{J}_{cb} \,, \tag{31.A.12}$$

and ω_μ^{ab} is a representation-independent field known as the *spin connection* that transforms as a covariant vector under general coordinate transformations. To satisfy the inhomogeneous local Lorentz transformation rule (31.A.7) we can take

$$\omega_\mu^{ab} = g^{\nu\lambda} e^a{}_\nu e^b{}_{\lambda;\mu} \,, \tag{31.A.13}$$

with the semi-colon again denoting an ordinary covariant derivative, constructed using the affine connection $\Gamma^\lambda_{\mu\nu}$. (This is antisymmetric in a and b because Eq. (31.A.4) gives $g^{\nu\lambda} e^a{}_\nu e^b{}_\lambda = \eta^{ab}$, a quantity with vanishing covariant derivative.) This is not the unique spin-connection for which Eq. (31.A.7) is satisfied; to it, we can add any field that is a covariant vector under general coordinate transformations and a tensor under local Lorentz transformations, a freedom of some importance in supergravity theories.

For any choice of spin connection there is a corresponding curvature tensor. From Eq. (31.A.7) it is straightforward to show that the quantity $\partial_\nu \Omega_\mu - \partial_\mu \Omega_\nu + [\Omega_\nu, \Omega_\mu]$ transforms homogeneously under local Lorentz

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

transformations

$$\partial_\nu \Omega_\mu - \partial_\mu \Omega_\nu + [\Omega_\nu, \Omega_\mu] \to D(\Lambda)\left(\partial_\nu \Omega_\mu - \partial_\mu \Omega_\nu + [\Omega_\nu, \Omega_\mu]\right) D^{-1}(\Lambda)\,. \tag{31.A.14}$$

Using Eqs. (31.A.11) and (31.A.12), this matrix can be expressed as

$$\partial_\nu \Omega_\mu - \partial_\mu \Omega_\nu + [\Omega_\nu, \Omega_\mu] = \tfrac{1}{2} i\, \mathcal{J}_{ab} R_{\mu\nu}{}^{ab}\,, \tag{31.A.15}$$

where

$$R_{\mu\nu}{}^{ab} \equiv \partial_\nu \omega_\mu^{ab} - \partial_\mu \omega_\nu^{ab} + \omega_\nu^{ac} \omega_{\mu\, c}{}^{b} - \omega_\mu^{ac} \omega_{\nu\, c}{}^{b}\,. \tag{31.A.16}$$

From Eq. (31.A.14) it follows that $R_{\mu\nu}{}^{ab}$ transforms as a tensor under local Lorentz transformations

$$R_{\mu\nu}{}^{ab} \to D(\Lambda)^a{}_c D(\Lambda)^b{}_d R_{\mu\nu}{}^{cd}\,. \tag{31.A.17}$$

It also obviously transforms as a tensor under general coordinate transformations

$$R_{\mu\nu}{}^{ab} \to \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} R_{\rho\sigma}{}^{ab}\,. \tag{31.A.18}$$

We can therefore form a coordinate tensor of fourth rank by writing

$$R_{\mu\nu}{}^{ab} = e^a{}_\kappa\, e^b{}_\lambda\, R_{\mu\nu}{}^{\kappa\lambda}\,. \tag{31.A.19}$$

The tensor $R_{\mu\nu}{}^{\kappa\lambda}$ constructed in this way is the Riemann–Christoffel curvature tensor corresponding to the particular spin connection ω_ν^{ab}.

Problems

1. Derive formulas (31.2.3)–(31.2.6) for components of the Einstein superfield.

2. Suppose that supersymmetry is unbroken. Show how to calculate the amplitude for emission of a gravitino of very low energy in a general scattering process in terms of the amplitude for this process without the gravitino.

3. Check that the supergravity action (31.6.11) is invariant under the local supersymmetry transformations (31.6.1)–(31.6.6) to all orders in G.

4. Calculate the change of the generalized D-component (31.6.40) under a general local supersymmetry transformation.

5. Calculate the fermionic part of the Lagrangian density (31.6.49).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

6. Consider the theory of a single chiral scalar superfield Φ interacting with supergravity, with modified Kahler potential $d(\Phi, \Phi^*) = \Phi^*\Phi$ and superpotential $f(\Phi) = M^2(\Phi + \beta)$, where M and β are constants. Find a value of β for which the classical field equations have a solution with flat spacetime. What is the value of ϕ for this solution?

References

1. P. Nath and R. Arnowitt, *Phys. Lett.* **56B**, 177 (1975); B. Zumino, in *Proceedings of the Conference on Gauge Theories and Modern Field Theories at Northeastern University, 1975*, R. Arnowitt and P. Nath, eds. (MIT Press, Cambridge, MA, 1976). This approach is described in detail by J. Wess and J. Bagger, *Supersymmetry and Supergravity*, 2nd edition (Princeton University Press, Princeton, NJ, 1992).
2. D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, *Phys. Rev.* **D13**, 3214 (1976); S. Deser and B. Zumino, *Phys. Lett.* **62B**, 335 (1976); S. Ferrara, J. Scherk, and P. van Nieuwenhuizen, *Phys. Rev. Lett.* **37**, 1035 (1976); S. Ferrara, F. Gliozzi, J. Scherk, and P. van Nieuwenhuizen, *Nucl. Phys.* **B117**, 333 (1976). These articles are reprinted in *Supersymmetry*, S. Ferrara, ed. (North Holland/World Scientific, Amsterdam/Singapore, 1987). Some of these results had been obtained by D. V. Volkov and V. A. Soroka, *JETP Lett.* **18**, 312 (1973); *Theor. Math. Phys.* **20**, 829 (1974). For a clear description of this approach, see P. West, *Introduction to Supersymmetry and Supergravity*, 2nd edition (World Scientific, Singapore, 1990).
3. S. Ferrara and B. Zumino, *Nucl. Phys.* **B134**, 301 (1978).

3*a*. M. T. Grisaru and H. N. Pendleton, *Phys. Lett.* **67B**, 323 (1977).

4. K. Stelle and P. C. West, *Phys. Lett.* **74B**, 330 (1978); S. Ferrara and P. van Nieuwenhuizen, *Phys. Lett.* **74B**, 333 (1978). These articles are reprinted in *Supersymmetry*, Ref. 2.
5. See, for example, S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Sec. 12.5.
6. See, for example, *Gravitation and Cosmology*, Ref. 5: Eq. (12.4.3).
7. See, for example, *Gravitation and Cosmology*, Ref. 5, Section 10.1.

7*a*. W. Nahm, *Nucl. Phys.* **B135**, 149 (1978).

8. S. Coleman and F. de Luccia, *Phys. Rev.* **D21**, 3305 (1980).
9. S. Weinberg, *Phys. Rev. Lett.* **48**, 1776 (1982).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

10. S. Deser and C. Teitelboim, *Phys. Rev. Lett.* **39**, 249 (1977); M. Grisaru, *Phys. Lett.* **73B**, 207 (1978); E. Witten, *Commun. Math. Phys.* **80**, 381 (1981).

11. W. Rarita and J. Schwinger, *Phys. Rev.* **60**, 61 (1941).

11*a*. L. Randall and R. Sundrum, hep-th/9810155, to be published; G. F. Giudice, M. Luty, R. Rattazzi, and H. Murayama, *JHEP* **12**, 027 (1998); A. Pomerol and R. Rattazzi, hep-ph/9903448, to be published; E. Katz, Y. Shadmi, and Y. Shirman, hep-ph/9906296, to be published.

11*b*. G. F. Giudice, M. Luty, R. Rattazzi, and H. Murayama, Ref. 11a.

12. K. Stelle and P. C. West, Ref. 4; S. Ferrara and P. van Nieuwenhuizen, Ref. 4; E. Cremmer, B. Julia, J. Scherk, S. Ferrara, L. Girardello, and P. van Nieuwenhuizen, *Phys. Lett.* **79B**, 231 (1978); *Nucl. Phys.* **B147**, 105 (1979) (reprinted in *Supersymmetry*, Ref. 2.); D. G. Boulware, S. Deser, and J. H. Kay, *Physica* **280**, 141 (1979); E. Cremmer, S. Ferrara, L. Girardello, and A. Van Proeyen, *Phys. Lett.* **116B**, 231 (1982); *Nucl. Phys.* **B212**, 413 (1983) (reprinted in *Supersymmetry*, Ref. 2). For this construction in two spacetime dimensions, see S. Deser and B. Zumino, *Phys. Lett.* **65B**, 369 (1976).

13. S. Deser and B. Zumino, *Phys. Rev. Lett.* **38**, 1433 (1977). This article is reprinted in *Supersymmetry*, Ref. 2. Also see D. Z. Freedman and A. Das, *Nucl. Phys.* **B120**, 221 (1977); P. K. Townsend, *Phys. Rev.* **D 15**, 2802 (1977).

13*a*. E. Cremmer, S. Ferrara, C. Kounnas, and D. V. Nanopoulos, *Phys. Lett.* **133B**, 61 (1983). For a review of models based on this idea, see A. B. Lahanas and D. V. Nanopoulos, *Phys. Rept.* **145**, 1 (1987).

14. H. P. Nilles, *Phys. Lett.* **115B**, 193 (1982); A. Chamseddine, R. Arnowitt, and P. Nath, *Phys. Rev. Lett.* **49**, 970 (1982); R. Barbieri, S. Ferrara, and C. A. Savoy, *Phys. Lett.* **119B**, 343 (1982); E. Cremmer, P. Fayet, and L. Girardello, *Phys. Lett.* **122B**, 41 (1983); L. Ibañez, *Phys. Lett.* **118B**, 73 (1982); H. P. Nilles, M. Srednicki, and D. Wyler, *Phys. Lett.* **120B**, 346 (1983); L. Hall, J. Lykken, and S. Weinberg, *Phys. Rev.* **D27**, 2359 (1983); L. Alvarez-Gaumé, J. Polchinski, and M. B. Wise, *Nucl. Phys.* **B221**, 495 (1983). The above articles are reprinted in *Supersymmetry*, Ref. 2. Also see S. Ferrara, D. V. Nanopoulos, and C. A. Savoy, *Phys. Lett.* **12B**, 214 (1983); J. M. Leon, M. Quiros, and M. Ramon Medrano, *Phys. Lett.* **127B**, 85 (1983); *Phys. Lett.* **129B**, 61 (1983); N. Ohta, *Prog. Theor. Phys.* **70**, 542 (1983); P. Nath, R. Arnowitt, and A. Chamseddine, *Phys. Lett.* **121B**, 33 (1983); J. Ellis, D. V. Nanopoulos, and K. Tamvakis, *Phys.*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

Lett. **121B**, 123 (1983). For a review, see H. P. Nilles, *Phys. Rept.* **110**, 1 (1984).

15. The breakdown of supersymmetry in a supergravity theory with an unobserved sector of chiral scalars seems to have been first proposed in an unpublished University of Budapest preprint by J. Polonyi (1977).
16. I. Affleck, M. Dine, and N. Seiberg, *Nucl. Phys.* **B256**, 557 (1985).
17. G. F. Giudice and A. Masiero, *Phys. Lett.* **B206**, 480 (1988). Also see J. A. Casas and C. Muñoz, *Phys. Lett.* **B306**, 288 (1993); J. E. Kim, hep-ph/9901204, to be published.
18. M. Dine and D. A. MacIntire, *Phys. Rev.* **D46**, 2594 (1992).
19. T. Banks, D. B. Kaplan, and A. Nelson, *Phys. Rev.* **D49**, 779 (1994); K. I. Izawa and T. Yanagida, *Prog. Theor. Phys.* **94**, 1105 (1995); A. Nelson, *Phys. Lett.* **B369**, 277 (1996).

19*a*. The cosmological implications of a wino that is lighter than the bino are discussed by T. Moroi and L. Randall, hep-ph/9906527, to be published.

20. V. Kaplunovsky and J. Louis, *Phys. Lett.* **B306**, 269 (1993); *Nucl. Phys.* **B422**, 57 (1994). Specific models of this sort have been proposed by P. Binétruy, M. K. Gaillard, and Y-Y. Wu, *Nucl. Phys.* **B493**, 27 (1997); *Phys. Lett.* **B412**, 288 (1997).
21. Theories with extra dimensions in which flavor-changing effects are suppressed have been suggested by L. Randall and R. Sundrum, Ref. 11a. Problems with such theories were pointed out by Z. Chacko, M. A. Luty, I. Maksymyk, and E. Pontón, hep-ph/9905390, to be published.
22. K. Hamaguchi, K.-I. Izawa, Y. Nomura, and T. Yanagida, hep-ph/9903207, to be published.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.010

32

Supersymmetry Algebras in Higher Dimensions

Ever since the ground-breaking work of Kaluza[1] and Klein,[2] theorists have from time to time tried to formulate a more nearly fundamental physical theory in spacetimes of higher than four dimensions. This approach was revived in superstring theories, which take their simplest form in 10 spacetime dimensions.[3] More recently, it has been suggested that the various versions of string theory may be unified in a theory known as *M theory*, which in one limit is approximately described by supergravity in 11 spacetime dimensions.[4] In this chapter we shall catalog the different types of supersymmetry algebra possible in higher dimensions, and use them to classify supermultiplets of particles.

32.1 General Supersymmetry Algebras

Our analysis of the general supersymmetry algebra in higher dimensions will follow the same logical outline as the work of Haag, Lopuszanski, and Sohnius[5] on supersymmetry algebras in four spacetime dimensions, described in Section 25.2. The proof of the Coleman–Mandula theorem in the appendix of Chapter 24 makes it clear that the list of possible bosonic symmetry generators is essentially the same in $d > 2$ spacetime dimensions as in four spacetime dimensions: in an S-matrix theory of particles, there are only the momentum d-vector P^μ, a Lorentz generator $J^{\mu\nu} = -J^{\nu\mu}$ (with μ and ν here running over the values $1, 2, \ldots, d-1, 0$), and various Lorentz scalar 'charges.' (In some theories there are topologically stable extended objects such as closed strings, membranes, etc., in addition to particles, which make possible other conserved quantities, to which we will return in Section 32.3.) The anticommutators of the fermionic symmetry generators with each other are bosonic symmetry generators, and therefore must be a linear combination of P^μ, $J^{\mu\nu}$, and various conserved scalars. This puts severe limits on the Lorentz transformation properties of the fermionic generators, and on the superalgebra to which they belong.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

We will first prove that the general fermionic symmetry generator must transform according to the fundamental spinor representations of the Lorentz group, which are reviewed in the appendix to this chapter, and not in higher spinor representations, such those obtained by adding vector indices to a spinor. As we saw in Section 25.2, the proof for $d = 4$ by Haag, Lopuszanski, and Sohnius made use of the isomorphism of $SO(4)$ to $SU(2) \times SU(2)$, which has no analog in higher dimensions. Here we will use an argument of Nahm,[6] which is actually somewhat simpler and applies in any number of dimensions.

Since the Lorentz transform of any fermionic symmetry generator is another fermionic symmetry generator, the fermionic symmetry generators furnish a representation of the homogeneous Lorentz group $O(d-1,d)$ (or, strictly speaking, of its covering group $Spin(d-1,1)$). Assuming that there are at most a finite number of fermionic symmetry generators, they must transform according to a finite-dimensional representation of the homogeneous Lorentz group. All of these representations can be obtained from the finite-dimensional *unitary* representations of the corresponding orthogonal group $O(d)$ (actually $Spin(d)$) by setting $x^d = ix^0$. So let us first consider the transformation of the fermionic generators under $O(d)$. For d even or odd, we can find $d/2$ or $(d-1)/2$ Lorentz generators J_{d1}, J_{23}, $J_{45}, \ldots$, which all commute with each other, and classify fermionic generators Q according to the values σ_{d1}, σ_{23}, ... that they destroy:

$$[J_{d1}\,,\,Q] = -\sigma_{d1}Q\,, \qquad [J_{23}\,,\,Q] = -\sigma_{23}Q\,, \qquad [J_{45}\,,\,Q] = -\sigma_{45}Q, \ldots\,. \tag{32.1.1}$$

Since the finite-dimensional representations of $O(d)$ are all unitary, the σs are all real.

Let us focus on one of these quantum numbers, $\sigma_{d1} \equiv w$ and refer to any fermionic or bosonic operator O as having *weight* w if

$$[J_{d1}\,,\,O] = -w\,O\,, \tag{32.1.2}$$

or, in terms of the Minkowski component $J_{01} = iJ_{d1}$,

$$[J_{01}\,,\,O] = -iw\,O\,. \tag{32.1.3}$$

The reason for concentrating on this particular quantum number is that it has the special property of being the same for an operator and its Hermitian adjoint. This is because J_{01} must be represented on Hilbert space (though not on field variables or symmetry generators) by a Hermitian operator, so that (remembering that w is real) the Hermitian adjoint of Eq. (32.1.3) is

$$-[J_{01}\,,\,O^*] = +iwO^*\,, \tag{32.1.4}$$

so O^* has the same weight as O.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Now consider the anticommutator $\{Q, Q^*\}$ of any fermionic symmetry generator Q with its Hermitian adjoint. According to the Coleman–Mandula theorem, it is at most a linear combination of P_μ, $J_{\mu\nu}$, and scalars. To calculate the weights of the components of P_μ, we recall the commutation relation (2.4.13)

$$i[P_\mu, J_{\rho\sigma}] = \eta_{\mu\rho} P_\sigma - \eta_{\mu\sigma} P_\rho \,,$$

which shows that $P_0 \pm P_1$ has weight $w = \pm 1$, while the other components $P_2, P_3, \ldots, P_{d-1}$ all have weight zero. In the same way, the commutation relation (2.4.12) of the $J_{\mu\nu}$ with each other show that $J_{0i} \pm J_{1i}$ with $i = 2, 3, \ldots d-1$ have weight $w = \pm 1$, the J_{ij} with both i and j between 2 and $d-1$ have weight zero, J_{10} has weight zero, and of course all scalars have weight zero. We conclude then that all bosonic symmetry generators have weight ± 1 or 0 and the anticommutator $\{Q, Q^*\}$ must be a linear combination of operators with such weights. If Q has weight w then $\{Q, Q^*\}$ has weight $2w$, and it is manifestly non-zero for any non-zero Q, so each fermionic generator can only have weight $\pm 1/2$. (Weight zero is excluded by the connection between spin and statistics — fermionic operators can only be constructed from odd numbers of operators with half-integer weights.) Going back to the Euclidean formalism, since the commutators of the particular $O(d)$ generator J_{01} with all generators Q in a representation of $O(d)$ are given by Eq. (32.1.2) with $w = \pm 1/2$, and there is nothing special about the 01 plane, $O(d)$ invariance requires that the same is true for all $O(d)$ generators J_{ij}, so that all the σs in Eq. (32.1.1) are $\pm 1/2$. The only irreducible representations of the homogeneous Lorentz group with all σs equal to $\pm 1/2$ are the fundamental spinor representations, so Q must belong to some direct sum of these representations.

We can also use this approach to show that the fermionic generators Q all commute with the d-momentum P_μ. For this purpose, note that the double commutator of a momentum operator $P_0 \pm P_1$ of weight ± 1 with any fermionic generator Q would have weight either $\pm 5/2$ if Q has weight $\pm 1/2$ or weight $\pm 3/2$ if Q has weight $\mp 1/2$, and since we have found that there are no fermionic symmetry generators of weight $\pm 3/2$ or $\pm 5/2$, these double commutators must all vanish:

$$[P_0 \pm P_1, [P_0 \pm P_1, Q]] = 0 \,.$$

It follows then that

$$[P_0 \pm P_1, [P_0 \pm P_1, \{Q, Q^*\}]] = -2\{Q_\pm, Q_\pm^*\} \,,$$

where

$$Q_\pm \equiv [P_0 \pm P_1, Q] \,.$$

Now, $\{Q, Q^*\}$ is at most a linear combination of Js, Ps, and scalar

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

symmetry generators. The commutators of $P_0 \pm P_1$ with the Ps and scalar symmetry generators vanish, while the commutators of $P_0 \pm P_1$ with the Js are linear combinations of Ps, which commute with the other $P_0 \pm P_1$, so the double commutator $[P_0 \pm P_1, [P_0 \pm P_1, \{Q, Q^*\}]]$ must vanish and therefore $\{Q_\pm, Q_\pm^*\} = 0$, which implies that $Q_\pm = 0$. Since *all* members of the representation of the Lorentz group provided by the Qs thus commute with P_0 and P_1, Lorentz invariance implies that all Qs commute with all Ps, as was to be shown.

There is an important corollary that since the Lorentz generators $J_{\mu\nu}$ do not commute with the momentum operators, they cannot appear on the right-hand side of the anticommutation relations. For the moment let us label the Qs as Q_n, where n runs over the labels for the different (not necessarily inequivalent) irreducible spinor representations among the Qs, now *including* their adjoints Q^*, and also over the index labelling members of these representations. The general anticommutation relation is then of the form

$$\{Q_n\,,\, Q_m\} = \Gamma^\mu_{nm} P_\mu + Z_{nm}\,, \tag{32.1.5}$$

where the Γ^μ_{nm} are c-number coefficients and the Z_{nm} are conserved scalar symmetry generators, which commute with the P_μ and $J_{\mu\nu}$. We now want to show that the Z_{nm} are *central charges* of the supersymmetry algebra — that is, that they commute with the Q_ℓ and each other as well as with the P_μ and $J_{\mu\nu}$ and all other symmetry generators.

To prove this for $d \geq 4$, note that for a given Z_{nm} to be non-zero, since it is a scalar all of the σs in Eq. (32.1.1) must be opposite for Q_n and Q_m. Consider another fermionic symmetry generator Q_ℓ, for which the σs of Eq. (32.1.1) are not all the same as those of either Q_n or Q_m. (For $d \geq 4$ there is always such a Q_ℓ in each set of Qs forming an irreducible spinor representation of $O(d)$.) We apply the super-Jacobi identity

$$[Q_\ell\,,\, \{Q_m\,,\, Q_n\}] + [Q_m\,,\, \{Q_n\,,\, Q_\ell\}] + [Q_n\,,\, \{Q_\ell\,,\, Q_m\}] = 0\,. \tag{32.1.6}$$

The anticommutators $\{Q_n\,,\, Q_\ell\}$ and $\{Q_\ell\,,\, Q_m\}$ are operators that have some σs non-zero, so they can only be linear combinations of Ps rather than Zs, and so must commute with all Qs. This leaves just

$$0 = [Q_\ell\,,\, \{Q_m\,,\, Q_n\}] = [Q_\ell\,,\, Z_{mn}]\,. \tag{32.1.7}$$

Thus in each set of Qs forming an irreducible spinor representation of $O(d)$ there is at least one that commutes with the given Z_{mn}. But Z_{mn} is a Lorentz scalar, so it must then commute with all Qs. It follows then immediately from Eq. (32.1.5) that they also commute with each other.

The fermionic generators must form a representation (perhaps trivial) of the algebra $\mathscr{A}$ consisting of *all* scalar bosonic symmetry generators. It follows then by precisely the same argument used in Section 25.2 that

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

the central charges Z_{mn} furnish an invariant Abelian subalgebra of $\mathscr{A}$. The Coleman–Mandula theorem tells us that $\mathscr{A}$ must be a direct sum of a compact semi-simple Lie algebra, which by definition contains no invariant Abelian subalgebras, together with $U(1)$ generators, so the Z_{mn} must be $U(1)$ generators, which commute with all other bosonic symmetry generators, not just with each other.

To obtain more detailed information about the structure of the anti-commutation relations (32.1.5), we must be more specific about the Lorentz transformation and reality properties of the fermionic symmetry generators Q_n. These are very different for spacetimes of even and odd dimensionality.

Odd Dimensionality

The appendix to this chapter shows that for odd spacetime dimensions d there is just one fundamental spinor representation of the Lorentz algebra, by matrices $\mathscr{J}_{\mu\nu}$ given in terms of Dirac matrices by Eq. (32.A.2), so we must label the fermionic generators as $Q_{\alpha r}$, where α is a $2^{(d-1)/2}$-valued Dirac index, and $r = 1, 2, \ldots, N$ labels different spinors in the case of N-extended supersymmetry. With this notation, the Lorentz transformation properties of the Qs imply that

$$[J_{\mu\nu}\,,\, Q_{\alpha r}] = -\sum_{\beta}(\mathscr{J}_{\mu\nu})_{\alpha\beta}\, Q_{\beta r}\,, \tag{32.1.8}$$

so that the anticommutators of these generators have the transformation rule

$$[J_{\mu\nu}\,,\, \{Q_{\alpha r}\,,\, Q_{\beta s}\}] = -\sum_{\bar{\alpha}}(\mathscr{J}_{\mu\nu})_{\alpha\bar{\alpha}}\{Q_{\bar{\alpha} r}\,,\, Q_{\beta s}\} - \sum_{\bar{\beta}}(\mathscr{J}_{\mu\nu})_{\beta\bar{\beta}}\{Q_{\alpha r}\,,\, Q_{\bar{\beta} s}\}\,.$$

Recalling the Lorentz transformation rule (2.4.13) for the momentum operator P_λ, we see that the matrix Γ^λ_{rs} and the operator Z_{rs} in Eq. (32.1.5) (with Dirac indices now suppressed) must satisfy the conditions

$$\mathscr{J}_{\mu\nu}(\Gamma_\lambda)_{rs} + (\Gamma_\lambda)_{rs}\mathscr{J}^{\mathrm{T}}_{\mu\nu} = -i\,(\Gamma_\mu)_{rs}\eta_{\nu\lambda} + i\,(\Gamma_\nu)_{rs}\eta_{\mu\lambda}\,, \tag{32.1.9}$$

$$\mathscr{J}_{\mu\nu}Z_{rs} + Z_{rs}\mathscr{J}^{\mathrm{T}}_{\mu\nu} = 0\,. \tag{32.1.10}$$

But Eq. (32.A.38) gives $\mathscr{J}^{\mathrm{T}}_{\mu\nu} = -\mathscr{C}^{-1}\mathscr{J}_{\mu\nu}\mathscr{C}$, so Eqs. (32.1.9) and (32.1.10) may be expressed as the requirement that $(\Gamma_\mu)_{rs}\mathscr{C}^{-1}$ satisfies the same commutation relation (32.A.32) with $\mathscr{J}_{\mu\nu}$ as γ_μ, while $Z_{rs}\mathscr{C}^{-1}$ commutes with $\mathscr{J}_{\mu\nu}$. For odd d the matrices satisfying these conditions are unique up to multiplication with constants, so we can conclude that

$$\Gamma^\lambda_{\alpha r\,\beta s} = i\,g_{rs}\,(\gamma^\lambda\mathscr{C})_{\alpha\beta} \tag{32.1.11}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

and

$$Z_{\alpha r\,\beta s} = \mathscr{C}_{\alpha\beta}\, z_{rs}\,, \tag{32.1.12}$$

with the factor i inserted in Eq. (32.1.11) for later convenience. With Dirac indices suppressed, the anticommutation relations (32.1.5) now read

$$\{Q_r\,,\, Q_s^{\mathrm{T}}\} = i\, g_{rs}\gamma^\lambda \mathscr{C}\, P_\lambda + z_{rs}\mathscr{C}\,. \tag{32.1.13}$$

Both $\Gamma^\lambda_{\alpha r\,\beta s}$ and $Z_{\alpha r\,\beta s}$ are symmetric under interchange of α and r with β and s, while Eqs. (32.A.30) and (32.A.31) (with $d = 2n+1$) show that: $\gamma^\lambda\mathscr{C}$ is symmetric and $\mathscr{C}$ is symmetric for $d = 1 \pmod 8$; $\gamma^\lambda\mathscr{C}$ is symmetric and $\mathscr{C}$ is antisymmetric for $d = 3 \pmod 8$; $\gamma^\lambda\mathscr{C}$ is antisymmetric and $\mathscr{C}$ is antisymmetric for $d = 5 \pmod 8$; and $\gamma^\lambda\mathscr{C}$ is antisymmetric and $\mathscr{C}$ is symmetric for $d = 7 \pmod 8$. It follows that: g_{rs} is symmetric and z_{rs} is symmetric for $d = 1 \pmod 8$; g_{rs} is symmetric and z_{rs} is antisymmetric for $d = 3 \pmod 8$; g_{rs} is antisymmetric and z_{rs} is antisymmetric for $d = 1 \pmod 8$; and g_{rs} is antisymmetric and z_{rs} is symmetric for $d = 1 \pmod 8$.

The complex conjugate of the matrices $\mathscr{J}_{\mu\nu}$ is given by Eq. (32.A.37). By taking the Hermitian adjoint of Eq. (32.1.8), we see that $\sum_\beta (\mathscr{C}\beta)_{\alpha\beta} Q^*_{\beta r}$ has the same Lorentz transformation properties as any $Q_{\alpha s}$, and therefore must be a linear combination of them

$$\sum_\beta (\mathscr{C}\beta)_{\alpha\beta} Q^*_{\beta r} = \sum_s \mathscr{S}_{rs} Q_{\alpha s}\,. \tag{32.1.14}$$

Taking the Hermitian adjoint of this equation and using Eqs. (32.A.28) and (32.A.29) with $d = 2n+1$ yields

$$\mathscr{S}^*\mathscr{S} = (-1)^a \cdot 1\,, \qquad a = (d-1)(d-3)/8\,. \tag{32.1.15}$$

For $d = 1 \pmod 8$ and $d = 3 \pmod 8$ the spinor representation of the Lorentz algebra is real, and we can choose a basis for the fermionic generators such that $\mathscr{S} = 1$. In contrast, for $d = 5 \pmod 8$ and $d = 7 \pmod 8$ the spinor representation of the Lorentz algebra is pseudoreal, and it is evidently impossible to choose a basis with $\mathscr{S} \propto 1$. By taking the determinant of Eq. (32.1.15), we see that in this case $\mathrm{Det}\,(-1) > 0$, so for $d = 5 \pmod 8$ and $d = 7 \pmod 8$ there must be an even number N of fermionic generators. In this case we can choose a basis in which $\mathscr{S} = \Omega$, where Ω is the real antisymmetric block-diagonal matrix

$$\Omega = \begin{pmatrix} e & 0 & 0 & \cdots \\ 0 & e & 0 & \cdots \\ 0 & 0 & e & \cdots \\ & \cdots & & \end{pmatrix}\,, \qquad e = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\,. \tag{32.1.16}$$

We can deduce the reality and positivity properties of g_{rs} and z_{rs} by

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

using Eq. (32.1.14) to rewrite the anticommutation relation (32.1.13) as

$$\{Q_r\,,\,Q_s^\dagger\} = i(g\mathscr{S}^{\mathrm{T}})_{rs}\gamma_\lambda\mathscr{C}(\mathscr{C}\beta)^{\mathrm{T}-1}P^\lambda + (z\mathscr{S}^{\mathrm{T}})_{rs}\mathscr{C}(\mathscr{C}\beta)^{\mathrm{T}-1}\,.$$

Eqs. (32.A.12), (32.A.16), and (32.A.30) with $d = 2n+1$ show that $\beta^{\mathrm{T}} = -\beta$ and $\mathscr{C}\mathscr{C}^{\mathrm{T}-1} = (-1)^{(d-1)(d+1)/8}\cdot 1$, so

$$\{Q_r\,,\,Q_s^\dagger\} = -(-1)^{(d-1)(d+1)/8}\Big[i(g\mathscr{S}^{\mathrm{T}})_{rs}\gamma_\lambda\beta P^\lambda + (z\mathscr{S}^{\mathrm{T}})_{rs}\beta)\Big]\,. \qquad (32.1.17)$$

Recalling that $\gamma_0 = i\beta$, we note that the operator matrix $-i\gamma_\lambda\beta\,P^\lambda$ is positive, and positive-definite aside from the vacuum state. By considering a state of sufficiently large momentum so that the central charge term in Eq. (32.1.17) may be neglected, we conclude that the matrix $(-1)^{(d-1)(d+1)/8}g\mathscr{S}^{\mathrm{T}}$ is positive and Hermitian. Then considering arbitrary momenta, we also find that the array of operators $(z\mathscr{S}^{\mathrm{T}})_{rs}$ is Hermitian. (For non-zero central charges there is a lower bound on the mass, analogous to Eq. (25.5.22), which will not be given here.) From the Hermiticity of $g\mathscr{S}^{\mathrm{T}}$ we have

$$g^\dagger = \mathscr{S}^{\mathrm{T}\dagger-1}g\mathscr{S}^{\mathrm{T}} = (-1)^a\mathscr{S}g\mathscr{S}^{\mathrm{T}}. \qquad (32.1.18)$$

We are now in a position by an appropriate choice of basis to put the anticommutation relations in a convenient canonical form.

For $d = 1$ (mod 8) we have g and z symmetric and $(-1)^a = +1$, so if we choose a basis in which $\mathscr{S} = 1$ then g is real and the individual z_{rs} are Hermitian operators. We may introduce new Qs without changing $\mathscr{S} = 1$ by multiplying the old Qs by any real matrix $\mathscr{A}$, with the result that g is changed to $\mathscr{A}g\mathscr{A}^{\mathrm{T}}$. Since g is a positive matrix for $d = 1$ (mod 8), by a well-known theorem[6a] we may choose $\mathscr{A}$ to make $g = 1$.

For $d = 3$ (mod 8) we have g symmetric, z antisymmetric, and $(-1)^a = +1$, so if we choose a basis in which $\mathscr{S} = 1$ then g is real and the individual z_{rs} are anti-Hermitian operators. As in the case of $d = 1$ (mod 8), we may further adapt the basis so that $g = 1$.

For $d = 5$ (mod 8) we have g antisymmetric, so, with the choice $\mathscr{S} = \Omega$, Eq. (32.1.18) reads $g^* = -\Omega g\Omega$, where Ω is the standard antisymmetric matrix (32.1.16). Here we may introduce new Qs while keeping $\mathscr{S} = \Omega$ by multiplying the old ones by any matrix $\mathscr{B}$ with $\mathscr{B}^* = -\Omega\mathscr{B}\Omega$, with the effect that g is changed to $\mathscr{B}g\mathscr{B}^{\mathrm{T}}$. Since $(-1)^a = -1$, $g\Omega$ is positive, and so in this way we may arrange that $g = -\Omega$. Also, z is antisymmetric and $z\Omega$ is Hermitian, so $z^* = -\Omega\, z\,\Omega$.

For $d = 7$ (mod 8) we again have g antisymmetric, but now $(-1)^a = +1$, so by the same method as for $d = 5$ (mod 8), we may choose a basis in which $g = +\Omega$. Also, z is now symmetric and $z\Omega$ is again Hermitian, so now $z^* = +\Omega\, z\,\Omega$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Even Dimensionality

The appendix to this chapter shows that for even spacetime dimensions d there are two inequivalent fundamental spinor representations of the Lorentz algebra by matrices $\mathscr{J}^{\pm}_{\mu\nu}$ given in terms of Dirac matrices by Eqs. (32.A.22), (32.A.2), and (32.A.17). Therefore here we must label the fermionic generators as $Q^{\pm}_{\alpha r}$, where α is a $2^{d/2}$-valued Dirac index, r labels different Qs belonging to equivalent representations of the Lorentz algebra in the case of extended supersymmetry, and

$$\sum_{\beta}(\gamma_{d+1})_{\alpha\beta}\, Q^{\pm}_{\beta r} = \pm Q^{\pm}_{\alpha r}\,, \tag{32.1.19}$$

where $\gamma_{d+1} \equiv i^{d/2-1}\gamma_1 \cdots \gamma_{d-1}\gamma_0$. With this notation, the Lorentz transformation properties of the Qs imply that

$$[J_{\mu\nu}\,,\, Q^{\pm}_{\alpha r}] = -\sum_{\beta}(\mathscr{J}^{\pm}_{\mu\nu})_{\alpha\beta}\, Q^{\pm}_{\beta r}\,, \tag{32.1.20}$$

where $\mathscr{J}^{\pm}_{\mu\nu}$ are the matrices (32.A.22). Taking account of Eq. (32.1.19) and the relation $\mathscr{C}^{-1}\gamma_{d+1}\mathscr{C} = (-1)^{d/2}\gamma_{d+1}$, the same arguments that gave the anticommutation relations (32.1.13) for odd d now give

$$\{Q^{\pm}_{r}\,,\, Q^{\mp(-1)^{d/2}\,\mathrm{T}}_{s}\} = i\, g^{\pm}_{rs}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\gamma^{\lambda}\mathscr{C}\, P_{\lambda}\,, \tag{32.1.21}$$

$$\{Q^{\pm}_{r}\,,\, Q^{\pm(-1)^{d/2}\,\mathrm{T}}_{s}\} = z^{\pm}_{rs}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\mathscr{C}\,. \tag{32.1.22}$$

Eqs. (32.A.30) and (32.A.31) show that $\mathscr{C}\gamma^{\lambda}$ is symmetric for $d = 0 \pmod 8$ and $d = 2 \pmod 8$ and antisymmetric for $d = 4 \pmod 8$ and $d = 6 \pmod 8$, while $\mathscr{C}$ is symmetric for $d = 0 \pmod 8$ and $d = 6 \pmod 8$ and antisymmetric for $d = 2 \pmod 8$ and $d = 4 \pmod 8$. Hence Eq. (32.1.21) requires the symmetry properties

$$g^{\pm}_{rs} = \begin{cases} g^{\mp(-1)^{d/2}}_{sr} & d = 0, 2 \pmod 8 \\ -g^{\mp(-1)^{d/2}}_{sr} & d = 4, 6 \pmod 8 \end{cases}\,, \tag{32.1.23}$$

while Eq. (32.1.22) requires that

$$z^{\pm}_{rs} = \begin{cases} z^{\pm(-1)^{d/2}}_{sr} & d = 0, 6 \pmod 8 \\ -z^{\pm(-1)^{d/2}}_{sr} & d = 2, 4 \pmod 8 \end{cases}\,. \tag{32.1.24}$$

In particular, $z^{\pm}$ is symmetric for $d = 0 \pmod 8$, $g^{\pm}$ is symmetric for $d = 2 \pmod 8$, $z^{\pm}$ is antisymmetric for $d = 4 \pmod 8$, and $g^{\pm}$ is antisymmetric for $d = 6 \pmod 8$.

Taking the Hermitian adjoint of Eq. (32.1.20) and using Eq. (32.A.25) shows that $\mathscr{C}\beta Q^{\pm *}_{r}$ has the same Lorentz transformation properties as the

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

operators $Q_s^{\mp(-1)^{d/2}}$ and is therefore a linear combination of them:

$$\mathscr{C}\beta\, Q_r^{\pm *} = \sum_s \mathscr{S}_{rs}^{\pm}\, Q_s^{\mp(-1)^{d/2}} . \tag{32.1.25}$$

Taking the Hermitian adjoint of this equation and using Eqs. (32.A.28) and (32.A.29) with $d = 2n$ gives

$$\mathscr{S}^{\pm *}\mathscr{S}^{\mp(-)d/2} = (-1)^a \cdot 1 , \qquad a = d(d-2)/8 . \tag{32.1.26}$$

For $d = 0 \pmod 8$ and $d = 4 \pmod 8$ Eq. (32.1.25) relates one irreducible representation to another, and we may choose bases in which $\mathscr{S}^{\pm} = 1$ for $d = 0 \pmod 8$ and $\mathscr{S}^{\pm} = \pm 1$ for $d = 4 \pmod 8$. For $d = 2 \pmod 8$ Eq. (32.1.25) relates real representations to themselves, and we may choose bases with $\mathscr{S}^{\pm} = 1$. For $d = 6 \pmod 8$ Eq. (32.1.25) relates pseudoreal representations to themselves; the determinant of Eq. (32.1.26) shows that there must be an even number of Q^+ and an even number (not necessarily the same!) of Q^-, and we may choose bases with $\mathscr{S}^{\pm} = \Omega^{\pm}$, where $\Omega^{\pm}$ are standard real antisymmetric matrices of the form (32.1.16).

We can deduce the reality and positivity properties of $g_{rs}^{\pm}$ and the reality properties of $z_{rs}^{\pm}$ by using Eq. (32.1.25) to rewrite the anticommutation relations (32.1.21) and (32.1.22) in the form

$$\{Q_r^{\pm}, Q_s^{\pm\dagger}\} = \left(\frac{1 \pm \gamma_{d+1}}{2}\right) \gamma_\lambda \beta\, P^\lambda \mathscr{C}(\mathscr{C}\beta)^{\mathrm{T}-1}\left(g^{\pm}\mathscr{S}^{\mp\mathrm{T}}\right)_{rs} , \tag{32.1.27}$$

$$\{Q_r^{\pm}, Q_s^{\mp\dagger}\} = \left(\frac{1 \pm \gamma_{d+1}}{2}\right) \mathscr{C}(\mathscr{C}\beta)^{\mathrm{T}-1}\left(z^{\pm}\mathscr{S}^{\mp\mathrm{T}}\right)_{rs} . \tag{32.1.28}$$

We again use the relations $\mathscr{C}\mathscr{C}^{\mathrm{T}-1} = (-1)^{d(d+2)/8}$, $\beta^{\mathrm{T}} = -\beta$, and $\gamma_0 = i\beta$, and conclude that $(-1)^{d(d+2)/8} g^{\pm}\mathscr{S}^{\pm\mathrm{T}}$ is Hermitian and positive, while $(z^+\mathscr{S}^{-\mathrm{T}})^\dagger = z^-\mathscr{S}^{+\mathrm{T}}$. For $d = 0 \pmod 8$ we can adopt a basis with $\mathscr{S}^{\pm} = 1$, $g^{\pm} = 1$, and $z^{+\dagger} = z^-$; for $d = 2 \pmod 8$ we can adopt a basis with $\mathscr{S}^{\pm} = 1$, $g^{\pm} = -1$, and $z^{+\dagger} = z^-$; for $d = 4 \pmod 8$ we can adopt a basis with $\mathscr{S}^{\pm} = \pm 1$, $g^{\pm} = \mp 1$, and $z^{+\dagger} = -z^-$; and for $d = 6 \pmod 8$ we can adopt a basis with $\mathscr{S}^{\pm} = \Omega^{\pm}$, $g^{\pm} = \Omega^{\pm}$, and $(z^+\Omega^-)^\dagger = z^-\Omega^+$.

To summarize, in appropriate bases the anticommutation relations and reality and symmetry conditions are as follows:[6b]

$d = 0 \pmod 8$

$$\{Q_r^{\pm}, Q_s^{\mp\,\mathrm{T}}\} = i\,\delta_{rs}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\gamma^\lambda \mathscr{C}\, P_\lambda , \tag{32.1.29}$$

$$\{Q_r^{\pm}, Q_s^{\pm\,\mathrm{T}}\} = z_{rs}^{\pm}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\mathscr{C} , \tag{32.1.30}$$

$$\mathscr{C}\beta\, Q_r^{\pm *} = Q_r^{\mp} , \qquad z_{rs}^{\pm} = z_{sr}^{\pm} = (z_{rs}^{\mp})^* . \tag{32.1.31}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

$d = 1 \pmod 8$

$$\{Q_r\,,\, Q_s^{\mathrm{T}}\} = i\,\delta_{rs}\gamma^{\lambda}\mathscr{C}\,P_{\lambda} + z_{rs}\mathscr{C}\,, \tag{32.1.32}$$

$$\mathscr{C}\beta Q_r^{*} = Q_r\,, \qquad z_{rs} = z_{sr} = z_{rs}^{*}\,. \tag{32.1.33}$$

$d = 2 \pmod 8$

$$\{Q_r^{\pm}\,,\, Q_s^{\pm\,\mathrm{T}}\} = -i\,\delta_{rs}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\gamma^{\lambda}\mathscr{C}\,P_{\lambda}\,, \tag{32.1.34}$$

$$\{Q_r^{\pm}\,,\, Q_s^{\mp\,\mathrm{T}}\} = z_{rs}^{\pm}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\mathscr{C}\,, \tag{32.1.35}$$

$$\mathscr{C}\beta\, Q_r^{\pm *} = Q_r^{\pm}\,, \qquad z_{rs}^{\pm} = -z_{rs}^{\pm *} = -z_{sr}^{\mp}\,. \tag{32.1.36}$$

$d = 3 \pmod 8$

$$\{Q_r\,,\, Q_s^{\mathrm{T}}\} = i\,\delta_{rs}\gamma^{\lambda}\mathscr{C}\,P_{\lambda} + z_{rs}\mathscr{C}\,, \tag{32.1.37}$$

$$\mathscr{C}\beta Q_r^{*} = Q_r\,, \qquad z_{rs} = -z_{sr} = -z_{rs}^{*}\,. \tag{32.1.38}$$

$d = 4 \pmod 8$

$$\{Q_r^{\pm}\,,\, Q_s^{\mp\,\mathrm{T}}\} = \mp i\,\delta_{rs}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\gamma^{\lambda}\mathscr{C}\,P_{\lambda}\,, \tag{32.1.39}$$

$$\{Q_r^{\pm}\,,\, Q_s^{\pm\,\mathrm{T}}\} = z_{rs}^{\pm}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\mathscr{C}\,, \tag{32.1.40}$$

$$\mathscr{C}\beta\, Q_r^{\pm *} = \pm Q_r^{\mp}\,, \qquad z_{rs}^{\pm} = z_{rs}^{\mp *} = -z_{sr}^{\pm}\,. \tag{32.1.41}$$

$d = 5 \pmod 8$

$$\{Q_r\,,\, Q_s^{\mathrm{T}}\} = -i\,\Omega_{rs}\gamma^{\lambda}\mathscr{C}\,P_{\lambda} + z_{rs}\mathscr{C}\,, \tag{32.1.42}$$

$$\mathscr{C}\beta Q_r^{*} = \sum_s \Omega_{rs} Q_s\,, \qquad z_{rs} = -z_{sr}\,, \qquad z^{*} = -\Omega z \Omega\,. \tag{32.1.43}$$

$d = 6 \pmod 8$

$$\{Q_r^{\pm}\,,\, Q_s^{\pm\,\mathrm{T}}\} = i\,\Omega_{rs}^{\pm}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\gamma^{\lambda}\mathscr{C}\,P_{\lambda}\,, \tag{32.1.44}$$

$$\{Q_r^{\pm}\,,\, Q_s^{\mp\,\mathrm{T}}\} = z_{rs}^{\pm}\left(\frac{1 \pm \gamma_{d+1}}{2}\right)\mathscr{C}\,, \tag{32.1.45}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

$$\mathscr{C}\beta\, Q_r^{\pm *} = \sum_s \Omega_{rs}^{\pm} Q_s^{\pm}\,, \qquad z_{rs}^{\pm *} = \Omega^{\pm} z_{rs}^{\pm} \Omega^{\mp} = z_{sr}^{\mp}\,. \tag{32.1.46}$$

$d = 7 \pmod 8$

$$\{Q_r\,,\, Q_s^{\mathrm{T}}\} = i\,\Omega_{rs}\gamma^{\lambda}\mathscr{C}\,P_{\lambda} + z_{rs}\mathscr{C}\,, \tag{32.1.47}$$

$$\mathscr{C}\beta Q_r^{*} = \sum_s \Omega_{rs} Q_s\,, \qquad z_{rs} = z_{sr}\,, \qquad z^{*} = +\Omega z \Omega\,. \tag{32.1.48}$$

Inspection of these anticommutation relations reveals that in the absence of central charges they are invariant under groups of linear transformations on the fermionic generators, of the form $Q_r \to \sum_s V_{rs} Q_s$ for d odd and $Q_r^{\pm} \to \sum_s V_{rs}^{\pm} Q_s^{\pm}$ for d even. In order to preserve the relations (32.1.29)–(32.1.48) it is necessary that the Vs should satisfy the conditions:

$d = 0$ and $d = 4 \pmod 8$

$$V^{\pm} V^{\mp \mathrm{T}} = 1\,, \qquad V^{\pm *} = V^{\mp}\,. \tag{32.1.49}$$

$d = 1$ and $d = 3 \pmod 8$

$$V V^{\mathrm{T}} = 1\,, \qquad V^{*} = V\,. \tag{32.1.50}$$

$d = 2 \pmod 8$

$$V^{\pm} V^{\pm \mathrm{T}} = 1\,, \qquad V^{\pm} = V^{\pm *}\,. \tag{32.1.51}$$

$d = 5 \pmod 8$

$$V \Omega V^{\mathrm{T}} = \Omega\,, \qquad V^{*} = -\Omega V \Omega\,. \tag{32.1.52}$$

$d = 6 \pmod 8$

$$V^{\pm} V^{\pm \mathrm{T}} = 1\,, \qquad V^{\pm *} = -\Omega^{\pm} V^{\pm} \Omega^{\pm}\,. \tag{32.1.53}$$

$d = 7 \pmod 8$

$$V \Omega V^{\mathrm{T}} = \Omega\,, \qquad V^{*} = -\Omega V \Omega\,. \tag{32.1.54}$$

These matrices form the groups:

$d = 0$ and $d = 4 \pmod 8$ $\qquad U(N)$.

$d = 1$ and $d = 3 \pmod 8$ $\qquad O(N)$.

$d = 2 \pmod 8$ $\qquad O(N_+) \times O(N_-)$.

$d = 5$ and $d = 7 \pmod 8$ $\qquad USp(N) \quad N$ even .

$d = 6 \pmod 8$ $\qquad USp(N_+) \times USp(N_-) \quad N_{\pm}$ even .

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Here N is the number of fundamental spinor representations among the Qs for d odd and the number of fundamental spinor representations of each chirality among the Qs for $d = 0 \pmod 8$ and $d = 4 \pmod 8$. For $d = 2 \pmod 8$ and $d = 6 \pmod 8$ the numbers of fundamental spinor representations among the Qs need not be the same for each chirality, and so these are denoted N_+ and N_-.

32.2 Massless Multiplets

We will now consider how the supersymmetry algebras constructed in the previous section may be used to construct supermultiplets of massless particle states in $d \geq 4$ spacetime dimensions. The momentum operator P^μ commutes with all fermionic symmetry generators, so we can work in the one-particle subspace of Hilbert space where P^μ has a definite lightlike eigenvector p^μ, which may be taken in a direction with $p^1 = p^0$ and all other spatial components of p^μ zero. Just as in the case of four spacetime dimensions, discussed in Section 2.5, these one-particle states are classified according to the finite-dimensional representations they provide of the little group, the subgroup of the homogeneous Lorentz group that leaves p^μ invariant. The little group contains combined boosts along directions perpendicular to $\mathbf{p}$ and rotations in planes in which $\mathbf{p}$ lies, such as the transformations (2.5.6) for four spacetime dimensions, but these form an invariant Abelian subgroup and therefore in a finite-dimensional representation must be represented by the unit operator. With this subgroup omitted, the reduced little group in d dimensions is $O(d-2)$, consisting of rotations in planes orthogonal to $\mathbf{p}$. We therefore classify massless particle states according to the representations they provide of $O(d-2)$ and of the automorphism groups described at the end of the previous section.

These representations are more complicated than those we have dealt with in four spacetime dimensions, where the reduced little group is $O(2)$ and the representations are one-dimensional, characterized by a single number, the helicity. Nevertheless it is useful to label the representations of the reduced little group $O(d-2)$ with a 'spin,' defined as the *maximum* absolute value of the eigenvalue of any generator J_{ij} in the representation.

It is widely believed that there are no consistent quantum field theories involving massless particles with spin greater than 2. It is known[7] that *soft* massless particles with spin $j > 1/2$ can only interact with conserved currents carrying spin j. For $j = 1$ these are the currents of ordinary conserved scalars, like electric charge; for $j = 3/2$ they are the one or

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

more supercurrents associated with supersymmetry; for $j = 2$ there is a single current, the energy-momentum tensor; but for $j > 5/2$ there is no conserved current with which a soft massless particle could interact. We may derive stringent limits on the dimensions in which supersymmetry is possible by adopting a prohibition against there being more than one type of massless particle of spin 2, or any massless particles whatever of spin greater than 2.

Let us return for a moment to the classification of operators used in Section 32.1, according to a weight equal to the value of the $O(d)$ generator J_{d1} that they destroy. (Recall that $J_{01} = iJ_{d1}$.) The fermionic supersymmetry generators have weights $1/2$ and $-1/2$, so the anticommutator of any of these generators with its Hermitian adjoint can only have weight $+1$ or -1, respectively, and therefore must be respectively proportional to the operator $P^0 + P^1$ or $P^0 - P^1$. But we are working in a subspace of Hilbert space in which the operator $P^0 - P^1$ vanishes, so in this subspace the fermionic supersymmetry generators of weight $-1/2$ all vanish. To classify the one-particle states we therefore have available just half of the supersymmetry generators, the 2^{n-1} generators with weight $\sigma_{d1} = +1/2$.

We can further divide the remaining supersymmetry generators into two classes, those in which $\sigma_{23} = 1/2$ or $\sigma_{23} = -1/2$ as well as $\sigma_{d1} = +1/2$. Since the operator $P^0 + P^1$ has $\sigma_{23} = 0$, the fermionic supersymmetry generators of each class anticommute with each other, though not necessarily with their adjoints or with generators of the other class.

Now consider a representation of the little group $O(d-2)$ with spin j, and consider any state $|\lambda\rangle$ that is an eigenstate of J_{23} with eigenvalue $\lambda > 0$ and is annihilated by all supersymmetry generators with $\sigma_{23} = -1/2$. (Any state that has the maximum eigenvalue j for J_{23} is of this type, but in general there may be other such states.) We may form states with $J_{23} = \lambda - k/2$ by acting on $|\lambda\rangle$ with k fermionic generators having $\sigma_{23} = +1/2$ as well as $\sigma_{d1} = +1/2$. (It can be shown that none of these states vanishes, because acting on them with the adjoints of k of these fermionic generators gives back the state $|\lambda\rangle$.) If there is a total of $\mathcal{N}$ fermionic supersymmetry generators of all types, then there are $\mathcal{N}/4$ of them with $\sigma_{23} = +1/2$ and $\sigma_{d1} = +1/2$, and since these operators all anticommute the number of states formed in this way with $J_{23} = \lambda - k/2$ will be given by the binomial coefficient

$$\binom{\mathcal{N}/4}{k}, \tag{32.2.1}$$

which when summed from $k = 0$ to the maximum value $k = \mathcal{N}/4$ gives a total of $2^{\mathcal{N}/4}$ components. The minimum eigenvalue of J_{23} obtained in this way is $\lambda - \mathcal{N}/8$, reached by multiplying the state $|\lambda\rangle$ by $k = \mathcal{N}/4$ supersymmetry generators. Taking $\lambda = j$, we see that to avoid having

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

eigenvalues of J_{23} greater than $+2$ or less than -2, we must have $j \leq 2$ and $j - \mathcal{N}/8 \geq -2$, which requires a total number of fermionic generators $\mathcal{N}$ no greater than 32.

Further, for $\mathcal{N} = 32$ supersymmetry generators there can be at most a single supermultiplet of massless particles formed in this way by acting on the state $|2\rangle$ with products of supersymmetry generators having $\sigma_{23} = +1/2$ and $\sigma_{d1} = +1/2$. These states have eigenvalues for any generator of the little group running in steps of $1/2$ from -2 to $+2$. It is only for $\mathcal{N} < 32$ that there can be 'matter' supermultiplets, supermultiplets that do not contain the graviton.

A single fundamental spinor representation in $2n$ or $2n+1$ dimensions has 2^n components, so in order to have no more than 32 fermionic generators we must have $n \leq 5$. The spacetime dimensionality can thus be no larger than $d = 11$ and in this case must have $N = 1$. Supergravity in 11 dimensions is of special interest, because it may be the 'low-energy' limit of a fundamental theory known as *M theory*,[4] which is also believed to yield various string theories in other limits. We will now work out the spin content of $N = 1$ supersymmetry in $d = 11$ dimensions in detail, as an example of how this can be done by enlightened counting.

We can construct all the states of the massless multiplet for $d = 11$ by acting on an eigenstate $|2\rangle$ of J_{23} having eigenvalue 2 with products of $k = 0, 1, \ldots, 8$ supersymmetry generators having $\sigma_{23} = +1/2$ and $\sigma_{2n-1\,2n} = +1/2$. According to Eq. (32.2.1), we obtain one state each with $J_{23} = \pm 2$, eight states each with $J_{23} = \pm 3/2$, twenty-eight states each with $J_{23} = \pm 1$, fifty-six states each with $J_{23} = \pm 1/2$, and seventy states with $J_{23} = 0$.

For $d = 11$ the spin 2 graviton representation of the little group $O(9)$ is a symmetric traceless tensor with $9 \times 10/2 - 1 = 44$ independent components: there is one $2 \pm i3$, $2 \pm i3$ component with $J_{23} = \pm 2$; seven $2 \pm i3$, k components with $J_{23} = \pm 1$; and twenty-eight k, ℓ components with $J_{23} = 0$. (Here k and ℓ run over the seven values 4, 5, ..., 10. We do not count the $2 + i3$, $2 - i3$ component because it is related to the k, ℓ components by the tracelessness condition in this representation.)

There is also a single spin 3/2 gravitino representation. This consists of a spinor ψ_i with an extra nine-vector index i, subject to an irreducibility condition $\sum_i \gamma_i \psi_i = 0$ which excludes spin $1/2$ components, and so has $9 \times 16 - 16 = 128$ independent components.

By subtracting the number of components with each value of J_{23} contained in the graviton and gravitino states from those formed acting on $|2\rangle$ with supersymmetry generators, we see that we need one or more additional states having a total of $28 - 7 = 21$ components with $J_{23} = \pm 1$ and $70 - 28 = 42$ components with $J_{23} = 0$. The only representations of orthogonal groups that have no eigenvalues of J_{ij}s other than ± 1 and 0

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

are the antisymmetric tensors. An antisymmetric tensor $T_{i_1\cdots i_p}$ of rank p in nine dimensions has

$$\begin{pmatrix} 7 \\ p \end{pmatrix} \text{ components } T_{k_1\cdots k_p} \text{ with } J_{23} = 0\,,$$

$$\begin{pmatrix} 7 \\ p-1 \end{pmatrix} \text{ components } T_{2\pm i3\,k_2\cdots k_p} \text{ with } J_{23} = \pm 1\,,$$

$$\begin{pmatrix} 7 \\ p-2 \end{pmatrix} \text{ components } T_{2+i3\,2-i3\,k_3\cdots k_p} \text{ with } J_{23} = 0\,,$$

where $k_1,\ldots,k_p$ run over the seven values $4, 5, \ldots, 10$. For $O(9)$ the only independent antisymmetric tensors are of rank $p = 0$, 1, 2, 3, and 4. The antisymmetric tensor of rank 4 has 35 components with $J_{23} = \pm 1$, which is more than needed, so it must be excluded. Any combination of $p = 0$, $p = 1$, and $p = 2$ tensors with the twenty-one needed components with $J_{23} = \pm 1$ would have too many components with $J_{23} = 0$ (147 for 21 one-forms and no two-forms; 120 for 14 one-forms and 1 two-forms; 53 for 7 one-forms and 2 two-forms; and 66 for no one-forms and 3 two-forms), so we must include at least one three-form. The antisymmetric tensor with rank $p = 3$ has just 21 components with $J_{23} = \pm 1$ and 42 components with $J_{23} = 0$, which is just what is needed. We conclude that *the unique massless particle multiplet for $N = 1$ supersymmetry in $d = 11$ contains a graviton, a gravitino, and a particle whose states transform under the little group as a single antisymmetric tensor of rank 3.*

There is a richer variety of possibilities for $d = 10$. Here there are two ways to have $\mathcal{N} = 32$ generators: the fermionic generators can comprise two 16-component Weyl spinors of the same chirality, with an automorphism group $O(2)$, or two of opposite chirality, with no automorphism group. For $d = 10$ it is also possible to have a single Weyl fermionic generator, with just $\mathcal{N} = 16$ independent components. These three possibilities play an important role in modern superstring theories — they represent the massless particle spectrum of three kinds of superstring: type IIA for 16 generators of each chirality; type IIB for 32 generators of the same chirality; and the heterotic superstring for 16 generators of just one chirality.

The type IIA case of $d = 10$ and opposite chirality is just like the case of $d = 11$, except that the irreducible representations of the little group $O(9)$ for $d = 11$ break up into separate irreducible representations of the little group $O(8)$ for $d = 10$. Thus the $O(9)$ graviton multiplet is decomposed into an $O(8)$ graviton with $8 \times 9/2 - 1 = 35$ components, an $O(8)$ vector with 8 components, and a scalar with 1 component; the $O(9)$ gravitino multiplet is decomposed into $O(8)$ gravitinos of each chirality with $(16 \times 8 - 16)/2 = 56$ components each and $O(8)$ spinors of each

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

chirality with 8 components each; and the $O(9)$ three-form is decomposed into a $O(8)$ three-form with 56 components and an $O(8)$ two-form with 28 components.

In the type IIB case of $d = 10$ with $N_+ = 2$ and $N_- = 0$ we must classify states according to the representation of the little group $O(8)$ and a quantum number q that labels the representations of the automorphism group $O(2)$, under which the supersymmetry generators transform as a 2-vector. Since there is only one graviton, it must have $q = 0$. Acting on these states with a supersymmetry generator gives *two* gravitinos with $q = \pm 1$ and 56 components each; acting with another supersymmetry generator gives 2 two-form tensors with $q = \pm 2$ and 28 components each; acting with another supersymmetry generator gives 2 Weyl spinors with $q = \pm 3$ and 8 components each; and acting with another supersymmetry generator gives 2 scalars with $q = \pm 4$ together with a self-dual four-form with $q = 0$ and 35 components.

In the heterotic case of $d = 10$ with a single Weyl fermionic generator there are just $\mathcal{N} = 16$ independent components. In this case there is a graviton supermultiplet consisting of a graviton transforming under $O(8)$ as a symmetric traceless tensor with 35 independent components; a gravitino with 56 independent components; an $O(8)$ two-form with 28 independent components; a Weyl spinor with 8 components; and a scalar. (This graviton supermultiplet is constructed by acting with supersymmetry generators on one state $|2\rangle$, *six* states $|1\rangle$, and one state $|0\rangle$, giving altogether $8 \times 2^4 = 128 = 35+56+28+8+1$ components.) Here we also have the possibility of gauge supermultiplets that contain no particles having values greater than 1 or less than -1 for the eigenvalue of any J_{ij}. These gauge supermultiplets are formed by acting with supersymmetry generators on a state $|1\rangle$, and contain one gauge particle belonging to the vector representation of $O(8)$, with 8 components, and one particle transforming as a fundamental Weyl spinor of $O(8)$, also with 8 components.

32.3 *p*-Branes

In some theories in addition to particles there are stable extended objects, either of infinite extent or stabilized by 'wrapping' around a topologically non-trivial spacetime. The study of supersymmetry and supergravity in higher-dimensional theories of this sort has opened up remarkable opportunities for the construction of string theories and supersymmetric field theories in lower-dimensional spacetime and for the proof of equivalencies among these theories,[4,8] which are beyond the scope of this book. The feature of these extended objects that concerns us here is that they can carry conserved bosonic quantities other than those allowed by the Coleman–

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Mandula theorem. These new conserved quantities may appear on the right-hand side of the anticommutation relations of supersymmetry, along with the momentum operator and ordinary conserved quantities.[9]

In the cases that have been studied so far, the new conserved bosonic quantities are all *forms* – antisymmetric tensors. For instance, an object of spatial dimensionality p (known as a 'p-brane') in a spacetime of dimensionality d is described by specifying the d spacetime coordinates $x^\mu(\sigma,t)$ (generally in overlapping patches covering the object) as functions of the time t and of a set of p coordinates σ^r that parameterize positions on this object. If the manifold $x^\mu = x^\mu(\sigma,t)$ at a given time is topologically non-trivial, in the sense that it cannot be continuously deformed to a single point, then it may have a non-vanishing value for the topologically invariant integral*

$$I^{\mu_1\mu_2\cdots\mu_p} = \int d\sigma^1\, d\sigma^2 \cdots d\sigma^p \sum_{r_1=1}^{p}\sum_{r_2=1}^{p}\cdots\sum_{r_p=1}^{p} \epsilon^{r_1 r_2 \cdots r_p} \times \frac{\partial x^{\mu_1}(\sigma,t)}{\partial\sigma^{r_1}}\frac{\partial x^{\mu_2}(\sigma,t)}{\partial\sigma^{r_2}}\cdots\frac{\partial x^{\mu_p}(\sigma,t)}{\partial\sigma^{r_p}} \,. \tag{32.3.1}$$

The invariance of such integrals under small changes in the functions $x^\mu(\sigma,t)$ shows in particular that they are invariant under spacetime translations, and hence may appear along with P^μ and central charges on the right-hand side of the anticommutation relations for the supersymmetry generators.[10] The calculation of the coefficients of such tensors on the right-hand side of the anticommutation relations is analogous to the Olive–Witten calculation of the scalar central charge Z_{rs} in $N = 2$ supersymmetry theories in four spacetime dimensions, discussed in Section 27.9. We will make no attempt in this section to evaluate these coefficients or to survey the other non-topological p-forms that may appear in the anticommutation relations,[11] but will simply consider the effects on the supersymmetry algebra of including conserved antisymmetric tensors that commute with the momentum operators.

* To see that this integral is topologically invariant, note that the effect of an infinitesimal change $\delta x^\mu(\sigma,t)$ in the function $x^\mu(\sigma,t)$ is to change $I^{\mu_1\mu_2\cdots\mu_p}$ by the amount

$$\delta I^{\mu_1\mu_2\cdots\mu_p} = \sum_{n=1}^{p}\sum_{r_1=1}^{p}\sum_{r_2=1}^{p}\cdots\sum_{r_p=1}^{p}\int d\sigma^1\, d\sigma^2\cdots d\sigma^p \frac{\partial}{\partial\sigma^{r_n}}\left[\epsilon^{r_1 r_2\cdots r_p} \times \frac{\partial x^{\mu_1}}{\partial\sigma^{r_1}}\frac{\partial x^{\mu_2}}{\partial\sigma^{r_2}}\cdots\frac{\partial x^{\mu_{n-1}}}{\partial\sigma^{r_{n-1}}}\delta x^{\mu_n}\frac{\partial x^{\mu_{n+1}}}{\partial\sigma^{r_{n+1}}}\cdots\frac{\partial x^{\mu_p}}{\partial\sigma^{r_p}}\right],$$

which vanishes when the integral is taken over a compact manifold. It also vanishes if the integral is over all σ, provided $\delta x^\mu(\sigma,t)$ is constrained to vanish rapidly when $\sigma^r \to \infty$.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

It is important that this possibility does not affect the key result that the supersymmetry generators always belong to the fundamental spinor representations of the Lorentz group. This is because a totally antisymmetric tensor in Euclidean coordinates can have at most one spacetime index equal to 1 and at most one spacetime index equal to d, and therefore its 'weight' defined by Eq. (32.1.2) can only be ± 1 or 0. Just as before, this means that the weight of a supersymmetry generator can only be $\pm 1/2$; Lorentz invariance then implies that all the σs defined by Eq. (32.1.1) are $\pm 1/2$, which is only possible if the supersymmetry generators belong to a fundamental spinor representation of $O(d-1,1)$. Also, because the new terms in the anticommutators of supersymmetry generators commute with momentum, the same argument that was given in Section 32.1 shows again that the supersymmetry generators also commute with momentum.

Lorentz invariance tells us that for non-zero values of the p-form 'charges,' the anticommutation relations (32.1.13) and (32.1.21)–(32.1.22) can only take the forms (in the same notation as in Section 32.1):

d odd

$$\{Q_r\,,\,Q_s^{\mathrm{T}}\} = g_{rs}\gamma^\lambda \mathscr{C} P_\lambda + \sum_p z_{rs}^{\mu_1\mu_2\cdots\mu_p}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p}\mathscr{C}\,. \tag{32.3.2}$$

d even

$$\{Q_r^\pm\,,\,Q_s^{\mp(-1)^{d/2}\,\mathrm{T}}\} = \left(\frac{1\pm\gamma_{d+1}}{2}\right) \times \left[g_{rs}^\pm\gamma^\lambda\mathscr{C}P_\lambda + \sum_{\text{odd } p} z_{rs}^{\mu_1\mu_2\cdots\mu_p\,\pm}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p}\mathscr{C}\right], \tag{32.3.3}$$

$$\{Q_r^\pm\,,\,Q_s^{\pm(-1)^{d/2}\,\mathrm{T}}\} = \left(\frac{1\pm\gamma_{d+1}}{2}\right) \times \sum_{\text{even } p} z_{rs}^{\mu_1\mu_2\cdots\mu_p\,\pm}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p}\mathscr{C}\,. \tag{32.3.4}$$

(Recall that $\mathscr{C}$ appears in the anticommutation relations because $\mathscr{J}_{\mu\nu}^{\mathrm{T}} = -\mathscr{C}^{-1}\mathscr{J}_{\mu\nu}\mathscr{C}$; the $Q_r^\pm$ are supersymmetry generators in the case of even d for which $\gamma_{d+1}Q_r^\pm = \pm Q_r^\pm$; and $\mathscr{C}^{-1}\gamma_{d+1}\mathscr{C} = (-1)^{d/2}\gamma_{d+1}$.) For even d we have

$$\epsilon^{\mu_1\mu_2\cdots\mu_d}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p} \propto \gamma_{d+1}\gamma_{\mu_{p+1}}\gamma_{\mu_{p+2}}\cdots\gamma_{\mu_d}\,,$$

while for d odd

$$\epsilon^{\mu_1\mu_2\cdots\mu_d}\gamma_{\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p} \propto \gamma_{\mu_{p+1}}\gamma_{\mu_{p+2}}\cdots\gamma_{\mu_d}\,.$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Thus p-branes and $d-p$ branes make the same contributions in Eqs. (32.3.1)–(32.3.3) for any d, so that we can restrict p to run only over values from 0 to $d/2$ for d even and from 0 to $(d-1)/2$ for d odd.

The symmetry of anticommutators is reflected in symmetry conditions on the p-brane central charges z^p_{rs} in Eqs. (32.3.2)–(32.3.4). Eqs. (32.A.15) and (32.A.30) give

$$\gamma_\mu^{\mathrm{T}} = (-1)^n \mathscr{C}^{-1} \gamma_\mu \mathscr{C} \,, \qquad \mathscr{C}^{\mathrm{T}} = (-1)^{n(n+1)/2} \mathscr{C} \,, \tag{32.3.5}$$

for both $d = 2n$ and $d = 2n+1$. These give the antisymmetrized products $\gamma_{[\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p]}$ the symmetry property

$$\begin{aligned}\gamma_{[\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p]}\mathscr{C} &= (-1)^{pn}(-1)^{n(n+1)/2}\left[\gamma_{[\mu_p}\gamma_{\mu_{p-1}}\cdots\gamma_{\mu_1]}\mathscr{C}\right]^{\mathrm{T}} \\ &= (-1)^{pn}(-1)^{n(n+1)/2}(-1)^{p(p-1)/2}\left[\gamma_{[\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p]}\mathscr{C}\right]^{\mathrm{T}} .\end{aligned} \tag{32.3.6}$$

It follows immediately that for odd d,

$$z^{\mu_1\mu_2\cdots\mu_p}_{rs} = (-1)^{pn}(-1)^{n(n+1)/2}(-1)^{p(p-1)/2} z^{\mu_1\mu_2\cdots\mu_p}_{sr} \,, \tag{32.3.7}$$

while for even d,

$$z^{\mu_1\mu_2\cdots\mu_p\,\pm}_{rs} = (-1)^{pn}(-1)^{n(n+1)/2}(-1)^{p(p-1)/2} z^{\mu_1\mu_2\cdots\mu_p}_{sr}\,(-1)^n(-1)^p{}^{\mp} \,. \tag{32.3.8}$$

Consider for instance the important case of $N = 1$ supersymmetry in $d = 11$ spacetime dimensions, which is one version of the popular M-theory generalization of string theories. Eq. (32.3.8) shows that the single p-form central charge $z^{\mu_1\mu_2\cdots\mu_p}$ vanishes unless

$$-(-1)^p(-1)^{p(p-1)/2} = +1 \,, \tag{32.3.9}$$

which is satisfied only for p equal to 1, 2, and 5. The value $p = 1$ is realized by the momentum operator itself, which arises from particles as well as extended objects. The other possibilities, $p = 2$ and $p = 5$, arise in theories with 2-branes and 5-branes, respectively. Note that there can be no other independent tensor central charges, such as a 1-form arising from 1-branes, because the number of independent components in P^μ and in a 2-form and a 5-form is

$$11 + \binom{11}{2} + \binom{11}{5} = 528 \,,$$

while the number of independent components in an anticommutator of two 32-component fundamental spinors is $32 \times 33/2 = 528$.

Just as the 0-form electric charge is the source of a 1-form gauge field $A_\mu(x)$, so also a p-form conserved quantity $z^{\mu_1\mu_2\cdots\mu_p}$ may serve as the source of a $p+1$-form gauge field $A_{\mu_1\mu_2\cdots\mu_{p+1}}$ of the sort discussed in Section 8.8. In fact, such gauge fields do appear in supergravity theories. For instance,

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

as remarked in the previous section, the $N = 1$ supergravity theory in $d = 11$ spacetime dimensions includes a massless particle whose states are in the 3-form representation of the $O(9)$ little group, and therefore must be described by a 3-form gauge field $A_{\mu\nu\rho}(x)$. The study of solutions of this supergravity theory shows that there are two-branes[12] that indeed do provide sources for $A_{\mu\nu\rho}(x)$. Also, as noted in Section 8.8, this gauge theory is equivalent to one with a $(d - p - 2 = 6)$-form gauge field, which can have the 5-form $z^{\mu_1 \cdots \mu_5}$ as a source, and there are indeed 5-brane solutions which provide such sources for this 6-form gauge field.[13] The $N = 1$ eleven-dimensional supersymmetry algebra does in fact receive contributions from these 2-branes and 5-branes.[11]

Appendix Spinors in Higher Dimensions

This appendix describes the fundamental spinor representations of the Lie algebra of the Lorentz group $O(d-1,1)$ in any number d of spacetime dimensions. These are obtained from the corresponding Clifford algebra, consisting of an irreducible set of finite matrices γ_μ that satisfy the anticommutation relations

$$\{\gamma_\mu\,,\,\gamma_\nu\} = 2\eta_{\mu\nu}\,, \tag{32.A.1}$$

where $\eta_{\mu\nu}$ is diagonal, with elements $+1$ on the diagonal except for $\eta_{00} = -1$, where x^0 is the time-component. From these we can construct matrices

$$\mathcal{J}_{\mu\nu} \equiv \frac{1}{4i}[\gamma_\mu\,,\,\gamma_\nu] = -\mathcal{J}_{\nu\mu}\,, \tag{32.A.2}$$

that satisfy the commutation relations (2.4.12) of the Lorentz group generators

$$i\,[\mathcal{J}_{\mu\nu}, \mathcal{J}_{\rho\sigma}] = \eta_{\nu\rho}\mathcal{J}_{\mu\sigma} - \eta_{\mu\rho}\mathcal{J}_{\nu\sigma} - \eta_{\sigma\mu}\mathcal{J}_{\rho\nu} + \eta_{\sigma\nu}\mathcal{J}_{\rho\mu}\,. \tag{32.A.3}$$

As we will see, although Eq. (32.A.2) always gives a representation of the Lorentz algebra, it is not always an irreducible representation.

We must now distinguish between the cases of even and odd dimensionality.

Even Dimensions: $d = 2n$

To construct a convenient specific representation for the gamma matrices in $d = 2n$ dimensions, we introduce n matrices

$$a_u \equiv \frac{1}{2}\left(\gamma_{2u-1} + i\,\gamma_{2u}\right) \qquad u = 1, 2, \ldots, n\,, \tag{32.A.4}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

and take $\gamma_1, \ldots, \gamma_{2n}$ as Hermitian, it being understood that as usual

$$\gamma_{2n} \equiv -i\gamma_0 \,. \tag{32.A.5}$$

These have the anticommutation relations

$$\{a_u\,,\, a_v^\dagger\} = \delta_{uv}\,, \qquad \{a_u\,,\, a_v\} = \{a_u^\dagger\,,\, a_v^\dagger\} = 0\,. \tag{32.A.6}$$

We introduce a vector $|0\rangle$ in the representation space of the γs, defined by the condition

$$a_u^\dagger |0\rangle = 0\,, \tag{32.A.7}$$

and define the basis vectors

$$|s_1\ s_2\ \cdots s_n\rangle = a_1^{s_1} a_2^{s_2} \cdots a_n^{s_n} |0\rangle\,. \tag{32.A.8}$$

Because $a_u^2 = 0$, the operator a_u raises the value of s_u to $+1$ if $s_u = 0$ and annihilates the vector if $s_u = +1$ (as well as yielding a sign $(-1)^S$ where $S \equiv \sum_{v<u} s_v$), so all s_u take only the values 0 and $+1$, and the vectors therefore span a space of dimensionality 2^n. In this basis, the matrices a_u take the form

$$a_u = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes 1 \cdots \otimes 1\,, \tag{32.A.9}$$

with the last 2×2 matrix in the uth place. Taking the Hermitian and anti-Hermitian parts then gives the gamma matrices

$$\gamma_{2u-1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes 1 \cdots \otimes 1\,, \tag{32.A.10}$$

$$\gamma_{2u} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \otimes 1 \cdots \otimes 1\,. \tag{32.A.11}$$

(Note that this does not give the same representation of the gamma matrices in four spacetime dimensions as was introduced in Section 5.4, and has been used throughout this book.)

The representation (32.A.10)–(32.A.11) gives the Euclidean γs the simple reality and symmetry properties

$$\gamma_i^* = \gamma_i^{\mathrm{T}} = \begin{cases} \gamma_i & \text{for } i \text{ odd} \\ -\gamma_i & \text{for } i \text{ even} \end{cases}\,, \tag{32.A.12}$$

where $i = 1, 2, \ldots, 2n$. This can be expressed as a similarity relation

$$\mathscr{C}^{-1} \gamma_i \mathscr{C} = (-1)^n \gamma_i^{\mathrm{T}} = (-1)^n \gamma_i^*\,, \tag{32.A.13}$$

where $\mathscr{C}$ is the matrix

$$\mathscr{C} \equiv \gamma_2 \gamma_4 \cdots \gamma_{2n}\,. \tag{32.A.14}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Taking account of the factor $-i$ in Eq. (32.A.5), we can write this in terms of Minkowskian components as

$$\gamma_\mu^* = -\beta\gamma_\mu^{\mathrm{T}}\beta = -(-1)^n(\mathscr{C}\beta)^{-1}\gamma_\mu(\mathscr{C}\beta)\,, \tag{32.A.15}$$

where

$$\beta \equiv \gamma_{2n} = -i\gamma_0\,. \tag{32.A.16}$$

In any even dimension we may define a matrix γ_{2n+1} that plays a role similar to that of γ_5 in four dimensions. We take

$$\gamma_{2n+1} \equiv i^n\gamma_1\gamma_2\cdots\gamma_{2n}\,. \tag{32.A.17}$$

The phase here is chosen so that

$$\gamma_{2n+1}^2 = 1\,. \tag{32.A.18}$$

From the anticommutation relations (32.A.1), it follows immediately that γ_{2n+1} anticommutes with the other gamma matrices

$$\{\gamma_{2n+1}\,,\,\gamma_\mu\} = 0 \quad \text{for } \mu = 1, 2, \ldots, 2n-1, 0\,. \tag{32.A.19}$$

It is straightforward to check that γ_{2n+1} is real and symmetric

$$\gamma_{2n+1}^\dagger = \gamma_{2n+1}^* = \gamma_{2n+1}^{\mathrm{T}} = \gamma_{2n+1}\,. \tag{32.A.20}$$

From Eq. (32.A.19) we see that γ_{2n+1} commutes with the generators (32.A.2) of the $O(2n-1,1)$ algebra:

$$[\gamma_{2n+1}\,,\,\mathscr{J}_{\mu\nu}] = 0\,, \tag{32.A.21}$$

so that the $\mathscr{J}_{\mu\nu}$ cannot furnish an *irreducible* representation of the algebra of $O(2n-1,1)$. Instead, we may define a pair of 'Weyl' irreducible representations by projecting out the subspaces with $\gamma_{2n+1} = \pm 1$:

$$\mathscr{J}_{\mu\nu}^{\pm} \equiv \mathscr{J}_{\mu\nu}\left(\frac{1 \pm \gamma_{2n+1}}{2}\right)\,. \tag{32.A.22}$$

From Eq. (32.A.15) and the relation $(\mathscr{C}\beta)^{-1}\gamma_{2n+1}\mathscr{C}\beta = -(-1)^n\gamma_{2n+1}$ we find that the complex conjugate and transpose of the Weyl Lorentz generators are

$$(\mathscr{J}_{\mu\nu}^{\pm})^* = -(\mathscr{C}\beta)^{-1}\mathscr{J}_{\mu\nu}^{\mp(-1)^n}(\mathscr{C}\beta)\,, \tag{32.A.23}$$

$$(\mathscr{J}_{\mu\nu}^{\pm})^T = -\mathscr{C}^{-1}\mathscr{J}_{\mu\nu}^{\pm(-1)^n}\mathscr{C}\,. \tag{32.A.24}$$

Hence, for n even, the Weyl irreducible representations are equivalent to complex conjugates of each other, while for n odd each is equivalent to its

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

own complex conjugate.* For n odd, we still have to decide whether the Weyl representations are *real*, which would mean that there is a matrix $\mathscr{S}$ such that

$$-(\mathscr{S}\mathscr{J}^{\pm}_{\mu\nu}\mathscr{S}^{-1})^* = \mathscr{S}\mathscr{J}^{\pm}_{\mu\nu}\mathscr{S}^{-1}\,, \tag{32.A.25}$$

or *pseudoreal*, in which case there is no such $\mathscr{S}$. Using Eq. (32.A.23), the condition (32.A.25) may be written as the requirement that $\mathscr{S}^{-1}\mathscr{S}^*(\mathscr{C}\beta)^{-1}$ commutes with all $\mathscr{J}^{\pm}_{\mu\nu}$. Since the matrices $\mathscr{J}^{\pm}_{\mu\nu}$ form an irreducible set, this would require that $\mathscr{S}^{-1}\mathscr{S}^*(\mathscr{C}\beta)^{-1}$ be proportional to the unit matrix

$$\mathscr{C}\beta = \alpha\mathscr{S}^{-1}\mathscr{S}^*\,, \tag{32.A.26}$$

with α some constant. For this to be possible, we must have

$$\mathscr{C}\beta\,(\mathscr{C}\beta)^* = |\alpha|^2\cdot 1\,. \tag{32.A.27}$$

But $\mathscr{C}\beta = \gamma_2\gamma_4\cdots\gamma_{2n-2}$, and since all γ_i with i even are imaginary, we have

$$\mathscr{C}\beta(\mathscr{C}\beta)^* = (-1)^{n-1}(\gamma_2\gamma_4\cdots\gamma_{2n-2})^2 = (-1)^a\cdot 1\,, \tag{32.A.28}$$

where

$$a = (n-1)+(n-2)+\cdots+1 = n(n-1)/2\,. \tag{32.A.29}$$

Hence the Weyl representations can only be real for $n = 1 \pmod 4$ and must be pseudoreal for $n = 3 \pmod 4$.

We also note for use in Section 32.1 that

$$\mathscr{C}^* = (-1)^n\mathscr{C}\,,\qquad \mathscr{C}^{\mathrm{T}} = (-1)^{n(n+1)/2}\mathscr{C}\,,\qquad \mathscr{C}^{-1} = (-1)^{n(n-1)/2}\mathscr{C}\,, \tag{32.A.30}$$

and therefore Eq. (32.A.13) gives

$$(\mathscr{C}\gamma_\mu)^{\mathrm{T}} = (-1)^{n(n-1)/2}\mathscr{C}\gamma_\mu\,. \tag{32.A.31}$$

The γ_μ form a vector, in the sense that

$$[\mathscr{J}_{\mu\nu},\,\gamma_\rho] = -i\gamma_\mu\eta_{\nu\rho} + i\gamma_\nu\eta_{\mu\rho}\,, \tag{32.A.32}$$

and they have normal parity, in the sense that

$$\beta\,\gamma_0\,\beta = +\gamma_0\,,\qquad \beta\,\gamma_i\,\beta = -\gamma_i\quad\text{for}\quad i = 1,\ldots,2n-1\,. \tag{32.A.33}$$

The anticommutation relation (32.A.1) prevents us from constructing new tensors by taking symmetric products of γs, but it allows us to construct

* We say that one representation of the Lorentz algebra by matrices $\mathscr{L}_{\mu\nu}$ (such as $\mathscr{J}_{\mu\nu}$ or $\mathscr{J}^+_{\mu\nu}$ or $\mathscr{J}^-_{\mu\nu}$) is the complex conjugate of another representation by matrices $\mathscr{L}'_{\mu\nu}$ if $\mathscr{L}'_{\mu\nu} = -\mathscr{L}^*_{\mu\nu}$. The minus sign is included because the matrices that represent elements of the Lorentz group in the neighborhood of the identity are of the form $1+\frac{1}{2}i\,\omega^{\mu\nu}\mathscr{L}_{\mu\nu}$, with $\omega^{\mu\nu}$ real infinitesimals.

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

antisymmetric tensors of rank up to $2n$

$$\gamma_{[\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p]}\,, \tag{32.A.34}$$

where the square brackets indicate antisymmetrization, and $p \leq 2n$. Each has a number of independent spacetime components given by the binomial coefficient $\begin{pmatrix} 2n \\ p \end{pmatrix}$, so the total number of matrices of this type is

$$\sum_{p=0}^{2n} \begin{pmatrix} 2n \\ p \end{pmatrix} = 2^{2n}\,. \tag{32.A.35}$$

None of these vanishes (as can be seen by calculating their squares) and they all have different Lorentz and/or parity transformation rules and are therefore linearly independent, so any $2^n \times 2^n$ matrix may be written as a linear combination of the 2^{2n} antisymmetric tensors (32.A.34).

Odd Dimensions: $d = 2n + 1$

Now let us consider an odd spacetime dimensionality $d = 2n + 1$. We can easily find a set of $2n + 1$ Dirac $n \times n$ matrices satisfying the anticommutation relations (32.A.1): we simply use the same γ_μ with $\mu = 1, 2, \ldots 2n - 1, 0$ as for $d = 2n$, and add the matrix γ_{2n+1} defined by Eq. (32.A.17). According to Eqs. (32.A.18) and (32.A.19) these gamma matrices satisfy the anticommutation relations (32.A.1), with μ and ν running over the values $1, 2, \ldots, 2n - 1, 0, 2n + 1$, and again $\gamma_0 = i\gamma_{2n}$.

Unlike the case of even dimensionality, we cannot here find any nontrivial matrix that commutes with all the Lorentz generators, because Eqs. (32.A.17) and (32.A.18) show that the product of the $2n + 1$ gamma matrices is trivial:

$$\gamma_1\gamma_2\cdots\gamma_{2n}\gamma_{2n+1} = i^{-n}\cdot 1\,. \tag{32.A.36}$$

The Lorentz generators (32.A.2) with μ and ν running over the values $1, 2, \ldots, 2n - 1, 0, 2n + 1$ therefore furnish an irreducible representation of the Lorentz group by themselves. To test their reality properties, we note that γ_{2n+1} is real and symmetric and satisfies $(\mathscr{C}\beta)^{-1}\gamma_{2n+1}\mathscr{C}\beta = -(-1)^n\gamma_{2n+1}$, so Eq. (32.A.15) applies for $\mu = 2n + 1$ as well as for $\mu = 1, 2, \ldots, 2n - 1, 0$. The Lorentz generators therefore satisfy

$$\mathscr{J}^*_{\mu\nu} = -(\mathscr{C}\beta)^{-1}\mathscr{J}_{\mu\nu}\mathscr{C}\beta\,, \tag{32.A.37}$$

$$\mathscr{J}^{\mathrm{T}}_{\mu\nu} = -\mathscr{C}^{-1}\mathscr{J}_{\mu\nu}\mathscr{C}\,, \tag{32.A.38}$$

so in each odd dimension the fundamental spinor representation is either real or pseudoreal. Exactly the same argument as in the case where $d = 2n$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

tells us that the spinor representations for $d = 2n + 1$ are again real or pseudoreal according to whether the sign $(-1)^a$ in Eq. (32.A.28) is positive or negative, and therefore according to Eq. (32.A.29) they are real for $n = 0 \pmod 4$ and $n = 1 \pmod 4$ and pseudoreal for $n = 2 \pmod 4$ and $n = 3 \pmod 4$.

We can again construct antisymmetric tensors (32.A.34), now of rank p up to $2n + 1$, but only half of these are independent, because they obey relations

$$\epsilon^{\mu_1\mu_2\cdots\mu_{2n+1}}\gamma_{[\mu_1}\gamma_{\mu_2}\cdots\gamma_{\mu_p]} \propto \gamma^{[\mu_{p+1}}\gamma^{\mu_2}\cdots\gamma^{\mu_{2n+1}]}\,, \tag{32.A.39}$$

with $\epsilon^{\mu_1\mu_2\cdots\mu_{2n+1}}$ as usual totally antisymmetric. (For $d = 2n$ no such relations are possible, because the left- and right-hand sides of Eq. (32.A.39) have opposite parity, but this argument does not apply for $d = 2n + 1$, where $\epsilon^{\mu_1\mu_2\cdots\mu_{2n+1}}$ has even spatial parity.) The total number of independent matrices of the form (32.A.34) is now given by

$$\sum_{p=0}^{n}\begin{pmatrix} 2n+1 \\ p \end{pmatrix} = 2^{2n}\,, \tag{32.A.40}$$

so any $2n \times 2n$ matrix may be written as a linear combination of the $n+1$ independent antisymmetric tensors (32.A.34) with $0 \leq p \leq n + 1$.

Finally, we note that for either $d = 2n$ or $d = 2n + 1$, the $O(d-1,1)$ Dirac and Lorentz algebras are related to the corresponding $O(d)$ algebras by setting

$$\gamma_{2n} \equiv -i\gamma_0\,, \qquad \mathcal{J}_{i\,2n} \equiv -i\mathcal{J}_{i0}\,, \tag{32.A.41}$$

so that

$$\{\gamma_i\,,\,\gamma_j\} = 2\delta_{ij}\,, \tag{32.A.42}$$

and

$$\mathcal{J}_{ij} = \frac{1}{4i}[\gamma_i\,,\,\gamma_j] = -\mathcal{J}_{ji}\,, \tag{32.A.43}$$

with i and j running from 1 to d. It follows from Eq. (32.A.42) that for $i \neq j$, $\mathcal{J}_{ij}^2 = 1/4$, so the eigenvalues of each $\mathcal{J}_{ij}$ are limited to $\pm 1/2$. To be more specific, in the fundamental spinor representation the generators of the Cartan subalgebra are represented by

$$\mathcal{J}_{2u-1\,2u} = \frac{1}{2}[a_u\,,\,a_u^*] = a_u a_u^* - \frac{1}{2}\,, \tag{32.A.44}$$

for which the basis vectors (32.A.8) are eigenvectors, with

$$\mathcal{J}_{2u-1\,2u}\left|s_1\ s_2\ \cdots s_n\right\rangle = \left(s_u - \frac{1}{2}\right)\left|s_1\ s_2\ \cdots s_n\right\rangle\,. \tag{32.A.45}$$

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

The difference between dimensionalities $d = 2n$ and $d = 2n + 1$ is that for $d = 2n$ we have two fundamental spinor representations in which the eigenvalue $(-2\sigma_1)(-2\sigma_2)\cdots(-2\sigma_n)$ of γ_{2n+1} is constrained to be $+1$ or -1, while for $d = 2n + 1$ there is one fundamental spinor representation with no such limitation on the σ_u.

It was the limitation of the eigenvalues of each $\mathcal{J}_{ij}$ to $\pm 1/2$ that in Section 32.1 identified the fundamental spinor representations as the only possible representations of the Lorentz algebra that could be furnished by fermionic symmetry generators. Indeed, from this condition we could have inferred that the $O(d)$ generators may be represented in a basis of the form (32.A.8), with $s_u = \sigma_u + 1/2$, and carried out the derivations of this appendix in reverse order, using Eqs. (32.A.4)–(32.A.7) (along with Eq. (32.A.17) for d odd) to express the Lorentz generators in terms of a set of γ_μ satisfying the anticommutation relations (32.A.1).

Problems

1. Classify the massless particle multiplets for each allowed kind of supersymmetry in six spacetime dimensions, when all central charges vanish.

2. Suppose that it were possible to have massless particles for all spins up to $j = 3$, but no higher. Taking into account the fact that massless particles of spin 2 do exist, what is the maximum spacetime dimensionality in which supersymmetry is possible? What is the maximum number of supersymmetry generators for each allowed value of the spacetime dimensionality?

3. Consider types IIA and IIB supersymmetry in ten spacetime dimensions. Assume that only scalar central charges appear in the extended supersymmetry anticommutation relations. Find a lower bound on particle masses in terms of these central charges. Describe the 'BPS' massive particle multiplets allowed for particles whose masses are at this lower bound.

4. List the possible independent scalar and/or tensor central charges for $N = 1$ supersymmetry in nine spacetime dimensions.

References

1. T. Kaluza, *Sitz. Preuss. Akad. Wiss.* **K1**, 966 (1921).
2. O. Klein, *Z. Phys.* **37**, 895 (1926); *Nature* **118**, 516 (1926).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

3. J. H. Schwarz, *Nucl. Phys.* **B46**, 61 (1972); R. C. Brower and K. A. Friedman, *Phys. Rev.* **D7**, 535 (1972).

4. The supergravity field theory in 11 spacetime dimensions was formulated by E. Cremmer, B. Julia, and J. Scherk, *Phys. Lett.* **76B**, 409 (1978). The idea that weakly coupled Type IIA string theories in 10 spacetime dimensions have an 11-dimensional origin is due to M. J. Duff, P. S. Howe, T. Inami, and K. Stelle, *Phys. Lett.* **B191**, 70 (1987). This was shown for strongly coupled Type IIA string theories in ten dimensions by P. K. Townsend, *Phys. Lett.* **B350**, 184 (1995). Connections between these theories and all the other ten-dimensional string theories were then pointed out by E. Witten, *Nucl. Phys.* **B445**, 85 (1995).

5. R. Haag, J. Lopuszanski, and M. Sohnius, *Nucl. Phys.* **B88**, 257 (1975).

6. W. Nahm, *Nucl. Phys.* **B135**, 149 (1978).

6*a*. See, for example, H. W. Turnbull and A. C. Aitkens, *An Introduction to the Theory of Canonical Matrices* (Dover Publications, New York, 1961).

6*b*. A useful summary is given by J. Strathdee, *Int. J. Mod. Phys.* **A2**, 273 (1987).

7. For these arguments in four spacetime dimensions, see S. Weinberg, *Phys. Rev.* **135**, B1049 (1964); *Phys. Rev.* **138**, B988 (1965); R. P. Feynman, unpublished; M. T. Grisaru and H. N. Pendleton, *Phys. Lett.* **67B**, 323 (1977). The arguments in higher spacetime dimensions are similiar.

8. See, for example, J. Hughes, J. Liu, and J. Polchinski, *Phys. Lett.* **B180**, 370 (1986); E. Bergshoeff, E. Sezgin, P. K. Townsend, *Phys. Lett.* **B189**, 75 (1987); A. Achúcarro, J. M. Evans, P. K. Townsend, and D. L. Wiltshire, *Phys. Lett.* **B198**, 441 (1987); P. K. Townsend, *Phys. Lett.* **B202**, 53 (1988); P. K. Townsend, in *Particles, Strings, and Cosmology: Proceedings of Workshop on Current Problems in Particle Theory 19 at Johns Hopkins University, March 1995* (World Scientific, Singapore, 1996); J. Maldacena, *Adv. Theor. Math. Phys.* **2**, 231 (1998). For reviews, see M. J. Duff, R. R. Khuri, and J.-X. Lu, *Phys. Rep.* **259**, 213 (1995); A. Giveon and D. Kutasov, 1998 preprint hep-th/9802067, to be published in *Rev. Mod. Phys.*

9. J. W. van Holten and A. Van Proeyen, *J. Phys. A: Math. Gen.* **15**, 3763 (1982).

10. J. A. de Azcárraga, J. P. Gauntlett, J. M. Izquierdo, and P. K. Townsend *Phys. Rev. Lett.* **63**, 2443 (1989).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

11. D. Sorokin and P. K. Townsend, *Phys. Lett.* **B412**, 265 (1997); J. P. Gauntlett, J. Gomis, and P. K. Townsend, *J. High Energy Phys.* **9801**, 003 (1998).

12. The covariant field equations for these 2-branes were given by E. Bergshoeff, E. Sezgin, P. K. Townsend, Ref. 8 and *Ann. Phys. (NY)* **185**, 330 (1988). The demonstration that these 2-branes are solutions of the supergravity field equations that provide sources for the 3-form gauge field is due to M. J. Duff and K. Stelle, *Phys. Lett.* **B253**, 113 (1991).

13. The demonstration that these 5-branes are solutions of the supergravity field equations that provide sources for the 6-form gauge field was given by R. Gueven, *Phys. Lett.* **B276**, 49 (1992). The covariant field equations for these 5-branes were given by M. Aganagic, J. Park, C. Popescu, and J. Schwarz, *Nucl. Phys.* **B496**, 191 (1997); P. S. Howe and E. Sezgin, *Phys. Lett.* **B394**, 62 (1997); P. Pasti, D. Sorokin, and M. Tonin, *Phys. Lett.* **B398**, 41 (1997); P. S. Howe, E. Sezgin, and P. C. West, *Phys. Lett.* **B399**, 49 (1997); I. Bandos, K. Lechner, A. Nurmagambetov, P. Pasti, D. Sorokin, and M. Tonin, *Phys. Rev. Lett.* **78**, 4332 (1997).

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198.011

Author Index

Where page numbers are given in *italics*, they refer to publications cited in lists of references.

Abreu, P. *245*
Achúcarro, A. *408*
Affleck, I. 266, 285, 286, *306*, *381*
Aganagic, M. *409*
Aharonov, Y. *23*
Aitkens, A. C. *408*
Akulov, V. P. 7, *23*
Alvarez-Gaumé, L. *245*, *380*
Amaldi, U. *243*
Antoniadis, I. *241*, *306*
Arkani-Hamed N. *241*, *246*
Arnowitt, R. *244*, *245*, *247*, *379*, *380*
Astbury, A. *245*
Axen, D. *245*

Babu, K. S. *244*, *246*, *247*
Bagger, J. *242*, *243*, *246*, *306*, *379*
Bais, F. A. *177*
Bandos, I. *409*
Banks, T. *242*, *381*
Barate, R. *245*
Barbieri, R. *244*, *380*
Barger, V. *243*
Beg, M. A. *23*
Belyaev, V. M. *247*
Bennett, J. R. J. *242*
Berezin, F. A. 86, *112*
Bergshoeff, E. *408*, *409*
Bessis, D. *176*
Binétruy, P. *381*
Bogomol'nyi, E. B. *177*
Boulware, D. 380
Braaten, E. *244*
Breitenlohner, P. *177*
Brignole, A. *245*
Brink, L. *177*
Brower, R. C. *408*
Buras, A. J. *244*, *245*

Capper, D. M. *316*
Carena, M. *245*
Casas, J. A. *381*
Casher, A. *23*
Chacko, Z. *381*
Chamseddine, A. H. *247*, *380*
Choudhury, S. R. *244*
Chun, E. J. *246*
Claudson, M. *245*
Cleaver, G. *246*
Cohen, A. G. *242*
Coleman, S. 2, 12, 13, 14, 15, 17, 19, 20, *23*, *177*, 330, *379*
Cordes, S. F. *306*
Cox, P. H. *244*
Cremmer, E. *380*, *408*
Cvetič, J. R. *246*

D'Adda, A. *177*
Dai, J. *244*
Das, A. *380*
Davis, A. C. 266, *306*
Dawson, S. *243*, *245*
de Azcárraga, J. A. *408*
de Boer, W. *243*
de Luccia, F. 330, *379*
De Rújula, A. *244*
de Wit, B. 158, 159, *177*
Delbourgo, R. *23*, *316*, *317*
Deser, S. *112*, *379*, *380*
Di Vecchia, P. *177*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Dicus, A. *242*, *244*, *246*
Dienes, K. R. *243*
Dimopoulos, S. 193, *241*, *242*, *243*, *244*, *245*, *246*, *247*
Dine, M. *176*, *244*, *245*, *246*, 266, 285, 286, *306*, *381*
Dixon, L. *242*
Donoghue, J. F. *243*
Dubovsky, S. L. *246*
Duff, M. *244*, *245*, *408*, *409*
Dugan, M. *244*
Dutta, B. *246*
Dvali, G. *241*, *242*, *246*
Dykstra, H. *244*
Dyson, F. J. *22*

Efthimiou, C. *176*
Ellis, J. 188, *242*, *243*, *244*, *245*, *246*, *247*, *380*
Espinosa, J. R. *246*
Evans, J. M. *408*
Everett, L. *246*

Farrar, G. *242*
Fayet, P. *176*, *177*, *242*, *380*
Feng, J. L. *242*
Ferrara, S. *23*, *54*, *112*, *176*, *242*, *244*, *305*, *306*, *316*, *379*, *380*
Feynman, R. P. *408*
Fischler, W. *176*, *245*
Freed, D. *112*
Freedman, D. Z. 158, 159, *177*, *379*, *380*
Friedman, K. A. *408*
Fujikawa, K. *316*
Furstmann, H. *243*

Gabbiani, F. *244*, *245*
Gabrielli, E. *244*, *245*
Gaillard, M. K. 203, *244*, *381*
Galperin, A. *306*
Gatto, R. *244*
Gauntlett, J. P. *408*, *409*
Gava, E. *306*
Gavela, M. B. *244*
Gell-Mann, M. 1, *23*, *24*, 152
Georgi, H. 193, *242*, *243*, *245*
Gervais, J.-L. 5, *23*
Gildener, E. *241*
Girardello, L. *176*, *178*, *306*, *380*
Giudice, G. F. 226, 228, *242*, *246*, *380*, *381*
Giunti, C. *243*
Giveon, A. *178*, *408*
Gliozzi, F. 7, *23*, *379*
Gol'fand, Yu. A. 7, *23*
Gomis, J. *112*, *409*
Gorbunov, D. S. *246*
Goto, T. *247*
Gràcia, X. *112*
Green, M. B. *23*
Greene, B. *176*
Greenberg O. W. *23*
Griffiths, P. *112*
Grimm, R. *177*
Grinstein, B. *244*
Grisaru, M. T. 45, *54*, *112*, *176*, *317*, *379*, *408*
Grossman, Y. *244*
Grünewald, M. *245*
Gueven, R. *409*
Gunion, J. F. *243*, *246*
Gursey, F. *22*

Haag, R. *24*, 25, 53, *54*, 382, 383, *408*
Haber, H. E. *243*, *246*
Hagelin, J. S. *242*, *244*, *247*
Hall, L. J. *243*, *244*, *380*
Halperin, A. *177*
Hamaguchi, K. *381*
Harada, K. *176*
Harris, J. *112*
Harris, P. G. *244*
He, X.-G. *244*
Heinemayer, S. *245*
Hisano, J. *242*, *247*
Hollik, W. *245*
Honerkamp, J. *316*
Hořava, P. *243*
Horsley, R. *177*
Howe, P. *177*, *408*, *409*
Hughes, J. 249, *306*, *408*

Ibáñez, L. E. *247*, *380*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Ibrahim, T. *244*
Iliopoulos, J. *112*, *176*
Inami, T. *408*
Intrilligator, K. *306*
Ivanov, E. A. *177*
Izawa, K. I. *381*
Izquierdo, J. M. *408*

Julia, B. *177*, *380*, *408*

Kac, V. G. 279, *306*
Kagan, A. *244*, *245*
Kaluza, T. 382, *407*
Kane, G. L. *243*
Kaplan, D. B. *242*, *381*
Kaplunovsky, V. *381*
Karlen, D. *245*
Katz, E. *380*
Kay, J. H. *380*
Kelley, S. *243*, *244*
Khuri, R. R. *408*
Kim, C. W. *243*
Kim, J. E. *246*, *381*
Kizuri, Y. *244*
Klein, O. 382, *407*
Kolb, E. W. *242*
Kolda, C. *242*, *243*, *244*, *246*
Kounnas, C. *380*
Kovacs, S. *178*
Krause, F. *316*
Kurosawa, K. *242*
Kutasov, D. *408*

Lahanas, A. B. *380*
Lang, W. *247*, *316*
Langacker, P. *243*, *246*, *247*
Le Yaouanc, A. *244*
Lechner, K. *409*
Lee, B. W. 203, *242*, *244*
Lee, U. W. *243*
Leibrandt, G. *316*
Leigh, R. *244*
Leon, J. M. *380*
Lepeintre, F. *242*
Lévy, M. *112*
Li, C. S. *244*
Likhtman, E. P. 7, *23*
Lindgren, O. *177*
Lindner, M. *244*
Liu, J. 249, *306*, *408*
Lopez, J. *244*
Lopuszanski, J. T. *24*, 25, 53, *54*, 382, 383, *408*
Louis, J. *381*
Low 152
Lu, J.-X. *408*
Lucas, V. *247*
Luo, M.-X. *243*
Luty, M. A. *246*, *380*, *381*
Lykken, J. *241*, *380*

MacIntire, D. A. *381*
Maksymyk, I. *381*
Maldacena, J. *408*
Mandelstam, S. *177*
Mandula, J. 2, 12, 13, 14, 15, 17, 19, 20, *23*
Manohar, A. *245*
March-Russell, J. *244*
Martin, S. P. *243*, *246*
Masiero, A. *242*, *244*, *245*, *381*
Matchev, K. T. *243*
McGlinn, W. D. *23*
McKellar, S. R. *244*
Michel, L. *23*
Misiak, M. *245*
Mohapatra, R. N. *242*
Montonen, C. 174, *177*
Moroi, T. *381*
Muñoz, C. *247*, *381*
Münz, M. *245*
Murayama, H. *247*, *380*

Nahm, W. *379*, 383, *408*
Nandi, S. *246*
Nanopoulos, D. V. *242*, *243*, *244*, *245*, *247*, *380*
Nappi, C. *245*
Nath, P. *244*, *247*, *379*, *380*
Ne'erman, Y. 1, *24*
Nelson, A. E. *242*, *246*, *381*
Neveu, A. 5, *23*, *177*
Nihei, T. *247*
Nilles, H. P. *176*, *242*, *243*, *246*, *380*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Nilsson, B. E. W. *177*
Nir, Y. *244*, *245*, *246*
Nomura, Y. *242*, *381*
Novikov, V. A. *306*
Nurmagambetov, A. *409*

O'Raifeartaigh, L. *23*, 83, *112*
Ogievetsky, V. I. *177*
Ohta, N. *380*
Olive, D. 7, *23*, 174, *177*, *242*
Oliver, L. *244*
Osborn, H. *177*, *178*
Oshimo, N. *244*
Ovrut, B. *245*

Paban, S. *244*, *245*
Pagels, H. *243*
Pais, A. *22*, *23*
Pakvasa, S. *244*
Palumbo, F. *176*
Panvini, R. S. *243*
Park, J. *409*
Parke, S. *177*
Partouche, H. *306*
Pasti, P. *409*
Pati, J. C. *247*
Peccei, R. D. *242*
Pendleton, H. N. 45, *54*, *379*, *408*
Pène, O. *244*
Pham, T. N. *244*
Piguet, O. *316*
Pokorski, S. *245*
Polchinski, J. *23*, *176*, *242*, *244*, 249, *306*, *380*, *408*
Polonsky, N. *242*, *243*, *246*
Polonyi, J. *380*
Pomerol, A. *242*, *246*, *380*
Pons, J. *112*
Pontón, E. *381*
Popescu, C. *409*
Porrati, M. *178*, *306*
Prasad, M. K. *177*
Primack, J. R. *177*, *243*

Quinn, H. *242*, *243*
Quiros, M. *245*, *380*

Raby, S. *176*, *242*, *243*, *245*, *246*, *247*
Radicatti, L. A. *22*
Ramon Medrano, M. *380*
Ramond, P. 5, *23*
Randall, L. *245*, *380*, *381*
Rarita, W. 333, *380*
Rattazzi, R. 226, 228, *244*, *246*, *380*
Raynal, J.-C. *244*
Ridolfi, G. *245*
Riotto, A. *242*
Robinson, J. *245*
Roček, M. *176*, *177*, *317*
Ross, G. *247*
Rossi, G. C. *306*
Roszkowski, L. *246*
Rudaz, S. *247*

Sakai, N. *176*, 243, *246*
Sakita, B. 5, *22*, *23*
Salam, A. *23*, 55, 59, 60, *112*, *176*, *242*, *316*
Samuel, S. *245*
Sarid, U. *243*
Savoy, C. A. *380*
Scherk, J. 7, *23*, *177*, *379*, *380*, *408*
Scheunert, M. *316*
Schlindwein, M. *316*
Schwarz, J. H. 5, *23*, *177*, *408*, *409*
Schwinger, J. *177*, 333, *380*
Seiberg, N. 148, *176*, *177*, 226, *244*, *246*, 266, 285, 286, 287, 291, 292, 299, 302, 303, *306*, 307, *381*
Sen, A. *178*
Sezgin, E. *408*, *409*
Shadmi, Y. *380*
Shifman, M. A. *306*
Shirman, Y. *245*, *246*, *380*
Siegel, W. *176*, *177*, *317*
Silvestrini, L. *244*, *245*
Smilga, A. V. 279, *306*
Smith, J. R. *241*
Smith, P. F. *242*
Sohnius, M. *24*, 25, 53, *54*, *177*, 382, 383, *408*
Solà, J. *243*
Sommerfield, C. M. *177*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Soper, D. E. *243*, *306*
Soroka, V. *379*
Sorokin, D. *409*
Srednicki, M. *242*, *245*, *380*
Stelle, K. S. *177*, *244*, *245*, *379*, *380*, *408*, *409*
Strathdee, J. *23*, 55, 59, 60, *112*, *176*, *242*, *316*, *408*
Sundrum, R. *380*, *381*
Susskind, L. *23*, *176*
Sutter, D. *243*, *244*
Symanzik, K. *176*

Takita, M. *247*
Tamvakis, K. *246*, *247*, *380*
Tanaka, T. *244*
Taylor, T. R. *306*
Teitelboim, C. *379*
Teplitz, V. L. *242*
Thomas, S. *245*, *246*
Tommasini, D. *242*
Tonin, M. *409*
Townsend, P. K. *177*, *380*, *408*, *409*
Troitsky, S. V. *246*
Turnbull, H. W. *408*

Vafa, C. *178*
Vainshtein, A. I. *306*
van der Waerden 30
van Holten, J. W. *408*
van Nieuwenhuizen, P. *242*, *379*, *380*
Van Proeyen, A. *380*, *408*
Vegas, F. J. *244*
Veneziano, G. *306*
Volkov, D. V. 7, *23*, *379*
Vysotsky, M. I. *247*
Wagner, C. E. M. *245*
Wali, K. C. *23*
Wang, J. *246*
Weiglein, G, *245*
Weiler, T. J. *243*
Weinberg, S. *23*, *176*, *241*, *242*, *243*, *244*, *245*, *246*, *379*, *380*, *408*
Wells, J. D. *246*
Wess, J. 6, 7, 8, *23*, 59, 83, *112*, *176*, *379*
West, P. C. *177*, *379*, *380*, *409*
Wigner, E. P. 1, 10, *22*
Wilczek, F. *242*, *243*, *244*, *245*, *246*, *247*
Wiltshire, D. L. *408*
Wise, M. B. *244*, *245*, *380*
Witten, E. *23*, *177*, *178*, *243*, 249, 278, 287, 291, 292, 294, 299, 302, 303, *305*, *306*, *408*
Wolfram, S. *242*
Wu, Y.-Y. *381*
Wyler, D. *243*, *380*

Yanagida, T. *246*, *247*, *381*
Yildiz, A. *244*
Yuan, T. C. *244*

Zaffaroni, A. *178*
Zakharov, V. I. *306*
Zee, A. *177*, *246*
Zhang, R.-J. *242*, *244*
Zumino, B. 6, 7, 8, *23*, *54*, 59, 83, 106, *112*, *176*, *379*, *380*
Zwanziger, D. *177*
Zwirner, F. *245*, *246*

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:48, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Subject Index

Ah terms, 199–201, 234, 341, 363
anomalies, 45, 101, 147, 181, 267–8, 294, 339
anomaly-mediated supersymmetry breaking, 337–41, 364
anti-de Sitter space, 330
auxiliary fields, 59, 78, 119, 133, 145, 323, 327–8, 350
axion, *see* Peccei–Quinn symmetry

Baker–Hausdorff formula, 115
baryon and lepton number non-conservation, 182–4, 235–40
baryon parity, 240
β angle, defined, 213
bino, 180, 222–6, 229, 238, 340, 364
$B\mu$ term, 199–201, 211, 232–3, 363
BPS states, 52–53, 166, 300

$\mathscr{C}$-matrix, 108, 402
Cartan subalgebra, 163, 280
Casimir constants C_1 and C_2, 267, 278–9
central charges, 30–1, 35–6, 163–5, 296, 385–6, 389–92, 398–401
  also see BPS states
Ch terms, 199–201, 234, 363, 374
charginos, 219–20
chiral superfields, 68–74, 346–7
Coleman–Mandula theorem, 2, 12–21, 382, 397–8
collapsed supermultiplet, 48, 51
complexification of symmetries, 106–7, 135
component fields
  of chiral superfields, 68–72
  of general superfields, 61–64
  of metric superfield, 322–3
conformal symmetry
  in 2 dimensions, 5
  in 4 dimensions, 21–2
connection matrix Γ, 114
cosmic abundances, 186–8
cosmological constant, 157, 348, 359, 371
covariant derivative D_μ, 343–4
CP-non-conservation, 204–9, 234
CP-odd scalar A, 214
current superfields, 89–90
curvature $R_{\mu\nu}{}^{ab}$ in vierbein formalism, 378

D-terms in action, 65, 72, 76–7, 87, 103–5, 347
de Wit–Freedman transformation, 158–60
de Sitter space, 330
Dirac matrices, xx–xxi, 401–7
dotted indices, xx, 30
duality, 174–5, 292–4
dyons, 167, 297, 300–2

Einstein–Hilbert action, 328
Einstein superfield, 326–7
Einstein tensor, 326, 332
electric dipole moments, *see* CP-non-conservation
electroweak symmetry breaking, 180–1, 219
energy-momentum tensor $T^{\mu\nu}$, 98–99, 324

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

extended supersymmetry, 36, 45–7, 160–75
also see Seiberg–Witten solution
external field translation symmetry, 150–1, 268

F-parameter of broken supersymmetry, 258–9, 336–7
$\mathscr{F}$-terms in action, 72, 77–8, 87, 125–6, 131, 141–4, 181–2, 185, 348
Fayet–Iliopoulos terms, 124, 144, 152–3
finiteness of $N = 4$ theories, 174
flat directions, 84–5, 154, 163
flavor-changing processes, 198–201, 218

gauge-mediated supersymmetry breaking, 220–35
gauge superfield W_α, 125–7, 130–1
gauginos, 119, 143–4, 189, 281, 339, 363–4, 374–5
gaugino condensation, 284–5, 366
also see gluinos, winos, binos
Gell-Mann–Low equations, *see* renormalization group equations
generalized gauge transformations, 116–7
generalized supersymmetric quantum chromodynamics, 267, 270–4
gluino, 180, 222–6, 229, 237, 340, 364
goldstino, 85, 147–8, 234, 251, 256, 259, 262–4
graded Lie algebras, 25–9
also see supersymmetry algebras
gravitino, 45, 188, 196–8, 234, 320–6, 333–7, 349, 356, 359, 368
gravity-mediated supersymmetry breaking, 355–75

Haag–Lopuszanski–Sohnius theorem, 30–40
heterotic supergravity, 397
hidden sector, 195, 355–7, 365
hierarchy problem, 179, 185, 194, 355–6
Higgs fields and superfields, 180–1, 210–9, 226–32
higgsinos, 181, 183, 190–2
holomorphy arguments, 150, 226–7, 266–76
hypermultiplets, 46–7, 167, 300

instantons, 131–2, 268, 270, 284, 295

Jacobi and super-Jacobi identities, 26

Kahler manifolds, 106–7
Kahler metric, 105, 350
modified, 352–3
Kahler potential, 73, 102–6, 349
modified, 352–3

left-chiral superfields, *see* chiral superfields
lepton number non-conservation, *see* baryon and lepton number non-conservation
lepton superfields, 180
linear superfields, 75, 90
little group, 393
local supersymmetry, 93–4, 341–3

M theory, 382, 395, 400–1
magnetic monopoles, 52, 166–7, 294, 300–2
Majorana potential, 10
Majorana spinors, xviii, 58, 81, 107–11
masses
for broken supersymmetry, 145–8, 193–4
for unbroken supersymmetry, 136–40
sum rule, 147, 192
also see charginos, gauginos, Higgs fields, neutralinos, sleptons, squarks, gravitino
messenger superfields, 220–2, 224
metric superfield H_μ, 319–26
minimal supersymmetric standard model, 198–240
modular fields, 287–8, 365
monodromy, 299–304

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

Montonen–Olive duality, *see* duality
μ-term, 185, 210, 220, 232–3, 361, 368
multiplication rules, 64, 71–72, 346–7

naive dimensional analysis, 208
naturalness bounds, 186
neutralinos, 219–30
next-to-lightest supersymmetric particle, 234–5
non-linear σ-models, 106–7
non-renormalization theorems, 148–55, 269, 313–6
no-scale theories, 354

on shell symmetries, 79, 169
O'Raifeartaigh models, 83–5, 252–6

p branes, 397–401
parity
  in Wess–Zumino model, 83
  of component fields 66, 74
  of superfields, 66, 74
  of supersymmetry generators, 41–42
Peccei–Quinn symmetry, 185, 214, 232, 361–2, 365, 368, 373
photino, 277
  also see bino and wino
potential of scalar fields, 79, 84–5, 134, 144, 163, 352–4, 362, 371–2
potential superfields, 88–9, 308–9, 359, 361
prepotential, 291
proton decay, *see* baryon and lepton number non-conservation

quantum chromodynamics, *see* generalized supersymmetric quantum chromodynamics, gluino, squark
quark superfields, 180

R-symmetry, 36–7, 40, 45, 74–5, 84, 161, 171–3, 182, 267–8, 284–5, 290–2, 392
  current $\mathscr{R}^\mu$, 100–1
  R-parity, 184, 236
Rarita–Schwinger equation, 333–5, 344
reduced coupling constants, 208
renormalizability conditions, 76
renormalization group equations, 149, 173–4, 188–92, 269, 279, 339
right-chiral superfields, *see* chiral superfields

scale non-invariance superfield X, 97, 324, 338
Seiberg–Witten solution, 287–305
short supermultiplets, *see* BPS states
$SL(2,\mathbb{Z})$ symmetry, 296
sleptons, 180, 189, 223–4, 229, 363, 374
space inversion, *see* parity
sparticles, 43
  also see gauginos, squarks, sleptons, higgsinos, charginos, neutralinos
spin, in higher dimensions, 393, 401–7
spin connection ω_μ^{ab}, 343–4, 377
squarks, 180, 186, 189, 223–4, 229, 363, 374
strings, 4–7, 175, 183, 240, 365, 382, 396–7
$SU(4)$ symmetry, in nuclear physics, 1, 10
$SU(6)$ symmetry, in quark model, 1, 8–12
sum rules for supersymmetry current, 256–66
superconformal symmetry
  in 2 dimensions, 6
  in 4 dimensions, 38–40
supercurrent Θ^μ, 90, 96–102, 321, 323–5
superderivatives $\mathscr{D}_\alpha$, defined, 64–5
superfields, 59–68
  field equations, 96–7
  also see chiral superfields, current superfields, linear superfields, potential superfields, gauge superfields, Higgs superfields, quark superfields, lepton superfields, messenger superfields, metric superfield, supercurrent, Einstein superfield, scale-non-invariance superfield

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198

supergraphs, 307–16
supergravity, 45, 318–9, 395
  for weak fields, 319–43
  to all orders, 343–55
  also see gravitino, anomaly-mediated supersymmetry breaking, gravity-mediated supersymmetry breaking, heterotic supergravity, types IIA and IIB supergravity
supermultiplets
  massive particles in 4 dimensions, 48–52
  massless particles in 4 dimensions, 43–7
  massless particles in higher dimensions, 393–7
superpotential, 73, 82, 149–52, 266–76
superpropagators, *see* supergraphs
superspace, 60
  field equations 87–9
  integrals 86
  also see superfields, superderivatives
superstrings, *see* strings
supersymmetry
  algebras, xx, 29–40, 382–93
  current $S^{\mu}(x)$, 90–3, 140–1, 248, 324
  gauge theories, 101–76, 276–87, 353
  generators, *see* algebras
  history, 1, 4–8
  also see extended supersymmetry, supergravity, minimal supersymmetric standard model
supersymmetry breaking
  partial breaking, 249–50
  soft breaking, 155–7, 199–201
  spontaneous breaking, 7–8, 80, 134–6, 153–5, 194–5, 248–56, 276–87
  also see gauge-mediated supersymmetry breaking, gravity-mediated supersymmetry breaking, anomaly-mediated supersymmetry breaking
surface tension, 330–1

types IIA and IIB supergravity, 396–7

unconventional symmetries, 1–4
unification of strong and electroweak couplins, 188–92, 221
unitarity gauge, 136

vacuum energy density, 33, 249–51, 257–8, 264–6, 329–31, 333
vierbein, 319, 375–6

weight, defined, 383
Wess–Zumino gauge, 117–9, 159–60
  for supergravity, 323, 341
Wess–Zumino model, 6–7, 58–9, 82–3, 252
Weyl group, 282
Weyl transformation, 331–3, 351
Wilsonian Lagrangian, 149–52, 268
winding number ν, 132, 268, 270, 284, 295
wino, 180, 222–6, 229, 238, 340, 364
Witten index, 251–6, 277–285

X superfield, *see* scale-non-invariance superfield

zino, *see* bino and wino
$\mathbb{Z}_8$ symmetry, 295, 298

Downloaded from https://www.cambridge.org/core. University of Science and Technology of China, on 21 Mar 2020 at 13:29:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139644198