QUANTUM FIELD THEORY

A Modern Introduction

MICHIO KAKU

Department of Physics
City College of the
City University of New York

New York Oxford
OXFORD UNIVERSITY PRESS
1993

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland Madrid

and associated companies in

Berlin Ibadan

Copyright © 1993 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.,
200 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stroed in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Kaku, Michio. Quantum field theory: a modern introduction/
Michio Kaku.
p. cm. Includes bibliographical references and index.
ISBN 0-19-507652-4
ISBN 0-19-509158-2 (pbk.)
1. Quantum field theory. 2. Gauge fields (Physics)
3. Standard model (Nuclear physics)
I. title QC174.45.K34 1993 530.1'43—dc20 92-27704

9 8 7 6 5 4 3

Printed in the United States of America
on acid-free paper

This book is dedicated to my parents.

Preface

In the 1960s, there were a number of classic books written on quantum field theory. Because of the phenomenal experimental success of quantum electrodynamics (QED), quantum field theory became a rigorous body of physical knowledge, as established as nonrelativistic quantum mechanics.

In the 1970s and 1980s, because of the growing success of gauge theories, it was clear that a typical 1-year course in quantum field theory was rapidly becoming obsolete. A number of advanced books appeared on various aspects of gauge theories, so often a 1-year course on quantum field theory became disjoint, with one book on QED being the basis of the first semester and one of several books on various aspects of gauge theories being the basis of the second semester.

Today, because of the success of the Standard Model, it is necessary to consolidate and expand the typical 1-year quantum field theory course. There is obviously a need for a book for the 1990s, one that presents this material in a coherent fashion and uses the Standard Model as its foundation in the same way that earlier books used QED as their foundation. Because the Standard Model is rapidly becoming as established as QED, there is a need for a textbook whose focus is the Standard Model.

As a consequence, we have divided the book into three parts, which can be used in either a two- or a three-semester format:

I: Quantum Fields and Renormalization

II: Gauge Theory and the Standard Model

III: Non preturbative Methods and Unification

Part I of this book summarizes the development of QED. It provides the foundation for a first-semester course on quantum field theory, laying the basis for perturbation theory and renormalization theory. (However, one may also use it in the last semester of a three-semester course on quantum mechanics, treating it as the relativistic continuation of a course on nonrelativistic quantum mechanics.

In this fashion, students who are not specializing in high-energy physics will find Part I particularly useful, since perturbation theory and Feynman diagrams have now penetrated into all branches of quantum physics.)

In Part II, the Standard Model is the primary focus. This can be used as the basis of a second semester course on quantum field theory. Particular attention is given to the method of path integrals and the phenomenology of the Standard Model. This chapter is especially geared to students wanting an understanding of high-energy physics, where a working knowledge of the Standard Model is a necessity. It is hoped that the student will finish this section with an appreciation of the overwhelming body of experimental evidence pointing to the correctness of the Standard Model.

Because the experiments necessary to go beyond the Standard Model are rapidly becoming prohibitively expensive and time consuming, we are also aware that the development of physics into the next decade may become increasingly theoretical, and therefore we feel that an attempt should be made to explore the various theories that take us beyond the Standard Model.

Part III of this book, therefore, is geared to the students who wish to pursue more advanced material and can be used in one of two ways. A lecturer may want to treat a few of the chapters in Part III at the end of a typical two semester course on quantum field theory. Or, Part III can be used as the basis of a third semester course. We are providing a variety of topics so that the lecturer may pick and choose the chapters that are most topical and are of interest. We have written Part III to leave as much discretion as possible for the lecturer in using this material.

The approach that we have taken in our book differs from that taken in other books in several ways:

First, we have tried to consolidate and streamline, as much as possible in a coherent fashion, a large body of information in one book, beginning with QED, leading to the Standard Model, and ending on supersymmetry.

Second, we have emphasized the role of group theory, treating many of the features of quantum field theory as the byproduct of the Lorentz, Poincaré, and internal symmetry groups. Viewed in this way, many of the rather arbitrary and seemingly contrived conventions of quantum field theory are seen as a consequence of group theory. Group theory, especially in Part III, plays an essential role in understanding unification.

Third, we have presented three distinct proofs of renormalization theory. Most books, if they treat renormalization theory at all, only present one proof. However, because of the importance of renormalization theory to today's research, the serious student may find that a single proof of renormalization is not enough. The student may be ill prepared to handle research when renormalization theory is developed from an entirely different approach. As a consequence, we have presented three different proofs of renormalization theory so that the student can become fluent in at least two different methods. We have presented the original

Dyson/Ward proof in Chapter 7. In Part II, we also present two different proofs based on the BPHZ method and the renormalization group.

Fourth, we should caution the reader that experimental proof of nonperturbative quark confinement or of supersymmetry is absolutely nonexistent. However, since the bulk of current research in theoretical high-energy physics is focused on the material covered in Part III, this section should give the student a brief overview of the main currents in high-energy physics. Moreover, our attitude is to treat nonperturbative field theory and supersymmetries as useful theoretical "laboratories" in which to test many of our notions about quantum field theory. We feel that these techniques in Part III, if viewed as a rich, productive laboratory in which to probe the limits of field theory, will yield great dividends for the serious student.

We have structured the chapters so that they can be adapted in many different ways to suit different needs. In Part I, for example, the heart of the canonical quantization method is presented in Chapters 3–6. These chapters are essential for building a strong foundation to quantum field theory and Feynman diagrams. Although path integral methods today have proven more flexible for gauge theories, a student will have a much better appreciation for the rigor of quantum field theory by reading these chapters. Chapters 2 and 7, however, can be skipped by the student who either already understands the basics of group theory and renormalization, or who does not want to delve that deeply into the intricacies of quantum field theory.

In Part II, the essential material is contained in Chapters 8–11. In these chapters, we develop the necessary material to understand the Standard Model, that is, path integrals, gauge theory, spontaneous symmetry breaking, and phenomenology. This forms the heart of this section, and cannot be omitted. However, Chapters 12–14 should only be read by the student who wants a much more detailed presentation of the subtleties of quantum field theory (BRST, anomalies, renormalization group, etc.).

In Part III, there is great freedom to choose which material to study, depending on the person's interests. We have written Part III to give the greatest flexibility to different approaches in quantum field theory. For those want an understanding of quark confinement and nonperturbative methods, Chapters 15–17 are essential. The student wishing to investigate Grand Unified Theories should study Chapter 18. However, the student who wishes to understand some of the most exciting theoretical developments of the past decade should read Chapters 19–21.

Because of the wide and often confusing range of notations and conventions found in the literature, we have tried to conform, at least in the early chapters, to those appearing in Bjorken and Drell, Itzykson and Zuber, and Cheng and Li. We also choose our metric to be $g_{\mu\nu} = (+, -, -, -)$.

We have also included 311 exercises in this book, which appear after each chapter. We consider solving these exercises essential to an understanding of the

material. Often, students complain that they understand the material but cannot do the problems. We feel that this is a contradiction in terms. If one cannot do the exercises, then one does not really fully understand the material.

In writing this book, we have tried to avoid two extremes. We have tried to avoid giving an overly tedious treatise of renormalization theory and the obscure intricacies of Feynman graphs. One is reminded of being an apprentice during the Middle Ages, where the emphasis was on mastering highly specialized, arcane techniques and tricks, rather than getting a comprehensive understanding of the field.

The other extreme is a shallow approach to theoretical physics, where many vital concepts are deleted because they are considered too difficult for the student. Then the student receives a superficial introduction to the field, creating confusion rather than understanding. Although students may prefer an easier introduction to quantum field theory, ultimately it is the student who suffers. The student will be totally helpless when confronted with research. Even the titles of the high-energy preprints will be incomprehensible.

By taking this intermediate approach, we hope to provide the student with a firm foundation in many of the current areas of research, without overwhelming the student in an avalanche of facts. We will consider the book a success if we have been able to avoid these extremes.

New York M. K.
July 1992

Acknowledgments

I would especially like to thank J. D. Bjorken, who has made countless productive and useful suggestions in every chapter of this book. This book has greatly benefited from his careful reading and critical comments, which have significantly strengthened the presentation in a number of important places.

I would like to thank Dr. Bunji Sakita and Dr. Joseph Birman for their constant support during the writing of this book. I would like to thank A. Das, S. Samuel, H. Chang, D. Karabali, R. Ray, and C. Lee, for reading various chapters of the book and making valuable comments and numerous corrections that have greatly enhanced this book.

I would also like to thank my editor Jeffrey Robbins, who has skillfully guided the passage of the three books that I have written for him. I would also like to thank the National Science Foundation and CUNY–FRAP for partial support.

Contents

I Quantum Fields and Renormalization

1. Why Quantum Field Theory? 3

1.1 Historical Perspective 3
1.2 Strong Interactions 6
1.3 Weak Interactions 8
1.4 Gravitational Interaction 9
1.5 Gauge Revolution 11
1.6 Unification 14
1.7 Action Principle 16
1.8 From First to Second Quantization 21
1.9 Noether's Theorem 23
1.10 Exercises 30

2. Symmetries and Group Theory 33

2.1 Elements of Group Theory 33
2.2 $SO(2)$ 35
2.3 Representations of $SO(2)$ and $U(1)$ 39
2.4 Representations of $SO(3)$ and $SU(2)$ 42
2.5 Representations of $SO(N)$ 45
2.6 Spinors 48
2.7 Lorentz Group 49
2.8 Representations of the Poincaré Group 53
2.9 Master Groups and Supersymmetry 56
2.10 Exercises 58

3. Spin-0 and $\frac{1}{2}$ Fields 61

3.1 Quantization Schemes 61
3.2 Klein–Gordon Scalar Field 63
3.3 Charged Scalar Field 69
3.4 Propagator Theory 72
3.5 Dirac Spinor Field 77
3.6 Quantizing the Spinor Field 86

3.7 Weyl Neutrinos 93
3.8 Exercises 95

4. Quantum Electrodynamics 99
4.1 Maxwell's Equations 99
4.2 Relativistic Quantum Mechanics 102
4.3 Quantizing the Maxwell Field 106
4.4 Gupta–Bleuler Quantization 112
4.5 C, P, and T Invariance 115
4.5.1 Parity 116
4.5.2 Charge Conjugation 117
4.5.3 Time Reversal 119
4.6 CPT Theorem 120
4.7 Exercises 123

5. Feynman Rules and LSZ Reduction 127
5.1 Cross Sections 127
5.2 Propagator Theory and Rutherford Scattering 134
5.3 LSZ Reduction Formulas 141
5.4 Reduction of Dirac Spinors 145
5.5 Time Evolution Operator 147
5.6 Wick's Theorem 151
5.7 Feynman's Rules 156
5.8 Exercises 159

6. Scattering Processes and the S Matrix 163
6.1 Compton Effect 163
6.2 Pair Annihilation 170
6.3 Møller Scattering 173
6.4 Bhabha Scattering 176
6.5 Bremsstrahlung 177
6.6 Radiative Corrections 184
6.7 Anomalous Magnetic Moment 189
6.8 Infrared Divergence 194
6.9 Lamb Shift 196
6.10 Dispersion Relations 199
6.11 Exercises 204

7. Renormalization of QED 209
7.1 The Renormalization Program 209
7.2 Renormalization Types 212
7.2.1 Nonrenormalizable Theories 213
7.2.2 Renormalizable Theories 215

7.2.3 Super-renormalizable Theories 216
7.2.4 Finite Theories 217
7.3 Overview of Renormalization in ϕ^4 Theory 218
7.4 Overview of Renormalization in QED 227
7.5 Types of Regularization 235
7.6 Ward–Takahashi Identities 243
7.7 Overlapping Divergences 247
7.8 Renormalization of QED 250
7.8.1 Step One 250
7.8.2 Step Two 251
7.8.3 Step Three 252
7.8.4 Step Four 254
7.9 Exercises 256

II Gauge Theory and the Standard Model

8. Path Integrals 261
8.1 Postulates of Quantum Mechanics 261
8.1.1 Postulate I 262
8.1.2 Postulate II 262
8.2 Derivation of the Schrödinger Equation 272
8.3 From First to Second Quantization 273
8.4 Generator of Connected Graphs 279
8.5 Loop Expansion 284
8.6 Integration over Grassmann Variables 285
8.7 Schwinger–Dyson Equations 288
8.8 Exercises 291

9. Gauge Theory 295
9.1 Local Symmetry 295
9.2 Faddeev–Popov Gauge Fixing 298
9.3 Feynman Rules for Gauge Theory 304
9.4 Coulomb Gauge 307
9.5 The Gribov Ambiguity 311
9.6 Equivalence of the Coulomb and Landau Gauge 314
9.7 Exercises 318

10. The Weinberg–Salam Model 321
10.1 Broken Symmetry in Nature 321
10.2 The Higgs Mechanism 326
10.3 Weak Interactions 333
10.4 Weinberg–Salam Model 335
10.5 Lepton Decay 338

10.6 R_ξ Gauge 342
10.7 't Hooft Gauge 345
10.8 Coleman–Weinberg Mechanism 348
10.9 Exercises 357

11. The Standard Model 363
11.1 The Quark Model 363
11.2 QCD 374
11.2.1 Spin-Statistics Problem 375
11.2.2 Pair Annihilation 376
11.2.3 Jets 376
11.2.4 Absence of Exotics 377
11.2.5 Pion Decay 378
11.2.6 Asymptotic Freedom 378
11.2.7 Confinement 378
11.2.8 Chiral Symmetry 379
11.2.9 No Anomalies 380
11.3 Jets 380
11.4 Current Algebra 384
11.5 PCAC and the Adler–Weisberger Relation 389
11.5.1 CVC 390
11.5.2 PCAC 391
11.5.3 Adler–Weisberger Relation 393
11.6 Mixing Angle and Decay Processes 396
11.6.1 Purely Leptonic Decays 397
11.6.2 Semileptonic Decays 397
11.6.3 Nonleptonic Decays 398
11.7 GIM Mechanism and Kobayashi–Maskawa Matrix 399
11.8 Exercises 403

12. Ward Identities, BRST, and Anomalies 407
12.1 Ward–Takahashi Identity 407
12.2 Slavnov–Taylor Identities 411
12.3 BRST Quantization 412
12.4 Anomalies 414
12.5 Non-Abelian Anomalies 419
12.6 QCD and Pion Decay into Gamma Rays 420
12.7 Fujikawa's Method 424
12.8 Exercises 429

13. BPHZ Renormalization of Gauge Theories 431
13.1 Counterterms in Gauge Theory 431
13.2 Dimensional Regularization of Gauge Theory 436

13.3 BPHZ Renormalization 441
13.4 Forests and Skeletons 447
13.5 Does Quantum Field Theory Really Exist? 451
13.6 Exercises 456

14. QCD and the Renormalization Group 459
14.1 Deep Inelastic Scattering 459
14.2 Parton Model 463
14.3 Neutrino Sum Rules 467
14.4 Product Expansion at the Light-Cone 470
14.5 Renormalization Group 476
14.6 Asymptotic Freedom 483
14.7 Callan–Symanzik Relation 485
14.8 Minimal Subtraction 488
14.9 Scale Violations 491
14.10 Renormalization Group Proof 494
14.10.1 Step One 496
14.10.2 Step Two 497
14.10.3 Step Three 497
14.11 Exercises 499

III Nonperturbative Methods and Unification

15. Lattice Gauge Theory 505
15.1 The Wilson Lattice 505
15.2 Scalars and Fermions on the Lattice 508
15.3 Confinement 512
15.4 Strong Coupling Approximation 514
15.5 Monte Carlo Simulations 517
15.6 Hamiltonian Formulation 521
15.7 Renormalization Group 523
15.8 Exercises 524

16. Solitons, Monopoles, and Instantons 529
16.1 Solitons 529
16.1.1 Example: ϕ^4 531
16.1.2 Example: Sine–Gordon Equation 533
16.1.3 Example: Nonlinear O(3) Model 536
16.2 Monopole Solutions 539
16.3 't Hooft–Polyakov Monopole 543
16.4 WKB, Tunneling, and Instantons 545
16.5 Yang–Mills Instantons 554
16.6 θ Vacua and the Strong CP Problem 559
16.7 Exercises 566

17. Phase Transitions and Critical Phenomena 571
17.1 Critical Exponents 571
17.2 The Ising Model 575
17.2.1 XYZ Heisenberg Model 580
17.2.2 IRF and Vertex Models 580
17.3 Yang–Baxter Relation 581
17.4 Mean-Field Approximation 584
17.5 Scaling and the Renormalization Group 588
17.5.1 Step One 593
17.5.2 Step Two 596
17.5.3 Step Three 596
17.5.4 Step Four 597
17.6 ϵ Expansion 597
17.7 Exercises 605

18. Grand Unified Theories 609
18.1 Unification and Running Coupling Constants 609
18.2 $SU(5)$ 611
18.3 Anomaly Cancellation 612
18.4 Fermion Representation 613
18.5 Spontaneous Breaking of $SU(5)$ 619
18.6 Hierarchy Problem 622
18.7 $SO(10)$ 622
18.8 Beyond GUT 627
18.8.1 Technicolor 627
18.8.2 Preons or Subquarks 627
18.8.3 Supersymmetry and Superstrings 628
18.9 Exercises 628

19. Quantum Gravity 633
19.1 Equivalence Principle 633
19.2 Generally Covariant Action 638
19.3 Vierbeins and Spinors in General Relativity 640
19.4 GUTs and Cosmology 642
19.5 Inflation 647
19.6 Cosmological Constant Problem 649
19.7 Kaluza–Klein Theory 650
19.8 Generalization to Yang–Mills Theory 652
19.9 Quantizing Gravity 657
19.10 Counterterms in Quantum Gravity 658
19.11 Exercises 660

20. Supersymmetry and Supergravity 663
20.1 Supersymmetry 663
20.2 Supersymmetric Actions 665
20.3 Superspace 669
20.4 Supersymmetric Feynman Rules 680
20.5 Nonrenormalization Theorems 682
20.6 Finite Field Theories 684
20.7 Super Groups 688
20.8 Supergravity 692
20.9 Exercises 696

21. Superstrings 699
21.1 Why Strings? 699
21.2 Points versus Strings 701
21.3 Quantizing the String 705
21.3.1 Gupta–Bleuler Quantization 705
21.3.2 Light-Cone Gauge 709
21.3.3 BRST Quantization 711
21.4 Scattering Amplitudes 712
21.5 Superstrings 717
21.6 Types of Strings 721
21.6.1 Type I 721
21.6.2 Type IIA 722
21.6.3 Type IIB 722
21.6.4 Heterotic String 722
21.7 Higher Loops 723
21.8 Phenomenology 726
21.9 Light-Cone String Field Theory 730
21.10 BRST Action 732
21.11 Exercises 736

Appendix 741
A.1 $SU(N)$ 741
A.2 Tensor Products 743
A.3 $SU(3)$ 747
A.4 Lorentz Group 749
A.5 Dirac Matrices 751
A.6 Infrared Divergences to All Orders 755
A.7 Dimensional Regularization 760

Notes 763
References 775
Index 779

Part I

Quantum Fields and Renormalization

Chapter 1
Why Quantum Field Theory?

Anyone who is not shocked by the quantum theory does not understand it.

—N. Bohr

1.1 Historical Perspective

Quantum field theory has emerged as the most successful physical framework describing the subatomic world. Both its computational power and its conceptual scope are remarkable. Its predictions for the interactions between electrons and photons have proved to be correct to within one part in 10^8. Furthermore, it can adequately explain the interactions of three of the four known fundamental forces in the universe. The success of quantum field theory as a theory of subatomic forces is today embodied in what is called the Standard Model. In fact, at present, there is no known experimental deviation from the Standard Model (excluding gravity).

This impressive list of successes, of course, has not been without its problems. In fact, it has taken several generations of the world's physicists working over many decades to iron out most of quantum field theory's seemingly intractable problems. Even today, there are still several subtle unresolved problems about the nature of quantum field theory.

The undeniable successes of quantum field theory, however, were certainly not apparent in 1927 when P.A.M. Dirac[1] wrote the first pioneering paper combining quantum mechanics with the classical theory of radiation. Dirac's union of nonrelativistic quantum mechanics, which was itself only 2 years old, with the special theory of relativity and electrodynamics would eventually lay the foundation of modern high-energy physics.

Breakthroughs in physics usually emerge when there is a glaring conflict between experiment and theory. Nonrelativistic quantum mechanics grew out of the inability of classical mechanics to explain atomic phenomena, such as black body

radiation and atomic spectra. Similarly, Dirac, in creating quantum field theory, realized that there were large, unresolved problems in classical electrodynamics that might be solved using a relativistic form of quantum mechanics.

In his 1927 paper, Dirac wrote: "...hardly anything has been done up to the present on quantum electrodynamics. The questions of the correct treatment of a system in which the forces are propagated with the velocity of light instead of instantaneously, of the production of an electromagnetic field by a moving electron, and of the reaction of this field on the electron have not yet been touched."

Dirac's keen physical intuition and bold mathematical insight led him in 1928 to postulate the celebrated Dirac electron theory. Developments came rapidly after Dirac coupled the theory of radiation with his relativistic theory of the electron, creating Quantum Electrodynamics (QED). His theory was so elegant and powerful that, when conceptual difficulties appeared, he was not hesitant to postulate seemingly absurd concepts, such as "holes" in an infinite sea of negative energy. As he stated on a number of occasions, it is sometimes more important to have beauty in your equations than to have them fit experiment.

However, as Dirac also firmly realized, the most beautiful theory in the world is useless unless it eventually agrees with experiment. That is why he was gratified when his theory successfully reproduced a series of experimental results: the spin and magnetic moment of the electron and also the correct relativistic corrections to the hydrogen atom's spectra. His revolutionary insight into the structure of matter was vindicated in 1932 with the experimental discovery of antimatter. This graphic confirmation of his prediction helped to erase doubts concerning the correctness of Dirac's theory of the electron.

However, the heady days of the early 1930s, when it seemed like child's play to make major discoveries in quantum field theory with little effort, quickly came to a halt. In some sense, the early successes of the 1930s only masked the deeper problems that plagued the theory. Detailed studies of the higher-order corrections to QED raised more problems than they solved. In fact, a full resolution of these question would have to wait several decades. From the work of Weisskopf, Pauli, Oppenheimer, and many others, it was quickly noticed that QED was horribly plagued by infinities. The early successes of QED were premature: they only represented the lowest-order corrections to classical physics. Higher-order corrections in QED necessarily led to divergent integrals.

The origin of these divergences lay deep within the conceptual foundation of physics. These divergences reflected our ignorance concerning the small-scale structure of space–time. QED contained integrals which diverged as $x \to 0$, or, in momentum space, as $k \to \infty$. Quantum field theory thus inevitably faced divergences emerging from regions of space–time and matter–energy beyond its regime of applicability, that is, infinitely small distances and infinitely large energies.

These divergences had their counterpart in the classical "self-energy" of the electron. Classically, it was known to Lorentz and others near the turn of the century that a complete description of the electron's self-energy was necessarily plagued with infinities. An accelerating electron, for example, would produce a radiation field that would act back on itself, creating absurd physical effects such as the breakdown of causality. Also, other paradoxes abounded; for example it would take an infinite amount of energy to assemble an electron.

Over the decades, many of the world's finest physicists literally brushed these divergent quantities under the rug by manipulating infinite quantities as if they were small. This clever sleight–of–hand was called *renormalization theory*, because these divergent integrals were absorbed into an infinite rescaling of the coupling constants and masses of the theory. Finally, in 1949, Tomonaga, Schwinger, and Feynman[2–4] penetrated this thicket of infinities and demonstrated how to extract meaningful physical information from QED, for which they received the Nobel Prize in 1965.

Ironically, Dirac hated the solution to this problem. To him, the techniques of renormalization seemed so abstruse, so artificial, that he could never reconcile himself with renormalization theory. To the very end, he insisted that one must propose newer, more radical theories that required no renormalization whatsoever.

Nevertheless, the experimental success of renormalization theory could not be denied. Its predictions for the anomalous magnetic moment of the electron, the Lamb shift, etc. have been tested experimentally to one part in 10^8, which is a remarkable degree of confirmation for the theory.

Although renormalized QED enjoyed great success in the 1950s, attempts at generalizing quantum field theory to describe the other forces of nature met with disappointment. Without major modifications, quantum field theory appeared to be incapable of describing all four fundamental forces of nature.[5] These forces are:

1. *The electromagnetic force*, which was successfully described by QED.
2. *The strong force*, which held the nucleus together.
3. *The weak force*, which governed the properties of certain decaying particles, such as the beta decay of the neutron.
4. *The gravitational force*, which was described classically by Einstein's general theory of relativity.

In the 1950s, it became clear that quantum field theory could only give us a description of one of the four forces. It failed to describe the other interactions for very fundamental reasons.

Historically, most of the problems facing the quantum description of these forces can be summarized rather succinctly by the following:

$$\begin{aligned} \alpha_{\text{em}} &\sim 1/137.0359895(61) \\ \alpha_{\text{strong}} &\sim 14 \\ G_{\text{weak}} &\sim 1.02 \times 10^{-5}/M_p^2 \\ &\sim 1.16639(2) \times 10^{-5}\text{GeV}^{-2} \\ G_{\text{Newton}} &\sim 5.9 \times 10^{-39}/M_p^2 \\ &\sim 6.67259(85) \times 10^{-1}\text{m}^3\text{kg}^{-1}\text{s}^{-2} \end{aligned} \tag{1.1}$$

where M_p is the mass of the proton and the parentheses represent the uncertainties.

Several crucial features of the various forces can be seen immediately from this chart. The fact that the coupling constant for QED, the "fine structure constant," is approximately 1/137 meant that physicists could successfully power expand the theory in powers of α_{em}. The power expansion in the fine structure constant, called "perturbation theory," remains the predominant tool in quantum field theory. The smallness of the coupling constant in QED gave physicists confidence that perturbation theory was a reliable approximation to the theory. However, this fortuitous circumstance did not persist for the other interactions.

1.2 Strong Interactions

In contrast to QED, the strongly interacting particles, the "hadrons" (from the Greek word *hadros*, meaning "strong"), have a large coupling constant, meaning that perturbation theory was relatively useless in predicting the spectrum of the strongly interacting particles. Unfortunately, nonperturbative methods were notoriously crude and unreliable. As a consequence, progress in the strong interactions was painfully slow.

In the 1940s, the first seminal breakthrough in the strong interactions was the realization that the force binding the nucleus together could be mediated by the exchange of π mesons:

$$\begin{aligned} \pi^- + p &\leftrightarrow n \\ \pi^+ + n &\leftrightarrow p \end{aligned} \tag{1.2}$$

Theoretical predictions by Yukawa[6] of the mass and range of the π meson, based on the energy scale of the strong interactions, led experimentalists to find the π

meson in their cosmic ray experiments. The π meson was therefore deduced to be the carrier of the nuclear force that bound the nucleus together.

This breakthrough, however, was tempered with the fact that, as we noted, the pion–nucleon coupling constant was much greater than one. Although the Yukawa meson theory as a quantum field theory was known to be renormalizable, perturbation theory was unreliable when applied to the Yukawa theory. Nonperturbative effects, which were exceedingly difficult to calculate, become dominant.

Furthermore, the experimental situation became confusing when so many "resonances" began to be discovered in particle accelerator experiments. This indicated again that the coupling constant of some unknown underlying theory was large, beyond the reach of conventional perturbation theory. Not surprisingly, progress in the strong interactions was slow for many decades for these reasons. With each newly discovered resonance, physicists were reminded of the inadequacy of quantum field theory.

Given the failure of conventional quantum field theory, a number of alternative approaches were investigated in the 1950s and 1960s. Instead of focusing on the "field" of some unknown constituent as the fundamental object (which is in principle unmeasurable), these new approaches centered on the S matrix itself. Borrowing from classical optics, Goldberger and his colleagues[7] assumed the S matrix was an analytic function that satisfied certain dispersion relations. Alternatively, Chew[8] assumed a type of "nuclear democracy"; that is, there were no fundamental particles at all. In this approach, one hoped to calculate the S matrix directly, without using field theory, because of the many stringent physical conditions that it satisfied.

The most successful approach, however, was the $SU(3)$ "quark" theory of the strongly interacting particles (the hadrons). Gell-Mann, Ne'eman, and Zweig,[9–11] building on earlier work of Sakata and his collaborators,[12,13] tried to explain the hadron spectrum with the symmetry group $SU(3)$.

Since quantum field theory was unreliable, physicists focused on the quark model as a strictly phenomenological tool to make sense out of the hundreds of known resonances. Composite combinations of the "up," "down," and "strange" quarks could, in fact, explain all the hadrons discovered up to that time. Together, these three quarks formed a representation of the Lie group $SU(3)$:

$$q_i = \begin{pmatrix} u \\ d \\ s \end{pmatrix} \tag{1.3}$$

The quark model could predict with relative ease the masses and properties of particles that were not yet discovered. A simple picture of the strong interactions was beginning to emerge: Three quarks were necessary to construct a *baryon*, such as a proton or neutron (or the higher resonances, such as the Λ, Ξ, Ω, etc.),

while a quark and an antiquark were necessary to assemble a meson, such as the π meson or the K meson:

$$\text{Hadrons} = \begin{cases} \text{Baryons} & = & q_i q_j q_k \\ \text{Mesons} & = & \bar{q}_i q_j \end{cases} \tag{1.4}$$

Ironically, one problem of the quark model was that it was too successful. The theory was able to make qualitative (and often quantitative) predictions far beyond the range of its applicability. Yet the fractionally charged quarks themselves were never discovered in any scattering experiment. Perhaps they were just a mathematical artifice, reflecting a deeper physical reality that was still unknown. Furthermore, since there was no quantum field theory of quarks, it was unknown what dynamical force held them together. As a consequence, the model was unable to explain why certain bound states of quarks (called "exotics") were not found experimentally.

1.3 Weak Interactions

Equation (1.1), which describes the coupling constants of the four fundamental forces, also reveals why quantum field theory failed to describe the weak interactions. The coupling constant for the weak interactions has the dimensions of inverse mass squared. Later, we will show that theories of this type are nonrenormalizable; that is, theories with coupling constants of negative dimension predict infinite amplitudes for particle scattering.

Historically, the weak interactions were first experimentally observed when strongly interacting particles decayed into lighter particles via a much weaker force, such as the decay of the neutron into a proton, electron, and antineutrino:

$$n \rightarrow p + e^- + \bar{\nu} \tag{1.5}$$

These light particles, such as the electron, its neutrino, the muon μ, etc., were called *leptons*:

$$\text{Leptons} = e^{\pm}, \nu, \mu^{\pm}, \text{etc.} \tag{1.6}$$

Fermi, back in the 1930s, postulated the form of the action that could give a reasonably adequate description of the lowest-order behavior of the weak interactions. However, any attempt to calculate quantum corrections to the Fermi theory floundered because the higher-order terms diverged. The theory was nonrenormalizable because the coupling constant had negative dimension.

Furthermore, it could be shown that the naive Fermi theory violated unitarity at sufficiently large energies.

The mystery of the weak interactions deepened in 1956, when Lee and Yang[14] theorized that parity conservation, long thought to be one of the fundamental principles of physics, was violated in the weak interactions. Their conjecture was soon proved to be correct by the careful experimental work of Wu and also Lederman and Garwin and their colleagues.[15,16]

Furthermore, more and more weakly interacting leptons were discovered over the next few decades. The simple picture of the electron, neutrino, and muon was shattered as the muon neutrino and the tau lepton were found experimentally. Thus, there was the unexplained embarrassment of three exact copies or "generations" of leptons, each generation acting like a Xerox copy of the previous one. (The solution of this problem is still unknown.)

There were some modest proposals that went beyond the Fermi action, which postulated the existence of a massive vector meson or W boson that mediated the weak forces. Buoyed by the success of the Yukawa meson theory, physicists postulated that a massive spin-one vector meson might be the carrier of the weak force. However, the massive vector meson theory, although it was on the right track, had problems because it was also nonrenormalizable. As a result, the massive vector meson theory was considered to be one of several phenomenological possibilities, not a fundamental theory.

1.4 Gravitational Interaction

Ironically, although the gravitational interaction was the first of the four forces to be investigated classically, it was the most difficult one to be quantized.

Using some general physical arguments, one could calculate the mass and spin of the gravitational interaction. Since gravity was a long-range force, it should be massless. Since gravity was always attractive, this meant that its spin must be even. (Spin-one theories, such as electromagnetism, can be both attractive and repulsive.) Since a spin-0 theory was not compatible with the known bending of starlight around the sun, we were left with a spin-two theory. A spin-two theory could also be coupled equally to all matter fields, which was consistent with the equivalence principle. These heuristic arguments indicated that Einstein's theory of general relativity should be the classical approximation to a quantum theory of gravity.

The problem, however, was that quantum gravity, as seen from Eq. (1.1), had a dimensionful coupling constant and hence was nonrenormalizable. This coupling constant, in fact, was Newton's gravitational constant, the first important universal physical constant to be isolated in physics. Ironically, the very success

of Newton's early theory of gravitation, based on the constancy of Newton's constant, proved to be fatal for a quantum theory of gravity.

Another fundamental problem with quantum gravity was that, according to Eq. (1.1), the strength of the interaction was exceedingly weak, and hence very difficult to measure. For example, it takes the entire planet earth to keep pieces of paper resting on a tabletop, but it only takes a charged comb to negate gravity and pick them up. Similarly, if an electron and proton were bound in a hydrogen atom by the gravitational force, the radius of the atom would be roughly the size of the known universe. Although gravitational forces were weaker by comparison to the electromagnetic force by a factor of about 10^{-40}, making it exceedingly difficult to study, one could also show that a quantum theory of gravity had the reverse problem, that its natural energy scale was 10^{19} Gev. Once gravity was quantized, the energy scale at which the gravitational interaction became dominant was set by Newton's constant G_N. To see this, let r be the distance at which the gravitational potential energy of a particle of mass M equals its rest energy, so that $G_N M^2/r = Mc^2$. Let r also be the Compton wavelength of this particle, so that $r = \hbar/Mc$. Eliminating M and solving for r, we find that r equals the Planck length, 10^{-33} cm, or 10^{19} GeV:

$$\begin{aligned} [\hbar G_N c^{-3}]^{1/2} &= 1.61605(10) \times 10^{-33}\,\text{cm} \\ [\hbar c G_N^{-1}]^{1/2} &= 1.221047(79) \times 10^{19}\,\text{GeV/c}^2 \end{aligned} \tag{1.7}$$

This is, of course, beyond the range of our instruments for the foreseeable future. So physicists were faced with the double problem: The classical theory of gravity was so weak that macroscopic experiments were difficult to perform in the laboratory, but the quantum theory of gravity dominated subatomic reactions at the incredible energy scale of 10^{19} GeV, which was far beyond the range of our largest particle accelerators.

Yet another problem arose when one tried to push the theory of gravity to its limits. Phenomenologically, Einstein's general relativity has proved to be an exceptionally reliable tool over cosmological distances. However, when one investigated the singularity at the center of a black hole or the instant of the Big Bang, then the gravitational fields became singular, and the theory broke down. One expected quantum corrections to dominate in those important regions of space–time. However, without a quantum theory of gravity, it was impossible to make any theoretical calculation in those interesting regions of space and time.

In summary, an enormous amount of information is summarized in Eq. (1.1). Some of the fundamental reasons why the development of quantum field theory was stalled in the 1950s are summarized in this chart.

1.5 Gauge Revolution

In the 1950s and 1960s, there was a large mass of experimental data for the strong and weak interactions that was patiently accumulated by many experimental groups. However, most of it could not be explained theoretically. There were significant strides taken experimentally, but progress in theory was, by contrast, painfully slow.

In 1971, however, a dramatic discovery was made by G. 't Hooft,[17] then a graduate student. He reinvestigated an old theory of Yang and Mills, which was a generalization of the Maxwell theory of light, except that the symmetry group was much larger. Building on earlier pioneering work by Veltman, Faddeev, Higgs, and others, 't Hooft showed that Yang–Mills gauge theory, even when its symmetry group was "spontaneously broken," was renormalizable. With this important breakthrough, it now became possible to write down renormalizable theories of the weak interactions, where the W bosons were represented as gauge fields.

Within a matter of months, a flood of important papers came pouring out. An earlier theory of Weinberg and Salam[18,19] of the weak interactions, which was a gauge theory based on the symmetry group $SU(2) \otimes U(1)$, was resurrected and given serious analysis. The essential point, however, was that because gauge theories were now known to be renormalizable, concrete numerical predictions could be made from various gauge theories and then checked to see if they reproduced the experimental data. If the predictions of gauge theory disagreed with the experimental data, then one would have to abandon them, no matter how elegant or aesthetically satisfying they were. Gauge theorists realized that the ultimate judge of any theory was experiment.

Within several years, the agreement between experiment and the Weinberg–Salam theory proved to be overwhelming. The data were sufficiently accurate to rule out several competing models and verify the correctness of the Weinberg–Salam model. The weak interactions went from a state of theoretical confusion to one of relative clarity within a brief period of time. The experimental discovery of the gauge bosons $W^{\pm}$ and Z in 1983 predicted by Weinberg and Salam was another important vindication of the theory.

The Weinberg–Salam model arranged the leptons in a simple manner. It postulated that the (left–handed) leptons could be arranged according to $SU(2)$ doublets in three separate generations:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}; \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}; \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.8}$$

The interactions between these leptons were generated by the intermediate vector bosons:

$$\text{Vector mesons}: \quad W_\mu^\pm, \qquad Z_\mu \tag{1.9}$$

(The remaining problem with the model is to find the Higgs bosons experimentally, which are responsible for spontaneous symmetry breaking, or to determine if they are composite particles.)

In the realm of the strong interactions, progress was also fairly rapid. The gauge revolution made possible Quantum Chromodynamics (QCD), which quickly became the leading candidate for a theory of the strong interactions. By postulating a new "color" $SU(3)$ symmetry, the Yang–Mills theory now provided a glue by which the quarks could be held together. [The $SU(3)$ color symmetry should not be confused with the earlier $SU(3)$ symmetry of Gell–Mann, Ne'eman, and Zweig, which is now called the "flavor" symmetry. Quarks thus have two indices on them; one index $a = u, d, s, c, t, b$ labels the flavor symmetry, while the other index labels the color symmetry.]

The quarks in QCD are represented by:

$$\begin{pmatrix} u^1 & u^2 & u^3 \\ d^1 & d^2 & d^3 \\ s^1 & s^2 & s^3 \\ c^1 & c^2 & c^3 \\ \vdots & \vdots & \vdots \end{pmatrix} \tag{1.10}$$

where the 1, 2, 3 index labels the color symmetry. QCD gave a plausible explanation for the mysterious experimental absence of the quarks. One could calculate that the effective $SU(3)$ color coupling constant became large at low energy, and hence "confined" the quarks permanently into the known hadrons. If this picture was correct, then the gluons condensed into a taffy-like substance that bound the quarks together, creating a string-like object with quarks at either end. If one tried to pull the quarks apart, the condensed gluons would resist their being separated. If one pulled hard enough, then the string might break and another bound quark–antiquark pair would be formed, so that a single quark cannot be isolated (similar to the way that a magnet, when broken, simply forms two smaller magnets, and not single monopoles).

The flip side of this was that one could also prove that the $SU(3)$ color coupling constant became small at large energies. This was called "asymptotic freedom," which was discovered in gauge theories by Gross, Wilczek, Politzer, and 't Hooft.[20–22] At high energies, it could explain the curious fact that the quarks acted as if they were described by a free theory. This was because the effective

coupling constants decreased in size with rising energy, giving the appearance of a free theory. This explained the fact that the quark model worked much better than it was supposed to.

Gradually, a small industry developed around finding nonperturbative solutions to QCD that could explain the confinement of quarks. On one hand, physicists showed that two-dimensional "toy models" could reproduce many of the features that were required of a quantum field theory of quarks, such as confinement and asymptotic freedom. Many of these features followed from the exact solution of these toy models. On the other hand, a compelling description of the four dimensional theory could be achieved through Wilson's lattice gauge theory,[23] which gave qualitative nonperturbative information concerning QCD. In the absence of analytic solutions to QCD, lattice gauge theories today provide the most promising approach to the still-unsolved problem of quark confinement.

Soon, both the electro–weak and QCD models were spliced together to become the Standard Model based on the gauge group $SU(3) \otimes SU(2) \otimes U(1)$. The Standard Model was more than just the sum of its parts. The leptons in the Weinberg–Salam model were shown to possess "anomalies" that threatened renormalizability. Fortunately, these potentially fatal anomalies precisely cancelled against the anomalies coming from the quarks. In other words, the lepton and quark sectors of the Standard Model cured each other's diseases, which was a gratifying theoretical success for the Standard Model. As a result of this and other theoretical and experimental successes, the Standard Model was rapidly recognized to be a first-order approximation to the ultimate theory of particle interactions.

The spectrum of the Standard Model for the left-handed fermions is schematically listed here, consisting of the neutrino ν, the electron e, the "up" and "down" quarks, which come in three "colors," labeled by the index i. This pattern is then repeated for the other two generations (although the top quark has not yet been discovered):

$$\begin{pmatrix} \nu \\ e \end{pmatrix} \begin{pmatrix} u^i \\ d^i \end{pmatrix}; \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} c^i \\ s^i \end{pmatrix}; \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \begin{pmatrix} t^i \\ b^i \end{pmatrix} \tag{1.11}$$

In the Standard Model, the forces between the leptons and quarks were mediated by the massive vector mesons for the weak interactions and the massless gluons for the strong interactions:

$$\begin{cases} \text{Massive vector mesons}: & W^{\pm}, Z \\ \text{Massless gluons}: & A^a_\mu \end{cases} \tag{1.12}$$

The weaknesses of the Standard Model, however, were also readily apparent. No one saw the theory as a fundamental theory of matter and energy. Containing at

least 19 arbitrary parameters in the theory, it was a far cry from the original dream of physicists: a single unified theory with at most *one* undetermined coupling constant.

1.6 Unification

In contrast to the 1950s, when physicists were flooded with experimental data without a theoretical framework to understand them, the situation in the 1990s may be the reverse; that is, the experimental data are all consistent with the Standard Model. As a consequence, without important clues coming from experiment, physicists have proposed theories beyond the Standard Model that cannot be tested with the current level of technology. In fact, even the next generation of particle accelerators may not be powerful enough to rule out many of the theoretical models currently being studied. In other words, while experiment led theory in the 1950s, in the 1990s theory may lead experiment.

At present, attempts to use quantum field theory to push beyond the Standard Model have met with modest successes. Unfortunately, the experimental data at very large energies are still absent, and this situation may persist into the near future. However, enormous theoretical strides have been made that give us some confidence that we are on the right track.

The next plausible step beyond the Standard Model may be the GUTs (Grand Unified Theories), which are based on gauging a single Lie group, such as $SU(5)$ or $SO(10)$ (Fig. 1.1).

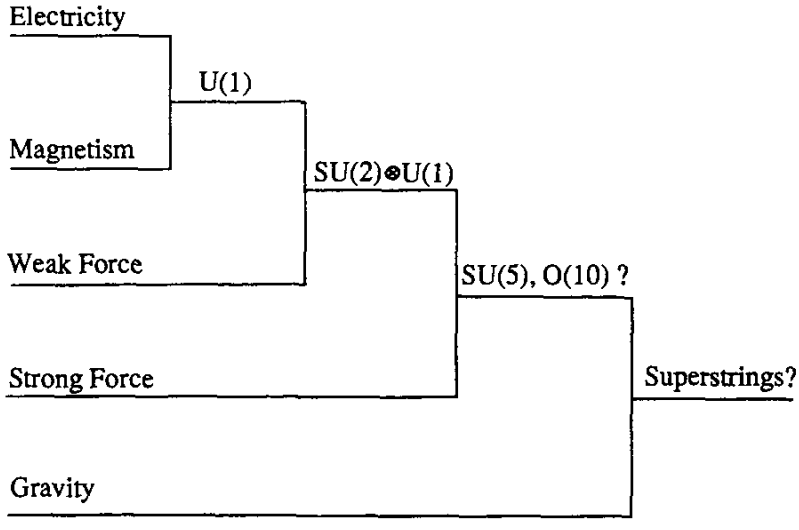


Figure 1.1. This chart shows how the various forces of nature, once thought to be fundamentally distinct, have been unified over the past century, giving us the possibility of unifying all known forces via quantum field theory.

According to GUT theory, the energy scale at which the unification of all three particle forces takes place is enormously large, about 10^{15} GeV, just below the Planck energy. Near the instant of the Big Bang, where such energies were found, the theory predicts that all three particle forces were unified by one GUT symmetry. In this picture, as the universe rapidly cooled down, the original GUT symmetry was broken down successively into the present–day symmetries of the Standard Model. A typical breakdown scheme might be:

$$O(10) \rightarrow SU(5) \rightarrow SU(3) \otimes SU(2) \otimes U(1) \rightarrow SU(3) \otimes U(1) \qquad (1.13)$$

Because quarks and electrons are now placed in the same multiplet, it also means that quarks can decay into leptons, giving us the violation of baryon number and the decay of the proton. So far, elaborate experimental searches for proton decay have been disappointing. The experimental data, however, are good enough to rule out minimal $SU(5)$; there are, however, more complicated GUT theories that can accomodate longer proton lifetimes.

Although GUT theories are a vast improvement over the Standard Model, they are also beset with problems. First, they cannot explain the three generations of particles that have been discovered. Instead, we must have three identical GUT theories, one for each generation. Second, it still has many arbitrary parameters (e.g., Higgs parameters) that cannot be explained from any simpler principle. Third, the unification scale takes place at energies near the Planck scale, where we expect gravitational effects to become large, yet gravity is not included in the theory. Fourth, it postulates that there is a barren "desert" that extends for twelve orders of magnitude, from the GUT scale down to the electro-weak scale. (The critics claim that there is no precedent in physics for such an extrapolation over such a large range of energy.) Fifth, there is the "hierarchy problem," meaning that radiative corrections will mix the two energy scales together, ruining the entire program.

To solve the last problem, physicists have studied quantum field theories that can incorporate "supersymmetry," a new symmetry that puts fermions and bosons into the same multiplet. Although not a single shred of experimental data supports the existence of supersymmetry, it has opened the door to an entirely new kind of quantum field theory that has remarkable renormalization properties and is still fully compatible with its basic principles.

In a supersymmetric theory, once we set the energy scale of unification and the energy scale of the low-energy interactions, we never have to "retune" the parameters again. Renormalization effects do not mix the two energy scales. However, one of the most intriguing properties of supersymmetry is that, once it is made into a local symmetry, it necessarily incorporates quantum gravity into the spectrum. Quantum gravity, instead of being an unpleasant, undesirable feature

of quantum field theory, is necessarily an integral part of any theory with local supersymmetry.

Historically, it was once thought that all successful quantum field theories required renormalization. However, today supersymmetry gives us theories, like the $SO(4)$ super Yang–Mills theory and the superstring theory, which are *finite to all orders in perturbation theory*, a truly remarkable achievement. For the first time since its inception, it is now possible to write down quantum field theories that require no renormalization whatsoever. This answers, in some sense, Dirac's original criticism of his own creation.

Only time will tell if GUTs, supersymmetry, and the superstring theory can give us a faithful description of the universe. In this book, our attitude is that they are an exciting theoretical laboratory in which to probe the limits of quantum field theory. And the fact that supersymmetric theories can improve and in some cases solve the problem of ultraviolet divergences without renormalization is, by itself, a feature of quantum field theory worthy of study.

Let us, therefore, now leave the historical setting of quantum field theory and begin a discussion of how quantum field theory gives us a quantum description of point particle systems with an infinite number of degrees of freedom. Although the student may already be familiar with the foundations of classical mechanics and the transition to the quantum theory, it will prove beneficial to review this material from a slightly different point of view, that is, systems with an infinite number of degrees of freedom. This will then set the stage for quantum field theory.

1.7 Action Principle

Before we begin our discussion of field theory, for notational purposes it is customary to choose units so that:

$$\begin{aligned} c &= 1 \\ \hbar &= 1 \end{aligned} \tag{1.14}$$

(We can always do this because the definition of c and $\hbar = h/2\pi$ depends on certain conventions that grew historically in our understanding of nature. Imposing $c = 1$, for example, means that seconds and centimeters are to be treated on the same footing, such that exactly 299,792,458 meters is equivalent to 1 sec. Thus, the second and the centimeter are to be treated as if they were expressions of the same unit. Likewise, setting $\hbar = 1$ means that the erg $\times$ sec. is now dimensionless, so the erg and the second are inverses of each other. This also means that the gram

is inversely related to the centimeter. With these conventions, the only unit that survives is the centimeter, or equivalently, the gram.)

In classical mechanics, there are two equivalent formulations, one based on Newton's equations of motion, and the other based on the action principle. At first, these two formalisms seem to have little in common. The first depends on iterating infinitesimal changes sequentially along a particle's path. The second depends on evaluating all possible paths between two points and selecting out the one with the minimum action. One of the great achievements of classical mechanics was the demonstration that the Newtonian equations of motion were equivalent to minimizing the action over the set of all paths:

$$\text{Equation of motion} \leftrightarrow \text{Action principle} \tag{1.15}$$

However, when we generalize our results to the quantum realm, this equivalence breaks down. The Heisenberg Uncertainty Principle forces us to introduce probabilities and consider all possible paths taken by the particle, with the classical path simply being the most likely. Quantum mechanically, we know that there is a finite probability for a particle to deviate from its classical equation of motion. This deviation from the classical path is very small, on the scale determined by Planck's constant. However, on the subatomic scale, this deviation becomes the dominant aspect of a particle's motion. In the microcosm, motions that are in fact forbidden classically determine the primary characteristics of the atom. The stability of the atom, the emission and absorption spectrum, radioactive decay, tunneling, etc. are all manifestations of quantum behavior that deviate from Newton's classical equations of motion.

However, even though Newtonian mechanics fails within the subatomic realm, it is possible to generalize the action principle to incorporate these quantum probabilities. The action principle then becomes the only framework to calculate the probability that a particle will deviate from its classical path. In other words, the action principle is elevated into one of the foundations of the new mechanics.

To see how this takes place, let us first begin by describing the simplest of all possible classical systems, the nonrelativistic point particle. In three dimensions, we say that this particle has three degrees of freedom, each labeled by coordinates $q^i(t)$. The motion of the particle is determined by the Lagrangian $L(q^i, \dot{q}^i)$, which is a function of both the position and the velocity of the particle:

$$L = \frac{1}{2}m(\dot{q}^i)^2 - V(q) \tag{1.16}$$

where $V(q)$ is some potential in which the particle moves. Classically, the motion of the particle is determined by minimizing the action, which is the integral of the

Lagrangian:

$$S = \int_{t_1}^{t_2} L(q^i, \dot{q}^i)\, dt \tag{1.17}$$

We can derive the classical equations of motion by minimizing the action; that is, the classical Newtonian path taken by the particle is the one in which the action is a minimum:

$$\delta S = 0 \tag{1.18}$$

To calculate the classical equations of motion, we make a small variation in the path of the particle $\delta q^i(t)$, keeping the endpoints fixed; that is, $\delta q^i(t_1) = \delta q^i(t_2) = 0$. To calculate δS, we must vary the Lagrangian with respect to both changes in the position and the velocity:

$$\delta S = \int_{t_1}^{t_2} dt \left(\frac{\delta L}{\delta q^i} \delta q^i + \frac{\delta L}{\delta \dot{q}^i} \delta \dot{q}^i \right) = 0 \tag{1.19}$$

We can integrate the last expression by parts:

$$\delta S = \int dt \left\{ \delta q^i \left[\frac{\delta L}{\delta q^i} - \frac{d}{dt} \frac{\delta L}{\delta \dot{q}^i} \right] + \frac{d}{dt} \left(\delta q^i \frac{\delta L}{\delta \dot{q}^i} \right) \right\} = 0 \tag{1.20}$$

Since the variation vanishes at the endpoints of the integration, we can drop the last term. Then the action is minimized if we demand that:

$$\frac{\delta L}{\delta q^i} = \frac{d}{dt} \frac{\delta L}{\delta \dot{q}^i} \tag{1.21}$$

which are called the Euler–Lagrange equations.

If we now insert the value of the Lagrangian into the Euler–Lagrange equations, we find the classical equation of motion:

$$m \frac{d^2 q^i}{dt^2} = -\frac{\partial V(q)}{\partial q^i} \tag{1.22}$$

which forms the basis of Newtonian mechanics.

Classically, we also know that there are two different formulations of Newtonian mechanics, the Lagrangian formulation, where the position q^i and velocity $\dot{q}^i$ of a point particle are the fundamental variables, and the Hamiltonian formulation, where we choose the position q^i and the momentum p^i to be the independent variables.

To make the transition from the Lagrangian to the Hamiltonian formulation of classical mechanics, we first define the momentum as follows:

$$p^i = \frac{\delta L}{\delta \dot{q}^i} \tag{1.23}$$

For our choice of the Lagrangian, we find that $p^i = m\dot{q}^i$. The definition of the Hamiltonian is then given by:

$$H(q^i, p^i) = p^i \dot{q}^i - L(q^i, \dot{q}^i) \tag{1.24}$$

With our choice of Lagrangian, we find that the Hamiltonian is given by:

$$H(q^i, p^i) = \frac{p^{i2}}{2m} + V(q) \tag{1.25}$$

In the transition from the Lagrangian to the Hamiltonian system, we have exchanged independent variables from q^i, $\dot{q}^i$ to q^i, p^i. To prove that $H(q^i, p^i)$ is not a function of $\dot{q}^i$, we can make the following variation:

$$\begin{aligned} \delta H &= p^i \delta \dot{q}^i + \delta p^i \dot{q}^i - \frac{\delta L}{\delta q^i}\delta q^i - \frac{\delta L}{\delta \dot{q}^i}\delta \dot{q}^i \\ &= \dot{q}^i \delta p^i - \frac{\delta L}{\delta q^i}\delta q^i \\ &= \delta p^i \frac{\delta H}{\delta p^i} + \delta q^i \frac{\delta H}{\delta q^i} \end{aligned} \tag{1.26}$$

where we have used the definition of p^i to eliminate the dependence of the Hamiltonian on $\dot{q}^i$ and have used the chain rule for the last step. By equating the coefficients of the variations, we can make the following identification:

$$\dot{q}^i = \frac{\delta H}{\delta p^i}; \quad -\dot{p}^i = \frac{\delta H}{\delta q^i} \tag{1.27}$$

where we have used the equations of motion.

In the Hamiltonian formalism, we can also calculate the time variation of any field F in terms of the Hamiltonian:

$$\begin{aligned} \dot{F} &= \frac{\partial F}{\partial t} + \frac{\delta F}{\delta q^i}\dot{q}^i + \frac{\delta F}{\delta p^i}\dot{p}^i \\ &= \frac{\partial F}{\partial t} + \frac{\delta F}{\delta q^i}\frac{\delta H}{\delta p^i} - \frac{\delta F}{\delta p^i}\frac{\delta H}{\delta q^i} \end{aligned} \tag{1.28}$$

If we define the Poisson bracket as:

$$\{A, B\}_{\text{PB}} = \sum_i \left(\frac{\partial A}{\partial p^i} \frac{\partial B}{\partial q^i} - \frac{\partial A}{\partial q^i} \frac{\partial B}{\partial p^i} \right) \tag{1.29}$$

then we can write the time variation of the field F as:

$$\dot{F} = \frac{\partial F}{\partial t} + \{H, F\}_{\text{PB}} \tag{1.30}$$

At this point, we have derived the Newtonian equations of motion by minimizing the action, reproducing the classical result. This is not new. However, we will now make the transition to the quantum theory, which treats the action as the fundamental object, incorporating both allowed and forbidden paths. The Newtonian equations of motion then specify the most likely path, but certainly not the only path. In the subatomic realm, in fact, the classically forbidden paths may dominate over the classical one.

There are many ways in which to make the transition from classical mechanics to quantum mechanics. (Perhaps the most profound and powerful is the *path integral* method, which we present in Chapter 8.) Historically, however, it was Dirac who noticed that the transition from classical to quantum mechanics can be achieved by replacing the classical Poisson brackets with commutators:

$$\{A, B\}_{\text{PB}} \rightarrow \frac{i}{\hbar}[A, B] \tag{1.31}$$

With this replacement, the Poisson brackets between canonical coordinates are replaced by:

$$[p^i, q^k] = -i\hbar\delta^{ik} \tag{1.32}$$

Quantum mechanics makes the replacement:

$$p^i = -i\hbar \frac{\partial}{\partial q^i}; \quad E = i\hbar \frac{\partial}{\partial t} \tag{1.33}$$

Because p^i is now an operator, the Hamiltonian also becomes an operator, and we can now satisfy Hamilton's equations by demanding that they vanish when acting on some function $\psi(q^i, t)$:

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial q^{i2}} + V(q) \right) \psi = i\hbar \frac{\partial \psi}{\partial t} \tag{1.34}$$

This is the Schrödinger wave equation, which is the starting point for calculating the spectral lines of the hydrogen atom.

1.8 From First to Second Quantization

This process, treating the coordinates q^i and p^i as quantized variables, is called *first quantization*. However, the object of this book is to make the transition to *second quantization*, where we quantize *fields* which have an infinite number of degrees of freedom. The transition from quantum mechanics to quantum field theory arises when we make the transition from three degrees of freedom, described by x^i, to an infinite number of degrees of freedom, described by a field $\phi(x^i)$ and then apply quantization relations directly onto the fields.

Before we made the transition to quantum field theory in Chapter 3, let us discuss how we describe classical field theory, that is, classical systems with an infinite number of degrees of freedom. Let us consider a classical system, a series of masses arranged in a line, each connected to the next via springs (Fig. 1.2).

The Lagrangian of the system then contains two terms, the potential energy of each spring as well as the kinetic energy of the masses. The Lagrangian is therefore given by:

$$L = \sum_{r=1}^{N} \left[\frac{1}{2} m_r (\dot{x}_r)^2 - \frac{1}{2} k (x_r - x_{r+1})^2 \right] \tag{1.35}$$

Now let us assume that we have an infinite number of masses, each separated by a distance ϵ. In the limit $\epsilon \to 0$, the Lagrangian becomes:

$$\begin{aligned} L &= \sum_{r=1}^{N} \frac{\epsilon}{2} \left(\frac{m}{\epsilon} \dot{x}_r^2 - k \frac{(x_r - x_{r+1})^2}{\epsilon} \right) \\ &\to \int dx \, \frac{1}{2} \left[\mu \, \dot{\phi}(x)^2 - Y \left(\frac{\partial \phi(x)}{\partial x} \right)^2 \right] \end{aligned} \tag{1.36}$$

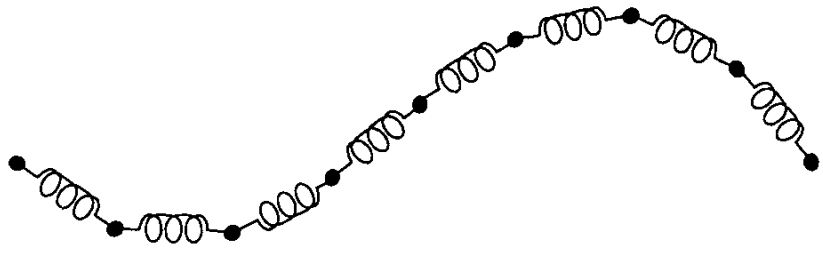


Figure 1.2. The action describing a finite chain of springs, in the limit of an infinite number of springs, becomes a theory with an infinite number of degrees of freedom.

where we have taken the limit via:

$$\begin{aligned} \epsilon &\rightarrow dx \\ \frac{m}{\epsilon} &\rightarrow \mu \\ k\epsilon &\rightarrow Y \\ x_r &\rightarrow \phi(x,t) \end{aligned} \tag{1.37}$$

where $\phi(x, t)$ is the displacement of the particle located at position x and time t, where μ is the mass density along the springs, and Y is the Young's modulus.

If we use the Euler–Lagrange equations to find the equations of motion, we arrive at:

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{\mu}{Y}\frac{\partial^2 \phi}{\partial t^2} = 0 \tag{1.38}$$

which is just the familiar wave equation for a one-dimensional system traveling with velocity $\sqrt{Y/\mu}$.

On one hand, we have proved the rather intuitively obvious statement that waves can propagate down a long, massive spring. On the other hand, we have made the highly nontrivial transition from a system with a finite number of degrees of freedom to one with an infinite number of degrees of freedom.

Now let us generalize our previous discussion of the Euler–Lagrange equations to include classical field theories with an infinite number of degrees of freedom. We begin with a Lagrangian that is a function of both the field $\phi(x)$ as well as its space–time derivatives $\partial_\mu \phi(x)$:

$$L\left(\phi(x), \partial_\mu \phi(x)\right) \tag{1.39}$$

where:

$$\partial_\mu = \left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x^i}\right) \tag{1.40}$$

The action is given by a four dimensional integral over a Lagrangian density $\mathscr{L}$:

$$\begin{aligned} L &= \int d^3x\, \mathscr{L}(\phi, \partial_\mu \varphi) \\ S &= \int d^4x\, \mathscr{L} = \int dt\, L \end{aligned} \tag{1.41}$$

integrated between initial and final times.

As before, we can retrieve the classical equations of motion by minimizing the action:

$$\delta S = 0 = \int d^4x \left(\frac{\delta \mathscr{L}}{\delta \phi} \delta \phi + \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} \delta \partial_\mu \phi \right) \tag{1.42}$$

We can integrate by parts, reversing the direction of the derivative:

$$\delta S = \int d^4x \left[\left(\frac{\delta \mathscr{L}}{\delta \phi} - \partial_\mu \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} \right) \delta \phi + \partial_\mu \left(\frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} \delta \phi \right) \right] \tag{1.43}$$

The last term vanishes at the endpoint of the integration; so we arrive at the Euler–Lagrange equations of motion:

$$\partial_\mu \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} = \frac{\delta \mathscr{L}}{\delta \phi} \tag{1.44}$$

The simplest example is given by the scalar field $\phi(x)$ of a point particle:

$$\mathscr{L} = \frac{1}{2} \left(\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2 \right) \tag{1.45}$$

Inserting this into the equation of motion, we find the standard Klein–Gordon equation:

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0 \tag{1.46}$$

where:

$$\partial^\mu = \left(\frac{\partial}{\partial t}, -\frac{\partial}{\partial x^i} \right) \tag{1.47}$$

One of the purposes of this book is to generalize the Klein–Gordon equation by introducing higher spins and higher interactions. To do this, however, we must first begin with a discussion of special relativity. It will turn out that invariance under the Lorentz and Poincaré group will provide the most important guide in constructing higher and more sophisticated versions of quantum field theory.

1.9 Noether's Theorem

One of the achievements of this formalism is that we can use the symmetries of the action to derive conservation principles. For example, in classical mechanics,

if the Hamiltonian is time independent, then energy is conserved. Likewise, if the Hamiltonian is translation invariant in three dimensions, then momentum is conserved:

$$\begin{aligned} \text{Time independence} &\rightarrow \text{Energy conservation} \\ \text{Translation independence} &\rightarrow \text{Momentum conservation} \\ \text{Rotational independence} &\rightarrow \text{Angular momentum conservation} \end{aligned} \tag{1.48}$$

The precise mathematical formulation of this correspondence is given by Noether's theorem. In general, an action may be invariant under either an internal, isospin symmetry transformation of the fields, or under some space–time symmetry. We will first discuss the isospin symmetry, where the fields ϕ^α vary according to some small parameter $\delta\epsilon^\alpha$.

The action varies as:

$$\begin{aligned} \delta S &= \int d^4x \left(\frac{\delta \mathscr{L}}{\delta \phi^\alpha} \delta\phi^\alpha + \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi^\alpha} \delta \partial_\mu \phi^\alpha \right) \\ &= \int d^4x \left(\partial_\mu \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi^\alpha} \delta\phi^\alpha + \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi^\alpha} \partial_\mu \delta\phi^\alpha \right) \\ &= \int d^4x \, \partial_\mu \left(\frac{\delta \mathscr{L}}{\delta \partial_\mu \phi^\alpha} \delta\phi^\alpha \right) \end{aligned} \tag{1.49}$$

where we have used the equations of motion and have converted the variation of the action into the integral of a total derivative. This defines the *current* $J^{\mu\alpha}$:

$$J^\mu_\alpha = \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi^\beta} \frac{\delta \phi^\beta}{\delta \epsilon^\alpha} \tag{1.50}$$

If the action is invariant under this transformation, then we have established that the current is conserved:

$$\partial_\mu J^\mu_\alpha = 0 \tag{1.51}$$

From this conserved current, we can also establish a conserved charge, given by the integral over the fourth component of the current:

$$Q_\alpha \equiv \int d^3x \, J^0_\alpha \tag{1.52}$$

Now let us integrate the conservation equation:

$$\begin{aligned} 0 &= \int d^3x\, \partial_\mu J^\mu_\alpha = \int d^3x\, \partial_0 J^0_\alpha + \int d^3x\, \partial_i J^i_\alpha \\ &= \frac{d}{dt}\int d^3x\, J^0_\alpha + \int dS_i\, J^i_\alpha = \frac{d}{dt} Q_\alpha + \text{surface term} \end{aligned} \tag{1.53}$$

Let us assume that the fields appearing in the surface term vanish sufficiently rapidly at infinity so that the last term can be neglected. Then:

$$\partial_\mu J^\mu_\alpha = 0 \rightarrow \frac{dQ_\alpha(t)}{dt} = 0 \tag{1.54}$$

In summary, the symmetry of the action implies the conservation of a current J^μ_α, which in turn implies a conservation principle:

$$\text{Symmetry} \rightarrow \text{Current conservation} \rightarrow \text{Conservation principle} \tag{1.55}$$

Now let us investigate the second case, when the action is invariant under the space–time symmetry of the Lorentz and Poincaré groups. Lorentz symmetry implies that we can combine familiar three-vectors, such as momentum and space, into four-vectors. We introduce the space–time coordinate x^μ as follows:

$$x^\mu \equiv (x^0, x^i) = (t, \mathbf{x}) \tag{1.56}$$

where the time coordinate is defined as $x^0 = t$.

Similarly, the momentum three-vector $\mathbf{p}$ can be combined with energy to form the four-vector:

$$p^\mu = \left(p^0, p^i\right) = (E, \mathbf{p}) \tag{1.57}$$

We will henceforth use Greek symbols from the middle of the alphabet μ, ν to represent four-vectors, and Roman indices from the middle of the alphabet i, j, k to represent space coordinates.

We will raise and lower indices by using the following *metric* $g_{\mu\nu}$ as follows:

$$A_\mu = g_{\mu\nu} A^\nu \tag{1.58}$$

where:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{1.59}$$

Now let us use this formalism to construct the current associated with making a translation:

$$x^\mu \rightarrow x^\mu + a^\mu \tag{1.60}$$

where a^μ is a constant. a^0 represents time displacements, and a^i represents space displacements. We will now rederive the result from classical mechanics that displacement in time (space) leads to the conservation of energy (momentum).

Under this displacement, a field $\phi(x)$ transforms as $\phi(x) \rightarrow \phi(x+a)$. For small a^μ, we can power expand the field in a power series in a^μ. Then the change in the field after a displacement is given by:

$$\delta\phi = \phi(x+a) - \phi(x) \sim \phi(x) + a^\mu \partial_\mu \phi(x) - \phi(x) = a^\mu \partial_\mu \phi(x) \tag{1.61}$$

Therefore, if we make a translation $\delta x^\mu = a^\mu$, then the fields transform, after making a Taylor expansion, as follows:

$$\begin{aligned} \delta\phi &= a^\mu \partial_\mu \phi \\ \delta\partial_\mu\phi &= a^\nu \partial_\nu \partial_\mu \phi \end{aligned} \tag{1.62}$$

The calculation of the current associated with translations proceeds a bit differently from the isospin case. There is an additional term that one must take into account, which is the variation of the Lagrangian itself under the space–time symmetry. The variation of our Lagrangian is given by:

$$\delta\mathscr{L} = a^\mu \partial_\mu \mathscr{L} = \frac{\delta\mathscr{L}}{\delta\phi}\delta\phi + \frac{\delta\mathscr{L}}{\delta\partial_\mu\phi}\delta\partial_\mu\phi \tag{1.63}$$

Substituting in the variation of the fields and using the equations of motion, we find:

$$\begin{aligned} \delta\mathscr{L} &= a^\mu \partial_\mu \mathscr{L} \\ &= \frac{\delta\mathscr{L}}{\delta\phi} a^\nu \partial_\nu \phi + \frac{\delta\mathscr{L}}{\delta\partial_\mu\phi} a^\nu \partial_\nu \partial_\mu \phi \\ &= a^\nu \left(\partial_\mu \frac{\delta\mathscr{L}}{\delta\partial_\mu\phi} \partial_\nu \phi + \frac{\delta\mathscr{L}}{\delta\partial_\mu\phi} \partial_\mu \partial_\nu \phi \right) \\ &= a^\nu \partial_\mu \left(\frac{\delta\mathscr{L}}{\delta\partial_\mu\phi} \partial_\nu \phi \right) \end{aligned} \tag{1.64}$$

Combining both terms under one derivative, we find:

$$\partial_\mu \left[\mathscr{L}\delta^\mu_\nu - \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} \partial_\nu \phi \right] a^\nu = 0 \tag{1.65}$$

This then defines the *energy–momentum tensor* $T^\mu{}_\nu$:

$$T^\mu{}_\nu = \frac{\delta \mathscr{L}}{\delta \partial_\mu \phi} \partial_\nu \phi - \delta^\mu_\nu \mathscr{L} \tag{1.66}$$

which is conserved:

$$\partial_\mu T^\mu{}_\nu = 0 \tag{1.67}$$

If we substitute the Klein–Gordon action into this expression, we find the energy–momentum tensor for the scalar particle:

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \mathscr{L} \tag{1.68}$$

Using the equations of motion, we can explicitly check that this energy–momentum tensor is conserved.

By integrating the energy–momentum tensor, we can generate conserved currents, as we saw earlier. As the name implies, the conserved charges corresponding to the energy–momentum tensor give us energy and momentum conservation. Let us define:

$$P^\mu = (E, P^i) \tag{1.69}$$

which combines the energy $P^0 = E$ and the momentum P^i into a single four-vector. We can show that energy and momentum are both conserved by making the following definition:

$$\begin{aligned} P^\mu &\equiv \int d^3x (T^\mu_0) \\ \frac{d}{dt} P^\mu &= 0 \end{aligned} \tag{1.70}$$

The conservation of energy–momentum is therefore a consequence of the invariance of the action under translations, which in turn corresponds to invariance under time and space displacements. Thus, we now have derived the result from classical Newtonian mechanics mentioned earlier in Eq. (1.48).

Next, let us generalize this discussion. We know from classical mechanics that invariance of the action under rotations generates the conservation of angular

momentum. Now, we would like to derive the Lorentz generalization of this result from classical mechanics.

Rotations in three dimensions are described by:

$$\delta x^i = a^{ij} x^i \tag{1.71}$$

where a^{ij} is an anti-symmetric matrix describing the rotation (i.e., $a^{ij} = -a^{ji}$).

Let us now construct the generalization of this rotation, the current associated with Lorentz transformations. We define how a four–vector x^μ changes under a Lorentz transformation:

$$\text{Lorentz transformation}: \quad \begin{cases} \delta x^\mu = \epsilon^\mu{}_\nu x^\mu \\ \delta\phi(x) = \epsilon^\mu{}_\nu x^\nu \partial_\mu \phi(x) \end{cases} \tag{1.72}$$

where $\epsilon^\mu{}_\nu$ is an infinitesimal, antisymmetric constant matrix (i.e., $\epsilon^{\mu\nu} = -\epsilon^{\nu\mu}$).

Repeating the same steps with this new variation, we have:

$$\begin{aligned} \delta\mathscr{L} &= \epsilon^\mu{}_\nu x^\nu \partial_\mu \mathscr{L} \\ &= \frac{\delta\mathscr{L}}{\delta\phi}\delta\phi + \frac{\delta\mathscr{L}}{\delta\partial_\rho\phi}\delta\partial_\rho\phi \\ &= \partial_\rho\left(\frac{\delta\mathscr{L}}{\delta\partial_\rho\phi}\delta\phi\right) \\ &= \partial_\rho\left(\frac{\delta\mathscr{L}}{\delta\partial_\rho\phi}\epsilon^\mu{}_\nu x^\nu \partial_\mu\phi\right) \end{aligned} \tag{1.73}$$

If we extract the coefficient of $\epsilon^\mu{}_\nu$ and put everything within the partial derivative ∂_ρ, we find:

$$\epsilon_{\mu\nu}\partial_\rho\left(\frac{\delta\mathscr{L}}{\delta\partial_\rho\phi}(\partial^\mu\phi x^\nu - \partial^\nu\phi x^\mu) - \delta^{\rho\mu}x^\nu\mathscr{L} + \delta^{\rho\nu}x^\mu\mathscr{L}\right) = 0 \tag{1.74}$$

This gives us the conserved current:

$$\begin{aligned} \mathscr{M}^{\rho,\mu\nu} &= T^{\rho\nu}x^\mu - T^{\rho\mu}x^\nu \\ \partial_\rho\mathscr{M}^{\rho,\mu\nu} &= 0 \end{aligned} \tag{1.75}$$

and the conserved charge:

$$\begin{aligned} M^{\mu\nu} &= \int d^3x\, \mathscr{M}^{0,\mu\nu} \\ \frac{d}{dt}M^{\mu\nu} &= 0 \end{aligned} \tag{1.76}$$

If we restrict our discussion to rotations in three dimensional space, then $\dot{M}^{ij} = 0$ corresponds to the conservation of angular momentum. If we take all the components of this matrix, however, we find that Lorentz transformations are an invariant of the action.

There is, however, a certain ambiguity in the definition of the energy–momentum tensor. The energy–momentum tensor is not a measurable quantity, but the integrated charges correspond to the physical energy and momentum, and hence are measurable.

We can exploit this ambiguity and add to the energy–momentum tensor a term:

$$\partial_\lambda E^{\lambda\mu\nu} \tag{1.77}$$

where $E^{\lambda\mu\nu}$ is antisymmetric in the first two indices:

$$E^{\lambda\mu\nu} = -E^{\mu\lambda\nu} \tag{1.78}$$

Because of this antisymmetry, this tensor satisfies trivially:

$$\partial_\lambda \partial_\mu E^{\lambda\mu\nu} = 0 \tag{1.79}$$

So we can make the replacement:

$$T^{\mu\nu} \rightarrow T^{\mu\nu} + \partial_\lambda E^{\lambda\mu\nu} \tag{1.80}$$

This new energy–momentum tensor is conserved, like the previous one. We can choose this tensor such that the new energy–momentum tensor is symmetric in μ and ν.

The addition of this extra tensor to the energy–momentum tensor does not affect the energy and the momentum, which are measurable quantities. If we take the integrated charge, we find that the contribution from $E^{\lambda\mu\nu}$ vanishes:

$$\begin{aligned} P^\mu &\rightarrow P^\mu + \int d^3x \partial_\lambda E^{\lambda 0\mu} \\ &= P^\mu + \int \partial_i E^{i0\mu} d^3x \\ &= P^\mu + \int_S E^{i0\mu} dS_i \\ &= P^\mu \end{aligned} \tag{1.81}$$

Thus, the physical energy and momentum are not affected as long as this tensor vanishes sufficiently rapidly at infinity.

The purpose of adding this new term to the energy–momentum tensor is that the original one was not necessarily symmetric. This is a problem, since the conservation of angular momentum requires a symmetric energy–momentum tensor. For example, if we take the divergence of $\mathscr{M}^{\rho\mu\nu}$, we find that it does not vanish in general, but equals $T^{\mu\nu} - T^{\nu\mu}$. However, we can always choose $E^{\lambda\mu\nu}$ such that the energy–momentum tensor is symmetric, so angular momentum is conserved.

Yet another reason for requiring a symmetric energy–momentum tensor is that in general relativity, the gravitational field tensor, which is symmetric, couples to the energy–momentum tensor. By the equivalence principle, the gravitational field couples equally to all forms of matter via its energy–momentum content. Hence, when we discuss general relativity in Chapter 19, we will need a symmetric energy–momentum tensor.

In summary, in this chapter we have made the transition from a classical system with a finite number of degrees of freedom to a classical field theory with an infinite number of degrees of freedom. Instead of a one-particle, classical description of a point particle in terms of coordinates, we now have a classical formalism defined in terms of fields $\phi(x)$.

In constructing field theories, we clearly see that the study of symmetries plays a crucial role. In fact, over the past two decades, physicists have come to the realization that symmetries are perhaps the most powerful tool that we have in describing the physical universe. As a consequence, it will prove beneficial to begin a more systematic discussion of symmetries and group theory. With this foundation, many of the rather strange and seemingly arbitrary conventions of quantum field theory begin to take on an elegant and powerful form.

Therefore, in the next chapter we will discuss various symmetries that have been shown experimentally to describe the fundamental particles of nature. Then in Chapter 3 we will begin a formal introduction to the quantum theory of systems with an infinite number of degrees of freedom.

1.10 Exercises

1. Show that the Poisson brackets obey the Jacobi identity:

$$\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0 \tag{1.82}$$

2. A transformation from the coordinates p and q to the new set $P = P(q, p, t)$ and $Q = Q(q, p, t)$ is called *canonical* if Hamilton's equations in Eq. (1.27) are satisifed with the new variables when we introduce a new Hamiltonian $\tilde{H}(Q, P, t)$. Show that the Poisson brackets between P and Q are the same

as those between p and q. Thus, the Poisson brackets of the coordinates are preserved under a canonical transformation.

3. Show that Poisson's brackets of two arbitrary functions A and B are invariant under a canonical transformation.

4. Since the action principle must be satisfied in the new coordinates, then $p\dot{q} - H$ must be equal to $P\dot{Q} - \tilde{H}$ up to a total derivative, that is, up to some arbitrary function $\dot{F}$. Show that, without losing any generality, we can take F to be one of four functions, given by: $F_1(q, Q, t)$, $F_2(q, P, t)$, $F_3(p, Q, t)$ or $F_4(p, P, t)$.

5. If we choose $F = F_1(q, Q, t)$, then prove that $p = \partial F_1/\partial q$, $P = -\partial F/\partial Q$, and $\tilde{H} = H + \dot{F}_1$.

6. What are the analogous relations for the other three F_i functions?

Chapter 2
Symmetries and Group Theory

...although the symmetries are hidden from us, we can sense that they are latent in nature, governing everything about us. That's the most exciting idea I know: that nature is much simpler than it looks. Nothing makes me more hopeful that our generation of human beings may actually hold the key to the universe in our hands—that perhaps in our lifetimes we may be able to tell why all of what we see in this immense universe of galaxies and particles is logically inevitable.

—S. Weinberg

2.1 Elements of Group Theory

So far, we have only described the broad, general principles behind classical field theory. In this chapter, we will study the physics behind specific models. We must therefore impose extra constraints on our Lagrangian that come from group theory. These groups, in turn, are extremely important because they describe the symmetries of the subatomic particles found experimentally in nature. Then in the next chapter, we will make the transition from classical field theory to the quantum theory of fields.

The importance of symmetries is seen when we write down the theory of radiation. When we analyze Maxwell fields, we find that they are necessarily relativistic. Therefore, we can also say that that quantum field theory arises out of the marriage of group theory (in particular the Lorentz and Poincaré groups) and quantum mechanics. Roughly speaking, we have:

$$\text{Quantum field theory} = \begin{cases} \text{Group theory} \\ \text{Quantum mechanics} \end{cases} \tag{2.1}$$

In fact, once the group structure of a theory (including the specific representations) are fixed, we find that the S matrix is essentially unique, up to certain

parameters specifying the interactions. More precisely, we will impose the following constraints in constructing a quantum field theory:

1. *We demand that all fields transform as irreducible representations of the Lorentz and Poincaré groups and some isospin group.*
2. *We demand that the theory be unitary and the action be causal, renormalizable, and an invariant under these groups.*

As simple as these postulates are, they impose enormous constraints on the theory. The first assumption will restrict the fields to be massive or massless fields of spin 0, 1/2, 1, etc. and will fix their isotopic representations. However, this constraint alone cannot determine the action, since there are invariant theories that are noncausal or nonunitary. (For example, there are theories with three or higher derivatives that satisfy the first condition. However, higher derivative theories have "ghosts," or particles of negative norm that violate unitarity, and hence must be ruled out.) The second condition, that the action obeys certain physical properties, then fixes the action up to certain parameters and representations, such as the various coupling constants found in the interaction.

Because of the power of group theory, we have chosen to begin our discussion of field theory with a short introduction to representation theory. We will find this detour to be immensely important; many of the curious "accidents" and conventions that occur in field theory, which often seem contrived and artificial, are actually byproducts of group theory.

There are three types of symmetries that will appear in this book.

1. *Space–time symmetries* include the Lorentz and Poincaré groups. These symmetries are noncompact, that is, the range of their parameters does not contain the endpoints. For example, the velocity of a massive particle can range from 0 to c, but cannot reach c.
2. *Internal symmetries* are ones that mix particles among each other, for example, symmetries like $SU(N)$ that mix N quarks among themselves. These internal symmetries rotate fields and particles in an abstract, "isotopic space," in contrast to real space–time. These groups are compact, that is, the range of their parameters is finite and contains their endpoints. For example, the rotation group is parametrized by angles that range between 0 and π or 2π. These internal symmetries can be either *global* (i.e., independent of space–time) or *local*, as in gauge theory, where the internal symmetry group varies at each point in space and time.
3. *Supersymmetry* nontrivially combines both space–time and internal symmetries. Historically, it was thought that space–time and isotopic symmetries

were distinct and could never be unified. "No-go theorems," in fact, were given to prove the incompatibility of compact and noncompact groups. Attempts to write down a nontrivial union of these groups with finite-dimensional unitary representations inevitably met with failure. Only recently has it become possible to unify them nontrivially and incorporate them into quantum field theories with supersymmetry, which manifest remarkable properties that were previously thought impossible. For example, certain supersymmetric theories are finite to all orders in perturbation theory, without the need for any renormalization. (However, since elementary particles with supersymmetry have yet to be discovered, we will only discuss this third class of symmetry later in the book in Chapters 20 and 21.)

2.2 $SO(2)$

We say that a collection of elements g_i form a *group* if they either obey or possess the following:

1. *Closure* under a multiplication operation; that is, if g_i and g_j are members of the group, then $g_i \cdot g_j$ is also a member of the group.
2. *Associativity* under multiplication; that is,

$$g_i \cdot (g_j \cdot g_k) = (g_i \cdot g_j) \cdot g_k \tag{2.2}$$

3. *An identity element* $\mathbf{1}$; that is, there exists an element $\mathbf{1}$ such that $g_i \cdot \mathbf{1} = \mathbf{1} \cdot g_i = g_i$.
4. *An inverse*; that is, every element g_i has an element g_i^{-1} such that $g_i \cdot g_i^{-1} = \mathbf{1}$.

There are many kinds of groups. A discrete group has a finite number of elements, such as the group of rotations that leave a crystal invariant. An important class of discrete groups are the parity inversion P, charge conjugation C, and time-reversal symmetries T. At this point, however, we are more interested in the *continuous* groups, such as the rotation and Lorentz group, which depend on a set of continuous angles.

To illustrate some of these abstract concepts, it will prove useful to take the simplest possible nontrivial example, $O(2)$, or rotations in two dimensions. Even the simplest example is surprisingly rich in content. Our goal is to construct the irreducible representations of $O(2)$, but first we have to make a few definitions. We know that if we rotate a sheet of paper, the length of any straight line on the paper is constant. If (x, y) describe the coordinates of a point on a plane, then

this means that, by the Pythagorean theorem, the following length is an invariant under a rotation about the origin:

$$\text{Invariant}: \quad x^2 + y^2 \tag{2.3}$$

If we rotate the plane through angle θ, then the coordinates (x', y') of the same point in the new system are given by:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{2.4}$$

We will abbreviate this by:

$$x^{i'} = O^{ij}(\theta)x^j \tag{2.5}$$

where $x^1 = x$ and $x^2 = y$. (For the rotation group, it makes no difference whether we place the index as a superscript, as in x^i, or as a subscript, as in x_i.)

For small angles, this can be reduced to:

$$\delta x = \theta y; \quad \delta y = -\theta x \tag{2.6}$$

or simply:

$$\delta x^i = \theta \epsilon^{ij} x^j \tag{2.7}$$

where ϵ^{ij} is antisymmetric and $\epsilon^{12} = -\epsilon^{21} = 1$. These matrices form a group; for example, we can write down the inverse of any rotation, given by $O^{-1}(\theta) = O(-\theta)$:

$$O(\theta)O(-\theta) = \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.8}$$

We can also prove associativity, since matrix multiplication is associative.

The fact that these matrices preserve the invariant length places restrictions on them. To find the nature of these restrictions, let us make a rotation on the invariant distance:

$$\begin{aligned} x'^i x'^i &= O^{ij} x^j \, O^{ik} x^k \\ &= x^j \left\{ O^{ij} O^{ik} \right\} x^k \\ &= x^j x^j \end{aligned} \tag{2.9}$$

that is, this is invariant if the O matrix is orthogonal;

$$O^{ij}O^{ik} = \delta^{jk} \tag{2.10}$$

or, more symbolically:

$$O^T \mathbf{1} O = \mathbf{1} \tag{2.11}$$

To take the inverse of an orthogonal matrix, we simply take its transpose. The unit matrix $\mathbf{1}$ is called the *metric* of the group.

The rotation group $O(2)$ is called the *orthogonal group* in two dimensions. The orthogonal group $O(2)$, in fact, can be defined as the set of all real, two-dimensional orthogonal matrices. Any orthogonal matrix can be written as the exponential of a single antisymmetric matrix τ:

$$O(\theta) = e^{\theta\tau} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(\theta\tau)^n \tag{2.12}$$

where:

$$\tau = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.13}$$

To see this, we note that the transpose of $e^{\theta\tau}$ is $e^{-\theta\tau}$:

$$O^T = \left(e^{\theta\tau}\right)^T = e^{-\theta\tau} = O^{-1} \tag{2.14}$$

Another way to prove this identity is simply to power expand the right-hand side and sum the series. We then see that the Taylor expansion of the cosine and sine functions re-emerge. After summing the series, we arrive at:

$$e^{\theta\tau} = \cos\theta\mathbf{1} + \tau\sin\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{2.15}$$

All elements of $O(2)$ are parametrized by one angle θ. We say that $O(2)$ is a one-parameter group; that is, it has dimension 1.

Let us now take the determinant of both sides of the defining equation:

$$\det\left(OO^T\right) = \det O \det O^T = (\det O)^2 = 1 \tag{2.16}$$

This means that the determinant of O is equal to ± 1. If we take $\det O = 1$, then the resulting subgroup is called $SO(2)$, or the special orthogonal matrices in two

dimensions. The rotations that we have been studying up to now are members of $SO(2)$. However, there is also the curious subset where $\det O = -1$. This subset consists of elements of $SO(2)$ times the matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.17}$$

This last transformation corresponds to a *parity* transformation:

$$\begin{aligned} x &\rightarrow x \\ y &\rightarrow -y \end{aligned} \tag{2.18}$$

A parity transformation P takes a plane and maps it into its mirror image, and hence it is a discrete, not continuous, transformation, such that $P^2 = 1$.

An important property of groups is that they are uniquely specified by their multiplication law. It is easy to show that these two dimensional matrices O^{ij} can be multiplied in succession as follows:

$$O^{ij}(\theta)O^{jk}(\theta') = O^{ik}(\theta + \theta') \tag{2.19}$$

which simply corresponds to the intuitively obvious notion that if we rotate a coordinate system by an angle θ and then by an additional angle θ', then the net effect is a rotation of $\theta + \theta'$. In fact, *any* matrix $D(\theta)$ (not necessarily orthogonal or even 2×2 in size) that has this multiplication rule:

$$D(\theta)D(\theta') = D(\theta + \theta'); \quad D(\theta) = D(\theta + 2\pi) \tag{2.20}$$

forms a *representation* of $O(2)$, because it has the same multiplication table.

For our purposes, we are primarily interested in the transformation properties of fields. For example, we can calculate how a field $\phi(x)$ transforms under rotations. Let us introduce the operator:

$$L \equiv i\epsilon^{ij}x^i\frac{\partial}{\partial x^j} = i(x^1\partial^2 - x^2\partial^1) \tag{2.21}$$

Let us define:

$$U(\theta) \equiv e^{i\theta L} \tag{2.22}$$

Then we define a *scalar* field as one that transforms under $SO(2)$ as:

$$\text{Scalar}: \quad U(\theta)\phi(x)U^{-1}(\theta) = \phi(x') \tag{2.23}$$

(To prove this equation, we use the fact that:

$$e^A\, B\, e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots \tag{2.24}$$

Then we reassemble these terms via a Taylor expansion to prove the transformation law.)

We can also define a *vector* field $\phi^i(x)$, where the additional i index also transforms under rotations:

$$\text{Vector}: \quad U(\theta)\phi^i(x)U^{-1}(\theta) = O^{ij}(-\theta)\phi^j(x') \tag{2.25}$$

[For this relation to hold, Eq. (2.21) must contain an additional term that rotates the vector index of the field.] Not surprisingly, we can now generalize this formula to include the transformation property of the most arbitrary field. Let $\phi^A(x)$ be an arbitrary field transforming under some representation of $SO(2)$ labeled by some index A. Then this field transforms as:

$$U(\theta)\phi^A(x)U^{-1}(\theta) = \mathscr{D}^{AB}(-\theta)\phi^B(x') \tag{2.26}$$

where $\mathscr{D}^{AB}$ is some representation, either reducible or irreducible, of the group.

2.3 Representations of $SO(2)$ and $U(1)$

One of the chief goals of this chapter is to find the irreducible representations of these groups, so let us be more precise. If g_i is a member of a group G, then the object $D(g_i)$ is called a *representation* of G if it obeys:

$$D(g_i)D(g_j) = D(g_i g_j) \tag{2.27}$$

for all the elements in the group. In other words, $D(g_i)$ has the same multiplication rules as the original group.

A representation is called *reducible* if $D(g_i)$ can be brought into block diagonal form; for example, the following matrix is a reducible representation:

$$D(g_i) = \begin{pmatrix} D_1(g_i) & 0 & 0 \\ 0 & D_2(g_i) & 0 \\ 0 & 0 & D_3(g_i) \end{pmatrix} \tag{2.28}$$

where D_i are smaller representations of the group. Intuitively, this means $D(g_i)$ can be split up into smaller pieces, with each piece transforming under a smaller representation of the same group.

The principal goal of our approach is to find all irreducible representations of the group in question. This is because the basic fields of physics transform as irreducible representations of the Lorentz and Poincaré groups. The complete set of finite-dimensional representations of the orthogonal group comes in two classes, the *tensors* and *spinors*. (For a special exception to this, see Exercise 10.)

One simple way of generating higher representations of $O(2)$ is simply to multiply several vectors together. The product $A^i B^j$, for example, transforms as follows:

$$\left(A^{i'} B^{j'}\right) = \left[O^{i'i}(\theta) O^{j'j}(\theta)\right] \left(A^i B^j\right) \tag{2.29}$$

This matrix $O^{i'i}(\theta) O^{j'j}(\theta)$ forms a representation of $SO(2)$. It has the same multiplication rule as $O(2)$, but the space upon which it acts is 2×2 dimensional. We call any object that transforms like the product of several vectors a *tensor*.

In general, a tensor $T^{ijk\cdots}$ under $O(2)$ is nothing but an object that transforms like the product of a series of ordinary vectors:

$$\text{Tensor}: \quad (T')^{i_1, i_2, \cdots} = O^{i_1, j_1} O^{i_2, j_2} \ldots T^{j_1, j_2, \cdots} \tag{2.30}$$

The transformation of $T^{ijk\cdots}$ is identical to the transformation of the product $x^i x^j x^k \cdots$. This product forms a representation of $O(2)$ because the following matrix:

$$O^{i_1, i_2 \cdots i_N; j_1, j_2, \cdots j_N}(\theta) \equiv O^{i_1, j_1}(\theta) O^{i_2, j_2}(\theta) \cdots O^{i_N, j_N}(\theta) \tag{2.31}$$

has the same multiplication rule as $SO(2)$.

The tensors that we can generate by taking products of vectors are, in general, reducible; that is, within the collection of elements that compose the tensor, we can find subsets that by themselves form representations of the group. By taking appropriate symmetric and antisymmetric combinations of the indices, we can extract irreducible representations (see Appendix).

A convenient method that we will use to create irreducible representations is to use two tensors under $O(2)$ that are actually constants: δ^{ij} and ϵ^{ij}, where the latter is the antisymmetric constant tensor and $\epsilon^{12} = -\epsilon^{21} = +1$.

Although they do not appear to be genuine tensors, it is easy to prove that they are. Let us hit them with the orthogonal matrix O^{ij}:

$$\begin{aligned} \delta^{i'j'} &= O^{i'i} O^{j'j} \delta^{ij} \\ \epsilon^{i'j'} &= O^{i'i} O^{j'j} \epsilon^{ij} \end{aligned} \tag{2.32}$$

We instantly recognize the first equation: it is just the definition of an orthogonal matrix, and so δ^{ij} is an invariant tensor. The second equation, however, is more difficult to see. Upon closer inspection, it is just the definition of the determinant of the O matrix, which is equal to one for $SO(2)$. Thus, both equations are satisfied by construction. Because the ϵ^{ij} transforms like a tensor only if the determinant of O is +1, we sometimes call it a *pseudotensor*. The pseudotensors pick up an extra minus one when they are transformed under parity transformations.

Using these two constant tensors, for example, we can immediately contract the tensor $A^i B^j$ to form two scalar combinations: $A^i B^i$ and $A^i \epsilon^{ij} B^j = A^1 B^2 - A^2 B^1$.

This process of symmetrizing and antisymmetrizing all possible tensor indices to find the irreducible representations is also aided by the simple identities:

$$\begin{aligned} \epsilon^{ij}\epsilon^{kl} &= \delta^{ik}\delta^{jl} - \delta^{il}\delta^{jk} \\ \epsilon^{ij}\epsilon^{jk} &= -\delta^{ik} \end{aligned} \tag{2.33}$$

Finally, we can show the equivalence between $O(2)$ and yet another formulation. We can take a complex object $u = a + ib$, and say that it transforms as follows:

$$u' = U(\theta)u = e^{i\theta}u \tag{2.34}$$

The matrix $U(\theta)$ is called a *unitary matrix*, because:

$$\text{Unitary matrix}: \quad U \times U^\dagger = 1 \tag{2.35}$$

The set of all one-dimensional unitary matrices $U(\theta) = e^{i\theta}$ defines a group called $U(1)$. Obviously, if we make two such transformations, we find:

$$e^{i\theta}e^{i\theta'} = e^{i\theta+i\theta'} \tag{2.36}$$

We have the same multiplication law as $O(2)$, even though this construction is based on a new space, the space of complex one-dimensional numbers. We thus say that:

$$SO(2) \sim U(1) \tag{2.37}$$

This means that there is a correspondence between the two, even though they are defined in two different spaces:

$$e^{\tau(\theta)} \leftrightarrow e^{i\theta} \tag{2.38}$$

To see the correspondence between $O(2)$ and $U(1)$, let us consider two scalar fields ϕ^1 and ϕ^2 that transform infinitesimally under $SO(2)$ as in Eq. (2.6):

$$\delta\phi^i = \theta\epsilon^{ij}\phi^j \tag{2.39}$$

which is just the usual transformation rule for small θ. Because $SO(2) \sim U(1)$, these two scalar fields can be combined into a single complex scalar field:

$$\phi = \frac{1}{\sqrt{2}}(\phi^1 + i\phi^2) \tag{2.40}$$

Then the infinitesimal variation of this field under $U(1)$ is given by:

$$\delta\phi = -i\theta\phi \tag{2.41}$$

for small θ. Invariants under $O(2)$ or $U(1)$ can be written as:

$$\frac{1}{2}\phi^i\phi^i = \phi^*\phi \tag{2.42}$$

2.4 Representations of $SO(3)$ and $SU(2)$

The previous group $O(2)$ was surprisingly rich in its representations. It was also easy to analyze because all its elements commuted with each other. We call such a group an *Abelian group*. Now, we will study *non-Abelian groups*, where the elements do not necessarily commute with each other. We define $O(3)$ as the group that leaves distances in three dimensions invariant:

$$\text{Invariant}: \quad x^2 + y^2 + z^2 \tag{2.43}$$

where $x^{i'} = O^{ij}x^j$. Generalizing the previous steps for $SO(2)$, we know that the set of 3×3, real, and orthogonal matrices $O(3)$ leaves this quantity invariant. The condition of orthogonality reduces the number of independent numbers down to $9 - 6 = 3$ elements. Any member of $O(3)$ can be written as the exponential of an antisymmetric matrix:

$$O = \exp\left(i\sum_{i=1}^{3}\theta^i\tau^i\right) \tag{2.44}$$

where τ^i has purely imaginary elements. There are only three independent antisymmetric 3×3 matrices, so we have the correct counting of independent degrees

of freedom. Therefore $O(3)$ is a three-parameter Lie group, parametrized by three angles.

These three antisymmetric matrices τ^i can be explicitly written as:

$$\tau^1 = \tau^x = -i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}; \quad \tau^2 = \tau^y = -i \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{2.45}$$

$$\tau^3 = \tau^z = -i \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.46}$$

By inspection, this set of matrices can be succinctly represented by the fully antisymmetric ϵ^{ijk} tensor as:

$$(\tau^i)^{jk} = -i\epsilon^{ijk} \tag{2.47}$$

where $\epsilon^{123} = +1$. These antisymmetric matrices, in turn, obey the following properties:

$$[\tau^i, \tau^j] = i\epsilon^{ijk}\tau^k \tag{2.48}$$

This is an example of a *Lie algebra* (not to be confused with the Lie group). The constants ϵ^{ijk} appearing in the algebra are called the *structure constants* of the algebra. A complete determination of the structure constants of any algebra specifies the Lie algebra, and also the group itself.

For small angles θ^i, we can write the transformation law as:

$$\delta x^i = \epsilon^{ijk}\theta^k x^j \tag{2.49}$$

As before, we will introduce the operators:

$$L^i \equiv i\epsilon^{ijk}x^j\partial^k \tag{2.50}$$

We can show that the commutation relations of L^i satisfy those of $SO(3)$. Let us construct the operator:

$$U(\theta^i) = e^{i\theta^i L^i} \tag{2.51}$$

Then a scalar and a vector field, as before, transform as follows:

$$\begin{aligned} U(\theta^k)\phi(x)U^{-1}(\theta^k) &= \phi(x') \\ U(\theta^k)\phi^i(x)U^{-1}(\theta^k) &= (O^{-1})^{ij}(\theta^k)\phi^j(x') \end{aligned} \tag{2.52}$$

For higher tensor fields, we must be careful to select out only the irreducible fields. The easiest way to decompose a reducible a tensor is to take various symmetric and anti-symmetric combinations of the indices. Irreducible representations can be extracted by using the two constant tensors, δ^{ij} and ϵ^{ijk}. In carrying out complicated reductions, it is helpful to know:

$$\begin{aligned} \epsilon^{ijk}\epsilon^{lmn} &= \delta^{il}\delta^{jm}\delta^{kn} - \delta^{il}\delta^{jn}\delta^{km} + \delta^{im}\delta^{jn}\delta^{kl} \\ &\quad -\delta^{im}\delta^{jl}\delta^{kn} + \delta^{in}\delta^{jl}\delta^{km} - \delta^{in}\delta^{jm}\delta^{kl} \\ \epsilon^{ijk}\epsilon^{klm} &= \delta^{il}\delta^{jm} - \delta^{im}\delta^{jl} \end{aligned} \tag{2.53}$$

More generally, we can use the method of Young Tableaux described in the Appendix to find more irreducible representations.

As in the case of $O(2)$, we can also find a relationship between $O(3)$ and a unitary group. Consider the set of all unitary, 2×2 matrices with unit determinant. These matrices form a group, called $SU(2)$, which is called the special unitary group in two dimensions. This matrix has $8 - 4 - 1 = 3$ independent elements in it. Any unitary matrix, in turn, can be written as the exponential of a Hermitian matrix H, where $H = H^\dagger$:

$$U = e^{iH} \tag{2.54}$$

Again, to prove this relation, simply take the Hermitian conjugate of both sides: $U^\dagger = e^{-iH^\dagger} = e^{-iH} = U^{-1}$.

Since an element of $SU(2)$ can be parametrized by three numbers, the most convenient set is to use the Hermitian *Pauli spin matrices*. Any element of $SU(2)$ can be written as:

$$U = e^{i\theta^i\sigma^i/2} \tag{2.55}$$

where:

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.56}$$

where σ^i satisfy the relationship:

$$\left[\frac{\sigma^i}{2}, \frac{\sigma^j}{2}\right] = i\epsilon^{ijk}\frac{\sigma^k}{2} \tag{2.57}$$

We now have exactly the same algebra as $SO(3)$ as in Eq. (2.48). Therefore, we can say:

$$SO(3) \sim SU(2) \tag{2.58}$$

To make this correspondence more precise, we will expand the exponential and then recollect terms, giving us:

$$e^{i\sigma^j\theta^j/2} = \cos(\theta/2) + i(\sigma^k n^k)\sin(\theta/2) \tag{2.59}$$

where $\theta^i = n^i\theta$ and $(n^i)^2 = 1$. The correspondence is then given by:

$$e^{i\tau^j\theta^j} \leftrightarrow e^{i\sigma^j\theta^j/2} \tag{2.60}$$

where the left-hand side is a real, 3×3 orthogonal matrix, while the right-hand size is a complex, 2×2 unitary matrix. (The isomorphism is only local, i.e., within a small neighborhood of each of the parameters. In general, the mapping is actually one-to-two, and not one-to-one.) Even though these two elements exist in different spaces, they have the same multiplication law. This means that there should also be a direct way in which vectors (x, y, z) can be represented in terms of these spinors. To see the precise relationship, let us define:

$$h(\mathbf{x}) = \boldsymbol{\sigma}\cdot\mathbf{x} = \begin{pmatrix} z & x - iy \\ x + iy & z \end{pmatrix} \tag{2.61}$$

Then the $SU(2)$ transformation:

$$h' = UhU^{-1} \tag{2.62}$$

is equivalent to the $SO(3)$ transformation:

$$\mathbf{x}' = \mathbf{O}\cdot\mathbf{x} \tag{2.63}$$

2.5 Representations of $SO(N)$

By now, the generalization to $O(N)$ should be straightforward. (The representations of $SU(N)$, which are important when we discuss the quark model, are discussed in the Appendix.)

The essential feature of a rotation in N dimensions is that it preserves distances. Specifically, the distance from the origin to the point x^i, by the Pythagorean Theorem, is given by $\sqrt{(x^i)^2}$. Therefore, $x^i x^i$ is an invariant, where an N-dimensional rotation is defined by: $x'^i = O^{ij} x^j$.

The number of independent elements in each member of $O(N)$ is N^2 minus the number of constraints arising from the orthogonality condition:

$$N^2 - \frac{1}{2}N(N+1) = \frac{1}{2}N(N-1) \tag{2.64}$$

This is exactly the number of independent antisymmetric, $N \times N$ matrices, that is, we can parametrize the independent components within $O(N)$ by either orthogonal matrices or by exponentiating antisymmetric ones, (i.e., $O = e^A$).

Any orthogonal matrix can thereby be parametrized as follows:

$$O = \exp\left(i \sum_{i=1}^{N(N-1)/2} \theta^i \tau^i \right) \tag{2.65}$$

where τ^i are linearly independent, antisymmetric matrices with purely imaginary elements. They are called the *generators* of the group, and θ^i are the rotation angles or the *parameters* of the group.

Finding representations of $O(N)$ is complicated by the fact that the multiplication table for the parameters of $O(N)$ are quite complicated. However, we know from the Baker–Campbell–Hausdorff theorem that:

$$e^A e^B = e^{A+B+(1/2)[A,B]+\cdots} \tag{2.66}$$

where the ellipsis represent multiple commutators of A and B. If e^A and e^B are close together, then these elements form a group as long as the commutators of A and B form an algebra.

If we take the commutator of two antisymmetric matrices, then it is easy to see that we get another antisymmetric matrix. This means that the algebra created by commuting all antisymmetric matrices is a closed one. We will represent the algebra as follows:

$$[\tau^i, \tau^j] = i f^{ijk} \tau^k \tag{2.67}$$

As before, we say that the f^{ijk} are the structure constants of the group. (It is customary to insert an i in the definition of τ, so that an i appears in the commutator.)

For arbitrary N, it is possible to find an exact form for the structure constants. Let us define the generator of $O(N)$ as:

$$(M^{ij})_{ab} = -i(\delta^i_a \delta^j_b - \delta^j_a \delta^i_b) \tag{2.68}$$

Since the matrix is antisymmetric in i, j, there are $N(N-1)/2$ such matrices, which is the correct number of parameters for $O(N)$. The indices a, b denote the various matrix entries of the generator. If we commute these matrices, the calculation is rather easy because it reduces to contracting over a series of delta functions:

$$[M^{ij}, M^{lm}] = i\left(-\delta^{il}M^{jm} + \delta^{jl}M^{im} + \delta^{im}M^{jl} - \delta^{mj}M^{il}\right) \tag{2.69}$$

To define the action of $SO(N)$ on the fields, let us define the operator:

$$L^{ij} \equiv i(x^i \partial^j - x^j \partial^i) \tag{2.70}$$

It is easy to check that L^{ij} satisfies the commutation relations of $SO(N)$. Now construct the operator:

$$U(\theta^{ij}) = e^{i\theta^{ij}L^{ij}} \tag{2.71}$$

where θ^{ij} is antisymmetric. The structure constants of the theory f^{ijk} can also be thought of as a representation of the algebra. If we define:

$$\left(\tau^i\right)^{jk} = f^{ijk} \tag{2.72}$$

then τ^i as written as a function of the structure constants also forms a representation of the generators of $O(N)$. We call this the *adjoint* representation. This also means that the structure constant f^{ijk} is a constant tensor, just like δ^{ij}.

For our purposes, we will often be interested in how fields transform under some representation of $O(N)$. Without specifying the exact representation, we can always write:

$$\delta\phi^i = i\theta^a \left(\tau^a\right)_{ij} \phi^j \tag{2.73}$$

This simply means that we are letting ϕ^i transform under some unspecified representation of the generators of $O(N)$, labeled by the indices i, j.

Finally, one might wonder whether we can find more identities such as $O(2) \sim U(1)$ and $O(3) \sim SU(2)$. With a little work, we can show:

$$SO(4) \sim SU(2) \otimes SU(2); \quad SO(6) \sim SU(4) \tag{2.74}$$

One may then wonder whether there are higher sequences of identities between $O(N)$ and $SU(M)$. Surprisingly, the answer is no. To see that any correspondence would be nontrivial can be checked by simply calculating the number of parameters in $O(N)$ and $SU(M)$, which are quite different. These "accidents" between Lie groups only occur at low dimensionality and are the exception, rather than the rule.

2.6 Spinors

In general, there are two major types of representations that occur repeatedly throughout physics. The first, of course, are tensors, which transform like the product of vectors. Irreducible representations are then found by taking their symmetric and antisymmetric combinations.

However, one of the most important representations of $O(N)$ are the *spinor representations*. To explain the spinor representations, let us introduce N objects Γ^i, which obey:

$$\begin{aligned} \{\Gamma^i, \Gamma^j\} &= 2\delta^{ij} \\ \{A, B\} &\equiv AB + BA \end{aligned} \tag{2.75}$$

where the brackets represent an anticommutator rather than a commutator. This is called a *Clifford algebra*. Then we can construct the spinor representation of the generators of $SO(N)$ as follows:

$$\text{Spinor representation}: \quad M^{ij} = \frac{i}{4}[\Gamma^i, \Gamma^j] \tag{2.76}$$

By inserting this value of M^{ij} into the definition of the algebra of $O(N)$, it satisfies the commutation relations of Eq. (2.69) and hence gives us a new representation of the group.

In general, we can find a spinorial matrix representation of $O(N)$ (for N even) that is $2^{N/2}$ dimensional and complex. The simplest spinor representation of $O(4)$, we will see, gives us the compact version of the celebrated Dirac matrices. For the odd orthogonal groups $O(N+1)$ (where N is even), we can construct the spinors Γ^i from the spinors for the group $O(N)$. We simply add a new element to the old

set:

$$\Gamma_{N+1} = \Gamma^1\Gamma^2\cdots\Gamma^N \tag{2.77}$$

Γ_{N+1} has the same anti-commutation relations as the other spinors, and hence we can form the generators of $O(N+1)$ out of the spinors of $O(N)$. As before, we can construct the transformation properties of the Γ^i. Let us define:

$$U(\theta^{ij}) = \exp\left(i\theta^{ij}M^{ij}\right) \tag{2.78}$$

where M^{ij} is constructed out of the spinors Γ^i.

Then it is easy to show that Γ^i satisfies the following identity:

$$U(\theta^{ij})\Gamma^k U^{-1}(\theta^{ij}) = (O^{-1})^{kl}(\theta^{ij})\Gamma^l \tag{2.79}$$

which proves that Γ^i transforms as a vector.

Finally, we should also mention that this spinorial representation of the group $O(N)$ is reducible. For example, we could have constructed two projection operators:

$$\begin{aligned} P_R &= \frac{1+\Gamma_{N+1}}{2} \\ P_L &= \frac{1-\Gamma_{N+1}}{2} \end{aligned} \tag{2.80}$$

which satisfy the usual properties of projection operators:

$$\begin{aligned} P_L^2 &= P_L \\ P_R^2 &= P_R \\ P_R P_L &= 0 \\ P_L + P_R &= 1 \end{aligned} \tag{2.81}$$

With these two projection operators, the group splits into two self-contained pieces. (When we generalize this construction to the Lorentz group, these will be called the "left-handed" and "right-handed" Weyl representations. They will allow us to describe neutrino fields.)

2.7 Lorentz Group

Now that we have completed our brief, warm-up discussion of some compact Lie groups, let us tackle the main problem of this chapter, finding the representations of the noncompact Lorentz and Poincaré groups.

We define the Lorentz group as the set of all 4×4 real matrices that leave the following invariant:

$$\text{Invariant}: \quad s^2 = c^2t^2 - x^i x^i = (x^0)^2 - (x^i)^2 = x^\mu g_{\mu\nu} x^\nu \tag{2.82}$$

The minus signs, of course, distinguish this from the group $O(4)$.

A Lorentz transformation can be parametrized by:

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \tag{2.83}$$

Inserting this transformation into the invariant, we find that the Λ matrices must satisfy:

$$g_{\mu\nu} = \Lambda^\rho{}_\mu g_{\rho\sigma} \Lambda^\sigma{}_\nu \tag{2.84}$$

which can be written symbolically as $g = \Lambda^T g \Lambda$. Comparing this with Eq. (2.11), we say that $g_{\mu\nu}$ is the *metric* of the Lorentz group. If the signs within the metric $g_{\mu\nu}$ were all the same sign, then the group would be $O(4)$. To remind ourselves that the signs alternate, we call the Lorentz group $O(3, 1)$:

$$\text{Lorentz group}: \quad O(3, 1) \tag{2.85}$$

where the comma separates the positive from the negative signs within the metric. In general, an orthogonal group that preserves a metric with M indices of one sign and N indices of another sign is denoted $O(M, N)$.

The minus signs in the metric create an important difference between $O(4)$ and the Lorentz group: The invariant distance s^2 can be negative or positive, while the invariant distance for $O(4)$ was always positive. This means that the x_μ plane splits up into distinct regions that cannot be connected by a Lorentz transformation. If x and y are two position vectors, then these regions can be labeled by the value of the invariant distance s^2:

$$\begin{aligned} (x-y)^2 > 0: &\quad \text{time-like} \\ (x-y)^2 = 0: &\quad \text{light-like} \\ (x-y)^2 < 0: &\quad \text{space-like} \end{aligned} \tag{2.86}$$

Excluding this crucial difference, the representations of the Lorentz group, for the most part, bear a striking resemblance with those of $O(4)$. For example, as before, we can introduce the operator $L^{\mu\nu}$ in order to define the action of the Lorentz group on fields:

$$L^{\mu\nu} = x^\mu p^\nu - x^\nu p^\nu = i(x^\mu \partial^\nu - x^\nu \partial^\mu) \tag{2.87}$$

where $p_\mu = i\partial_\mu$. As before, we can show that this generates the algebra of the Lorentz group:

$$[L^{\mu\nu}, L^{\rho\sigma}] = ig^{\nu\rho}L^{\mu\sigma} - ig^{\mu\rho}L^{\nu\sigma} - ig^{\nu\sigma}L^{\mu\rho} + ig^{\mu\sigma}L^{\nu\rho} \tag{2.88}$$

Let us also define:

$$U(\Lambda) = \exp\left(i\epsilon^{\mu\nu}L_{\mu\nu}\right) \tag{2.89}$$

where, infinitesimally, we have:

$$\Lambda^{\mu\nu} = g^{\mu\nu} + \epsilon^{\mu\nu} + \cdots \tag{2.90}$$

Then the action of the Lorentz group on a vector field ϕ^μ can be expressed as:

$$\text{Vector}: \quad U(\Lambda)\phi^\mu(x)U^{-1}(\Lambda) = (\Lambda^{-1})^\mu{}_\nu \phi^\nu(x') \tag{2.91}$$

(where $U(\Lambda)$ contains an additional piece which rotates the vector index of ϕ^μ). Let us parametrize the Λ^μ_ν of a Lorentz transformation as follows:

$$x' = \frac{x + vt}{\sqrt{1 - v^2/c^2}}; \quad y' = y; \quad z' = z; \quad t' = \frac{t + vx/c^2}{\sqrt{1 - v^2/c^2}} \tag{2.92}$$

We can make several observations about this transformation from a group point of view. First, the velocity v is a parameter of this group. Since the velocity varies as $0 \leq v < c$, where v is strictly less than the speed of light c, we say the Lorentz group is *noncompact*; that is, the range of the parameter v does not include the endpoint c. This stands in contrast to the group $O(N)$, where the parameters have finite range, include the endpoints, and are hence compact.

We say that the three components of velocity v^x, v^y, v^z are the parameters of Lorentz boosts. If we make the standard replacement:

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \cosh\phi, \quad \beta\gamma = \sinh\phi; \quad \beta = v/c \tag{2.93}$$

then this transformation can be written as:

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \cosh\phi & \sinh\phi & 0 & 0 \\ \sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.94}$$

Let us rewrite this in terms of $M^{\mu\nu}$. Let us define:

$$\begin{aligned} J^i &= \frac{1}{2}\epsilon^{ijk}M^{jk} \\ K^i &= M^{0i} \end{aligned} \tag{2.95}$$

Written out explicitly, this becomes:

$$K^x = K^1 = -i\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.96}$$

Then this Lorentz transformation can be written as:

$$e^{iK^x v^x} = \cosh\phi + i\sinh\phi K^x \tag{2.97}$$

Similarly, Lorentz boosts in the y and z direction are generated by exponentiating K^y and K^z:

$$K^y = K^2 = -i\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad K^z = K^3 = -i\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{2.98}$$

Unfortunately, it is easily checked that a boost in the x direction, followed by a boost in the y direction, does not generate another Lorentz boost. Thus, the three K matrices by themselves do not generate a closed algebra. To complete the algebra, we must introduce the generators of the ordinary rotation group $O(3)$ as

well:

$$J^x = J^1 = -i \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}; \quad J^y = J^2 = -i \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{2.99}$$

$$J^z = J^3 = -i \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.100}$$

The K and J matrices have the following commutation relations:

$$\begin{aligned} [K^i, K^j] &= -i\epsilon^{ijk} J^k \\ [J^i, J^j] &= i\epsilon^{ijk} J^k \\ [J^i, K^j] &= i\epsilon^{ijk} K^k \end{aligned} \tag{2.101}$$

Two pure Lorentz boosts generated by K^i taken in succession do *not* generate a third boost, but must generate a rotation generated by J^i as well. (Physically, this rotation, which arises after several Lorentz boosts, gives rise to the Thomas precession effect.)

By taking linear combinations of these generators, we can show that the Lorentz group can be split up into two pieces. We will exploit the fact that $SO(4) = SU(2) \otimes SU(2)$, and that the algebra of the Lorentz group is similar to the algebra of $SO(4)$ (modulo minus signs coming from the metric). By taking linear combinations of these generators, one can show that the algebra actually splits into two pieces:

$$\begin{aligned} A^i &= \frac{1}{2}(J^i + iK^i) \\ B^i &= \frac{1}{2}(J^i - iK^i) \end{aligned} \tag{2.102}$$

We then have $[A^i, B^j] = 0$, so the algebra has now split up into two distinct pieces, each piece generating a separate $SU(2)$.

If we change the sign of the metric so that we only have compact groups, then we have just proved that the Lorentz group, for our purposes, can be written as $SU(2) \otimes SU(2)$. This means that irreducible representations (j) of $SU(2)$, where $j = 0, 1/2, 1, 3/2$, etc., can be used to construct representations of the

Lorentz group, labeled by (j, j') (see Appendix). By simply pairing off two representations of $SU(2)$, we can construct all the representations of the Lorentz group. (The representation is spinorial when $j+j'$ is half-integral.) In this fashion, we can construct both the tensor and spinor representations of the Lorentz group.

We should also remark that not all groups have spinorial representations. For example, $GL(N)$, the group of all real $N \times N$ matrices, does not have any finite-dimensional spinorial representation. (This will have a great impact on the description of electrons in general relativity in Chapter 19.)

2.8 Representations of the Poincaré Group

Physically, we can generalize the Lorentz group by adding translations:

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu} \tag{2.103}$$

The Lorentz group with translations now becomes the Poincaré group. Because the Poincaré group includes four translations in addition to three rotations and three boosts, it is a 10-parameter group. In addition to the usual generator of the Lorentz group, we must add the translation generator $p_{\mu} = i\partial_{\mu}$.

The Poincaré algebra is given by the usual Lorentz algebra, plus some new relations:

$$\begin{aligned} [L_{\mu\nu}, P_{\rho}] &= -ig_{\mu\rho}P_{\nu} + ig_{\nu\rho}P_{\mu} \\ [P_{\mu}, P_{\nu}] &= 0 \end{aligned} \tag{2.104}$$

These relations mean that two translations commute, and that translations transform as a genuine vector under the Lorentz group.

To find the irreducible representations of a Lie group, we will often use the technique of simultaneously diagonalizing a subset of its generators. Let the *rank* of a Lie group be the number of generators that simultaneously commute among themselves. The rank and dimension of $O(N)$ and $SU(N)$ are given by:

Group	Dimension	Rank
$SO(N)$ (N even)	$(1/2)N(N-1)$	$N/2$
$SO(N)$ (N odd)	$(1/2)N(N-1)$	$(N-1)/2$
$SU(N)$	N^2-1	$N-1$

(2.105)

For example, the group $SO(3)$ has rank 1, so we can choose L_3 to be the generator to diagonalize. The group $SO(4)$, as we saw, can be re-expressed in terms of two $SU(2)$ subgroups, so that there are two generators that commute among themselves.

In addition, we have the *Casimir operators* of the group, that is, those operators that commute with all the generators of the algebra. For the group $O(3)$, we know, for example, that the Casimir operator is the sum of the squares of the generators: L_i^2. Therefore, we can simultaneously diagonalize L_i^2 and L_3. The representations of $SO(3)$ then correspond to the eigenvalues of these two operators.

For the Poincaré group, we first know that $P_\mu^2 = m^2$, or the mass squared, is a Casimir operator. Under Lorentz transformations, it transforms as a genuine scalar, and is hence invariant. Also, it is invariant under translations because all translations commute.

To find the other Casimir operator, we introduce:

$$W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} P_\nu L_{\rho\sigma} \tag{2.106}$$

which is called the Pauli–Lubanski tensor (where $\epsilon^{0123} = +1$). Then the square of this tensor is a Casimir operator as well:

$$\text{Casimir operators} = \{P_\mu^2,\, W_\mu^2\} \tag{2.107}$$

All physical states in quantum field theory can be labeled according to the eigenvalue of these two Casimir operators (since the Casimir commutes with all generators of the algebra). However, the physical significance of this new Casimir operator is not immediately obvious. It cannot correspond to spin, since our intuitive notion of angular momentum, which we obtain from nonrelativistic quantum mechanics, is, strictly speaking, lost once we boost all particles by a Lorentz transformation. The usual spin operator is no longer a Casimir operator of the Lorentz group.

To find the physical significance of W_μ^2, let us therefore to go the rest frame of a massive particle: $P^\mu = (m, 0)$. Inserting this into the equation for the Pauli–Lubanski tensor, we find:

$$\begin{aligned} W_i &= -\frac{1}{2} m \epsilon_{ijk0} J^{jk} \\ &= -m L_i \\ W_0 &= 0 \end{aligned} \tag{2.108}$$

where L_i is just the usual rotation matrix in three dimensions. Thus, in the rest frame of a massive particle, the Pauli–Lubanski tensor is just the spin generator.

Its square is therefore the Casimir of $SO(3)$, which we know yields the spin of the particle:

$$W_i^2 = m^2 s(s+1) \tag{2.109}$$

where s is the spin eigenstate of the particle.

In the rest frame of the massive particle, we have $2s+1$ components for a spin-s particle. This corresponds to a generalization of our intuitive understanding of spin coming from nonrelativistic quantum mechanics.

However, this analysis is incomplete because we have not discussed massless particles, where $P_\mu^2 = 0$. In general, the counting rule for massive spinning states breaks down for massless ones. For these particles, we have:

$$W_\mu^2|p\rangle = W_\mu P^\mu|p\rangle = P_\mu P^\mu|p\rangle = 0 \tag{2.110}$$

The only way to satisfy these three conditions is to have W_μ and P^μ be proportional to each other; that is, $W_\mu|p\rangle = hP_\mu|p\rangle = 0$ on a massless state $|p\rangle$. This number h is called the *helicity*, and describes the number of independent components of a massless state.

Using the definition of W_μ for a massless state, we can show that h can be written as:

$$h = \frac{\mathbf{J}\cdot\mathbf{P}}{|\mathbf{P}|} \tag{2.111}$$

Because of the presence of $\epsilon^{\mu\nu\rho\sigma}$ in the definition of W_μ, the Pauli–Lubanski vector is actually a pseudovector. Under a parity transformation, the helicity h therefore transforms into $-h$. This means that massless states have two helicity states, corresponding to a state where W_μ is aligned parallel to the momentum vector and also aligned antiparallel to the momentum vector. Regardless of the spin of a massless particle, the helicity can have only *two* values, h and $-h$. There is thus an essential difference between massless and massive states in quantum field theory.

It is quite remarkable that we can label all irreducible representations of the Poincaré group (and hence all the known fields in the universe) according to the eigenvalues of these Casimir operators. A complete list is given as follows in terms of the mass m, spin s, and helicity h:

$$\begin{cases} P_\mu^2 > 0: & |m, s\rangle,\ s = 0, 1/2, 1, 3/2, \cdots \\ P_\mu^2 = 0: & |h\rangle,\ h = \pm s \\ P_\mu^2 = 0: & s \text{ continuous} \\ P_\mu^2 < 0: & \text{tachyon} \end{cases} \tag{2.112}$$

In nature, the physical spectrum of states seems to be realized only for the first two categories with $P^0 > 0$. The other states, which have continuous spins or tachyons, have not been seen in nature.

2.9 Master Groups and Supersymmetry

Although we have studied the representations of $O(3, 1)$ by following our discussion of the representations of $O(4)$, the two groups are actually profoundly different. One nontrivial consequence of the difference between $O(4)$ and $O(3, 1)$ is summarized by the following:

No-Go Theorem: *There are no finite-dimensional unitary representations of noncompact Lie groups. Any nontrivial union of the Poincaré and an internal group yields an S matrix which is equal to 1.*

This theorem has caused a certain amount of confusion in the literature. In the 1960s, after the success of the $SU(3)$ description of quarks, attempts were made to construct Master Groups that could nontrivially combine both the Poincaré group and the "internal" group $SU(3)$:

$$\text{Master group} \supset P \otimes SU(3) \tag{2.113}$$

In this way, it was hoped to give a unified description of particle physics in terms of group theory. There was intense interest in groups like $SU(6, 6)$ or $\tilde{U}(12)$ that combined both the internal and space–time groups. Only later was the no-go theorem discovered, which seemed to doom all these ambitious efforts. Because of the no-go theorem, unitary representations of the particles were necessarily infinite dimensional: These groups possessed nonphysical properties, such as an infinite number of particles in each irreducible representation, or a continuous spectrum of masses for each irreducible representation. As a consequence, after a period of brief enthusiasm, the no-go theorem doomed all these naive efforts to build a Master Group for all particle interactions.

Years later, however, it was discovered that there was a loophole in the no-go theorem. It was possible to evade this no-go theorem (the most comprehensive version being the Coleman–Mandula theorem) because it made an implicit assumption: that the parameters θ_i of the Master Group were all c numbers. However, if the θ_i could be anticommuting, then the no-go theorem could be evaded.

This leads us to the super groups, and eventually to the superstring, which have revived efforts to build Master Groups containing all known particle interactions, including gravity. Although supersymmetry holds the promise of being a

fundamental symmetry of physics, we will study these theories not because they have any immediate application to particle physics, but because they provide a fascinating laboratory in which one can probe the limits of quantum field theory.

In summary, the essential point is that quantum field theory grew out of the marriage between quantum mechanics and group theory, in particular the Lorentz and Poincaré group. In fact, it is one of the axioms of quantum field theory that the fundamental fields of physics transform as irreducible representations of these groups. Thus, a study of group theory goes to the heart of quantum field theory. We will find that the results of this chapter will be used throughout this book.

2.10 Exercises

1. By a direct calculation, show that M^{ij} given in Eq. (2.68) and the spinor representation given in Eq. (2.76) do, in fact, satisfy the commutation relations of the Lorentz algebra in Eq. (2.88) if we use the Lorentz metric instead of Kronecker delta functions.

2. Prove that, under a proper orthochronous Lorentz transformation (see Appendix for a definition), the sign of the time t variable (if we are in the forward light cone) does not change. Thus, the sign of t is an invariant in the forward light cone, and we cannot go backwards in time by using rotations and proper orthochronous Lorentz boosts.

3. Show that the proper orthochronous Lorentz group is, in fact, a group. Do the other branches of the Lorentz group also form a group?

4. For $O(3)$, show that the dimensions of the irreducible tensor representations are all positive odd integers.

5. For $O(3)$, show that:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{7} \oplus \mathbf{5} \oplus \mathbf{5} \oplus \mathbf{3} \oplus \mathbf{3} \oplus \mathbf{3} \oplus \mathbf{1} \tag{2.114}$$

   (see Appendix).

6. Prove that:

$$e^A F(B) e^{-A} = F\left(e^A B e^{-A}\right) \tag{2.115}$$

   where A and B are operators, and F is an arbitrary function. (Hint: use a Taylor expansion of F.)

7. Prove Eq. (2.24).

8. Prove that the Pauli–Lubanski vector is a genuine Casimir operator.

9. Prove that an element of a Clifford algebra Γ^{μ} transforms as a vector under the Lorentz group; that is, verify Eq. (2.79).

10. For $SO(2)$, show that the spin eigenvalue can be continuous. (Hint: examine $\exp(i\alpha\theta)$ under a complete rotation in θ if α is fractional.)

11. Prove that the Lorentz group can be written as $SL(2, C)$, the set of complex, 2×2 matrices with unit determinant. Show that this representation (as well as the other representations we have discussed in this chapter) is not unitary, and hence satisfies the no-go theorem. Unitarity, however, is required to satisfy the conservation of probability. Does the nonunitarity of these representations violate this important principle?

12. In three dimensions, using the contraction of two ϵ^{ijk} constant tensors into Kronecker deltas in Eq. (2.53), prove that the curl of the curl of a vector A is given by $\nabla^2 \mathbf{A} - \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{A})$.

13. Prove that $SU(4)$ is locally isomorphic to $SO(6)$. (Hint: show the equivalence of their Lie algebras.)

14. Prove that there are two constants, δ^{ij} and $\epsilon^{ijklm\cdots}$, which transform as genuine tensors under $SO(N)$. To prove these constants are genuine tensors, act upon them with O^{ij}. Show that the tensor equation for δ^{ij} reduces to the definition of an orthogonal group. Prove that $\epsilon^{ijklm\cdots}$ satisfies:

$$O^{i_1 j_1} O^{i_2 j_2} \cdots O^{i_N j_N} \epsilon^{j_1 j_2 \cdots j_N} = A\, \epsilon^{i_1 i_2 \cdots i_N} \tag{2.116}$$

What is the constant A?

15. Prove that:

$$L_{\mu\nu}^2; \quad \epsilon^{\mu\nu\rho\sigma} L_{\mu\nu} L_{\rho\sigma} \tag{2.117}$$

are Casimir operators for the Lorentz group (but are not Casimir operators for the full Poincaré group).

16. Re-express:

$$\epsilon^{i_1 i_2 \cdots i_N} \epsilon^{j_1 j_2 \cdots j_N} \tag{2.118}$$

entirely in terms of delta functions for $N = 4$ and 5. (Hint: check that the antisymmetry properties of the $\epsilon^{\mu\nu\cdots}$ tensor are satisfied for the product of delta functions.)

Chapter 3
Spin-0 and $\frac{1}{2}$ Fields

It is more important to have beauty in one's equations than to have them fit experiment... because the discrepancy may be due to minor features that are not properly taken into account and that will get cleared up with further developments of the theory.... It seems that if one is working from the point of view of getting beauty in one's equations, and if one has really a sound insight, one is on a sure line of progress.

—P.A.M. Dirac

3.1 Quantization Schemes

In the previous chapters, we presented the classical theory of fields and also the symmetries they obey. In this chapter, we now make the transition to the quantum theory of fields.

Symbolically, we may write:

$$\lim_{N\to\infty} \text{Quantum mechanics} = \text{Quantum field theory} \tag{3.1}$$

where N is the number of degrees of freedom of the system. We will see that one important consequence of this transition is that quantum field theory describes *multiparticle* states, while ordinary quantum mechanics is based on a single-particle interpretation. We will find that second quantized systems are ideally suited to describing relativistic physics, since relativity introduces pair creation and annihilation and hence inevitably introduces multiparticle states.

In this chapter, we will develop the second quantization program for the irreducible representations of the Lorentz group for fields with spin 0 and $\frac{1}{2}$. We stress, however, that a number of different types of quantization schemes have been proposed over the decades, each with their own merits and drawbacks:

1. The most direct method is the *canonical quantization* program, which we will develop in this chapter. Canonical quantization closely mimics the development of quantum mechanics; that is, time is singled out as a special coordinate and manifest Lorentz invariance is sacrificed. The advantage of canonical quantization is that it quantizes only physical modes. Unitarity of the system is thus manifest. At the level of QED, the canonical quantization method is not too difficult, but the canonical quantization of more complicated theories, such as non-Abelian gauge theories, is often prohibitively tedious.

2. The *Gupta–Bleuler* or covariant quantization method will also be mentioned in this chapter. Contrary to canonical quantization, it maintains full Lorentz symmetry, which is a great advantage. The disadvantage of this approach is that ghosts or unphysical states of negative norm are allowed to propagate in the theory, and are eliminated only when we apply constraints to the state vectors.

3. The *path integral method* is perhaps the most elegant and powerful of all quantization programs. One advantage is that one can easily go back and forth between many of the other quantization programs to see the relationships between them. Although some of the conventions found in various quantization programs may seem a bit bizarre or contrived, the path integral approach is based on simple, intuitive principles that go to the very heart of the assumptions of quantum theory. The disadvantage of the path integral approach is that functional integration is a mathematically delicate operation that may not even exist in Minkowski space.

4. The Becchi–Rouet–Stora–Tyupin (BRST) approach is one of the most convenient and practical covariant approaches used for gauge theories. Like the Gupta-Bleuler quantization program, negative norm states or ghosts are allowed to propagate and are eliminated by applying the BRST condition onto the state vectors. All the information is contained in a single operator, making this a very attractive formalism. The BRST approach can be easily expressed in terms of path integrals.

5. Closely related to the BRST method is the Batalin–Vilkovisky (BV) quantization program, which has proved powerful enough to quantize the most complicated actions so far proposed, such as those found in string and membrane theories. The formalism is rather cumbersome, but it remains the only program that can quantize certain complex actions.

6. *Stochastic quantization* is yet another quantization program that preserves gauge invariance. One postulates a fictitious fifth coordinate, such that the physical system eventually settles down to the physical solution as the fifth coordinate evolves.

3.2 Klein–Gordon Scalar Field

Let us begin our discussion by quantizing the simplest possible relativistic field theory, the free scalar field. The theory was proposed independently by six different physicists.[1–6] The Lagrangian is given by:

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 \tag{3.2}$$

Historically, the quantization of the Klein–Gordon equation caused much confusion. Schrödinger, even before he postulated his celebrated nonrelativistic equation, considered this relativistic scalar equation but ultimately discarded it because of problems with negative probability and negative energy states. In any fully relativistic equation, we must obey the "mass-shell condition" $p_\mu^2 = E^2 - \mathbf{p}^2 = m^2$. This means that the energy is given by:

$$E = \pm\sqrt{\mathbf{p}^2 + m^2} \tag{3.3}$$

The energy can be negative, which is quite disturbing. Even if we banish the negative energy states by fiat, we find that interactions with other particles will reduce the energy and create negative energy states. This means that all positive energy states will eventually collapse into negative energy states, destabilizing the entire theory. One can show that even if we prepare a wave packet with only states of positive energy, interactions will inevitably introduce negative energy states. We will see, however, that the solution of these problems with negative probability and negative energy can be resolved once one quantizes the theory.

The canonical quantization program begins with fields ϕ and their conjugate momentum fields π, which satisfy equal time commutation relations among themselves. Then the time evolution of these quantized fields is governed by a Hamiltonian. Thus, we closely mimic the dynamics found in ordinary quantum mechanics. We begin by singling out time as a special coordinate and then defining the canonical conjugate field to ϕ:

$$\pi(\mathbf{x}, t) = \frac{\delta \mathscr{L}}{\delta \dot{\phi}(\mathbf{x}, t)} = \dot{\phi}(\mathbf{x}, t) \tag{3.4}$$

We can introduce the Hamiltonian as:

$$\mathscr{H} = \pi\dot{\phi} - \mathscr{L} = \frac{1}{2}\left[\pi^2 + (\nabla\phi)^2 + m^2\phi^2\right] \tag{3.5}$$

Then the transition from classical mechanics to quantum field theory begins when we postulate the commutation relations between the field and its conjugate momentum:

$$[\phi(\mathbf{x}, t), \pi(\mathbf{y}, t)] = i\delta^3(\mathbf{x} - \mathbf{y}) \tag{3.6}$$

(The right-hand side is proportional to $\hbar$, which we omit. This is the point where the quantum principle begins to emerge from the classical theory.) All other commutators (i.e., between π and itself, and ϕ and itself) are set equal to zero. [Although this expression looks non-relativistic, notice that x^μ and y^μ are separated by a space-like distance, $(x - y)^2 < 0$, which is preserved under a Lorentz transformation.]

Much of what follows is a direct consequence of this commutation relation. There are an infinite number of ways in which to satisfy this relationship, but our strategy will be to find a specific Fourier representation of this commutation relation in terms of plane waves. When these plane-wave solutions are quantized in terms of harmonic oscillators, we will be able to construct the multiparticle Hilbert space and also find a specific operator representation of the Lorentz group in terms of oscillators.

We first define the quantity:

$$k \cdot x \equiv k_\mu x^\mu = (Et - \mathbf{p} \cdot \mathbf{x}) \tag{3.7}$$

We want a decomposition of the scalar field where the energy k^0 is positive, and where the Klein–Gordon equation is explicitly obeyed. In momentum space, the operator $\partial_\mu^2 + m^2$ becomes $k^2 - m^2$. Therefore, we choose:

$$\phi(x) = \frac{1}{(2\pi)^{3/2}} \int d^4k \; \delta(k^2 - m^2)\theta(k_0)\left[A(k)e^{-ik\cdot x} + A^\dagger(k)e^{ik\cdot x}\right] \tag{3.8}$$

where θ is a step function [$\theta(k_0) = +1$ if $k_0 > 0$ and $\theta(k_0) = 0$ otherwise], and where $A(k)$ are operator-valued Fourier coefficients. It is now obvious that this field satisfies the Klein–Gordon equation. If we hit this expression with $(\partial_\mu^2 + m^2)$, then this pulls down a factor of $k^2 - m^2$, which then cancels against the delta function.

We can simplify this expression by integrating out dk^0 (which also breaks manifest Lorentz covariance). To perform the integration, we need to re-express the delta function. We note that a function $f(x)$, which satisfies $f(a) = 0$, obeys the relation:

$$\delta(f(x)) = \frac{\delta(x - a)}{|f'(a)|} \tag{3.9}$$

for x near a. (To prove this relation, simply integrate both sides over x, and then change variables. This generates a Jacobian, which explains the origin of f'.)

Since $k^2 = m^2$ has two roots, we find:

$$\delta(k^2 - m^2) = \frac{\delta(k^0 - \sqrt{\mathbf{k}^2 + m^2})}{2k^0} + \frac{\delta(k^0 + \sqrt{\mathbf{k}^2 + m^2})}{2|k^0|} \tag{3.10}$$

Putting this back into the integral, and using only the positive value of k^0, we find:

$$\begin{aligned} \int d^4k\ \delta(k^2 - m^2)\theta(k_0) &= \int d^3k \int_0^\infty \frac{dk^0}{2k^0}\,\delta\left(k_0 - \sqrt{\mathbf{k}^2 + m^2}\right) \\ &= \int \frac{d^3k}{2\omega_k}; \qquad \omega_k = \sqrt{\mathbf{k}^2 + m^2} \end{aligned} \tag{3.11}$$

Now let us insert this expression back into the Fourier decomposition of $\phi(x)$. We now find:

$$\phi(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left[a(k)e^{-ikx} + a^\dagger(k)e^{ikx}\right] \tag{3.12}$$

$$= \int d^3k\ \left[a(k)e_k(x) + a^\dagger(k)e_k^*\right] \tag{3.13}$$

$$\pi(x) = \int d^3k\ i\omega_k \left[-a(k)e_k + a^\dagger(k)e_k^*\right] \tag{3.14}$$

where:

$$e_k(x) = \frac{e^{-ik\cdot x}}{\sqrt{(2\pi)^3 2\omega_k}} \tag{3.15}$$

where $A(k) = \sqrt{2\omega_k}a(k)$ and where k^0 appearing in $k \cdot x$ is now equal to ω_k. We can also invert these relations, solving for the Fourier modes $a(k)$ in terms of the original scalar field:

$$\begin{aligned} a(k) &= i \int d^3x\ e_k^*(x) \overset{\leftrightarrow}{\partial}_0 \phi(x) \\ a^\dagger(k) &= -i \int d^3x\ e_k(x) \overset{\leftrightarrow}{\partial}_0 \phi(x) \end{aligned} \tag{3.16}$$

where:

$$A \overset{\leftrightarrow}{\partial} B \equiv A\partial B - (\partial A)B \tag{3.17}$$

Because the fields satisfy equal-time canonical commutation relations, the Fourier modes must also satisfy commutation relations:

$$[a(k), a^\dagger(k')] = \delta^3(\mathbf{k} - \mathbf{k}') \tag{3.18}$$

and all other commutators are zero. To prove that this commutation relation is consistent with the original commutator, let us insert the Fourier expansion into the equal-time commutator:

$$\begin{aligned}
\left[\phi(\mathbf{x}, t), \pi(\mathbf{x}', t)\right] &= \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \int \frac{i\omega_{k'} d^3k'}{\sqrt{(2\pi)^3 2\omega_{k'}}} \Big[a(k)e^{-ik\cdot x} \\
&\quad + a^\dagger(k)e^{ik\cdot x}, -a(k')e^{-ik'\cdot x'} + a^\dagger(k')e^{ik'\cdot x'}\Big] \\
&= \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \int \frac{d^3k'}{\sqrt{(2\pi)^3 2\omega_{k'}}} i\omega_{k'} \\
&\quad \times \delta^3(\mathbf{k} - \mathbf{k}') \left(e^{-ik\cdot x + ik'\cdot x'} + e^{ik\cdot x - ik'\cdot x'}\right) \\
&= i \int \frac{d^3k}{2(2\pi)^3} \left(e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} + e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\right) \\
&= i\delta^3(\mathbf{x} - \mathbf{x}')
\end{aligned} \tag{3.19}$$

Thus, this is a consistent choice for the commutators. Now we can calculate the Hamiltonian in terms of these Fourier modes:

$$\begin{aligned}
H &= \frac{1}{2} \int d^3x \left[\pi^2 + \partial_i\phi\partial_i\phi + m^2\phi^2\right] \\
&= \frac{1}{2} \int d^3k\, \omega_k \left[a(k)a^\dagger(k) + a^\dagger(k)a(k)\right] \\
&= \int d^3k\, \omega_k \left[a^\dagger(k)a(k) + \frac{1}{2}\right]
\end{aligned} \tag{3.20}$$

Similarly, we can calculate the momentum $\mathbf{P}$:

$$\begin{aligned}
\mathbf{P} &= -\int \pi \nabla\phi\, d^3x \\
&= \frac{1}{2} \int d^3k\, \mathbf{k} \left[a^\dagger(k)a(k) + a(k)a^\dagger(k)\right]
\end{aligned}$$

$$= \int d^3k \, \mathbf{k} \left[a^\dagger(k)a(k) + \frac{1}{2} \right] \tag{3.21}$$

(We caution that both the energy and momentum are actually divergent because of the factor $\frac{1}{2}$ appearing in the infinite sum. We will clarify this important point shortly.)

With these expressions, it is now easy to check that the operators P^μ and $M^{\mu\nu}$ generate translations and Lorentz rotations, as they should:

$$\begin{aligned} i[P_\mu, \phi] &= \partial_\mu \phi \\ i[M^{\mu\nu}, \phi] &= (x^\mu \partial^\nu - x^\nu \partial^\mu)\phi \end{aligned} \tag{3.22}$$

If we exponentiate the generators of translations and Lorentz rotations, we can calculate how the field $\phi(x)$ transforms under the Poincaré group. Let us define:

$$U(\Lambda, a) = \exp\left(-\frac{i}{2}\epsilon_{\mu\nu} M^{\mu\nu} + i a_\mu P^\mu \right) \tag{3.23}$$

where $\Lambda_{\mu\nu} = g_{\mu\nu} + \epsilon_{\mu\nu} + \cdots$. Then it is straightforward to show:

$$U(\Lambda, a)\phi(x)U^{-1}(\Lambda, a) = \phi(\Lambda x + a) \tag{3.24}$$

This demonstrates that $\phi(x)$ transforms as a scalar field under the Poincaré group.

Now that we have successfully shown how to quantize the Klein–Gordon field, we must now calculate the eigenstates of the Hamiltonian to find the spectrum of states. Let us now define the "vacuum" state as follows:

$$a(k)|0\rangle = 0 \tag{3.25}$$

By convention, we call $a(k)$ an "annihilation" operator. We define a one-particle state via the "creation" operator as a Fock space:

$$a^\dagger(k)|0\rangle = |k\rangle \tag{3.26}$$

The problem with this construction, however, is that the energy associated with the vacuum state is formally infinite because of the presence of $1/2$ in the sum in Eq. (3.20). We will simply drop this infinite term, since infinite shifts in the Hamiltonian cannot be measured. Dropping the zero point energy in the expression for harmonic oscillators has a simple counterpart in x space. The zero-point energy emerged when we commuted creation and annihilation operators past each other. Dropping the zero-point energy is therefore equivalent to moving all creation operators to the left and annihilation operators to the right. This operation,

in x space, can be accomplished by "normal ordering." Since the product of two or more fields at the same point is formally divergent, we can remove this divergence by the normal ordering operation, which corresponds to moving the part containing the creation operators to the left and the annihilation operators to the right. If we decompose $\phi = \phi^+ + \phi^-$, where $-$ (+) represent the creation (annihilation) part of an operator with negative (positive) frequency, then we define:

$$: \phi_1 \phi_2 :\equiv \phi_1^+ \phi_2^+ + \phi_1^- \phi_2^+ + \phi_1^- \phi_2^- + \phi_2^- \phi_1^+ \tag{3.27}$$

Then, by applying the normal ordering to the definition of the Hamiltonian, we can simply drop the $\frac{1}{2}$ appearing in Eq. (3.20). From now on, we assume that when two fields are multiplied at the same point in space–time, they are automatically normal ordered.

Once we have normal ordered the operators, we now have an explicitly positive Hamiltonian. In this fashion, we have been able to handle the question of negative energy states for the Klein–Gordon theory. (More subtleties concerning negative energy states will be discussed when we analyze the Dirac equation.)

One essential point in introducing these creation and annihilation operators is that we can write down the N-particle Fock space:

$$|k_1, k_2, \cdots, k_N\rangle = a^\dagger(k_1) a^\dagger(k_2) \cdots a^\dagger(k_N)|0\rangle \tag{3.28}$$

This is the chief distinguishing feature between first and second quantization. In first quantization, we quantized the x_i corresponding to a single particle. First quantized systems were hence inherently based on single-particle dynamics. In the second quantized formalism, by contrast, we quantize *multiparticle states*. To count how many particles we have of a certain momentum, we introduce the "number" operator:

$$N = \int d^3k \, a^\dagger(k) a(k) \tag{3.29}$$

The advantage of this number operator is that we can now calculate how many particles there are of a certain momentum. For example, let $|n(k)\rangle$ equal a state consisting of $n(k)$ identical particles with momentum k:

$$|n(k)\rangle = \frac{a^\dagger(k)^{n(k)}}{\sqrt{n(k)!}}|0\rangle \tag{3.30}$$

It is easy to show [by commuting $a(k)$ to the right, until they annihilate on the vacuum] that:

$$N|n(k)\rangle = n(k)|n(k)\rangle \tag{3.31}$$

that is, N simply counts the number of states there are at momentum k. Not surprisingly, a multiparticle state, consisting of many particles of different momenta, can be represented as a Fock space:

$$|n(k_1)n(k_2)\cdots n(k_m)\rangle = \prod_{i=1}^{m} \frac{\left(a^\dagger(k_i)\right)^{n(k_i)}}{\sqrt{n(k_i)!}}|0\rangle \tag{3.32}$$

Then the number operator N acting on this multiparticle state just counts the number of particles present:

$$N|n(k_1)n(k_2)\cdots n(k_m)\rangle = \left(\sum_{i=1}^{m} n(k_i)\right)|n(k_1)n(k_2)\cdots n(k_m)\rangle \tag{3.33}$$

Finally, it is essential to notice that the norm of these multiparticle states is positive. If we define $\langle k| \equiv \langle 0|a(k)$ and set $\langle 0|0\rangle = 1$, then the norm is given by $\langle k|k'\rangle = +\delta^3(\mathbf{k}-\mathbf{k}')$. The norm is positive because the appropriate sign appears in the commutation relation, Eq. (3.18). If the sign of the commutator had been reversed and the norm were negative, then we would have a negative norm state, or "ghost" state, which would give us negative probabilities and would violate unitarity. (For example, we would not be able to write the completeness statement $\mathbf{1} = \sum_n |n\rangle\langle n|$ which is used in unitarity arguments.) To preserve unitarity, it is essential that a physical theory be totally free of ghost states (or that they cancel completely). We will encounter this important question of ghosts repeatedly throughout this book.

3.3 Charged Scalar Field

We can generalize our discussion of the Klein–Gordon field by postulating the existence of several scalar fields. In particular, we can arrange two independent scalar fields into a single complex field:

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{3.34}$$

The action then becomes:

$$\mathscr{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi \tag{3.35}$$

If we insert this decomposition into the action, then we find the sum of two independent actions for ϕ_1 and ϕ_2.

The quantization of this action proceeds as before by calculating the conjugate field and postulating the canonical commutation relations. The conjugate field is given by:

$$\pi = \frac{\delta \mathscr{L}}{\delta \dot{\phi}} = \dot{\phi}^\dagger \tag{3.36}$$

The commutation relations now read:

$$[\phi(\mathbf{x}, t), \pi(\mathbf{y}, t)] = i\delta^3(\mathbf{x} - \mathbf{y}) \tag{3.37}$$

We can always decompose this field in terms of its Fourier components:

$$\phi_i(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left(a_i(k)e^{-ik\cdot x} + a_i^\dagger(k)e^{ik\cdot x}\right) \tag{3.38}$$

Then the canonical commutation relations can be satisfied if the Fourier components obey the following commutation relations:

$$[a_i(k), a_j^\dagger(k')] = \delta^3(\mathbf{k} - \mathbf{k}')\delta_{ij} \tag{3.39}$$

All other commutators vanish. We could also choose the decomposition:

$$\begin{aligned} a(k) &= \frac{1}{\sqrt{2}}[a_1(k) + ia_2(k)]; \quad a^\dagger(k) = \frac{1}{\sqrt{2}}\left[a_1^\dagger(k) - ia_2^\dagger(k)\right] \\ b(k) &= \frac{1}{\sqrt{2}}[a_1(k) - ia_2(k)]; \quad b^\dagger(k) = \frac{1}{\sqrt{2}}\left[a_1^\dagger(k) + ia_2^\dagger(k)\right] \end{aligned} \tag{3.40}$$

For these operators, the new commutation relations read:

$$[a(k), a^\dagger(k')] = [b(k), b^\dagger(k')] = \delta^3(\mathbf{k} - \mathbf{k}') \tag{3.41}$$

All other commutators are zero. Now let us construct the symmetries of the action and the corresponding Noether currents. The action is symmetric under the following transformation:

$$\phi \to e^{i\theta}\phi; \quad \phi^\dagger \to e^{-i\theta}\phi^\dagger \tag{3.42}$$

which generates a $U(1)$ symmetry. Written out in components, we find, as in the previous chapter, the following $SO(2)$ transformation:

$$\begin{pmatrix} \phi_1' \\ \phi_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \tag{3.43}$$

This symmetry generates a Noether current, which equals:

$$J_\mu = i\phi^\dagger \partial_\mu \phi - i\partial_\mu \phi^\dagger \phi \tag{3.44}$$

Now let us calculate the charge Q corresponding to this current in terms of the quantized operators:

$$\begin{aligned} Q &= \int d^3x\, i(\phi^\dagger \dot{\phi} - \dot{\phi}^\dagger \phi) \\ &= \int d^3k \left[a^\dagger(k)a(k) - b^\dagger(k)b(k)\right] \\ &= N_a - N_b \end{aligned} \tag{3.45}$$

where the number operator for the a and b oscillator is given by:

$$N_a = \int d^3k\, a^\dagger(k)a(k); \quad N_b = \int d^3k\, b^\dagger(k)b(k) \tag{3.46}$$

Historically, this conserved current caused a certain amount of confusion. If J^0 is considered to be the probability density of the wave function, then it can be negative, and hence negative probabilities creep into the theory. In fact, Schrödinger originally studied this equation as a candidate for the theory of the electron but abandoned it because of these negative probabilities. As a consequence, he later went on to write another equation that did not suffer from this problem, the celebrated nonrelativistic Schrödinger equation.

However, in 1934 Pauli and Weisskopf[7] finally gave the correct quantum interpretation of these negative probabilities.

First, because of the crucial minus sign appearing in front of the b oscillators, we will find it convenient to redefine the current J_μ as the current corresponding to the *electric charge*, rather than probability density, so that the a oscillators correspond to a positively charged particle and the b oscillators correspond to a negatively charged one. In this way, we can construct the quantum theory of charged scalar particles, where the minus sign appearing in the current is a desirable feature, rather than a fatal illness of the theory. In the next chapter, we will show how to couple this theory to the Maxwell field and hence rigorously show how this identification works.

Second, we will interpret the $b^\dagger$ oscillator as the creation operator for a new state of matter, antimatter. It was Dirac who originally grappled with these new states found in any relativistic theory and deduced the fact that a new form of matter, with opposite charge, must be given serious physical consideration. The discovery of the antielectron gave graphic experimental proof of this conjecture.

Third, we no longer have a simple, single-particle interpretation of the $\phi(x)$, which now contains both the matter and antimatter fields. Thus, we must abandon the strict single-particle interpretation for ϕ and reinterpret it as a field. We will see this unusual feature emerging again when we discuss the Dirac equation.

3.4 Propagator Theory

Now that we have defined the canonical commutation relations among the particle fields, we are interested in how these particles actually move in space–time. To define a propagator, and also anticipate interactions, let us modify the Klein–Gordon equation to include a source term $J(x)$:

$$(\partial_\mu^2 + m^2)\phi(x) = J(x); \partial_\mu^2 \equiv \partial_\mu \partial^\mu \tag{3.47}$$

To solve this equation, we use the standard theory of Green's functions. We first define a propagator that satisfies:

$$(\partial_\mu^2 + m^2)\Delta_F(x - y) = -\delta^4(x - y) \tag{3.48}$$

Then the solution of the interacting ϕ field is given by:

$$\phi(x) = \phi_0(x) - \int d^4x \, \Delta_F(x - y)J(y) \tag{3.49}$$

where $\phi_0(x)$ is any function that satisfies the Klein–Gordon equation without any source term. If we hit both sides of this expression with $(\partial_\mu^2 + m^2)$, then we find that it satisfies the original Klein–Gordon equation in the presence of a source term. As we know from the theory of Green's functions, the way to solve this equation is to take the Fourier transform:

$$\Delta_F(x - y) = \int \frac{d^4k}{(2\pi)^4} e^{-ik(x-y)} \Delta_F(k) \tag{3.50}$$

If we hit both sides of this equation with $(\partial_\mu^2 + m^2)$, then we can solve for $\Delta(k)$:

$$\Delta_F(k) = \frac{1}{k^2 - m^2} \tag{3.51}$$

At this point, however, we realize that there is an ambiguity in this equation. The integral over d^4k cannot be performed on the real axis, because the denominator diverges at $k_\mu^2 = m^2$. This same ambiguity, of course, occurs even in the classical

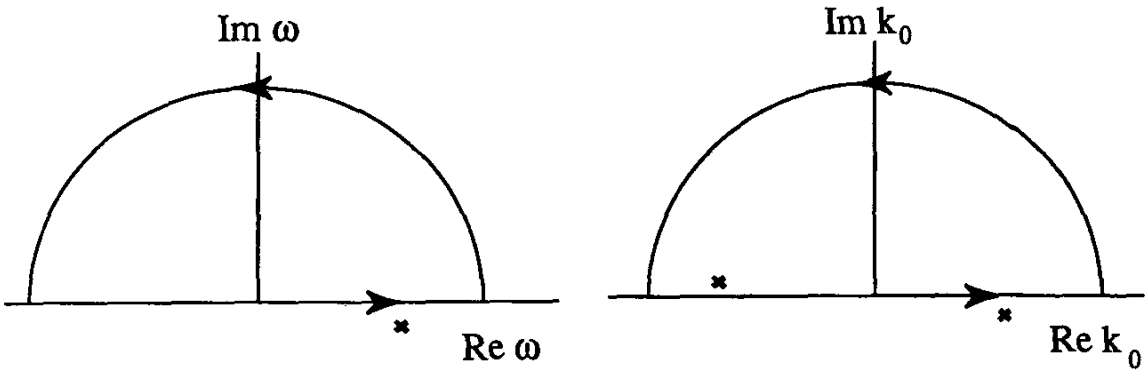


Figure 3.1. Contour integration for the Green's function. The contour on the left gives us the nonrelativistic retarded Green's functions, while the contour on the right gives us the Feynman prescription for a relativistic Green's function.

theory of wave equations, and is not specific to the Lorentz covariant theory. The origin of this ambiguity lies not in the mathematics, but in the physics, in the fact that we have yet to fix our boundary conditions.

For example, consider the Green's function for the Schrödinger equation:

$$\left(i\frac{\partial}{\partial t} - H_0\right) G_0(x - x') = \delta^4(x - x') \tag{3.52}$$

If we take the Fourier transform of this equation and solve for the Green's function, we find:

$$G_0(x - x') = \int \frac{d^4p}{(2\pi)^4} \frac{1}{\omega - \mathbf{p}^2/2m} e^{-ip\cdot(x-x')} \tag{3.53}$$

where $p^\mu = (\omega, \mathbf{p})$. This expression also suffers from an ambiguity, because the integration over ω is divergent.

Let us take the convention that we integrate over the real ω axis as in Figure 3.1, so that we integrate above the singularity. This can be accomplished by inserting a factor of $i\epsilon$ into the denominator, replacing $\omega - p^2/2m$ with $\omega - p^2/2m + i\epsilon$. Then the ω integration can be performed.

We simply convert the integration over the real axis into a contour integration over a complex variable ω. We are going to add the contour integral over the upper half plane (which vanishes) such that:

$$\begin{aligned} G_0(x - x') &= \int \frac{d^3p}{(2\pi)^3} e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}')} \int_{-\infty}^{\infty} \frac{d\omega}{(2\pi)} \frac{e^{-i\omega(t-t')}}{\omega - p^2/2m + i\epsilon} \\ &= -i \int \frac{d^3p}{(2\pi)^3} e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}') - i(p^2/2m)(t-t')} \theta(t - t') \end{aligned} \tag{3.54}$$

where the θ function has been written as:

$$\begin{aligned}\theta(t-t') &= -\lim_{\epsilon\to 0}\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{e^{-i\omega(t-t')}\,d\omega}{\omega+i\epsilon} \\ &= \begin{cases} 1 & \text{if } t > t' \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{3.55}$$

which equals 1 for $t > t'$, and vanishes otherwise. (To prove this last relation, extend the contour integral into a semicircle in the complex ω plane, closing the contour in the lower half plane when $t > t'$. Then the contour integral picks up the pole at $\omega = -i\epsilon$.)

To see this a bit more explicitly, let us define:

$$\phi_p(x) \equiv \frac{e^{-ip\cdot x}}{(2\pi)^{3/2}} \tag{3.56}$$

Then the Green's function can be written as:

$$G_0(x-x') = -i\theta(t-t')\int d^3p\ \phi_p(x)\phi_p^*(x') \tag{3.57}$$

In this way, the $+i\epsilon$ insertion has selected out the *retarded* Green's function, which obeys the usual concept of causality. Taking the $-i\epsilon$ prescription would have given us the *advanced* Green's function, which would violate causality.

Finally, taking the d^3p integration (which is simply a Gaussian integral), we find the final result for the Green's function:

$$G_0(x-x') = -i\left(\frac{m}{2\pi i(t-t')}\right)^{3/2}\exp\left\{\frac{im|\mathbf{x}-\mathbf{x}'|^2}{2(t-t')}\right\}\theta(t-t') \tag{3.58}$$

which is just the Green's function found in ordinary quantum mechanics.

Now that we have seen how various prescriptions for $i\epsilon$ give us various boundary conditions, let us apply this knowledge to the relativistic case and choose the following, unorthodox prescription:

$$\Delta_F(k) = \frac{1}{k^2-m^2+i\epsilon} \tag{3.59}$$

To see how this $i\epsilon$ prescription modifies the boundary conditions, we will find it useful to decompose this as:

$$\frac{1}{k^2-m^2+i\epsilon} = \frac{1}{2k^0}\left(\frac{1}{k^0-\sqrt{m^2+\mathbf{k}^2}+i\epsilon}+\frac{1}{k^0+\sqrt{m^2+\mathbf{k}^2}-i\epsilon}\right) \tag{3.60}$$

The integral over k^0 now picks up contributions from both terms. Performing the integration as before, we find:

$$\begin{aligned}\Delta_F(x-x') &= -i\theta(t-t')\int \frac{d^3k}{(2\pi)^3 2\omega_k} e^{-ik\cdot(x-x')} \\ &\quad -i\theta(t'-t)\int \frac{d^3k}{(2\pi)^3 2\omega_k} e^{ik\cdot(x-x')} \\ &= -i\theta(t-t')\int \frac{d^3k}{(2\pi)^3 2\omega_k} \phi_p(x)\phi_p^*(x') \\ &\quad -i\theta(t'-t)\int \frac{d^3k}{(2\pi)^3 2\omega_k} \phi_p(x)^*\phi_p(x') \qquad (3.61)\end{aligned}$$

We see the rather unusual feature of this prescription: positive energy solutions are carried forward in time, but negative energy solutions are carried *backwards* in time.

In classical physics, the usual solutions of the Maxwell theory give us retarded and advanced waves, and we eliminate the advanced waves by a choice of boundary conditions. However, in the quantum theory we are encountering a new type of propagator that, classically, makes no sense, with negative energy solutions going backwards in time. This propagator never appears in classical physics because it is complex and is hence forbidden.

Quantum mechanically, negative energy solutions are an inherent problem with any relativistic theory. Even if we ban them at the beginning, quantum interactions will inevitably re-create them later. However, as we saw in the previous section, these negative energy states can be reinterpreted. Feynman's approach to this problem was to assume that these negative energy states, because they are going backwards in time, appear as a new form of matter with positive energy going forwards in time, antimatter. Although matter going backwards in time seems to contradict causality, this poses no problem because one can show that, experimentally, a system where matter is going backwards in time is indistinguishable (if we reverse certain quantum numbers such as charge) from antimatter going forwards in time. For example, an electron placed in an electric field may move to the right; however, if it is moving backwards in time, it appears to move to the left. However, this is indistinguishable experimentally from a positively charged electron moving forwards in time to the left. In this way, we can interpret this theory as one in which everything (matter plus antimatter) has positive energy. (We will discuss this new reinterpretation further when we analyze the Dirac theory.)

The previous expression for the Green's function was written in terms of plane waves ϕ_p. However, we can replace the plane wave ϕ_p by the quantum field $\phi(x)$ if we take the vacuum expectation value of the product of fields. From the

previous equation Eq. (3.61), we easily find:

$$i\Delta_F(x-x') = \langle 0|T\phi(x)\phi(x')|0\rangle \tag{3.62}$$

where T is called the time-ordered operator, defined as:

$$T\phi(x)\phi(x') = \begin{cases} \phi(x)\phi(x') & \text{if } t > t' \\ \phi(x')\phi(x) & \text{if } t' > t \end{cases} \tag{3.63}$$

T makes sure that the operator with the latest time component always appears to the left. This equation for Δ_F is our most important result for propagators. It gives us a bridge between the theory of propagators, in which scattering amplitudes are written in terms of $\Delta_F(x-x')$, and the theory of operators, where everything is written in terms of quantum field $\phi(x)$. The previous expression will be crucial to our discussion when we calculate the S matrix for QED.

Finally, we remark that our theory must obey the laws of causality. For our purposes, we will define microscopic causality as the statement that information cannot travel faster than the speed of light. For field theory, this means that $\phi(x)$ and $\phi(y)$ cannot interact with each other if they are separated by space-like distances. Mathematically, this means that the commutator between these two fields must vanish for space-like separations.

Repeating the earlier steps, we can show that this commutator equals:

$$\begin{aligned} [\phi(x),\phi(y)] &= i\Delta(x-y) \\ &= \int \frac{d^4k}{(2\pi)^3}\delta(k^2-m^2)\epsilon(k_0)e^{-ik(x-y)} \\ &= \frac{i}{4\pi r}\frac{\partial}{\partial r}\begin{cases} J_0(m\sqrt{t^2-r^2}) & t > r \\ 0 & -r < t < r \\ -J_0(m\sqrt{t^2-r^2}) & t < -r \end{cases} \end{aligned} \tag{3.64}$$

where $\epsilon(k)$ equals $+1(-1)$ for positive (negative) k, $t = x^0 - y^0$, $r = |\mathbf{x}-\mathbf{y}|$, and J_0 is the Bessel function. (To prove this, convert the integral to radial coordinates, and then perform the k^0 and $|\mathbf{k}|$ integrations.)

With this explicit form for the commutator, we can easily show that, for space-like separations, we have:

$$\Delta(x-y) = 0 \text{ if } (x-y)^2 < 0 \tag{3.65}$$

This shows that our construction obeys microscopic causality.

3.5 Dirac Spinor Field

After Dirac considered the relativistic theory of radiation in 1927, he set out the next year to construct the relativistic theory of electrons. One severe limitation was the problem of negative probabilities. He started with the observation that the nonrelativistic Schrödinger equation did not have negative probabilities because it was linear in time, while the Klein–Gordon equation, being quadratic in time, did have negative probabilities.

Therefore Dirac tried to find a wave equation that was linear in time but still satisfied the relativistic mass-shell constraint:

$$p_\mu p^\mu = E^2 - \mathbf{p}^2 = m^2 \tag{3.66}$$

Dirac's original idea was to take the "square root" of the energy equation. In this way, he stumbled onto the spinorial representation of the Lorentz group discussed in Chapter 2. He began with a first-order equation:

$$i\frac{\partial \psi}{\partial t} = \left(-i\alpha_i \nabla^i + \beta m\right)\psi \tag{3.67}$$

where α_i and β are now constant matrices, not ordinary c numbers, which act on ψ, a column matrix.

By squaring the operator in front of the ψ field, we want to recover the mass-shell condition:

$$\begin{aligned} -\frac{\partial^2}{\partial t^2}\psi &= (-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + \beta m)^2\,\psi \\ &= (-\nabla^2 + m^2)\psi \end{aligned} \tag{3.68}$$

This is only possible if we demand that the matrices satisfy:

$$\begin{aligned} \{\alpha_i, \alpha_k\} &= 2\delta_{ik} \\ \{\alpha_i, \beta\} &= 0 \\ \alpha_i^2 &= \beta^2 = 1 \end{aligned} \tag{3.69}$$

To make the equations more symmetrical, we can then define $\gamma^0 = \beta$ and $\gamma^i = \beta\alpha^i$. Multiplying the wave equation by β, we then have the celebrated Dirac equation[8]:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \tag{3.70}$$

where the γ^μ matrices satisfy:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \tag{3.71}$$

It is no accident that such a relativistic construction is possible. After all, in the previous chapter we studied spinor representations of $O(N)$ in Section 2.6 by defining a Clifford algebra, which is precisely the algebra formed by the γ^μ. Thus, what we are really constructing is the spin $\frac{1}{2}$ representation of the Lorentz group, that is, the spinors.

To calculate the behavior of this equation under the Lorentz group, let us define how spinors transform under some representation $S(\Lambda)$ of the Lorentz group:

$$\psi'(x') = S(\Lambda)\psi(x) \tag{3.72}$$

Then the Dirac equation transforms as follows:

$$[iS^{-1}(\Lambda)\gamma^\mu S(\Lambda)\partial_\mu - m]\psi = [i\gamma^\mu(\Lambda)^\nu_\mu\partial_\nu - m]\psi = 0 \tag{3.73}$$

where we have multiplied the transformed Dirac equation by $S^{-1}(\Lambda)$ on the left, and we have taken into account the transformed $\partial'_\mu = (\Lambda)^\nu_\mu\partial_\nu$. In order for the equation to be Lorentz covariant, we must therefore have the following relation:

$$S(\Lambda)\gamma^\mu S^{-1}(\Lambda) = (\Lambda^{-1})^\mu_\nu\gamma^\nu \tag{3.74}$$

which we first encountered in Section 2.6. To find an explicit representation for $S(\Lambda)$, let us introduce the following matrix:

$$\sigma_{\mu\nu} = \frac{i}{2}[\gamma_\mu, \gamma_\nu] \tag{3.75}$$

In Chapter 2, we saw that $(i/4)[\Gamma_\mu, \Gamma_\nu]$ are the generators of $O(N)$ in the spinor representation. Thus, the $\sigma_{\mu\nu}/2$ are the generators of the Lorentz group in this representation.

Thus, we can write a new Lorentz group generator that is the sum of the old generator $L_{\mu\nu}$ (which acts on the space–time coordinate) plus a new piece that also generates the Lorentz group but in the spinor representation:

$$M_{\mu\nu} = L_{\mu\nu} + \frac{1}{2}\sigma_{\mu\nu} \tag{3.76}$$

The $\sigma_{\mu\nu}$ also obey the following relation:

$$[\gamma^\mu, \sigma_{\alpha\beta}] = 2i(\delta^\mu_\alpha\gamma_\beta - \delta^\mu_\beta\gamma_\alpha) \tag{3.77}$$

which shows that the Dirac matrices transform as vectors under the spinor representation of the Lorentz group.

In terms of this new matrix, we can find an explicit representation of the $S(\Lambda)$ matrix:

$$S(\Lambda) = e^{-(i/4)\sigma_{\mu\nu}\omega^{\mu\nu}} \tag{3.78}$$

Now that we know how spinors transform under the Lorentz group, we would like next to construct invariants under the group. Let us take the Hermitian conjugate of the Dirac equation:

$$\psi^\dagger(i\gamma^{\dagger\mu}\overleftarrow{\partial}_\mu + m) = 0 \tag{3.79}$$

We will show shortly that there exists a representation of the Dirac matrices that satisfies:

$$\begin{aligned}(\gamma^0)^\dagger &= \gamma^0 \\ (\gamma^i)^\dagger &= -\gamma^i\end{aligned} \tag{3.80}$$

where γ^i is anti-Hermitian and γ^0 is Hermitian. This can also be written as $\gamma^{\mu\dagger} = \gamma_\mu$. (It may be puzzling that we did not take a representation that was completely Hermitian. However, as we mentioned earlier, there are no finite-dimensional unitary representations of the Lorentz group. If a purely Hermitian representation of the Dirac matrices could be found, then we could construct the generators of the Lorentz group out of them that would be unitary, violating this theorem. Thus, we are forced to take non-Hermitian representations.)

Now let us define:

$$\bar{\psi} \equiv \psi^\dagger\gamma^0 \tag{3.81}$$

If we hit the conjugated equation of motion with γ^0, we can replace the $\gamma^\dagger$ with γ matrices, leaving us with:

$$\bar{\psi}(i\gamma^\mu\overleftarrow{\partial}_\mu + m) = 0 \tag{3.82}$$

Under a Lorentz transformation, this new field $\bar{\psi}$ obeys:

$$\begin{aligned}\bar{\psi}'(x') &= \bar{\psi}(x)\gamma^0 S(\Lambda)^\dagger\gamma^0 \\ &= \bar{\psi}(x)S^{-1}(x)\end{aligned} \tag{3.83}$$

This is just what we need to form invariants and covariant tensors. For example, notice that $\bar{\psi}\psi$ is an invariant under the Lorentz group:

$$\bar{\psi}'(x')\psi'(x') = \bar{\psi}(x)S^{-1}(\Lambda)S(\Lambda)\psi = \bar{\psi}(x)\psi(x) \tag{3.84}$$

Similarly, $\bar{\psi}\gamma^\mu\psi$ is a genuine vector under the Lorentz group. We find:

$$\bar{\psi}'(x')\gamma^\mu\psi'(x') = \bar{\psi}(x)S^{-1}\gamma^\mu S\psi = \bar{\psi}(x)\Lambda^\mu{}_\nu\gamma^\nu\psi(x) \tag{3.85}$$

where we have used the fact that:

$$S^{-1}\gamma^\mu S = \Lambda^\mu{}_\nu\gamma^\nu \tag{3.86}$$

which is nothing but the statement that the γ^μ transform as vectors under the spinor representation of the Lorentz group, as we saw in Chapter 2. (To prove this formula, take an infinitesimal Lorentz transformation. Then $S^{-1}\gamma^\mu S$ becomes proportional to the commutator between $\sigma_{\lambda\rho}$ and γ^μ, which is gives just another gamma matrix. If we then exponentiate this process for finite transformations, we find the previous equation, as desired.)

In the same manner, it is also straightforward to show that $\bar{\psi}\sigma^{\mu\nu}\psi$ transforms as a genuine antisymmetric second-rank tensor under the Lorentz group. To find other Lorentz tensors that can be represented as bilinears in the spinors, let us follow our discussion of Chapter 2 and introduce the matrix:

$$\gamma_5 = \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = -\frac{i}{4!}\epsilon_{\mu\nu\sigma\rho}\gamma^\mu\gamma^\nu\gamma^\sigma\gamma^\rho \tag{3.87}$$

where $\epsilon^{\mu\nu\sigma\rho} = -\epsilon_{\mu\nu\sigma\rho}$ and $\epsilon^{0123} = +1$. Because γ_5 transforms like $\epsilon^{\mu\nu\rho\sigma}$, it is a pseudoscalar; that is, it changes sign under a parity transformation. Thus, $\bar{\psi}\gamma_5\psi$ is a pseudoscalar.

In fact, the complete set of bilinears, their transformation properties, and the number of elements within each tensor are given by:

Scalar :	$\bar{\psi}\psi$	[1]
Vector :	$\bar{\psi}\gamma^\mu\psi$	[4]
Tensor :	$\bar{\psi}\sigma^{\mu\nu}\psi$	[6]
Pseudovector :	$\bar{\psi}\gamma_5\gamma^\mu\psi$	[4]
Pseudoscalar :	$\bar{\psi}\gamma_5\psi$	[1]

(3.88)

There is a total of 16 independent components in this table. We can show that the following 16 matrices are linearly independent;

$$\Gamma^A = \{I, \gamma_\mu, \sigma_{\mu\nu}, \gamma_5\gamma_\mu, \gamma_5\} \tag{3.89}$$

where $(\Gamma^A)^2 = \pm 1$. To show that this set of 16 matrices forms a complete set, let us assume, for the moment, that a relation exists among them, so that:

$$\sum_A c_A \Gamma^A = 0 \tag{3.90}$$

where c_A are numbers. Then multiply this by Γ^B and take the trace. If $\Gamma^B = I$, we find that $c_I = 0$. If $\Gamma^B \neq I$, then we use the fact that that there exists Γ_C not equal to unity such that $\Gamma^A\Gamma^B = \Gamma^C$ if $A \neq B$. Taking the trace, we find that $c_B = 0$. Since B was arbitrary, this means that all coefficients are zero; so these 16 matrices must be linearly independent

Because γ^μ transforms as a vector under the Lorentz group, the following Lagrangian is invariant under the Lorentz group:

$$\mathscr{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \tag{3.91}$$

This, in turn, is the Lagrangian corresponding to the Dirac equation. Variations of this equation by ψ or by $\bar{\psi}$ will generate the two versions of the Dirac equation.

Up to now, we have not said anything specific about the representation of the Dirac matrices themselves. In fact, a considerable number of identities can be derived for these matrices in four dimensions without ever mentioning a specific representation, such as:

$$\begin{aligned}
\gamma^\mu\gamma_\mu &= 4 \\
\gamma^\rho\gamma^\mu\gamma_\rho &= -2\gamma^\mu \\
\gamma^\rho\gamma^\mu\gamma^\nu\gamma_\rho &= 4g^{\mu\nu} \\
\gamma^\rho\gamma^\mu\gamma^\nu\gamma^\sigma\gamma_\rho &= -2\gamma^\sigma\gamma^\nu\gamma^\mu
\end{aligned} \tag{3.92}$$

Some trace operations can also be defined:

$$\begin{aligned}
\mathrm{Tr}(\gamma^5\gamma^\mu) &= \mathrm{Tr}\,\sigma^{\mu\nu} = \mathrm{Tr}\,\gamma^\mu\gamma^\nu\gamma^5 = 0 \\
\mathrm{Tr}(\gamma^\mu\gamma^\nu) &= 4g^{\mu\nu} \\
\mathrm{Tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) &= 4(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho}) \\
\mathrm{Tr}(\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) &= 4i\epsilon^{\mu\nu\rho\sigma}
\end{aligned} \tag{3.93}$$

In particular, this means:

$$\mathrm{Tr}(\not{a}\not{b}\not{c}\not{d}) = 4[(a\cdot b)(c\cdot d) - (a\cdot c)(b\cdot d) + (a\cdot d)(b\cdot c)] \tag{3.94}$$

where $\not{a} \equiv a_\mu \gamma^\mu$.

It is often convenient to find an explicit representation of the Dirac matrices. The most common representation of these matrices is the *Dirac representation*:

$$\begin{aligned} \gamma^0 &= \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}; \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \\ \beta &= \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}; \quad \alpha^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix} \end{aligned} \tag{3.95}$$

where σ^i are the familiar Pauli spin matrices. Then the spinor ψ is a complex-valued field with four components describing a massive, spin $\frac{1}{2}$ field.

Now let us try to decompose $\psi(x)$ into plane waves in order to begin canonical quantization. To do this, we need to find a set of independent basis spinors for ψ. We will make the obvious choice:

$$u_1(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \quad u_2(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}; \quad v_1(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \quad v_2(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.96}$$

The trick is to act upon these spinors with $S(\Lambda)$ in order to boost them up to momentum p. The momentum-dependent spinors are given by:

$$\begin{aligned} u_\alpha(p) &= S(\Lambda)u_\alpha(0) \\ v_\alpha(p) &= S(\Lambda)v_\alpha(0) \end{aligned} \tag{3.97}$$

which can be shown to obey:

$$\begin{aligned} (\gamma\cdot p - m)u(p) &= 0 \\ (\gamma\cdot p + m)v(p) &= 0 \\ \bar{u}(p)(\gamma\cdot p - m) &= 0 \\ \bar{v}(p)(\gamma\cdot p + m) &= 0 \end{aligned} \tag{3.98}$$

The Lorentz transformation matrix $S(\Lambda)$ is not difficult to construct if we set all rotations to zero, leaving us with only Lorentz boosts. Then the only generators

we have are the K generators, which in turn are proportional to σ^i. Specifically, we have:

$$S(\Lambda) = \begin{pmatrix} \cosh(\phi/2) & \boldsymbol{\sigma}\cdot\mathbf{n}\sinh(\phi/2) \\ \boldsymbol{\sigma}\cdot\mathbf{n}\sinh(\phi/2) & \cosh(\phi/2) \end{pmatrix}$$

$$= \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \mathbf{1} & \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m} \\ \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m} & \mathbf{1} \end{pmatrix}$$

where $\cosh(\phi/2) = [(E+m)/2m]^{1/2}$ and $\sinh(\phi/2) = [(E-m)/2m]^{1/2}$.

Applying $S(\Lambda)$ to the independent spinor basis, we easily find:

$$u_1(p) = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 1 \\ 0 \\ \frac{p_z}{E+m} \\ \frac{p_+}{E+m} \end{pmatrix}, \quad u_2(p) = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 0 \\ 1 \\ \frac{p_-}{E+m} \\ \frac{-p_z}{E+m} \end{pmatrix} \tag{3.99}$$

$$v_1(p) = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_z}{E+m} \\ \frac{p_+}{E+m} \\ 1 \\ 0 \end{pmatrix}, \quad v_2(p) = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_-}{E+m} \\ \frac{-p_z}{E+m} \\ 0 \\ 1 \end{pmatrix} \tag{3.100}$$

where $p_\pm = p_x \pm ip_y$.

Because of the particular decomposition we have chosen, the u spinors correspond to electrons with positive energy particles (moving forwards in time), while the v spinors correspond to electrons with negative energy (moving backwards in time).

Next, we would like to describe spinors of definite spin. In many experiments, we can produce polarized beams of electrons; so it becomes important to understand how to incorporate projection operators that can select definite spin.

This is not as simple as one might suspect, since the intuitive concept of spin is rooted in our notion of the rotation group, which is only a subgroup of the Lorentz group. Hence, the naive concept of spin and its eigenfunctions no longer applies for boosted systems.

In the rest frame, however, we know that the spin of a system can be described by a three-vector $\mathbf{s}$ that points in a certain direction; so we may introduce the four-vector s_μ which, in its rest frame, reduces to $s_\mu = (0, \mathbf{s})$. Then, by demanding that this transform as a four-vector, we can boost this spin vector by a Lorentz transformation. Since we define $\mathbf{s}^2 = 1$, this means that $s_\mu^2 = -1$. In the rest frame,

we have $p_\mu = (m, 0)$; so we also have $p_\mu s^\mu = 0$, which must also hold in any boosted frame by Lorentz invariance. Thus, we now have two Lorentz-invariant conditions on the spin four-vector:

$$\begin{aligned} s_\mu^2 &= -1 \\ p_\mu s^\mu &= 0 \end{aligned} \tag{3.101}$$

Next, we would like to define a projection operator that selects out states of definite spin. Again, we will define the Lorentz-invariant projection operator by first examining the rest frame. At rest, we know that the operator $\boldsymbol{\sigma} \cdot \mathbf{s}$ serves as an operator that determines the spin of a system:

$$\begin{aligned} \boldsymbol{\sigma} \cdot \mathbf{s} u_\alpha(0) &= u_\alpha(0) \\ \boldsymbol{\sigma} \cdot \mathbf{s} v_\alpha(0) &= -v_\alpha(0) \end{aligned} \tag{3.102}$$

For a spin $\frac{1}{2}$ system, the projection operator at rest can be written as:

$$P(\mathbf{s}) = \frac{1 \pm \boldsymbol{\sigma} \cdot \mathbf{s}}{2} \tag{3.103}$$

where the + refers to the u spinor, and the $-$ refers to the v spinor. Our goal is to write a boosted version of this expression. Let us define the projection operator:

$$P(s) \equiv \frac{1 + \gamma_5 \not{s}}{2} \tag{3.104}$$

It is easy to show that, in the rest frame, this projection reduces to:

$$P(s) = \frac{1}{2} \begin{pmatrix} 1 + \boldsymbol{\sigma} \cdot \mathbf{s} & 0 \\ 0 & 1 - \boldsymbol{\sigma} \cdot \mathbf{s} \end{pmatrix} \tag{3.105}$$

Therefore, this operator reduces to the previous one, so this is the desired expression. The new eigenfunctions now have a spin s associated with them: $u(k, s)$. They satisfy:

$$\begin{aligned} P(s)u(k, s) &= u(k, s) \\ P(s)v(k, s) &= v(k, s) \\ P(-s)u(k, s) &= P(-s)v(k, s) = 0 \end{aligned} \tag{3.106}$$

These spinors are quite useful for practical calculations because they satisfy certain *completeness relations*. Any four-spinor can be written in terms of linear

combinations of the four $u_\alpha(0)$ and $v_\beta(0)$ because they span the space of four-spinors. If we boost these spinors with $S(\Lambda)$, then $u_\alpha(p)$ and $v_\beta(p)$ span the space of all four-spinors satisfying the Dirac equation.

Likewise, $u_\alpha^T(0)v_\beta(0)$, etc. have 16 independent elements, which in turn span the entire space of 4×4 matrices. Thus, $\bar{u}_\alpha(p)v_\beta(p)$, etc. span the space of all 4×4 matrices that also satisfy the Dirac equation.

We first normalize our spinors with the following conventions:

$$\begin{aligned} \bar{u}(p,s)u(p,s) &= 1 \\ \bar{v}(p,s)v(p,s) &= -1 \end{aligned} \tag{3.107}$$

With these normalizations, we can show that these spinors obey certain completeness relations:

$$\sum_s u_\alpha(p,s)\bar{u}_\beta(p,s) - v_\alpha(p,s)\bar{v}_\beta(p,s) = \delta_{\alpha\beta} \tag{3.108}$$

For the particular representation we have chosen, we find:

$$u_\alpha(p,s)\bar{u}_\beta(p,s) = \left(\frac{\not{p}+m}{2m}\frac{1+\gamma_5\not{s}}{2}\right)_{\alpha\beta} \tag{3.109}$$

and:

$$v_\alpha(p,s)\bar{v}_\beta(p,s) = -\left(\frac{m-\not{p}}{2m}\frac{1+\gamma_5\not{s}}{2}\right)_{\alpha\beta} \tag{3.110}$$

If we sum over the helicity s, we have two projection operators:

$$\begin{aligned} [\Lambda_+(p)]_{\alpha\beta} &= \sum_{\pm s} u_\alpha(p,s)\bar{u}_\beta(p,s) = \left(\frac{\not{p}+m}{2m}\right)_{\alpha\beta} \\ [\Lambda_-(p)]_{\alpha\beta} &= -\sum_{\pm s} v_\alpha(p,s)\bar{v}_\beta(p,s) = \left(\frac{-\not{p}+m}{2m}\right)_{\alpha\beta} \end{aligned} \tag{3.111}$$

These projection operators satisfy:

$$\Lambda_\pm^2 = \Lambda_\pm; \quad \Lambda_+\Lambda_- = 0; \quad \Lambda_+ + \Lambda_- = 1 \tag{3.112}$$

Because of the completeness relations, $\Lambda_\pm$ has a simple interpretation: It projects out the positive or negative energy solution.

3.6 Quantizing the Spinor Field

So far, we have only discussed the classical theory. To second quantize the Dirac field, we first calculate the momentum canonically conjugate to the spinor field:

$$\pi(x) = \frac{\delta \mathscr{L}}{\delta \dot{\psi}(x)} = i\psi^\dagger \tag{3.113}$$

Let us decompose the spinor field into its Fourier moments:

$$\begin{aligned}
\psi(x) &= \int \sqrt{\frac{m}{k_0}} \frac{d^3k}{\sqrt{(2\pi)^3}} \sum_{\alpha=1,2} \left[b_\alpha(k) u_\alpha(k) e^{-ikx} + d_\alpha^\dagger(k) v_\alpha(k) e^{ikx} \right] \\
\bar{\psi}(x) &= \int \sqrt{\frac{m}{k_0}} \frac{d^3k}{\sqrt{(2\pi)^3}} \sum_{\alpha=1,2} \left[b_\alpha^\dagger(k) \bar{u}_\alpha(k) e^{ikx} + d_\alpha(k) \bar{v}_\alpha(k) e^{-ikx} \right]
\end{aligned} \tag{3.114}$$

In terms of particles and antiparticles, this particular decomposition gives the following physical interpretation:

$$\psi(x) = \begin{cases} b(p)u(p)e^{-ip\cdot x} & \text{Annihilates positive energy electron} \\ d^\dagger(p)v(p)e^{+ip\cdot x} & \text{Creates positive energy positron} \end{cases} \tag{3.115}$$

(Having $d^\dagger$ create a positive energy positron can be viewed as annihilating a negative energy electron.)

Repeating the steps we took for the scalar particle, we invert these equations and solve for the Fourier moments in terms of the fields themselves:

$$\begin{aligned}
b_\alpha(k) &= \int d^3x \; \bar{U}_k^\alpha(x) \gamma^0 \psi(x) \\
b_\alpha^\dagger(k) &= \int d^3x \; \bar{\psi}(x) \gamma^0 U_k^\alpha(x) \\
d_\alpha(k) &= \int d^3x \; \bar{\psi}(x) \gamma^0 V_k^\alpha(x) \\
d_\alpha^\dagger(k) &= \int d^3x \; \bar{V}_k^\alpha(x) \gamma^0 \psi(x)
\end{aligned} \tag{3.116}$$

where:

$$U_k(x) \equiv \sqrt{\frac{m}{k_0(2\pi)^3}} u(k) e^{-ik\cdot x}$$

$$V_k(x) \equiv \sqrt{\frac{m}{k_0(2\pi)^3}} v(k) e^{ik\cdot x} \tag{3.117}$$

Now let us insert the Fourier decomposition back into the expression for the Hamiltonian:

$$\begin{aligned} H &= \int d^3x \, (\pi \dot{\psi} - \mathscr{L}) \\ &= \int d^3x \, \bar{\psi}(i\gamma^0 \partial_0 \psi) \\ &= \int d^3k \, k_0 \sum_\alpha [b_\alpha^\dagger(k) b_\alpha(k) - d_\alpha(k) d_\alpha^\dagger(k)] \end{aligned} \tag{3.118}$$

Here we encounter a serious problem. We find that the energy of the Hamiltonian can be negative. There is, however, an important way in which this minus sign can be banished. Let us define the canonical equal-time commutation relations of the fields and conjugate fields with *anticommutators*, instead of commutators:

$$\{\psi_i(\mathbf{x}, t), \psi_j^\dagger(\mathbf{y}, t)\} = \delta^3(\mathbf{x} - \mathbf{y})\delta_{ij} \tag{3.119}$$

In order to satisfy the canonical anticommutation relations, the Fourier moments must themselves obey anticommutation relations given by:

$$\begin{aligned} \{b_\alpha(k), b_{\alpha'}^\dagger(k')\} &= \delta_{\alpha,\alpha'} \delta^3(\mathbf{k} - \mathbf{k}') \\ \{d_\alpha(k), d_{\alpha'}^\dagger(k')\} &= \delta_{\alpha,\alpha'} \delta^3(\mathbf{k} - \mathbf{k}') \end{aligned} \tag{3.120}$$

Now, if we normal order the Hamiltonian, we must also drop the infinite zero point energy, and hence:

$$\begin{aligned} H &= \int d^3k \, k_0 \sum_\alpha [b_\alpha^\dagger(k) b_\alpha(k) + d_\alpha^\dagger(k) d_\alpha(k)] \\ \mathbf{P} &= \int d^3k \, \mathbf{k} \sum_\alpha [b_\alpha^\dagger(k) b_\alpha(k) + d_\alpha^\dagger(k) d_\alpha(k)] \end{aligned} \tag{3.121}$$

Thus, the use of anticommutation relations and normal ordering nicely solves the problem of the Hamiltonian with negative energy eigenvalues.

Furthermore, the $d^\dagger$ operators can be interpreted as creation operators for antimatter (or annihilation operators for negative energy electrons). In fact, this was Dirac's original motivation for postulating antimatter in the first place. To see how this interpretation of these new states emerges, we first notice that the Dirac Lagrangian is invariant under:

$$\psi \to e^{i\Lambda} \psi; \quad \bar{\psi} \to \bar{\psi} e^{-i\Lambda} \tag{3.122}$$

Therefore, there should be a conserved current associated with this symmetry. A direct application of Noether's theorem yields:

$$J^\mu = \bar{\psi}\gamma^\mu\psi, \quad \partial_\mu J^\mu = 0 \tag{3.123}$$

which is conserved, if we use the Dirac equation.

Classically, the conserved charge is positive definite since it is proportional to $\psi^\dagger\psi$. This was, in fact, an improvement over the classical Klein–Gordon equation, where the charge could be negative. However, once we quantize the system, the Dirac charge can also be negative. The quantized charge associated with this current is given by:

$$\begin{aligned} Q &= \int d^3x\, J^0 = \int d^3x\ :\psi^\dagger\psi: \\ &= \int d^3k \sum_\alpha [b_\alpha^\dagger(k)b_\alpha(k) - d_\alpha^\dagger(k)d_\alpha(k)] \end{aligned} \tag{3.124}$$

This quantity can be negative, and hence cannot be associated with the probability density. However, we can, as in the Klein–Gordon case, interpret this as the current associated with the coupling to electromagnetism; so Q corresponds to the electric charge. In this case, the minus sign in Q is a desirable feature, because it means that $d^\dagger$ is the creation operator of antimatter, that is, a positron with opposite charge to the electron.

Again, this also means that we have to abandon the simple-minded interpretation of ψ as a single-electron wave function, since it now describes both the electron and the antielectron. The anticommutation relations also reproduce the Pauli Exclusion Principle found in quantum mechanics. Because $d_\alpha^\dagger(k)d_\alpha^\dagger(k) = 0$, only one particle can occupy a distinct energy state with definite spin. Thus, a multiparticle state is given by:

$$\prod_{i=1}^{N} d_{\alpha_i}^\dagger(k_i) \prod_{j=1}^{M} b_{\alpha_j}^\dagger(k_j)|0\rangle \tag{3.125}$$

with only one particle in any given quantum state. This is the first example of the *spin-statistics theorem*, that field theories defined with integer spin are quantized with commutators and are called *bosons*, while theories with half-integral spins are quantized with anticommutators and are called *fermions*. The existence of two types of statistics, one based on commutators (i.e., Bose–Einstein statistics) and one based on anticommutators (i.e., Fermi–Dirac), has been experimentally observed in a wide variety of physical situations and has been applied to explain the behavior of low-temperature systems and even white dwarf stars. Repeating the same steps that we used for the Klein–Gordon field, we can also compute

the energy–momentum tensor and the angular momentum tensor from Noether's theorem. It is easy to show:

$$\begin{aligned} T^{\mu\nu} &= i\bar{\psi}\gamma^{\mu}\partial^{\nu}\psi \\ \mathscr{M}^{\lambda\mu\nu} &= i\bar{\psi}\gamma^{\lambda}\left(x^{\mu}\partial^{\nu} - x^{\nu}\partial^{\mu} - \frac{i}{2}\sigma^{\mu\nu}\right)\psi \end{aligned} \tag{3.126}$$

(The Lagrangian term in the energy–momentum tensor can be dropped since the Dirac Lagrangian is zero if the equations of motion are obeyed.)

The conserved angular momentum tensor is therefore:

$$\begin{aligned} M^{\mu\nu} &= \int \mathscr{M}^{0\mu\nu}\, d^3x \\ &= \int d^3x\, i\psi^{\dagger}\left(x^{\mu}\partial^{\nu} - x^{\nu}\partial^{\mu} - \frac{i}{2}\sigma^{\mu\nu}\right)\psi \end{aligned} \tag{3.127}$$

The crucial difference between these equations and the Klein–Gordon case is that the angular momentum tensor contains an extra piece, proportional to $\sigma^{\mu\nu}$, which represents the fact that the theory has nontrivial spin $\frac{1}{2}$.

It is then easy to complete this discussion by calculating how a quantized spinor field transforms under the Poincaré group:

$$\begin{aligned} i[P_{\mu}, \psi] &= \partial_{\mu}\psi \\ i[M^{\mu\nu}, \psi] &= \left(x^{\mu}\partial^{\nu} - x^{\nu}\partial^{\mu} - \frac{i}{2}\sigma^{\mu\nu}\right)\psi \end{aligned} \tag{3.128}$$

With these operator identities, we can confirm that the quantized spinor field transforms as a spin $\frac{1}{2}$ field under the Poincaré group:

$$U(\Lambda, a)\psi_{\alpha}(x)U^{-1}(\Lambda, a) = S^{-1}(\Lambda)_{\alpha\beta}\psi_{\beta}(\Lambda x + a) \tag{3.129}$$

As we mentioned earlier, one of our fundamental assumptions about quantum field theory is that it must be causal. Not surprisingly, the spin-statistics theorem is also intimately tied to the question of microcausality, that is, that no signals can propagate faster than the speed of light. From our field theory perspective, microcausality can be interpreted to mean that the commutator (anticommutator) of two boson (fermion) fields vanishes for spacelike separations:

$$\begin{aligned} [\phi(x), \phi(y)] &= 0 \quad \text{for } (x-y)^2 < 0 \\ \{\psi(x), \bar{\psi}(y)\} &= 0 \quad \text{for } (x-y)^2 < 0 \end{aligned} \tag{3.130}$$

To demonstrate the spin-statistics theorem, let us quantize bosons with anticommutators and arrive at a contradiction. Repeating earlier steps, we find, for large separations:

$$\begin{aligned}\langle 0|\{\phi(x),\phi(y)\}|0\rangle &= \int \frac{d^3k}{(2\pi)^3 2\omega_k}\left(e^{-ik\cdot(x-y)}+e^{ik\cdot(x-y)}\right)\\ &\sim \frac{\exp[-m\sqrt{|\mathbf{x}-\mathbf{y}|^2-(x^0-y^0)^2}]}{|\mathbf{x}-\mathbf{y}|^2-(x^0-y^0)^2}\end{aligned} \tag{3.131}$$

which clearly violates our original assumption of microcausality. (Likewise, one can prove that fermions quantized with commutators violates microcausality.)

Historically, anticommutators and antimatter were introduced by Dirac, who was troubled that his theory seemed riddled with negative energy states. Since physical systems prefer the state of lowest energy, there is a finite probability that all the electrons in nature would decay into these negative energy states, thereby creating a catastrophe. To solve the problem of negative energy states, Dirac was led to postulate a radically new interpretation of the vacuum (which is consistent with Feynman's interpretation, which we have chosen in this chapter). He postulated that the vacuum consisted of an infinite sea of filled negative energy states. Ordinary matter did not suddenly radiate an infinite amount of energy and cascade down to negative energy because the negative energy sea was completely filled. By the anticommutation relations, only one electron can occupy a negative energy state at a time; so an electron could not decay into the negative energy sea if it was already filled. In this way, electrons of positive energy could not cascade in energy down to negative energy states.

However, once in a while an electron may be knocked out of the negative energy sea, creating a "hole." This hole would act as if it were a particle. Dirac noticed that the absence of an electron of charge $-|e|$ and negative energy $-E$ is equivalent to the presence of a particle of positive charge $+|e|$ and positive energy $+E$. This hole then had positive charge and the same mass as the electron. All particles therefore had positive energy: Both the original positive energy electron as well as the absence of a negative energy electron possessed positive charge E (Fig. 3.2).

Dirac postulated that this hole would correspond to a new state of matter, an antielectron (although he initially considered the possibility that the hole was a proton). The vacuum was now elevated to an infinite storehouse of negative energy matter.

Dirac's hole theory meant that a new physical process was possible, pair production, where matter appeared out of the empty vacuum. Photons could knock an electron out of its negative energy sea, leaving us with an electron and its hole, that is, an electron and an antielectron.

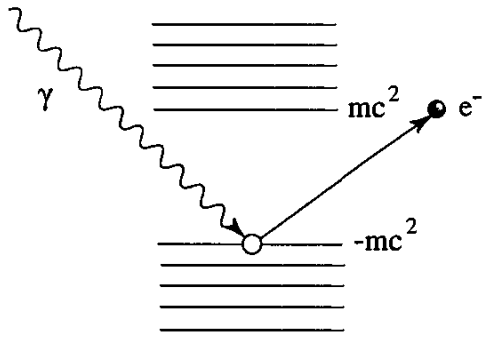


Figure 3.2. Dirac's "hole" picture. When a photon kicks an electron out of the infinite negative energy sea, it leaves a "hole" that behaves as if it had positive energy and positive charge (i.e., an anti-electron). This is pair production.

At first, Dirac's theory of an infinite sea of filled negative energy states was met with extreme skepticism. In the *Handbuch der Physik*, Pauli wrote: "Thus γ ray photons (at least two in order to satisfy the laws of conservation of energy and momentum) must be able to transform, by themselves, into an electron and an antielectron. We do not believe, therefore, that this explanation can be seriously considered." When Pauli's discouraging article appeared, Anderson had already observed the antielectron in cloud chamber photographs, verifying Dirac's conjecture. When confronted with the undeniable experimental verification of antimatter, Pauli later revised his opinion of Dirac's theory and made his famous remark, "... with his fine instinct for physical realities, he started his argument without knowing the end of it."

The interpretation that we have chosen in this chapter (that negative energy electrons going backwards in time are equivalent to positive energy antielectrons going forwards in time) is equivalent to Dirac's infinite negative energy sea. In fact, when we subtracted off an infinite constant in the Hamiltonian, this can be interpreted as subtracting off the energy of Dirac's infinite sea of negative energy states.

To see how these positive and negative energy states move in time, let us define the evolution of a wave function via a source as follows:

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = J(x) \tag{3.132}$$

To solve this equation, we introduce the Dirac propagator by:

$$(i\gamma^\mu \partial_\mu - m)S_F(x-y) = \delta^4(x-y) \tag{3.133}$$

Then the solution to the wave equation is given by:

$$\psi(x) = \psi_0(x) + \int d^4y \; S_F(x-y)J(y) \tag{3.134}$$

where ψ_0 solves the homogeneous Dirac equation.

An explicit representation of the Dirac propagator can be obtained by using the Fourier transform:

$$S_F(x-y) = \int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot(x-y)} \frac{\gamma^\mu p_\mu + m}{p^2 - m^2 + i\epsilon} \tag{3.135}$$

which satisfies:

$$S_F(x-y) = (i\gamma^\mu \partial_\mu + m)\Delta_F(x-y) \tag{3.136}$$

where $\Delta_F(x-y)$ is the Klein–Gordon propagator.

As before, we can solve for the propagator by integrating over k^0. The integration is identical to the one found earlier for the Klein–Gordon equation, except now we have additional factors of $\not{p} + m$ in the numerator.

Integrating out the energy, we can write the Green's function in terms of plane waves. The result is almost identical to the expression found for the Klein–Gordon propagator, except for the insertion of gamma matrices:

$$\begin{aligned} S_F(x-x') &= -i\int \frac{d^3p}{(2\pi)^3}\frac{m}{E}\left[\Lambda_+(p)e^{-ip\cdot(x-x')}\theta(t-t')\right. \\ &\quad \left. +\Lambda_-(p)e^{ip\cdot(x-x')}\theta(t'-t)\right] \\ &= \int d^3p \left[-i\theta(t-t')\sum_{r=1}^{2}\psi_p^r(x)\bar{\psi}_p^r(x')\right. \\ &\quad \left. +i\theta(t'-t)\sum_{r=3}^{4}\psi_p^r(x)\bar{\psi}_p^r(x')\right] \end{aligned} \tag{3.137}$$

where:

$$\psi_p^r(x) \equiv \sqrt{\frac{m}{E}}(2\pi)^{-3/2} w_r(\mathbf{p}) e^{-i\epsilon^r p\cdot x} \tag{3.138}$$

where $\epsilon^r = (1, 1, -1, -1)$ and $w_1 = u_1$, $w_2 = u_2$, $w_3 = v_1$, and $w_4 = v_2$. Written in this fashion, the states with positive energy propagate forward in time, while the states with negative energy propagate backwards in time.

As in the Klein–Gordon case, we can now replace the plane waves ψ_p with the quantized spinor field $\psi(x)$ by taking the vacuum expectation value of the spinor fields:

$$iS_F(x-y)_{\alpha\beta} = \langle 0|T\psi_\alpha(x)\bar{\psi}_\beta(y)|0\rangle \tag{3.139}$$

which is one of the most important results of this section. We will use this expression repeatedly in our discussion of scattering matrices.

3.7 Weyl Neutrinos

In the previous chapter, we saw that the spinorial representation of the Lorentz group is actually reducible if we introduce projection operators P_L and P_R. Normally, we are not concerned with this because this spinorial representation is irreducible under the full Poincaré group for massive states.

However, there is a situation when the spinorial representation becomes reducible even under the Poincaré group, and that is when the fermion is massless. For example, we can take an imaginary representation of the γ matrices, which gives us *Majorana* spinors[9] (see Appendix). For our purposes, what is more interesting is taking the *Weyl* representation,[10] which gives us a representation of neutrinos.

If we take the representation:

$$\gamma^0 = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix}; \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}; \quad \gamma_5 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \tag{3.140}$$

then in this representation, we can write down two *chiral* operators:

$$P_R = \frac{1+\gamma_5}{2} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

$$P_L = \frac{1-\gamma_5}{2} = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$$

To see how these projection operators affect the electron field, let us split ψ as follows:

$$\psi = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} \tag{3.141}$$

Then we have:

$$\frac{1+\gamma_5}{2}\psi = \begin{pmatrix} \psi_R \\ 0 \end{pmatrix}$$

$$\frac{1-\gamma_5}{2}\psi = \begin{pmatrix} 0 \\ \psi_L \end{pmatrix}$$

Because P_L and P_R commute with the Lorentz generators:

$$[P_{L,R}, \sigma^{\mu\nu}] = 0 \tag{3.142}$$

it means that the four-component spinor ψ actually spans a reducible representation of the Lorentz group. Contained within the complex Dirac representation are two distinct chiral representations of the Lorentz group. Although ψ_L and ψ_R separately form an irreducible representation of the Lorentz group, they do not form a representation of the Poincaré group for massive particles. To find an irreducible representation of the Poincaré group, we must impose $m = 0$.

The reason that these chiral fermions must be massless is because the mass term $m\bar{\psi}\psi$ in the Lagrangian is not invariant under these two separate Lorentz transformations. Because:

$$\bar{\psi}\psi = \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L \tag{3.143}$$

mass terms in the action necessarily mix these two distinct representations of the Lorentz group. Thus, this representation forces us to have massless fermions; that is, this is a theory of massless neutrinos.

The theory of massless neutrinos is therefore invariant under the following chiral transformation:

$$\begin{aligned} \psi &\rightarrow e^{i\gamma_5\Lambda}\psi \\ \bar{\psi} &\rightarrow \bar{\psi}e^{i\gamma_5\Lambda} \end{aligned} \tag{3.144}$$

This symmetry is violated by mass terms. In other words, the spinor representation of the Poincaré group (for zero-mass particles) is reducible, and we have the freedom to choose two-component rather than four-component spinor representations.

This will have important phenomenological implications later on when we consider the quark model in the limit of small quark masses. Then we can use the power of chiral symmetry in this limit to extract a large number of nontrivial relations among S matrix elements, called sum rules. In addition, the actual values of the masses of the quarks then give us a handle as to size of the violation of these chiral sum rules.

In summary, canonical quantization gives a rigorous formulation of a second quantized field theory capable of describing multiparticle states. There are other, more elegant quantization programs, but the canonical one is perhaps the most rigorous. We also saw that the Dirac equation emerged from a spinorial representation of the Lorentz group, which was developed in the previous chapter. One of the successes of the second quantized approach is that we have a physical interpretation for the negative energy states that inevitably occur in any relativistic formulation. In the next chapter, we will quantize a spin-one field and couple it to the Dirac electron theory, giving us quantum electrodynamics.

3.8 Exercises

1. Prove that the anticommutation relation of the Dirac spinors in Eq. (3.119) is satisfied if the harmonic oscillator states obey the anticommutation relation in Eq. (3.120).

2. Prove that:

$$\begin{aligned} \mathrm{Tr}(\not{a}_1 \not{a}_2 \cdots \not{a}_{2n}) &= a_1 \cdot a_2 \,\mathrm{Tr}(\not{a}_3 \cdots \not{a}_{2n}) - a_1 \cdot a_3 \,\mathrm{Tr}(\not{a}_2 \not{a}_4 \cdots \not{a}_{2n}) \\ &\quad + \cdots + a_1 \cdot a_{2n} \,\mathrm{Tr}(\not{a}_2 \cdots \not{a}_{2n-1}) \end{aligned} \tag{3.145}$$

3. Consider the 16 matrices Γ_A, where $A = S, T, V, P, A$. Show that $(\Gamma^A)^2 = \pm 1$. Show that each is traceless except for the scalar. Given Γ_A and Γ_B $(A \neq B)$, show that there exists Γ_C (not equal to unity) such that:

$$\Gamma_A \Gamma_B = \Gamma_C \tag{3.146}$$

4. By inserting the Fourier decomposition of the fields in Eq. (3.64), prove explicitly that:

$$\begin{aligned} [\phi(x), \phi(y)] &= i\Delta(x-y) \\ &= \int \frac{d^4x}{(2\pi)^3} \delta(k^2 - m^2)\epsilon(k_0) e^{-ik(x-y)} \end{aligned} \tag{3.147}$$

   Then perform the integration over k, leaving us with a Bessel function. Then show that the commutator vanishes outside the light cone, thereby establishing the causality of the system.

5. Prove Eq. (3.22) by explicitly performing the commutation relations.

6. Prove that Eqs. (3.108) and (3.109) are obeyed by explicit computation.

7. Prove the following formula, due to Fierz:

$$\sum_{D=S,V,T,A,P} c_D (\Gamma_D)_{\alpha\beta} (\Gamma_D)_{\gamma\delta} = \sum_{B=S,V,T,A,P} \tilde{c}_B (\Gamma_B)_{\alpha\delta} (\Gamma_B)_{\gamma\beta} \tag{3.148}$$

   where:

$$\begin{pmatrix} \tilde{c}_S \\ \tilde{c}_V \\ \tilde{c}_T \\ \tilde{c}_A \\ \tilde{c}_P \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & 4 & 12 & -4 & 1 \\ 1 & -2 & 0 & -2 & -1 \\ \frac{1}{2} & 0 & -2 & 0 & \frac{1}{2} \\ -1 & -2 & 0 & -2 & 1 \\ 1 & -4 & 12 & 4 & 1 \end{pmatrix} \begin{pmatrix} c_S \\ c_V \\ c_T \\ c_A \\ c_P \end{pmatrix} \tag{3.149}$$

(Hint: use the fact that the Γ_A matrices form a complete set of 16 matrices. Then treat the above expression as a matrix equation in order to power expand it in terms of the Γ_A.)

8. Use the Fierz identity to re-express $(\bar{\psi}_1 A\psi_2)(\bar{\psi}_3 B\psi_4)$ in terms of Dirac bilinears $(\bar{\psi}_1 C\psi_4)(\bar{\psi}_3 D\psi_2)$. Express the matrices C and D in terms of the matrices A and B.

9. The appearance of γ^0 within $\bar{\psi}\psi = \psi^\dagger\gamma^0\psi$ seems to violate Lorentz invariance, since γ^0 is manifestly non-invariant and transforms as the zeroth component of a vector. So why is $\bar{\psi}\psi$ still a Lorentz invariant? Furthermore, $\bar{\psi}\sigma^{\mu\nu}\psi$ clearly forms a finite dimensional representation of the Lorentz group. But this seems to be a contradiction of our no-go theorem. Is this so? If not, then why not?

10. Let $u_i(p)$ be spinors which satisfy the Dirac equation. Then prove the Gordon formula:

$$\bar{u}(p_2)\gamma^\mu u(p_1) = \bar{u}(p_2)\left(\frac{p_2^\mu + p_1^\mu}{2m} + \frac{i\sigma^{\mu\nu}q_\nu}{2m}\right)u(p_1) \tag{3.150}$$

where q_μ is the momentum transfer.

11. We define brackets to mean summing over all antisymmetric combinations of indices (see (A.12)):

$$\gamma^{\mu_1\mu_2\cdots\mu_N} \equiv \frac{1}{N!}\gamma^{[\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_N]} \tag{3.151}$$

In an arbitrary number of space–time dimensions, prove that:

$$\gamma^\mu\gamma^{\mu_1\mu_2\cdots\mu_N} = \gamma^{\mu\mu_1\mu_2\cdots\mu_N} + \sum_{i=1}^{N}(-1)^{i+1}g^{\mu\mu_i}\gamma^{\mu_1\cdots\hat{\mu}_i\cdots\mu_N} \tag{3.152}$$

where $\hat{\mu}_i$ means that the μ_i index is to be deleted in the sum. Prove:

$$\begin{aligned}\gamma^{\mu\nu}\gamma^{\mu_1\cdots\mu_N} &= \gamma^{\mu\nu\mu_1\cdots\mu_N} + \sum_i(-1)^{i+1}g^{\mu_i[\mu}\gamma^{\nu]\mu_1\cdots\hat{\mu}_i\cdots\mu_N} \\ &+ \sum_i(-1)^{i+j}g^{\mu_i\mu_j;\mu\nu}\gamma^{\mu_1\cdots\hat{\mu}_i\cdots\hat{\mu}_j\cdots\mu_N}\end{aligned} \tag{3.153}$$

where $g^{\mu\nu;\rho\sigma} = g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}$.

12. Based on the previous problem, derive a formula for:

$$\gamma^{\nu_1\nu_2\cdots\nu_N}\gamma^{\mu_1\mu_2\cdots\mu_M} \tag{3.154}$$

Prove it by induction.

13. Prove, by direct computation, that the Hamiltonian and the charge can be written in terms of Dirac harmonic oscillators as in Eqs. (3.118) and (3.124).

14. The Feynman propagator in x space can actually be computed analytically. Set $m = 0$, use radial coordinates, and show that the Feynman propagator in x space can be written in terms of Bessel functions.

15. Consider a unitary transformation U, with $H' = UHU^\dagger$ and $\psi' = U\psi$, which changes the Dirac equation to:

$$i\frac{\partial \psi'}{\partial t} = H'\psi' \tag{3.155}$$

Let U be given by:

$$U = \sqrt{\frac{m + |E|}{2|E|}} + \frac{\beta \boldsymbol{\alpha} \cdot \mathbf{p}}{\sqrt{2|E|(m + |E|)}} \tag{3.156}$$

Show that U removes all coupling between the positive and negative energy parts of the Dirac equation. This is an example of the Foldy–Wouthuysen transformation.

16. Prove that the asymptotic behavior of the anticommutator appearing in Eq. (3.131) is correct. Repeat the same calculation to show that spinors quantized with commutators also violate the spin-statistics theorem.

17. Prove that the derivative of a theta function gives us a Dirac delta function:

$$\frac{\partial}{\partial t}\theta(t - t') = \delta(t - t') \tag{3.157}$$

From this, prove that:

$$(\partial_\mu^2 + m^2)T\phi(x)\phi(x') = -i\,\delta^4(x - x') \tag{3.158}$$

Show that this equation is compatible with the expression for Δ_F in terms of the time-ordered product of two scalar fields.

18. The Feynman propagator $\Delta_F(x - y)$ can be expressed as the vacuum expectation value of the time-ordered product of two fields. The time-ordering operator T appears to violate Lorentz invariance, since x^0 is singled out. Why is the expression still Lorentz invariant?

Chapter 4
Quantum Electrodynamics

It was found that this equation gave the particle a spin of half a quantum. And also gave it a magnetic moment. It gave us the properties that one needed for an electron. That was really an unexpected bonus for me, completely unexpected.

—P.A.M. Dirac

4.1 Maxwell's Equations

Now that we have successfully quantized the free Dirac electron, we would like to discuss the question of coupling the Dirac electron to a spin-one Maxwell field A_μ. The resulting theory will be called *quantum electrodynamics*, which is perhaps the most successful physical theory ever proposed. After several decades of confusion, false starts, and frustration, QED has emerged as one of the cornerstones of the quantum theory.

Our discussion of the massless, spin-one field begins with the classical equations of Maxwell:

$$\begin{aligned} \operatorname{div}\mathbf{E} &= \rho \\ \operatorname{curl}\mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} &= \mathbf{j} \\ \operatorname{div}\mathbf{B} &= 0 \\ \operatorname{curl}\mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0 \end{aligned} \tag{4.1}$$

The source, in turn, obeys a conservation equation:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\mathbf{j} = 0 \tag{4.2}$$

Because the divergence of a curl is equal to zero, and because the curl of the gradient is equal to zero, we can replace the magnetic and electric field with

potentials A^0 and $\mathbf{A}$ as follows:

$$\mathbf{E} = -\nabla A^0 - \frac{\partial \mathbf{A}}{\partial t}; \quad \mathbf{B} = \text{curl}\,\mathbf{A} \tag{4.3}$$

The fact that these equations did not transform according to the standard Galilean transformation led to the discovery of special relativity. To see this relativistic invariance more clearly, let us define:

$$\begin{aligned} A^\mu &= (A^0, \mathbf{A}) \\ j^\mu &= (\rho, \mathbf{j}) \end{aligned} \tag{4.4}$$

Then the current conservation equation can be written:

$$\partial_\mu j^\mu = 0 \tag{4.5}$$

and Maxwell's equations can be summarized as:

$$\partial_\mu F^{\mu\nu} = j^\nu \tag{4.6}$$

where:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{4.7}$$

or:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix} \tag{4.8}$$

and:

$$F^{0i} = -E^i; \quad F^{ij} = -\epsilon^{ijk} B^k \tag{4.9}$$

We can now derive Maxwell's equations by writing down the following action:

$$\mathscr{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2) \tag{4.10}$$

If we insert this into the Euler-Lagrange equations of motion, we find that the equations of motion are given by:

$$\partial_\mu F^{\mu\nu} = 0 \tag{4.11}$$

which is just the classical Maxwell's equation with zero source.

A key consequence of this construction is that the Maxwell theory is invariant under a *local symmetry*, that is, one whose parameters are dependent on space–time:

$$\delta A_\mu = \partial_\mu \Lambda(x) \tag{4.12}$$

(A transformation whose parameters are constants is called a *global transformation*, like the isospin and Lorentz transformations discussed earlier.) If we apply successive gauge transformations, we find that they form a group with the simple addition law:

$$\Lambda_3 = \Lambda_1 + \Lambda_2 \tag{4.13}$$

This is the same group law we found for $U(1)$; so we see that Maxwell's equations are locally invariant under $U(1)$.

Under this transformation, the Maxwell tensor is an invariant:

$$\delta F_{\mu\nu} = 0 \tag{4.14}$$

so the Lagrangian is also invariant.

This also means that there is a large redundancy associated with the theory. The equations for A_μ are identical to the equations for $A'_\mu = A_\mu + \partial_\mu \Lambda$.

We also note that the naive energy–momentum tensor associated with Maxwell's theory has the wrong properties. It is neither symmetric, nor is it gauge invariant. A naive application of Noether's theorem gives us an energy–momentum tensor that equals:

$$T^{\mu\nu} = -F^{\mu\lambda}\partial^\nu A_\lambda + \frac{1}{4} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma} \tag{4.15}$$

which is not symmetric. This means that there is no conserved angular momentum tensor. Worse, it is not even gauge invariant, since it is not written entirely in terms of the Maxwell tensor $F_{\mu\nu}$.

However, since the energy–momentum tensor is not a directly measurable quantity, we are free to add another tensor to it:

$$T^{\mu\nu} \rightarrow T^{\mu\nu} + \partial_\lambda (F^{\mu\lambda} A^\nu) \tag{4.16}$$

The resulting energy–momentum tensor is conserved, symmetric, and gauge invariant:

$$T^{\mu\nu} = F^{\mu\rho}F^{\nu}_{\rho} + \frac{1}{4}g^{\mu\nu}F_{\rho\sigma}F^{\rho\sigma} \tag{4.17}$$

It is important to note that the addition of this extra term does not affect the integrated charges, which are directly measurable. To see what conserved charges are associated with this energy–momentum tensor, we find:

$$\begin{aligned} T^{00} &= \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2) \\ T^{i0} &= (\mathbf{E} \times \mathbf{B})^i \end{aligned} \tag{4.18}$$

which we recognize as the energy density and the Poynting vector. Thus, this new energy–momentum tensor is a physically acceptable quantity and compatible with gauge invariance.

Gauge invariance is thus a guide to calculating the physical properties of field theory. As we shall see throughout this book, it is also absolutely important to maintain gauge invariance for QED, for several reasons:

1. The proof of renormalization, that we can extract a finite S matrix order by order from the quantum theory, is crucially dependent on gauge invariance.
2. The proof of unitarity, that there are no ghost states with negative norm in the theory, also depends on gauge invariance. (The longitudinal vibration modes of the Maxwell field are negative norm states, which can be eliminated by choosing a gauge, such as the Coulomb gauge.)
3. The proof that the theory is Lorentz invariant after we have fixed the gauge in a nonrelativistic fashion requires the use of gauge symmetry.

4.2 Relativistic Quantum Mechanics

The problem facing us now is to write down the action for the Dirac theory coupled to the Maxwell theory, creating the quantum theory of electrodynamics. The most convenient way is to use the electron current as the source for the Maxwell field. The electron current is given by $\bar{\psi}\gamma^{\mu}\psi$, and hence we propose the coupling:

$$eA_{\mu}\bar{\psi}\gamma^{\mu}\psi \tag{4.19}$$

We saw earlier that this current emerged because of the invariance of the Dirac action under the symmetry $\psi \rightarrow \exp(i\Lambda)\psi$. Let us promote Λ into a local gauge parameter, so that Λ is a function of space–time. We want an action invariant under:

$$\begin{aligned} \psi(x) &\rightarrow e^{ie\Lambda(x)}\psi(x) \\ A_\mu &\rightarrow A_\mu - \partial_\mu \Lambda(x) \end{aligned} \tag{4.20}$$

The problem with this transformation is that $\Lambda(x)$ is a function of space–time; therefore, the derivative of a spinor $\partial_\mu \psi$ is not a covariant object. It picks up an extraneous term $\partial_\mu \Lambda(x)$ in its transformation. To eliminate this extraneous term, we introduce the *covariant derivative*:

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu + ieA_\mu \tag{4.21}$$

The advantage of introducing the covariant derivative is that it transforms covariantly under a gauge transformation:

$$\begin{aligned} D_\mu \psi &\rightarrow e^{ie\Lambda(x)} D_\mu \psi + (ie\partial_\mu \Lambda - ie\partial_\mu \Lambda)\psi \\ &\rightarrow e^{ie\Lambda(x)} D_\mu \psi \end{aligned} \tag{4.22}$$

This means that the following action is gauge invariant:

$$\mathscr{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{4.23}$$

which we obtain by simply replacing ∂_μ by D_μ.

The coupling of the electron to the Maxwell field reproduces the coupling proposed earlier. In the limit of velocities small compared to the speed of light, the Dirac equation should reduce to a modified version of the Schrödinger equation. We are hence interested in checking the correctness of the Dirac equation to lowest order, to see if we can reproduce the nonrelativistic results and corrections to them. The Dirac equation of motion, in the presence of an electromagnetic potential, now becomes:

$$i\frac{\partial \psi}{\partial t} = \left[\boldsymbol{\alpha} \cdot (-i\boldsymbol{\nabla} - e\mathbf{A}) + \beta m + eA^0\right]\psi \tag{4.24}$$

To find solutions of this equation, let us decompose this four-spinor into two smaller two-spinors:

$$\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix} = \begin{pmatrix} e^{-imt}\Phi \\ e^{-imt}\Psi \end{pmatrix} \tag{4.25}$$

Then the Dirac equation can be decomposed as the sum of two two-spinor equations:

$$\begin{aligned} i\frac{\partial \phi}{\partial t} &= \boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,\chi + eA^0\phi + m\phi \\ i\frac{\partial \chi}{\partial t} &= \boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,\phi + eA^0\chi - m\chi \end{aligned} \tag{4.26}$$

where $\boldsymbol{\pi} = \mathbf{p} - e\mathbf{A}$. Next, we will eliminate χ. For small fields, we can make the approximation that $eA^0 \ll 2m$. Then we can solve the second equation as:

$$\chi \sim \frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{2m}\phi \ll \phi \tag{4.27}$$

In this approximation, the Dirac equation can be expressed as a Schrödinger-like equation, but with crucial spin-dependent corrections:

$$\begin{aligned} i\frac{\partial \Phi}{\partial t} &= \left[\frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}{2m} + eA^0\right]\Phi \\ &= \left[\frac{(\mathbf{p}-e\mathbf{A})^2}{2m} - \frac{e}{2m}\boldsymbol{\sigma}\cdot\mathbf{B} + eA^0\right]\Phi \end{aligned} \tag{4.28}$$

where we have used the fact that :

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 = \boldsymbol{\pi}^2 - e\boldsymbol{\sigma}\cdot\mathbf{B} \tag{4.29}$$

The previous equation gives the first corrections to the Schrödinger equation in the presence of a magnetic and electric field. Classically, we know that the energy of a magnetic dipole in a magnetic field is given by the dot product of the magnetic moment with the field:

$$E = -\boldsymbol{\mu}\cdot\mathbf{B} = -\frac{e\hbar}{2mc}\boldsymbol{\sigma}\cdot\mathbf{B} \tag{4.30}$$

Since $\mathbf{S} = \hbar\boldsymbol{\sigma}/2$, the magnetic moment of the electron is therefore:

$$\boldsymbol{\mu} \equiv \frac{e}{mc}\frac{\hbar\boldsymbol{\sigma}}{2} = 2\left(\frac{e}{2mc}\right)\mathbf{S} \tag{4.31}$$

Thus, the Dirac theory predicted that the electron should have a magnetic moment twice what one might normally expect, that is, twice the Bohr magneton. This was perhaps the first major success of the Dirac theory of the electron. Historically, it gave confidence to physicists that the Dirac theory was correct, even if it seemed to have problems with negative energy states.

Yet another classic result that gave credibility to this relativistic formulation was the splitting of the spectral lines of the hydrogen atom. To solve for corrections to the Schrödinger atom, we set **B** to zero, and take A^0 to be the Coulomb potential. The Dirac equation is now written as:

$$E\psi = \left(-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + \beta m - \frac{Z\alpha}{r}\right)\psi = H\psi \tag{4.32}$$

where:

$$\alpha = e^2/4\pi$$

and:

$$H = \begin{pmatrix} m - \frac{Z\alpha}{r} & \boldsymbol{\sigma}\cdot\mathbf{p} \\ \boldsymbol{\sigma}\cdot\mathbf{p} & -m - \frac{Z\alpha}{r} \end{pmatrix} \tag{4.33}$$

In hindsight, it turns out to be convenient to introduce a judicious ansatz for the solution of the Dirac equation. First, as in the Schrödinger case, we want to separate variables, so that we take $\psi \sim f(r)Y(\theta,\phi)$, where the radial function is explicitly separated off. Second, because ∇^2 contains the Casimir operator L_i^2 in the usual Schrödinger formalism, we choose $Y(\theta,\phi)$ to be the standard spherical harmonics, that is, eigenfunctions of the angular momentum operators. For the spinning case, however, what appears is:

$$\mathbf{J} = \mathbf{L} + \mathbf{S} = \mathbf{L} + \boldsymbol{\sigma}/2 \tag{4.34}$$

that is, we have a combination of orbital spin **L** and intrinsic spin **S**. Thus, our eigenfunctions for the Dirac case must be labeled by eigenvalues j, l, m.

Based on these arguments, we choose as our ansatz:

$$\psi^l_{jm} = \begin{pmatrix} i[G_{lj}(r)/r]\phi^l_{jm} \\ [F_{lj}(r)/r]\left(\sigma\cdot\hat{\mathbf{r}}\right)\phi^l_{jm} \end{pmatrix} \tag{4.35}$$

Inserting this ansatz into Eq. (4.32) and factoring out the angular part, we find that the radial part of our eigenfunctions obey:

$$\begin{aligned} \left(E - m + \frac{Z\alpha}{r}\right)G_{lj}(r) &= -\frac{dF_{lj}(r)}{dr} \mp \left(j + \frac{1}{2}\right)\frac{F_{lj}(r)}{r} \\ \left(E + m + \frac{Z\alpha}{r}\right)F_{lj}(r) &= \frac{dG_{lj}}{dr} \mp \left(j + \frac{1}{2}\right)\frac{G_{lj}(r)}{r} \end{aligned} \tag{4.36}$$

where we use the $+(-)$ sign for $j = l + 1/2$ ($j = l - 1/2$). By power expanding this equation in r, this series of equations can be solved in terms of hypergeometric

functions. These equations, when power expanded, give us the energy eigenstates of the hydrogen atom[1,2]:

$$E_{nj} = m\left[1 + \left(\frac{Z\alpha}{n - (j+1/2) + \sqrt{(j+1/2)^2 - Z^2\alpha^2}}\right)^2\right]^{-1/2} \tag{4.37}$$

Experimentally, this formula correctly gave spin-dependent corrections to the Bohr formula to lowest order. In contrast to the usual Bohr formula, the energy is now a function of both the principal quantum number n as well as the total spin j. By power expanding, we can recover the usual nonrelativistic result to lowest order:

$$E_{nj} = m\left[1 - \frac{Z^2\alpha^2}{2n^2} - \frac{(Z^2\alpha^2)^2}{2n^4}\left(\frac{n}{j+1/2} - 3/4\right) + \cdots\right] \tag{4.38}$$

So far, we have only considered the Dirac electron interacting with a classical Coulomb potential. Therefore, it is not surprising that this formula neglects smaller quantum corrections in the hydrogen energy levels. In particular, the $2S_{1/2}$ and $2P_{1/2}$ levels have the same n and j values, so they should be degenerate according to the Dirac formula. However, experimentally these two levels are found to be split by a small amount, called the Lamb shift.

It was not until 1949, with the correct formulation of QED, that one could successfully calculate these small quantum corrections to the Dirac energy levels. The calculation of these quantum corrections in QED was one of the finest achievements of quantum field theory.

4.3 Quantizing the Maxwell Field

Because of gauge invariance, there are also complications when we quantize the theory. A naive quantization of the Maxwell theory fails for a simple reason: The propagator does not exist. To see this, let us write down the action in the following form:

$$\mathscr{L} = \frac{1}{2}A^\mu P_{\mu\nu}\partial^2 A^\nu \tag{4.39}$$

where:

$$P_{\mu\nu} = g_{\mu\nu} - \partial_\mu\partial_\nu/(\partial)^2 \tag{4.40}$$

The problem with this operator is that it is not invertible, and hence we cannot construct a propagator for the theory. In fact, this is typical of *any* gauge theory,

not just Maxwell's theory. This also occurs in general relativity and in superstring theory. The origin of the noninvertibility of this operator is because $P_{\mu\nu}$ is a projection operator, that is, its square is equal to itself:

$$P_{\mu\nu}P^{\nu\lambda} = P_{\mu}^{\lambda} \tag{4.41}$$

and it projects out longitudinal states:

$$\partial^{\mu}P_{\mu\nu} = 0 \tag{4.42}$$

The fact that $P_{\mu\nu}$ is a projection operator, of course, goes to the heart of why Maxwell's theory is a gauge theory. This projection operator projects out any states with the form $\partial_{\mu}\Lambda$, which is just the statement of gauge invariance.

The solution to this problem is that we must break this invariance by choosing a gauge. Because we have the freedom to add $\partial_{\mu}\Lambda$ to A_{μ}, we will choose a specific value of Λ which will break gauge invariance. We are guaranteed this degree of freedom if the variation $\delta A_{\mu} = \partial_{\mu}\Lambda$ can be inverted. There are several ways in which we can fix the gauge and remove this infinite redundancy. We can, for example, place constraints directly on the gauge field A_{μ}, or we may add the following term to the action:

$$-\frac{1}{2\alpha}\left(\partial_{\mu}A^{\mu}\right)^2 \tag{4.43}$$

where α is arbitrary. We list some common gauges:

Coulomb gauge :	$\nabla_i A_i = 0$
Axial gauge :	$A_3 = 0$
Temporal gauge :	$A_0 = 0$
Landau gauge :	$\partial_{\mu}A^{\mu} = 0$
Landau gauge :	$\alpha = 0$
Feynman gauge :	$\alpha = 1$
Unitary gauge :	$\alpha = \infty$

(4.44)

(Notice that there are two equivalent ways to represent the Landau gauge.) Each time we fix the constraint by restricting the gauge field A_{μ}, we must check that there exists a choice of Λ such that this gauge condition is possible. For example, if we set $A_3 = 0$, then we must show that:

$$A_3' = 0 = A_3 + \partial_3\Lambda \tag{4.45}$$

so that:

$$\Lambda = -\int^{x} dx^3 A_3 \tag{4.46}$$

In the Coulomb gauge, one extracts out the longitudinal modes of the field from the very start. To show that the gauge degree of freedom allows us to make this choice, we write down:

$$\nabla \cdot \mathbf{A}' = \nabla \cdot (\mathbf{A} + \nabla\Lambda) = 0 \tag{4.47}$$

Solving for Λ, we find:

$$\Lambda = -\frac{1}{\nabla^2}\nabla \cdot \mathbf{A} = -\int \frac{d^3x'}{4\pi|\mathbf{x} - \mathbf{x}'|}\nabla' \cdot A(x') \tag{4.48}$$

Likewise, the Landau gauge choice means that we can find a Λ such that:

$$\Lambda = -\frac{1}{\partial^2}\partial_\mu A^\mu \tag{4.49}$$

To begin the process of canonical quantization, we will take the Coulomb gauge in which only the physical states are allowed to propagate. Let us first calculate the canonical conjugate to the various fields. Since $\dot{A}_0$ does not occur in the Lagrangian, this means that A_0 does not appear to propagate, which is a sign that there are redundant modes in the action.

The other modes, however, have canonical conjugates:

$$\begin{aligned} \pi^0 &= \frac{\delta\mathscr{L}}{\delta\dot{A}_0} = 0 \\ \pi^i &= \frac{\delta\mathscr{L}}{\delta\dot{A}_i} = -\dot{A}^i - \partial_i A_0 = E^i \end{aligned} \tag{4.50}$$

We write the Lagrangian as:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}^2 = \frac{1}{2}F_{0i}^2 - \frac{1}{4}F_{ij}^2 \tag{4.51}$$

We now introduce the independent E_i field via a trick. We rewrite the action as:

$$\mathscr{L} = -\frac{1}{2}E_i^2 - E_i F_{0i} - \frac{1}{4}F_{ij}^2 \tag{4.52}$$

By eliminating E_i by its equation of motion, we find $E_i = -F_{0i}$. By inserting this value back into the Lagrangian, we find the original Lagrangian back again.

Then we can write the Lagrangian for QED as:

$$\begin{aligned}\mathscr{L} &= -\mathbf{E}\cdot\dot{\mathbf{A}} - A_0\nabla\cdot\mathbf{E} - \frac{1}{2}\mathbf{E}^2 - \frac{1}{4}F_{ij}^2 + \bar{\psi}(i\not{D} - m)\psi \\ &= -\mathbf{E}\cdot\dot{\mathbf{A}} - \frac{1}{2}\left(\mathbf{E}^2 + \mathbf{B}^2\right) + \mathscr{L}(\psi, \mathbf{A}) - A_0\left(\nabla\cdot\mathbf{E} - e\bar{\psi}\gamma^0\psi\right) \quad (4.53)\end{aligned}$$

A_0 is a Lagrange multiplier. If we solve for the equation of motion of this field, we find that there is the additional constraint:

$$\text{Gauss's Law}: \quad \nabla\cdot\mathbf{E} = \rho = e\bar{\psi}\gamma_0\psi \quad (4.54)$$

Thus, Gauss's Law emerges only after solving for the equation of motion of A_0. If we now count the independent degrees of freedom, we find we have only two degrees left, which correspond to the two independent transverse helicity states. Of the original four components of A_μ, we see that A_0 can be eliminated by its equation of motion, and that we can gauge away the longitudinal mode, therefore leaving us with $4-2$ degrees of freedom, which is precisely the two helicity states predicted in Section 2.8 from group-theoretical arguments alone for massless representations of the Poincaré group.

Intuitively, this means that a photon moving in the z direction can vibrate in the x and y direction, but not the z direction or the timelike direction. This corresponds to the intuitive understanding of transverse photons. In the Coulomb gauge, we can reduce all fields to their transverse components by eliminating their longitudinal components. Let us separate out the transverse and longitudinal parts as follows: $\mathbf{E} = \mathbf{E}_T + \mathbf{E}_L$, where $\nabla\cdot\mathbf{E}_T = 0$ and $\nabla\cdot\mathbf{E}_L = \rho$. Let us now solve for $\mathbf{E}_L$ in terms of ρ. Then we have:

$$\mathbf{E}_L = \nabla\frac{1}{\nabla^2}\rho \quad (4.55)$$

If we insert this back into the Lagrangian, then we find that all longitudinal contributions cancel, leaving only the transverse parts, except for the piece:

$$\begin{aligned}\frac{1}{2}\mathbf{E}_L^2 &= \frac{1}{2}\rho\frac{1}{\nabla^2}\rho \\ &= \frac{e^2}{8\pi}\int d^3x d^3y\,\frac{\psi^\dagger(\mathbf{x},t)\psi(\mathbf{x},t)\psi^\dagger(\mathbf{y},t)\psi(\mathbf{y},t)}{|\mathbf{x}-\mathbf{y}|} \quad (4.56)\end{aligned}$$

This last term is called the "instantaneous four-fermion Coulomb term," which seems to violate special relativity since this interaction travels instantly across space. However, we shall show at the end of this section that this term precisely cancels against another term in the propagator, so Lorentz symmetry is restored.

If we impose canonical commutation relations, we find a further complication. Naively, we might want to impose:

$$[A_i(\mathbf{x}, t), \pi^j(\mathbf{y}, t)] = -i\delta_{ij}\delta^3(\mathbf{x} - \mathbf{y}) \tag{4.57}$$

However, this cannot be correct because we can take the divergence of both sides of the equation. The divergence of A_i is zero, so the left-hand side is zero, but the right-hand side is not. As a result, we must modify the canonical commutation relations as follows:

$$[A_i(\mathbf{x}, t), \pi^j(\mathbf{y}, t)] = -i\tilde{\delta}_{ij}(\mathbf{x} - \mathbf{y}) \tag{4.58}$$

where the right-hand side must be transverse; that is:

$$\tilde{\delta}_{ij} = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \left(\delta_{ij} - \frac{k_i k_j}{\mathbf{k}^2}\right) \tag{4.59}$$

As before, our next job is to decompose the Maxwell field in terms of its Fourier modes, and then show that they satisfy the commutation relations. However, we must be careful to maintain the transversality condition, which imposes a constraint on the polarization vector. The decomposition is given by:

$$\mathbf{A}(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2k_0}} \sum_{\lambda=1}^{2} \boldsymbol{\varepsilon}^\lambda(k) \left[a^\lambda(k)e^{-ik\cdot x} + a^{\lambda\dagger}(k)e^{ik\cdot x}\right] \tag{4.60}$$

In order to preserve the condition that $\mathbf{A}$ is transverse, we take the divergence of this equation and set it to zero. This means that we must impose:

$$\begin{aligned} \boldsymbol{\varepsilon}^\lambda \cdot \mathbf{k} &= 0 \\ \boldsymbol{\varepsilon}^\lambda(k) \cdot \boldsymbol{\varepsilon}^{\lambda'}(k) &= \delta^{\lambda\lambda'} \end{aligned} \tag{4.61}$$

(The simplest way of satisfying these transversality conditions is to take the momentum along the z direction, and keep the polarization vector totally in the transverse directions, i.e., in the x and y directions. However, we will keep our discussion as general as possible.) By inverting these relations, we can solve for the Fourier moments in terms of the fields:

$$\begin{aligned} a^\lambda(k) &= i \int \frac{d^3x}{\sqrt{(2\pi)^3 2k_0}} e^{ik\cdot x} \overset{\leftrightarrow}{\partial}_0 \boldsymbol{\varepsilon}^\lambda(k) \cdot \mathbf{A}(x) \\ a^{\dagger\lambda}(k) &= -i \int \frac{d^3x}{\sqrt{(2\pi)^3 2k_0}} e^{-ik\cdot x} \overset{\leftrightarrow}{\partial}_0 \boldsymbol{\varepsilon}^\lambda(k) \cdot \mathbf{A}(x) \end{aligned} \tag{4.62}$$

In order to satisfy the canonical commutation relations among the fields, we must impose the following commutation relations among the Fourier moments:

$$[a^{\lambda}(k), a^{\dagger\lambda'}(k')] = \delta^{\lambda,\lambda'}\delta^3(\mathbf{k}-\mathbf{k}') \tag{4.63}$$

(An essential point is that the sign of the commutation relations gives us positive norm states. There are no negative norm states, or ghosts, in this construction in the Coulomb gauge.) Let us now insert this Fourier decomposition into the expression for the energy and momentum:

$$\begin{aligned} H &= \frac{1}{2}\int d^3x\,(\mathbf{E}^2+\mathbf{B}^2) \\ &= \sum_{\lambda=1}^{2}\int d^3k\;\omega[a^{\dagger\lambda}(k)a^{\lambda}(k)] \\ \mathbf{P} &= \int d^3x\,(:\mathbf{E}\times\mathbf{B}:) \\ &= \int d^3k\;\mathbf{k}\sum_{\lambda=1}^{2}a^{\lambda\dagger}(k)a^{\lambda}(k) \end{aligned} \tag{4.64}$$

After normal ordering, once again the energy is positive definite. Finally, we wish to calculate the propagator for the theory. Again, there is a complication because the field is transverse. The simplest way to construct the propagator is to write down the time-ordered vacuum expectation value of two fields. The calculation is almost identical to the one for scalar and spinor fields, except we have the polarization tensor to insert:

$$\begin{aligned} iD_F^{\mathrm{tr}}(x-x')_{\mu\nu} &= \langle 0|TA_{\mu}(x)A_{\nu}(x')|0\rangle \\ &= i\int\frac{d^4k}{(2\pi)^4}\frac{e^{-ik\cdot(x-x')}}{k^2+i\epsilon}\sum_{\lambda=1}^{2}\epsilon_{\mu}^{\lambda}(k)\epsilon_{\nu}^{\lambda}(k) \end{aligned} \tag{4.65}$$

The previous expression is not Lorentz invariant since we are dealing with transverse states. This is a bit troubling, until we realize that Green's functions are off-shell objects and are not measurable. However, these Lorentz violating terms should vanish in the full S matrix. To see this explicitly, let us choose a new orthogonal basis of four vectors, given by $\epsilon_{\mu}^{1}(k)$, $\epsilon_{\mu}^{2}(k)$, k^{μ}, and a new vector $\eta^{\mu} = (1,0,0,0)$. Any tensor can be power expanded in terms of this new basis. Therefore, the sum over polarization vectors appearing in the propagator can always be expanded in terms of the tensors $g_{\mu\nu}$, $\eta_{\mu}\eta_{\nu}$, $k_{\mu}k_{\nu}$, and $k_{\mu}\eta_{\nu}$. We can calculate the coefficients of this expansion by demanding that both sides of the

equation be transverse. Then it is easy to show:

$$\sum_{\lambda=1}^{2} \epsilon_\mu^\lambda(k)\epsilon_\nu^\lambda(k) = -g_{\mu\nu} - \frac{k_\mu k_\nu}{(k\cdot\eta)^2 - k^2} + \frac{(k\cdot\eta)(k_\mu\eta_\nu + k_\nu\eta_\mu)}{(k\cdot\eta)^2 - k^2} - \frac{k^2\eta_\mu\eta_\nu}{(k\cdot\eta)^2 - k^2} \tag{4.66}$$

Fortunately, the noninvariant terms involving η can all be dropped. The terms proportional to k_μ vanish when inserted into a scattering amplitude. This is because the propagator couples to two currents, which in turn are conserved by gauge invariance. (To see this, notice that the theory is invariant under $\delta A_\mu = \partial_\mu \Lambda$. In a scattering amplitude, this means that adding k_μ to the polarization vector ϵ_μ cannot change the amplitude. Thus, k_μ terms in the propagator do not couple to the rest of the diagram. This will be discussed more in detail when we study the Ward identities.)

If we drop terms proportional to k_μ in the propagator, we are left with:

$$D_F^{\text{tr}}(x - x')_{\mu\nu} = -g_{\mu\nu}\Delta_F(x - x'; m = 0) - \eta_\mu\eta_\nu \frac{\delta(t - t')}{4\pi|\mathbf{x} - \mathbf{x}'|} \tag{4.67}$$

The first term is what we want since it is covariant. The second term proportional to $\eta_\mu\eta_\nu$ is called the instantaneous Coulomb term. In any Feynman diagram, it occurs between two currents, creating $(\psi^\dagger\psi)\nabla^{-2}(\psi^\dagger\psi)$. This term precisely cancels the Coulomb term found in the Hamiltonian when we solved for $\mathbf{E}_L$ in Eq. (4.56).

As expected, we therefore find that although the Green's function is gauge dependent (and possesses terms that travel instantly across space), the S matrix is Lorentz invariant and causal.

4.4 Gupta–Bleuler Quantization

The advantage of the canonical quantization method in the Coulomb gauge is that we always work with transverse states. Thus, all states have positive norm:

$$\langle 0|a^\lambda(k)a^{\lambda'\dagger}(k')|0\rangle = \delta_{\lambda\lambda'}\delta^3(\mathbf{k} - \mathbf{k}') \tag{4.68}$$

However, the canonical quantization method, although it is guaranteed to yield a unitary theory, is cumbersome because Lorentz invariance is explicitly broken. For higher spin theories, the loss of Lorentz invariance multiplies the difficulty of any calculation by several times.

There is another method of quantization, called the Gupta–Bleuler[3,4] quantization method or covariant method, which keeps manifest Lorentz invariance and simplifies any calculation. There is, however, a price that must be paid, and that is the theory allows negative norm states, or ghosts, to propagate. The resulting theory is manifestly Lorentz invariant with the presence of these ghosts, but the theory is still self-consistent because we remove these ghost states by hand from the physical states of the theory. We begin by explicitly breaking gauge invariance by adding a noninvariant term into the action:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{2\alpha}(\partial_\mu A^\mu)^2 \tag{4.69}$$

for arbitrary α.

Then the action now reads:

$$\mathscr{L} \sim \frac{1}{2}A^\mu \tilde{P}_{\mu\nu}\partial^2 A^\nu \tag{4.70}$$

where:

$$\tilde{P}_{\mu\nu} = g_{\mu\nu} - \left(1 - \alpha^{-1}\right)\partial_\mu\partial_\nu/\partial^2 \tag{4.71}$$

Now that we have explicitly broken the gauge invariance, this operator is no longer a projection operator and hence can be inverted to find the propagator:

$$D_{\mu\nu} = -(\tilde{P}^{-1})_{\mu\nu}/\partial^2 = -\left[g_{\mu\nu} - (1-\alpha)\,\partial_\mu\partial_\nu/\partial^2\right]/\partial^2 \tag{4.72}$$

This propagator explicitly propagates ghost states that violate unitarity. The D_{00} component of the propagator occurs with the wrong sign, and hence represents a ghost state. For our purposes, we will take the gauge $\alpha = 1$, so that the equation of motion now reads:

$$\partial^2 A_\mu = 0 \tag{4.73}$$

In this gauge, we find that A_0 is no longer a redundant Lagrange multiplier, but a dynamical field and hence has a canonical conjugate to it. The conjugate field of A_μ is now a four-vector:

$$\pi_\mu = \frac{\delta\mathscr{L}}{\delta\dot{A}^\mu} = \dot{A}_\mu \tag{4.74}$$

Then the covariant canonical commutation relations read:

$$[A_\mu(x), \pi^\nu(x')] = i\delta_\mu^\nu\delta^3(\mathbf{x} - \mathbf{x}') \tag{4.75}$$

As usual, we can decompose the field in terms of the Fourier moments:

$$A_\mu(x) = \int \frac{d^3k}{2\omega_k} \left[a^\lambda(k)\epsilon_\mu^\lambda(k)e^{-ik\cdot x} + a^{\dagger\lambda}(k)\epsilon_\mu^\lambda(k)e^{ik\cdot x}\right] \tag{4.76}$$

The difference now is that the ϵ_μ vector is a genuine four-vector. In order for the canonical commutation relations to be satisfied, we necessarily choose the following commutation relations among the oscillators:

$$[a^\lambda(k), a^{\dagger\lambda'}(k')] = -g^{\lambda\lambda'}\delta^3(\mathbf{k} - \mathbf{k}') \tag{4.77}$$

The presence of the metric tensor in the commutation relation signals that the norm of the states may be negative; that is, a nonphysical, negative norm ghost is present in the theory. The norm of the state $a^{\dagger\lambda}(k)|0\rangle$ can now be negative. This is the price we pay for having a Lorentz covariant quantization scheme.

It is straightforward to prove that the propagator is now given by:

$$\langle 0|T\,A_\mu(x)A_\nu(y)|0\rangle = -ig_{\mu\nu}\Delta_F(x - y) \tag{4.78}$$

This can be proved by explicitly inserting the operator expression for $A_\mu(x)$. The important aspect of this propagator is that it contains the metric $g_{\mu\nu}$, which alternates in sign. Hence, it propagates ghost states. Since ghosts now propagate in the theory, we must be careful how we remove them. If we take the condition $\partial_\mu A^\mu|\Psi\rangle = 0$, we find that this condition is too stringent; it has no solutions at all. The Gupta–Bleuler formalism is based on the observation that a weaker condition is required:

$$\left(\partial_\mu A^\mu\right)^{(+)}|\Psi\rangle = 0 \tag{4.79}$$

where we only allow the destruction part of the constraint to act on physical states. In momentum space, this is equivalent to the condition that $k^\mu a_\mu(k)|\Psi\rangle = 0$. This guarantees that, although ghosts are allowed to circulate in the system, they are explicitly removed from all physical states of the theory.

(We can also quantize the massive vector field in much the same way. The quantization is almost identical to the one presented before, but now the counting of physical states is different. We recall from our discussion of the Poincaré group that a massless field only has two helicity components. However, the massive vector field has 3 components.)

4.5 C, P, and T Invariance

Although we have analyzed the behavior of quantum fields under continuous isospin and Lorentz symmetry, we must also investigate their behavior under discrete symmetries, such as that generated by parity, charge conjugation, and time-reversal symmetry:

$$\begin{aligned} P: &\quad \mathbf{x} \to -\mathbf{x} \\ C: &\quad e \to -e \\ T: &\quad t \to -t \end{aligned} \tag{4.80}$$

Classically, we know that the laws of physics are invariant under these transformations. For example, we find that both Newton's and Maxwell's equations are invariant under these transformations.

The easiest way in which to calculate how the A_μ field transforms is to examine Maxwell's equation, especially the source term j^μ. Under a parity transformation, the electric charge distribution ρ does not change, but $\mathbf{j} \to -\mathbf{j}$ because we are reversing the direction of the electric current. Thus:

$$\mathscr{P} j^\mu \mathscr{P}^{-1} = j_\mu \tag{4.81}$$

Under a charge conjugation, a positive electric current turns into a negative one, so that:

$$\mathscr{C} j^\mu \mathscr{C}^{-1} = -j^\mu \tag{4.82}$$

Then under a time reversal, once again ρ remains the same, but $\mathbf{j}$ reverses sign (because the current reverses direction), so:

$$\mathscr{T} j^\mu \mathscr{T}^{-1} = j_\mu \tag{4.83}$$

Then the transformation of A_μ can be found immediately. One way is to observe that $j^\mu A_\mu$ appears in the action and is an invariant. Then we can read off the transformation properties of A_μ. Another way in which to derive the transformation properties of A_μ is to use Maxwell's equations, $\partial_\mu F^{\mu\nu}(A) = j^\nu$. Knowing the transformation properties of ∂_μ and j^μ, then it is easy to solve for

the transformation properties of A_μ. We summarize our results as follows:

	∂_μ	j^μ	A_μ
P	∂^μ	j_μ	A^μ
C	∂_μ	$-j^\mu$	$-A_\mu$
T	$-\partial^\mu$	j_μ	A^μ

(4.84)

Similarly, we can read off the transformation properties of ψ, using the Dirac equation and also the fact that $j^\mu = e\bar{\psi}\gamma^\mu\psi$, although the calculation is considerably longer. With a bit of work, we can summarize how Dirac bilinear scalars, vectors, axial vectors, etc. transform under these discrete transformations, including $CPT = \Theta$:

	$\bar{\psi}\psi$	$\bar{\psi}\gamma_5\psi$	$\psi\gamma^\mu\psi$	$\bar{\psi}\gamma^\mu\gamma_5\psi$	$\bar{\psi}\sigma^{\mu\nu}\psi$
C	S	P	$-V^\mu$	A^μ	$-T^{\mu\nu}$
P	S	$-P$	V_μ	$-A_\mu$	$T_{\mu\nu}$
T	S	$-P$	V_μ	A_μ	$-T_{\mu\nu}$
Θ	S	P	$-V^\mu$	$-A^\mu$	$T^{\mu\nu}$

(4.85)

To prove this, let us examine each transformation separately.

4.5.1 *Parity*

When the Dirac field couples to the Maxwell field, we want the combination $j^\mu A_\mu$ appearing in the action to be parity conserving.

To find an explicit form for the operator $\mathscr{P}$ for the Dirac field, we will find it convenient to recall that any element of $O(3, 1)$ (which includes parity operations with $\det O = -1$) has the following effect on the Dirac matrix:

$$S(\Lambda)^{-1}\gamma^\mu S(\Lambda) = \Lambda^\mu_\nu \gamma^\nu \tag{4.86}$$

For our case, we can set the matrix Λ to be our parity operator if we specify that it reverses the sign of $\mathbf{x}$:

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{4.87}$$

which is equal to the metric tensor $g_{\mu\nu}$. Thus, we want a solution of the following equation:

$$\mathscr{P}^{-1}\gamma^{\nu}\mathscr{P} = g^{\nu\nu}\gamma^{\nu} = -(-1)^{\delta_{\nu,0}}\gamma^{\nu} \tag{4.88}$$

An explicit solution of this equation is simply given by:

$$\mathscr{P} = e^{i\phi}\gamma^{0} \tag{4.89}$$

where $e^{i\phi}$ is an irrelevant phase factor.

Thus, the action of the parity transformation on a spinor field is given by:

$$\text{Parity}: \quad \psi'(-\mathbf{x}, t) = S(\Lambda)\psi = e^{i\phi}\gamma^{0}\psi(\mathbf{x}, t) \tag{4.90}$$

This also means that the transformation of the $\bar{\psi}$ field is given by:

$$\bar{\psi}'(-\mathbf{x}, t) = \bar{\psi}(\mathbf{x}, t)\gamma^{0}e^{-i\phi} \tag{4.91}$$

From this, we can calculate how the various bilinear combinations transform under the parity operation in Eq. (4.85).

4.5.2 *Charge Conjugation*

Charge conjugation is easily studied by taking the Dirac equation and then reversing the sign of the electric charge. If we let ψ_c represent the Dirac field that has the opposite charge as ψ, then we have:

$$\begin{aligned} (i\,\partial\!\!\!/ - e\,A\!\!\!/ - m)\psi &= 0 \\ (i\,\partial\!\!\!/ + e\,A\!\!\!/ - m)\psi_c &= 0 \end{aligned} \tag{4.92}$$

In order to find the relationship between ψ with charge e and ψ_c with charge $-e$, let us take the complex conjugate and then the transpose of the first equation. Then we find:

$$\gamma^{\mu T}(-i\partial_{\mu} - eA_{\mu})\left(\gamma^{0T}\psi^{*}\right) = 0 \tag{4.93}$$

It can be shown that for any representation of the Dirac algebra, there exists a matrix C that satisfies:

$$C\gamma_\mu^T C^{-1} = -\gamma_\mu \tag{4.94}$$

Now let us compare the previous equation with the equation for the ψ_c field. We have an exact correspondence (up to a phase) if we set:

$$\psi_c = e^{i\phi} C(\gamma^{0T}\psi^*) = e^{i\phi} C\bar{\psi}^T \tag{4.95}$$

So far, we have not specified the representation of the Dirac matrices. There is more than one solution of the equation for C. In the Dirac representation, we find the following solution for the C matrix as:

$$C = i\gamma^2\gamma^0 = \begin{pmatrix} 0 & -i\sigma^2 \\ -i\sigma^2 & 0 \end{pmatrix} \tag{4.96}$$

which satisfies the following additional constraints:

$$-C = C^{-1} = C^T = C^\dagger \tag{4.97}$$

For us, however, the important feature of the C matrix is that it allows us to identify the particle–antiparticle structure of the Dirac field.

Applying the C matrix to the particle field, we obtain the antiparticle field:

$$\psi = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \quad \psi_c = e^{i\phi} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{4.98}$$

which justifies our earlier statement that the Dirac equation contains both the particle and antiparticle fields, with the charges as well as spins reversed.

Now let us compute how the current j^μ transforms under the charge conjugation operation:

$$j_c^\mu = \bar{\psi}_c\gamma^\mu\psi_c = \psi^T C\gamma_\mu C\bar{\psi}^T = \psi^T\gamma_\mu^T\bar{\psi}^T = -\bar{\psi}\gamma_\mu\psi \tag{4.99}$$

The last minus sign is important: Because ψ is an anticommuting field, we pick up an extra minus sign when we move one spinor past another.

Now let us try to determine how the combined Dirac and Maxwell system transforms under charge conjugation. The Dirac equation is left invariant if we

make the simultaneous change:

$$\begin{aligned} \psi &\rightarrow \psi_c = e^{i\phi} C \bar{\psi}^T \\ j^\mu &\rightarrow j_c^\mu = -j^\mu \\ A_\mu &\rightarrow -A_\mu \end{aligned} \tag{4.100}$$

(We have changed the sign of the Maxwell field, which simulates the change of the charge e that we made earlier.)

Thus, the combination $j_\mu A^\mu$ is invariant under charge conjugation, as desired.

4.5.3 *Time Reversal*

Finally, we analyze the effect of making the transformation $t \rightarrow -t$. We wish to find an explicit representation of the operator $\mathscr{T}$ in terms of harmonic oscillators. This can be done in several ways. We can represent the time-reversal operator as $S(\Lambda)$ acting on a spinor where $\Lambda_{\mu\nu} = -g_{\mu\nu}$. Or, we can write down the Dirac equation with a time reversal and try to retransform the equation back into the usual Dirac form.

Either way, we find the same result:

$$\mathscr{T}\psi(\mathbf{x}, t)\mathscr{T}^{-1} = e^{i\phi} T \psi(\mathbf{x}, -t) \tag{4.101}$$

where:

$$T\gamma^\mu T^{-1} = \gamma_\mu^T = \gamma^{\mu *} \tag{4.102}$$

An explicit representation of the T matrix is given by:

$$T = i\gamma^1\gamma^3 \tag{4.103}$$

where:

$$T = T^\dagger = T^{-1} = -T^* \tag{4.104}$$

We should also mention that the $\mathscr{T}$ operator is unusual because it is antiunitary. For example, consider the time evolution equation:

$$[H, \phi(x)] = -i\frac{\partial \phi}{\partial t} \tag{4.105}$$

Let us make a time-reversal transformation on this equation. If we reverse the sign of t in this equation, we find:

$$[H, \phi(x')] = +i\frac{\partial \phi}{\partial t} \tag{4.106}$$

However, this has the net effect of transforming $H \rightarrow -H$, which is illegal (because then we would have negative energy states). This difficulty did not happen for parity or charge conjugation because $\partial/\partial t$ did not change for those symmetries.

Correspondingly we wish that $\mathscr{T}$ would reverse the exponent appearing in the time evolution operator:

$$\mathscr{T} e^{iH(t_1-t_2)} \mathscr{T}^{-1} = e^{+iH(t_2-t_1)} \tag{4.107}$$

However, this is impossible if the Hamiltonian commutes with $\mathscr{T}$.

This means that the operator $\mathscr{T}$ must be antiunitary:

$$\langle \mathscr{T}\phi | \mathscr{T}\psi \rangle = \langle \psi | \phi \rangle \tag{4.108}$$

Or this operator contains yet another operator that can take the complex conjugate of any c number. If we postulate an operator that can reverse $i \rightarrow -i$, then $\mathscr{T}$ can commute with the Hamiltonian yet still reverse the sign of t.

4.6 CPT Theorem

In nature, these discrete symmetries are violated. Parity is maximally violated by the weak interactions, and the combination CP is violated in K meson decays. However, there is a remarkable theorem that states that any quantum field theory is invariant under the combined operation of CPT[5–7] under very general conditions.

The theorem states that the Hamiltonian $\mathscr{H}$ is invariant under CPT:

$$(CPT)\,\mathscr{H}(x)\,(CPT)^{-1} = \mathscr{H}(x') \tag{4.109}$$

if the following two conditions are met:

1. The theory must be local, possess a Hermitian Lagrangian, and be invariant under proper Lorentz transformations.

2. The theory must be quantized with commutators for integral spin fields and quantized with anticommutators for half-integral spin fields (i.e., the usual spin-statistics connection).

Thus, although quantum field theories are easily written down that violate these discrete symmetries separately, any quantum theory obeying these very general features must be invariant under CPT.

Various proofs of this powerful theorem have been proposed over the years. We will not review the results from axiomatic field theory, which are the most rigorous. Rather than give a detailed proof, we will show that the CPT theorem is satisfied for the spin 0, 1/2, and 1 fields that we have so far investigated.

We first note that it is easy to show that the CPT operation changes any quantum field theory obeying these two assumptions in the following way:

1. The coordinates change as follows:

$$\begin{aligned} x^\mu &\rightarrow -x^\mu \\ \partial_\mu &\rightarrow -\partial_\mu \end{aligned} \tag{4.110}$$

2. The Maxwell field transforms as follows:

$$\begin{aligned} \mathscr{P} A_\mu(x) \mathscr{P}^{-1} &= A^\mu(x') \\ \mathscr{C} A_\mu(x) \mathscr{C}^{-1} &= -A_\mu(x) \\ \mathscr{T} A_\mu(x) \mathscr{T}^{-1} &= A^\mu(x') \end{aligned} \tag{4.111}$$

where x' refers to the $\mathscr{P}$ or the $\mathscr{T}$ transformed variable. Thus, under the combined CPT, we have:

$$A_\mu(x) \rightarrow -A_\mu(-x) \tag{4.112}$$

3. A Dirac spinor transforms under CPT as:

$$\psi_\alpha \rightarrow -i\psi(-x)_\beta(\gamma_5\gamma_0)_{\beta\alpha} \tag{4.113}$$

4. A Dirac bilinear changes as follows:

$$\bar{\psi}(x) O \psi(x) \rightarrow (-1)^k \bar{\psi}(x') O \psi(x') \tag{4.114}$$

where k denotes the number of Lorentz indices appearing in the matrix O. (This assumed the spin-statistics connection, since we had to push one spinor past another.)

5. Any even-rank tensor transforms into its Hermitian conjugate, and all odd-rank tensors transform into the negative of their Hermitian conjugates.

6. All c numbers appearing in the theory are complex conjugated.

The sketch of the proof now proceeds as follows. First, we observe that the Lagrangian is invariant under CPT:

$$(CPT)\mathscr{L}(x)(CPT)^{-1} = \mathscr{L}(-x) \tag{4.115}$$

This is because the Lagrangian is a contraction of many tensors of different rank, but the sum of all the ranks must be even since $\mathscr{L}$ is a Lorentz scalar. By (5), this means that the Lagrangian is transformed into its Hermitian conjugate. But since the Lagrangian is Hermitian by the first assumption, we find that $\mathscr{L}$ must be CPT invariant.

One possible loophole in this construction is if the Lagrangian contains an infinite number of derivatives. Then condition (1) would be difficult to obey. This loophole is closed by the assumption that $\mathscr{L}$ is local. [Nonlocal theories containing terms like $\phi(x)\phi(y)$ can be power expanded as follows:

$$\phi(x)\phi(y) = \phi(x)e^{(y-x)^\mu \partial_\mu}\phi(x) \tag{4.116}$$

Nonlocal theories therefore contain an infinite number of derivatives, which are excluded from our discussion by the first assumption.]

Now that the Lagrangian is invariant, let us analyze the transformation property of the Hamiltonian, which is defined as:

$$\mathscr{H} = -\mathscr{L} + \sum_r : \pi_r(x)\dot{\phi}_r(x) : \tag{4.117}$$

where the sum r is over both bosons and fermions.

To calculate the transformation of the Hamiltonian under CPT, we recall the transformation properties of the boson and fermion fields under CPT:

$$\begin{aligned}
(CPT)\phi_r(x)(CPT) &= e^{i\theta_r}\phi_r(-x) \\
(CPT)\dot{\phi}_r(x)(CPT) &= -e^{i\theta_r}\phi_r(-x) \\
(CPT)\pi_r(x)(CPT) &= -e^{-i\theta_r}\pi_r(-x)
\end{aligned} \tag{4.118}$$

To show that the Hamiltonian is invariant, we observe that the equal-time commutation relations for the bosons and fermions are preserved if the spin-statistics relation holds:

$$\text{Boson}: \qquad [\pi_r(\mathbf{x}, t), \phi_s(\mathbf{x}', t)] = -i\delta^3(\mathbf{x} - \mathbf{x}')\delta_{rs}$$

$$\text{Fermion}: \qquad \{\pi_r(\mathbf{x}, t), \phi_s(\mathbf{x}', t)\} = i\delta^3(\mathbf{x} - \mathbf{x}')\delta_{rs} \tag{4.119}$$

Then we have all the identities necessary to show that:

$$(CPT)\mathscr{H}(x)(CPT)^{-1} = \mathscr{H}(-x) \tag{4.120}$$

which completes the proof.

In summary, we have seen how to quantize the Maxwell spin-one field via the canonical method. The formulation is a bit clumsy because we must quantize the theory in the Coulomb gauge to eliminate the ghost states with negative norm. The Gupta–Bleuler formulation is more convenient because it is Lorentz covariant, but we must apply the ghost-killing constraint on the states. The reason for this complication is that the Maxwell theory is a gauge theory with the group $U(1)$.

In Chapter 5, we will analyze how to calculate scattering cross sections for QED with quantum field theory.

4.7 Exercises

1. Prove that the radial functions G_{lj} and F_{lj} obey Eq. (4.36) starting with the Dirac equation for an electron in a Coulomb field.

2. There are no finite-dimensional unitary representations of the Lorentz group, but what about a vector field A_μ, which forms a finite-dimensional representation space for the Lorentz group. Does this violate the no-go theorem? Why not?

3. There are very few exactly solvable problems involving the classical Dirac equation. One of them is the electron in a Coulomb potential. Another is the electron in a constant magnetic field. Solve the Dirac equation in this case, and show that:

$$E = \sqrt{m^2 + p_z^2 + 2n|eB|} \tag{4.121}$$

where $n = 0, 1, 2, \ldots$.

4. Prove that:

$$\mathscr{P}b(p, s)\mathscr{P}^{-1} = b(-p, s); \quad \mathscr{P}d^\dagger(p, s)\mathscr{P}^{-1} = -d^\dagger(-p, s) \tag{4.122}$$

Prove that an explicit operator representation for $\mathscr{P}$ is given by:

$$\mathscr{P} = \exp\Big(-\frac{i\pi}{2}\int d^3p\Big[b^\dagger(p, s)b(p, s) - b^\dagger(p, s)b(-p, s)$$

$$+d^\dagger(p,s)d(p,s)+d^\dagger(p,s)d(-p,s)\Big]\Big) \tag{4.123}$$

5. Prove that:

$$\begin{aligned}\mathscr{C}b(p,s)\mathscr{C}^{-1} &= d(p,s)e^{i\phi}\\ \mathscr{C}d^\dagger(p,s)\mathscr{C}^{-1} &= b^\dagger(p,s)e^{i\phi}\end{aligned} \tag{4.124}$$

Prove that this operation, in turn, can be generated by the operator expression:

$$\mathscr{C} = \exp\left(\frac{i\pi}{2}\int d^3p\sum_s\left[b^\dagger(p,s)-d^\dagger(p,s)\right]\left[b(p,s)-d(p,s)\right]\right) \tag{4.125}$$

(for $\varphi = 0$).

6. Prove Eq. (4.66).

7. Show that:

$$(CPT)U(t,t_0)(CPT)^{-1} = U(-t,-t_0) \tag{4.126}$$

where U is the operator e^{-iHt} for constant H, so therefore:

$$(CPT)S(CPT)^{-1} = S^\dagger \tag{4.127}$$

where S is the S matrix.

8. Consider the magnetic field created by electrons both moving in a straight wire and circulating in a solenoid. Also consider the electric field created by a stationary electron. Draw the diagrams for these system when a C, P, and T transformation is applied to them. Show that we reproduce the results in Eq. (4.84).

9. Choose the $A_0 = 0$ gauge. Since A_0 is no longer a Lagrange multiplier, this means that Gauss's Law can no longer be applied; therefore, we cannot impose $\nabla \cdot \mathbf{E} = 0$. This means we cannot reduce the system to the Coulomb gauge. Resolve this paradox. [Hint: show that Gauss's Law commutes with the Hamiltonian, and then use Eqs. (1.30) and (1.31).]

10. Given a Lagrangian:

$$\mathscr{L} \sim \theta F^a_{\mu\nu}F^a_{\sigma\rho}\epsilon^{\mu\nu\sigma\rho} \tag{4.128}$$

what symmetries among C, P, T are broken?

11. Consider the coupling: $\bar{\psi}\sigma_{\mu\nu}\psi F^{\mu\nu}$. Take the nonrelativistic limit of this expression, and show how this relates to the magnetic moment of the electron.

12. Prove that the inversion of the operator in Eq. (4.71) gives us the propagator in Eq. (4.72).

13. Prove that the canonical commutation relation in Eq. (4.58) is satisfied if we take the commutation relations Eq. (4.63) among the harmonic oscillator states. Show that transversality is preserved.

14. It is often convenient to describe gauge theory in the mathematical language of *forms* (especially when working in higher dimensions). Let the infinitesimal dx^μ be antisymmetric under an operation we call $\wedge$; that is, $dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu$. Define the operator $d \equiv dx^\mu \partial_\mu$. Prove that d is nilpotent, that is, $d^2 = 0$. Define the one-form $A \equiv A_\mu dx^\mu$. Define a p-form as $\omega = \omega_{\mu_1\mu_2\cdots\mu_p} dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_p}$. Show that $dA = F$, where $F = F_{\mu\nu} dx^\mu \wedge dx^\nu$, where A is the vector potential and F is the Maxwell tensor.

15. In n dimensions, define the Hodge operator $*$, which generates a duality transformation:

$$*(dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_p}) = \frac{1}{(n-p)!} \epsilon_{\mu_1\mu_2\cdots\mu_n} dx^{\mu_{p+1}} \wedge dx^{\mu_{p+2}} \wedge \cdots \wedge dx^{\mu_n} \tag{4.129}$$

for $p < n$. Show that Maxwell's equations can now be summarized as:

$$dF = 0; \quad d * F = J \tag{4.130}$$

Show that the Bianchi identity is a consequence of $d^2 = 0$.

16. Define the operator δ as:

$$\delta \equiv (-1)^{np+n+1} * d * \tag{4.131}$$

Then show that $\delta^2 = 0$ and that the Laplacian is given by:

$$\Delta = (d + \delta)^2 = d\delta + \delta d \tag{4.132}$$

17. Prove that $T^{\mu\nu}$ in Eq. (4.17) is conserved. (Hint: use the Bianchi identity.)

Chapter 5
Feynman Rules and LSZ Reduction

The reason Dick's physics was so hard for ordinary people to grasp was that he did not use equations. The usual way theoretical physics was done since the time of Newton was to begin by writing down some equations and then to work hard calculating solutions of the equations.... He had a physical picture of the way things happen, and the picture gave him the solution directly with a minimum of calculation. It was no wonder that people who had spent their lives solving equations were baffled by him. Their minds were analytical; his was pictorial.

—F. Dyson on R. Feynman

5.1 Cross Sections

So far, our discussion has been rather formal, with no connection to experiment. This is because we have been concentrating on Green's functions, which are unphysical; that is, they describe the motion of "off-shell" particles where $p_\mu^2 \neq m^2$. However, the physical world that we measure in our laboratories is on-shell. To make the connection to experiment, we need to rewrite our previous results in terms of numbers that can be measured in the laboratory, such as decay rates of unstable particles and scattering cross sections. There are many ways in which to define the cross section, but perhaps the simplest and most intuitive way is to define it as the effective "size" of each particle in the target:

$$\text{Cross section} = \text{Effective size of target particle} \tag{5.1}$$

The cross section is thus the effective area of each target particle as seen by an incoming beam. Cross sections are often measured in terms of "barns." (One barn is 10^{-24} cm^2.) A nucleon is about one fermi, or 10^{-13} cm across. Its area is therefore about 10^{-26} cm^2, or about 0.01 barns. Thus, by giving the cross section

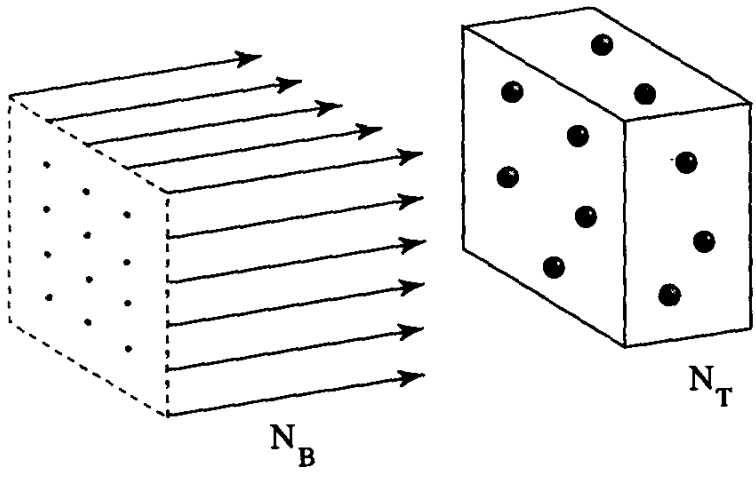


Figure 5.1. A target with N_T nuclei is bombarded with a beam with N_B particles. The cross section σ is the effective size in cm^2 of each nuclei as seen by the beam.

of a particle in a certain reaction, we can immediately calculate the effective size of that particle in relationship to a nucleon.

To calculate the cross section in terms of the rate of collisions in a scattering experiment, let us imagine a thin target with N_T particles in it, each particle with effective area σ or cross section. As seen from an incoming beam, the total amount of area taken up by these particles is therefore $N_T\sigma$. If we aim a beam of particles at the target with area A, then the chance of hitting one of these particles is equal to the total area that these target particles occupy ($N_T\sigma$) divided by the area A:

$$\text{Chance of hitting a particle} = \frac{N_T\sigma}{A} \tag{5.2}$$

Let us say we fire a beam containing N_B particles at the target. Then the number of particles in the beam that are absorbed or deflected is N_B times the chance of being hit. Thus, the number of scattering events is given by:

$$\text{Number of events} = N_B\frac{N_T\sigma}{A} \tag{5.3}$$

or simply:

$$\sigma = \left(\frac{\text{number of events}}{N_B N_T}\right) A \tag{5.4}$$

This reconfirms that the cross section has dimensions of area (Fig. 5.1).

In actual practice, a more convenient way of expressing the cross section is via the flux of the incoming beam, which is equal to ρv. If the beam is moving at velocity v toward a stationary target, then the number of particles in the beam N_B is equal to the density of the beam ρ times the volume. If the beam is a pulse

that is turned on for t seconds, then the volume of the beam is vtA. Therefore, $N_B = \rho vtA$. The cross section can therefore be written as:

$$\begin{aligned}\sigma &= \frac{\text{number of events}/t}{(\rho vtA)N_T/t}A \\ &= \frac{\text{number of events}/t}{\rho v} \\ &= \frac{\text{transition rate}}{\text{flux}}\end{aligned} \tag{5.5}$$

where we have normalized N_T to be 1 and the transition rate is the number of scattering events per second. The cross section is therefore equal to the transition rate divided by the flux of the beam. (More precisely, the cross section is equal to the transition probability per scatterer in the target and per unit incident flux.) This is usually taken as the starting point in discussions of the cross section, but unfortunately it is rather obscure and does not reveal its intuitive meaning as the effective size of the target particle.

(We will be calculating the cross section in collinear Lorentz frames, i.e., where the incoming beam and target move along the same axis. Two common collinear frames are the laboratory frame and the center-of-mass frame. If we make a Lorentz transformation to any other collinear frame, the cross section is an invariant, since a Lorentz contraction does not affect the cross section if we make a boost along this axis. However, the cross section is *not* a true Lorentz invariant. In fact, it transforms like an area under arbitrary Lorentz transformations.)

The next problem is to write the transition rate appearing in the cross section in terms of the S matrix. We must therefore calculate the probability that a collection of particles in some initial state i will decay or scatter into another collection of particles in some final state j. From ordinary nonrelativistic quantum mechanics, we know that the cross section σ can be calculated by analyzing the properties of the scattered wave. Using classical wave function techniques dating back to Rayleigh, we know that a plane wave e^{ikz} scattering off a stationary, hard target is given by:

$$e^{ikz} + \frac{f(\theta)}{r}e^{ikr} \tag{5.6}$$

where the term with e^{ikr} represents the scattered wave, which is expanding radially from the target. Therefore $|f(\theta)|^2$ is proportional to the probability that a particle scatters into an angle θ.

More precisely, the differential cross section is given by the square of $f(\theta)$:

$$\frac{d\sigma}{d\Omega} = |f(\theta)|^2 \tag{5.7}$$

where the solid angle differential is given by:

$$\int d\Omega = \int_0^{2\pi} d\phi \int_{-1}^{1} d\cos\theta \tag{5.8}$$

and the total cross section is given by:

$$\int d\Omega \, |f(\theta)|^2 = \int d\Omega \frac{d\sigma}{d\Omega} = \sigma \tag{5.9}$$

For our purposes, however, this formulation is not suitable because it is inherently nonrelativistic. To give a relativistic formulation, let us start at the beginning. We wish to describe the scattering process that takes us from an initial state consisting of a collection of free, asymptotic states at $t \to -\infty$ to a final state $|f\rangle$ at $t \to \infty$. To calculate the probability of taking us from the initial state to the final state, we introduce the S matrix:

$$\begin{aligned} S_{fi} &= \langle f|S|i\rangle \\ &= \delta_{fi} - i(2\pi)^4\delta^4(P_f - P_i)\mathscr{T}_{fi} \end{aligned} \tag{5.10}$$

where δ_{fi} symbolically represents the particles not interacting at all, and $\mathscr{T}_{fi}$ is called the transition matrix, which describes non-trivial scattering.

One of the fundamental constraints coming from quantum mechanics is that the S matrix is unitary:

$$\sum_f S_{fi}^* S_{fk} = \delta_{ik} \tag{5.11}$$

By taking the square of the S matrix, we can calculate the transition probabilities. The probability that the collection of states i will make the transition to the final states f is given by:

$$P_{fi} = S_{fi}^* S_{fi} \tag{5.12}$$

Likewise, the total probability that the initial states i will scatter into all possible final states f is given by:

$$P_{\text{total}} = \sum_f S_{fi}^* S_{fi} \tag{5.13}$$

Now we must calculate precisely what we mean by $\sum_f$. We begin by defining our states within a box of volume V:

$$|\mathbf{p}\rangle = \sqrt{(2\pi)^3 2E_p/V}\, a^\dagger(p)|0\rangle \quad \text{(bosons)}$$

$$|\mathbf{p}\rangle = \sqrt{(2\pi)^3(E_p/mV)}a^\dagger(p)|0\rangle \quad \text{(fermions)} \tag{5.14}$$

Our states are therefore normalized as follows:

$$\begin{aligned}\langle \mathbf{p}|\mathbf{p}'\rangle &= (2\pi)^3 2E_p\delta^3(\mathbf{p}-\mathbf{p}')/V \quad \text{(bosons)} \\ \langle \mathbf{p}|\mathbf{p}'\rangle &= (2\pi)^3\frac{E_p}{m}\delta^3(\mathbf{p}-\mathbf{p}')/V \quad \text{(fermions)}\end{aligned} \tag{5.15}$$

With this normalization, the unit operator (on single paricle states) can be expressed as:

$$\begin{aligned}1 &= \int \frac{Vd^3p}{(2\pi)^3 2E_p}|\mathbf{p}\rangle\langle\mathbf{p}| \quad \text{(bosons)} \\ 1 &= \int \frac{Vd^3p}{(2\pi)^3}\frac{m}{E_p}|\mathbf{p}\rangle\langle\mathbf{p}| \quad \text{(fermions)}\end{aligned} \tag{5.16}$$

To check our normalizations, we can let the number one act on an arbitrary state $|\mathbf{q}\rangle$, and we see that it leaves the state invariant. This means, however, that we have an awkward definition of the number of states at a momentum $\mathbf{p}$. With this normalization, we find that:

$$\langle \mathbf{p}|\mathbf{p}\rangle = (2\pi)^3 2E_p\delta^3(0)/V \tag{5.17}$$

which makes no sense. However, we will interpret this to mean that we are actually calculating particle densities inside a large but finite box of size L and volume V; that is, we define:

$$\delta^3(\mathbf{p}) = \lim_{L\to\infty}\left(\frac{1}{(2\pi)^3}\int\int\int_{-L/2}^{L/2} dxdydz\; e^{-i\mathbf{p}\cdot\mathbf{r}}\right) \tag{5.18}$$

This implies that we take the definition:

$$\delta^3(0) = \frac{V}{(2\pi)^3}$$

We will let the volume of the box V tend to infinity only at the end of the calculation. The origin of this problem is that we have been dealing with plane waves, rather than wave packets that are confined to a specific region of space and time. The price we pay for these nonlocalized plane waves is that we must carefully divide out infinite quantities proportional to the volume of space and time. (A more careful analysis would use wave packets that are completely

localized in space and time; i.e., they have an envelope that restricts most of the wave packet to a definite region of space and time. This analysis with wave packets is somewhat more complicated, but yields precisely the same results.)

Our task is now to calculate the scattering cross section $1+2 \to 3+4+\cdots$ and the rate of decay of a single particle $1 \to 2+3\cdots$.

We must now define how we normalize the sum over final states. We will integrate over all momenta of the various final states, and sum over all possible final states. For each final state, we will integrate over the final momentum in a Lorentz covariant fashion. We will use:

$$\int \frac{d^4p}{(2\pi)^4}\delta^4(p^2-m^2)\theta(p_0) = \int \frac{d^3p}{(2\pi)^3 2E_p} \tag{5.19}$$

The density of states dN_f, that is, the number of states within $\mathbf{p}$ and $\mathbf{p}+\delta\mathbf{p}$, is:

$$dN_f = \prod_{i=1}^{N_f} \frac{V d^3p}{(2\pi)^3} \tag{5.20}$$

As before, the differential cross section $d\sigma$ is the number of transitions per unit time per unit volume divided by the flux J of incident particles:

$$\begin{aligned} d\sigma &= \frac{\text{transitions per second per cm}^3}{\text{incident flux}} \\ &= \left(\frac{|S_{fi}|^2 dN_f}{VT}\right)\frac{1}{J} \end{aligned} \tag{5.21}$$

We also know that the transition rate per unit volume (within a momentum-space interval) is given by:

$$\begin{aligned} \text{Transition rate per cm}^3 &= \frac{|S_{fi}|^2 dN_f}{VT} \\ &= \frac{(2\pi)^8 |\mathscr{T}_{fi}|^2 \delta^4(P_f - P_i)\delta^4(0)}{VT} dN_f \\ &= (2\pi)^4 \delta^4(P_f - P_i)|\mathscr{T}_{fi}|^2 dN_f \end{aligned} \tag{5.22}$$

where $(2\pi)^4\delta^4(0) = VT$.

To calculate the incident flux, we will first take a collinear frame, such as the laboratory frame or center-of-mass frame. The incident flux J equals the product of the density of the initial state $(1/V)$ and the relative velocity $v = |v_1 - v_2|$, where $v_1 = |\mathbf{p}_1|/E_1$:

$$J = |v_1 - v_2|/V \tag{5.23}$$

In the center-of-mass frame, where $\mathbf{p}_1 = -\mathbf{p}_2$, we have:

$$\begin{aligned} V(2E_1)(2E_2)J &= (2E_1)(2E_2)|v_1 - v_2| = 2E_1 2E_2 \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right| \\ &= 4|\mathbf{p}_1 E_2 - \mathbf{p}_2 E_1| \\ &= 4|\mathbf{p}_1|(E_1 + E_2) = 4[(p_1 \cdot p_2)^2 - m_1^2 m_2^2]^{1/2} \end{aligned} \tag{5.24}$$

(The last step is a bit deceiving, since it appears as if the flux is a Lorentz invariant in all frames. The last equality only holds if the two particles are collinear. The last step is not necessarily true for arbitrary Lorentz frames in which the two particles are not collinear. To see this, we can also write the flux for a beam moving in the z direction as:

$$J \sim \epsilon_{xy\mu\nu} p_1^\mu p_2^\nu \tag{5.25}$$

The flux now transforms as the x, y component of an antisymmetric second-rank tensor; that is, as an inverse area. Since the cross section is an area as seen by a particle in the beam, we expect it should transform as an area under an arbitrary Lorentz transformation. From now on, we assume that all Lorentz frames are collinear, so we can drop this distinction.)

The final formula for bosons for the differential cross section for $1 + 2 \to 3 + 4 \cdots$ is therefore given by:

$$d\sigma = \frac{(2\pi)^4 |\mathscr{M}_{fi}|^2 \delta^4(P_f - P_i)}{4\left[(p_1 \cdot p_2)^2 - m_1^2 m_2^2\right]^{1/2}} \prod_{i=3}^{N} \frac{d^3 p_i}{(2\pi)^3 2E_{p_i}} \tag{5.26}$$

in a collinear frame where $\mathscr{T}_{fi} = \prod_{i=1}^{N} (2E_{p_i} V)^{-1/2} \mathscr{M}_{fi}$. Notice that all factors of V in the S matrix have precisely cancelled against other factors coming from dN_f and the flux.

[We should note that other normalization conventions are possible. For example, we can always change the fermion normalization such that the $2m$ factor appearing in Eqs. (3.109) and (3.110) disappears. Then, with this new normalization, Eq. (5.26) works for both fermions and bosons. The advantage of this is that we can then use Eq. (5.26) without having to make the distinction between boson and fermion normalizations.]

Finally, we will use this formalism to compute the probability of the decay of a single particle. The decay probability is given by:

$$\begin{aligned} P_{\text{total}} &= \sum_f \int dN_f |S_{fi}|^2 \\ &= \sum_f \int dN_f |\mathscr{T}_{fi}|^2 \left[(2\pi)^4 \delta^4(P_f - P_i)\right]^2 \end{aligned} \tag{5.27}$$

where we have taken $\delta_{fi} = 0$ for a decay process. The last expression, unfortunately, is singular because of the delta function squared. As before, however, we assume that all our calculations are being performed in a large but finite box of volume V over a large time interval T. We thus reinterpret one of the delta functions as:

$$\delta^4(0) = \frac{VT}{(2\pi)^4} \tag{5.28}$$

We now define the decay rate Γ of an unstable particle as the transition probability per unit volume of space and time:

$$\begin{aligned} \Gamma &= \frac{\text{transition probability}}{\text{sec} \times \text{volume}} \\ &= \frac{P_{\text{total}}}{VT(2E_i)} \end{aligned} \tag{5.29}$$

The final result for the decay rate is given by:

$$\Gamma = \frac{(2\pi)^4}{2E_i} \int \prod_{j=1}^{N_f} \frac{d^3p_j}{(2\pi)^3 2E_j} |\mathcal{M}_{fi}|^2 \delta^4(P_f - P_i) \tag{5.30}$$

The lifetime of the particle τ is then defined as the inverse of the decay rate:

$$\tau = \frac{1}{\Gamma} \tag{5.31}$$

5.2 Propagator Theory and Rutherford Scattering

Historically, calculations in QED were performed using two seemingly independent formulations. One formulation was developed by Schwinger[1] and Tomonaga[2] using a covariant generalization of operator methods developed in quantum mechanics. However, the formulation was exceedingly difficult to calculate with and was physically opaque. The second formulation was developed by Feynman[3] using the propagator approach. Feynman postulated a list of simple "rules" from which one could pictorially setup the calculation for scattering matrices of arbitrary complexity. The weakness of Feynman's graphical methods, however, was that they were not rigorously justified. Later, Dyson demonstrated the equivalence of these two formulations by deriving Feynman's rules from the interaction picture.

In this chapter, we first follow Feynman to show how the propagator method gives us a rapid, convenient method of calculating the lowest order terms in the

scattering matrix. Then we will develop the Lehmann–Symanzik–Zimmermann (LSZ) reduction formalism, in which one can develop the Feynman rules for diagrams of arbitrary complexity.

At this point, we should emphasize that the Green's functions that appear in the propagator approach are "off-shell"; that is, they do not satisfy the mass-shell condition $p_\mu^2 = m^2$. Neither do they obey the usual equations of motion. The Green's functions describe virtual particles, not physical ones. As we saw in the previous chapter, the Green's function develops a pole in momentum space at $p_\mu^2 = m^2$. However, there is no violation of cherished physical principles because the Green's functions are not measurable quantities. The only measurable quantity is the S matrix, where the external particles obey the mass-shell condition.

To begin calculating cross sections, let us review the propagator method in ordinary quantum mechanics, where we wish to solve the equation:

$$\left(i\frac{\partial}{\partial t} - H\right)\psi = 0 \tag{5.32}$$

We assume that the true Hamiltonian is split into two pieces: $H = H_0 + H_I$, where the interaction piece H_I is small. We wish to solve for the propagator $G(\mathbf{x}, t; \mathbf{x}', t')$:

$$\left(i\frac{\partial}{\partial t} - H_0 - H_I\right) G(\mathbf{x}, t; \mathbf{x}', t') = \delta^3(\mathbf{x} - \mathbf{x}')\delta(t - t') \tag{5.33}$$

If we could solve for the Green's function for the interacting case, then we can use Huygen's principle to solve for the time evolution of the wave. We recall that Huygen's principle says that the future evolution of a wave front can be determined by assuming that every point along a wave front is an independent source of an infinitesimal wave. By adding up the contribution of all these small waves, we can determine the future location of the wave front. Mathematically, this is expressed by the equation:

$$\psi(\mathbf{x}, t) = \int d^3x' G(\mathbf{x}, t; \mathbf{x}', t')\psi(\mathbf{x}', t'); \qquad t > t' \tag{5.34}$$

Our next goal, therefore, is to solve for the complete Green's function G, which we do not know, in terms of the free Green's function G_0, which is well understood. To find the propagator for the interacting case, we have to power expand in H_I. We will use the following formula for operators A and B:

$$\begin{aligned}\frac{1}{A+B} &= \frac{1}{A(1+A^{-1}B)} \\ &= \frac{1}{1+A^{-1}B}\frac{1}{A}\end{aligned}$$

$$\begin{aligned} &= (1 - A^{-1}B + A^{-1}BA^{-1}B + \cdots)A^{-1} \\ &= A^{-1} - A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} + \cdots \end{aligned} \tag{5.35}$$

Another way of writing this is:

$$\frac{1}{A+B} = A^{-1} - \frac{1}{A+B}BA^{-1} \tag{5.36}$$

Now let:

$$\begin{aligned} A &= -H_0 + i\frac{\partial}{\partial t} \\ B &= -H_I \\ G &= 1/(A+B) \\ G_0 &= 1/A \end{aligned} \tag{5.37}$$

Then we have the symbolic identities:

$$\begin{aligned} G &= G_0 + GH_IG_0 \\ G &= G_0 + G_0H_IG_0 + G_0H_IG_0H_IG_0 + \cdots \end{aligned} \tag{5.38}$$

More explicitly, we can recursively write this as:

$$\begin{aligned} G(\mathbf{x}, t; \mathbf{x}', t') &= G_0(\mathbf{x}, t; \mathbf{x}', t') + \int dt_1 \int d^3x_1\, G(\mathbf{x}, t; \mathbf{x}_1, t_1) \\ &\quad \times H_I(\mathbf{x}_1, t_1)G_0(\mathbf{x}_1, t_1; \mathbf{x}', t') \end{aligned} \tag{5.39}$$

If we power expand this expression, we find:

$$\begin{aligned} G(\mathbf{x}, t; \mathbf{x}', t') &= G_0(\mathbf{x}, t; \mathbf{x}', t') + \int dt_1 \int d^3x_1 \\ &\quad \times G_0(\mathbf{x}, t; \mathbf{x}_1, t_1)H_I(\mathbf{x}_1, t_1)G_0(\mathbf{x}_1, t_1; \mathbf{x}', t') \\ &\quad + \int_{-\infty}^{\infty} dt_1\, dt_2\, d^3x_1\, d^3x_2 G_0(\mathbf{x}, t; \mathbf{x}_1, t_1)H_I(\mathbf{x}_1, t_1) \\ &\quad \times G_0(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2)H_I(\mathbf{x}_2, t_2)G_0(\mathbf{x}_2, t_2; \mathbf{x}', t') + \cdots \end{aligned} \tag{5.40}$$

Using Huygen's principle, we can power expand for the time evolution of the wave function:

$$\psi(\mathbf{x}, t) = \psi_0(\mathbf{x}, t) + \int d^4x_1\, G_0(\mathbf{x}, t; \mathbf{x}_1, t_1)\psi_0(\mathbf{x}_1, t_1)$$

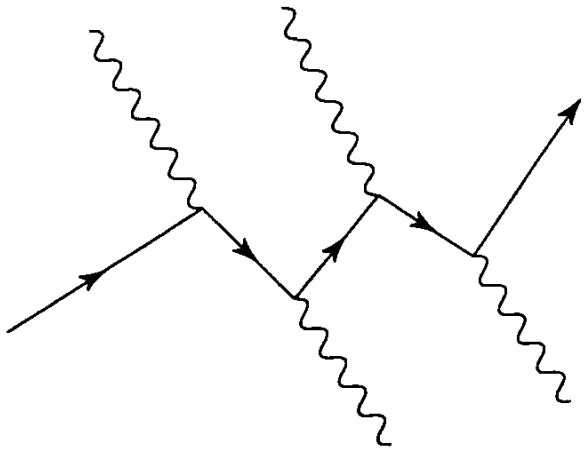


Figure 5.2. In the propagator approach, perturbation theory can be pictorially represented as a particle interacting with a background potential at various points along its trajectory.

$$+ \int d^4x_1\, d^4x_2\, G_0(\mathbf{x}, t; \mathbf{x}_1, t_1) H_I(\mathbf{x}_1, t_1) G_0(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2) \psi_0(\mathbf{x}_2, t_2)$$

$$+ \cdots + \int d^4x_1 \cdots d^4x_n\, G_0(\mathbf{x}, t; \mathbf{x}_1, t_1) H_I(\mathbf{x}_1, t_1) \cdots$$

$$\times\, G_0(\mathbf{x}_{n-1}, t_{n-1}; \mathbf{x}_n, t_n) \psi_0(\mathbf{x}_n, t_n) + \cdots \tag{5.41}$$

In Figure 5.2, we have a pictorial representation of this diagram.

Let us now solve for the S matrix in lowest order. We postulate that at infinity, there are free plane waves given by $\phi = e^{-ik\cdot x}/(2\pi)^{3/2}$. We want to calculate the transition probability that a wave packet starts out in a certain initial state i, scatters off the potential, and then re-emerges as another free plane wave, but in a different final state f. To lowest order, the transition probability can be calculated by examining Huygen's principle:

$$\psi(\mathbf{x}, t) = \phi_i(\mathbf{x}, t) + \int d^4x'\, G_0(\mathbf{x}, t; \mathbf{x}', t') H_I(\mathbf{x}', t') \phi_i(\mathbf{x}', t') + \cdots \tag{5.42}$$

To extract the S matrix, multiply this equation on the left by ϕ_j^* and integrate. The first term on the right then becomes δ_{ij}. Using the power expansion of the Green's function, we can express the Green's function G_0 in terms of these free fields. After integration, we find:

$$S_{fi} = \delta_{fi} + i \int d^4x\, \phi_f^*(x') H_I(x') \phi_i(x') + \cdots \tag{5.43}$$

Therefore, the transition matrix is proportional to the matrix element of the potential H_I. We will now generalize this exercise to the problem in question: the calculation in QED of the scattering of an electron due to a stationary Coulomb

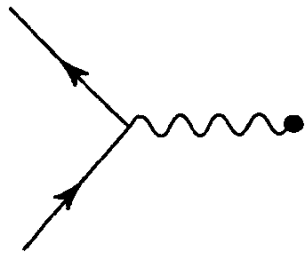


Figure 5.3. An electron scatters off a stationary Coulomb field in lowest order perturbation theory. This reproduces the Rutherford scattering cross section in the nonrelativistic limit.

potential. Our calculation should be able to reproduce the old Rutherford scattering amplitude to lowest order in the nonrelativistic limit and give higher-order quantum corrections to it. Our starting point is the Dirac electron in the presence of an external, classical Coulomb potential (Fig. 5.3).

The interacting Dirac equation reads:

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = e\not{A}(x)\psi(x) \tag{5.44}$$

Since we are only working to lowest order and are treating the potential A_μ as a classical potential, we can solve this equation using only propagator methods. The solution of this equation, as we have seen, is given by:

$$\psi(x) = \psi_0(x) + e\int d^4y\, S_F(x-y)\not{A}(y)\psi(y) \tag{5.45}$$

where ψ_0 is a solution of the free, homogeneous Dirac equation. To calculate the scattering matrix, it is convenient to insert the expansion of the Feynman propagator $S_F(x-x')$ in terms of the time-ordered function $\theta(t-t')$, as in Eq. (3.137). Then we find:

$$\psi(x) = \psi_i(x) - ie\int d^4y\, i\theta(t-t')\int d^3p \sum_{r=1}^{2} \bar{\psi}_p^r(x)\psi_p^r(y)\not{A}(y)\psi(y) \tag{5.46}$$

for $t \to \infty$. We now wish to extract from this expression the amplitude that the outgoing wave $\psi(x)$ will be scattered in the final state, given by $\psi_f(x)$. This is easily done by multiplying both sides of the equation by $\bar{\psi}_f$ and integrating over all space-time. The result gives us the S matrix to lowest order:

$$S_{fi} = \delta_{fi} - ie\int d^4x \bar{\psi}_f(x)\not{A}(x)\psi_i(x) \tag{5.47}$$

Now we insert the expression for the vector potential, which corresponds to an electric potential A_0 given by the standard $1/r$ Coulomb potential:

$$A_0(x) = -\frac{Ze}{4\pi|\mathbf{x}|} \tag{5.48}$$

whose Fourier transform is given by:

$$\int \frac{d^3x}{|\mathbf{x}|} e^{i\mathbf{q}\cdot\mathbf{x}} = \frac{4\pi}{|\mathbf{q}|^2} \tag{5.49}$$

Inserting the plane-wave expression for the fermion fields into the expression for the scattering matrix in Eq. (5.47) and performing the integration over x, we have:

$$\begin{aligned} S_{fi} &= \frac{iZe^2}{4\pi V}\int d^4x \sqrt{\frac{m}{E_f}}\bar{u}(p_f, s_f)e^{ip_f\cdot x}\left[\gamma^0\frac{1}{|\mathbf{x}|}\right]\sqrt{\frac{m}{E_i}}u(p_i, s_i)e^{-ip_i\cdot x_i} \\ &= \frac{iZe^2}{V}\sqrt{\frac{m^2}{E_f E_i}}\frac{\bar{u}(p_f, s_f)\gamma^0 u(p_i, s_i)}{|\mathbf{q}|^2}2\pi\delta(E_f - E_i) \end{aligned} \tag{5.50}$$

We recall that $Vd^3p_f/(2\pi)^3$ is the number of final states contained in the momentum interval d^3p_f. Multiply this by $|S_{fi}|^2$ and we have the probability of transition per particle into these states. [We recall that squaring the S matrix give us divergent quantities like $\delta(0)$, which is due to the fact that we have not rigorously localized the wave packets. We set $2\pi\delta(0) = T$, where we localize the scattering process in a box of size V and duration T.]

If we divide by T, this give us the rate R of transitions per unit time into this momentum interval. Finally, if we divide the rate of transitions by the flux of incident particles $|\mathbf{v}_i|/V$, this gives us the differential cross section:

$$d\sigma = |S_{fi}|^2\frac{Vd^3p_f}{(2\pi)^3}\frac{1}{VT|\mathbf{v}_i|/V} \tag{5.51}$$

To calculate the differential cross section per unit of solid angle, we must decompose the momentum volume element:

$$d^3p = d\Omega p^2\, dp \tag{5.52}$$

Using the fact that $p_f\, dp_f = E_f\, dE_f$, we have the result:

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{4Z^2\alpha^2m^2}{|\mathbf{q}|^4}\sum_{\text{spins}}\frac{1}{2}|\bar{u}(p_f, s_f)\gamma^0 u(p_i, s_i)|^2 \\ &= \frac{4Z^2\alpha^2m^2}{|\mathbf{q}|^4}\frac{1}{2}\mathrm{Tr}\left(\gamma^0\frac{\not{p}_i + m}{2m}\gamma^0\frac{\not{p}_f + m}{2m}\right) \end{aligned} \tag{5.53}$$

In the last step, we have used the fact that the summation of spins can be written as:

$$\begin{aligned}
\left|\bar{u}_f \Gamma u_i\right|^2 &= \left(\bar{u}_f \Gamma u_i\right)\left(u_i^\dagger \Gamma^\dagger \bar{u}_f^\dagger\right) \\
&= \left(\bar{u}_f \Gamma u_i\right)\left(\bar{u}_i \gamma_0 \Gamma^\dagger \gamma_0 u_f\right) \\
&= \bar{u}_{f,\alpha} \Gamma_{\alpha,\beta} u_{i,\beta} \bar{u}_{i,\gamma} \left(\gamma_0 \Gamma^\dagger \gamma_0\right)_{\gamma,\delta} u_{f,\delta} \\
&= \left(\frac{\not{p}_f + m}{2m}\right)_{\delta,\alpha} \Gamma_{\alpha,\beta} \left(\frac{\not{p}_i + m}{2m}\right)_{\beta,\gamma} \left(\gamma_0 \Gamma^\dagger \gamma_0\right)_{\gamma,\delta} \\
&= \mathrm{Tr}\left\{\frac{\not{p}_f + m}{2m} \Gamma \frac{\not{p}_i + m}{2m} \gamma_0 \Gamma^\dagger \gamma_0\right\} \qquad (5.54)
\end{aligned}$$

where we have use the fact that the sum over spins in Eq. (3.109) gives us:

$$\sum_{\text{spins}} u_\beta(p, s) \bar{u}_\alpha(p, s) = \left(\frac{\not{p}_i + m}{2m}\right)_{\beta\alpha} \qquad (5.55)$$

The last trace can be performed, since only the trace of even numbers of Dirac matrices survives:

$$\mathrm{Tr}\, \gamma^0 \not{p}_i \gamma^0 \not{p}_f = 4(2E_i E_f - p_i \cdot p_f) \qquad (5.56)$$

Finally, we need some kinematical information. If θ is the angle between p_f and p_i, then:

$$\begin{aligned}
p_i \cdot p_f &= m^2 + 2\beta^2 E^2 \sin^2(\theta/2) \\
|\mathbf{q}|^2 &= 4|\mathbf{p}|^2 \sin^2(\theta/2) \qquad (5.57)
\end{aligned}$$

We then obtain the Mott cross section[4]:

$$\frac{d\sigma}{d\Omega} = \frac{Z^2\alpha^2}{4|\mathbf{p}|^2 \sin^4(\theta/2)} \left[1 - \beta^2 \sin^2\left(\frac{\theta}{2}\right)\right] \qquad (5.58)$$

In the nonrelativistic limit, as $\beta \to 0$, we obtain the celebrated Rutherford scattering formula.

5.3 LSZ Reduction Formulas

Up to now, we have made the approximation that certain fields, such as the electromagnetic field, were classical. The development of scattering theory in this approximation was rather intuitive.

In this chapter, we would like to introduce a more convenient method, the LSZ reduction formalism,[5] from which we can derive scattering amplitudes to all orders in perturbation theory. The LSZ method gives us a simple derivation of Feynman's rules, which were originally derived from a more intuitive approach using propagator theory.

The LSZ approach begins with the physical S matrix, making as few assumptions as possible. We start by defining the "in" and "out" states, which are free particle states at asymptotic times; that is, $t = -\infty$ and ∞, respectively. We choose to distinguish them from the intermediate states, which are defined off-shell and are interacting. Our goal is to express the interacting S matrix, defined in terms of the unknown interacting field $\phi(x)$, in terms of these free asymptotic states. (We caution that in certain theories, such as QCD, the asymptotic states are bound and do not correspond to free states.)

The S matrix is defined as the matrix element of the transition from one asymptotic set of states to another. Let f denote a collection of free asymptotic states at $t = \infty$, while i refers to another collection of asymptotic states at $t = -\infty$. Then the S matrix describes the scattering of the i states into the f states:

$$S_{fi} = {}_{\text{out}}\langle f|i\rangle_{\text{in}} \tag{5.59}$$

We postulate the existence of an operator S that converts asymptotic states at $t = \infty$ to states at $t = -\infty$:

$$\begin{aligned} |f\rangle_{\text{in}} &= S|f\rangle_{\text{out}} \\ S_{fi} &= {}_{\text{out}}\langle f|S|i\rangle_{\text{out}} = {}_{\text{in}}\langle f|S|i\rangle_{\text{in}} \end{aligned} \tag{5.60}$$

For these asymptotic states, we also have asymptotic fields ϕ_{in} and ϕ_{out} that are free fields. Thus, we can use the machinery developed in Chapter 4 to describe these asymptotic free states. In particular, we can use Eqs. (3.15) and (3.16) to define the state vector as the vacuum state multiplied by creation operators:

$$\begin{aligned} |q\rangle_{\text{in}} &= a^{\dagger}_{\text{in}}(q)|0\rangle_{\text{in}} \\ &= -i\int d^3x\, e_q(x)\, \overleftrightarrow{\partial}_0\, \phi_{\text{in}}(x)|0\rangle_{\text{in}} \end{aligned} \tag{5.61}$$

The goal of the LSZ method is to reduce all expressions involving the full, interacting field $\phi(x)$ (which has matrix elements that cannot be easily computed) into simpler expressions involving the free, asymptotic fields ϕ_{in} and ϕ_{out}. The S matrix, written as a transition matrix, is useless to us at this moment. The goal of the LSZ approach, therefore, is to continue this process until we have gradually extracted out of the S matrix the entire set of fields contained within the asymptotic states. Then we can use the machinery developed in the previous chapter to manipulate and reduce these fields.

There is, however, a subtle point we should mention. Naively, one might expect that the interacting field $\phi(x)$, taken at infinitely negative or positive times, should smoothly approach the value of the free asymptotic fields, so that:

$$x^0 \to -\infty; \quad \phi(x) \to Z^{1/2}\phi(x)_{\text{in}} \tag{5.62}$$

where the factor $Z^{1/2}$ arises because of renormalization effects (which will be eliminated in Chapter 7). However, this naive assumption is actually incorrect. If we take this "strong" assumption, then it can be shown that the S matrix becomes trivial and no scattering takes place. We must therefore take the "weak" assumption, that the matrix elements of the two fields $\phi(x)$ and ϕ_{in} approach each other at infinitely negative times; that is:

$$x^0 \to -\infty; \qquad \langle f|\phi(x)|i\rangle \to Z^{1/2}\langle f|\phi_{\text{in}}(x)|i\rangle \tag{5.63}$$

For the moment, however, we will simply ignore the complications that arise due to this. We will return to the question of evaluating Z later.

Let us now take an arbitrary S matrix element for the scattering of m particles with momenta q_i into n particles with momenta p_j. We first extract the field $\phi_{\text{in}}(q_1)$ from the asymptotic "in" state:

$$\begin{aligned} {}_{\text{out}}\langle p_1, p_2 \cdots p_n|q_1, q_2, \cdots q_m\rangle_{\text{in}} &= {}_{\text{out}}\langle p_1, p_2, \cdots p_n|a^\dagger_{\text{in}}(q_1)|q_2, q_3, \cdots q_m\rangle_{\text{in}} \\ &= -i \lim_{t\to-\infty} \int d^3x\, e_{q_1}(x) \overleftrightarrow{\partial}_0 {}_{\text{out}}\langle p_1, p_2, \cdots, p_n|\phi(x)_{\text{in}}|q_2, q_3, \cdots, q_m\rangle_{\text{in}} \end{aligned} \tag{5.64}$$

Next, we wish to convert the three-dimensional integral $\int d^3x$ into a four dimensional one. We use the identity:

$$\left(\lim_{t\to\infty} - \lim_{t\to-\infty}\right) \int d^3x\, A(x,t) = \int_{-\infty}^{\infty} dt \frac{\partial}{\partial t} \int d^3x\, A(x,t) \tag{5.65}$$

Since we already have a term at $t \to -\infty$, we add and subtract the same term as $t \to \infty$. This therefore gives us an integral over four dimensional space–time,

plus a term at $t \to \infty$:

$$
\begin{aligned}
&{}_{\text{out}}\langle p_1, p_2, \cdots, p_n | q_1, q_2, \cdots, q_m \rangle_{\text{in}} = \\
&\quad + iZ^{-1/2} \int d^4x \, \partial_0 \left[e_{q_1}(x) \overleftrightarrow{\partial}_0 \, {}_{\text{out}}\langle p_1, p_2, \cdots | \phi(x) | q_2, \cdots \rangle_{\text{in}} \right] \\
&\quad - iZ^{-1/2} \lim_{t \to \infty} \int d^3x \left[e_{q_1}(x) \overleftrightarrow{\partial}_0 \, {}_{\text{out}}\langle p_1, p_2, \cdots | \phi(x) | q_2, \cdots \rangle_{\text{in}} \right]
\end{aligned}
\tag{5.66}
$$

This last term, in turn, can be written as the creation operator of an "out" state:

$$
\begin{aligned}
&-iZ^{-1/2} \lim_{t \to \infty} \int d^3x \left[e_{q_1}(x) \overleftrightarrow{\partial}_0 \, {}_{\text{out}}\langle p_1, p_2, \cdots | \phi(x) | q_2, \cdots \rangle_{\text{in}} \right] \\
&\quad = {}_{\text{out}}\langle p_1, p_2, \cdots | a^\dagger_{\text{out}}(q_1) | q_2, \cdots \rangle_{\text{in}}
\end{aligned}
\tag{5.67}
$$

The price of converting a three-dimensional integral to a four-dimensional one is that we have now generated a new term at $t \to \infty$, which is the matrix element of an "out" operator $a^\dagger_{\text{out}}(q_1)$. Because this "out" operator is an annihilation operator if it acts to the left, in general it gives us zero. The only exception is if there is, within the collection of "out" states, precisely the same state with momentum q_1. Thus, the matrix element vanishes unless, for example, the ith state with momentum p_i has exactly the same momentum as q_1:

$$
\begin{aligned}
&{}_{\text{out}}\langle p_1, p_2 \cdots | a^\dagger_{\text{out}}(q_1) | q_2, q_3, \cdots \rangle_{\text{in}} = \sum_{i=1}^{n} 2p_i^0 (2\pi)^3 \delta^3(p_i - q_1) \\
&\quad \times \, {}_{\text{out}}\langle p_1, \cdots \hat{p}_i \cdots, p_n | q_2, \cdots q_m \rangle_{\text{in}}
\end{aligned}
\tag{5.68}
$$

where the caret over a variable means that we delete that particular variable. This term is called a "disconnected graph," because one particle emerges unaffected by the scattering process, and is hence disconnected from the rest of the particles.

Now, we come to a key step in the calculation. So far, the expressions are noncovariant because of the presence of two time derivatives. We will now convert the time derivatives in the reduced matrix element into a fully covariant object, thus restoring Lorentz invariance. This is possible because the operator $\partial_0^2 - \partial_i^2 + m^2$ annihilates the plane wave $\exp(-iq_1 \cdot x)$. By integrating by parts, we will be able to convert the various time derivatives into a fully covariant ∂_μ^2:

$$
\int d^4x \, \partial_0 \left[e_{q_1}(x) \overleftrightarrow{\partial}_0 \, {}_{\text{out}}\langle p_1, \cdots | \phi(x) | q_2, \cdots \rangle_{\text{in}} \right]
$$

$$
\begin{aligned}
&= -\int d^4x \left[\partial_0^2 e_{q_1}(x)\right] {}_{\text{out}}\langle p_1, \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}} \\
&\quad + \int d^4x\, e_{q_1}(x)\, \partial_0^2\, {}_{\text{out}}\langle p_1, \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}} \\
&= \int d^4x \left[(-\partial_i^2 + m^2) e_{q_1}(x)\right] {}_{\text{out}}\langle p_1, \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}} \\
&\quad + \int d^4x\, e_{q_1}(x)\, \partial_0^2\, {}_{\text{out}}\langle p_1, \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}} \\
&= \int d^4x\, e_{q_1}(x)(\partial_\mu^2 + m^2)\, {}_{\text{out}}\langle p_1 \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}} \qquad (5.69)
\end{aligned}
$$

Collecting our results, we find:

$$
\begin{aligned}
&{}_{\text{out}}\langle p_1, p_2, \cdots, p_n | q_1, q_2, \cdots, q_m\rangle_{\text{in}} = \text{disconnected graph} \\
&\quad + iZ^{-1/2} \int d^4x\, e_{q_1}(x)\, (\partial_\mu^2 + m^2)\, {}_{\text{out}}\langle p_1, p_2, \cdots |\phi(x)|q_2, \cdots\rangle_{\text{in}}
\end{aligned} \qquad (5.70)
$$

We have now completed the first step of the LSZ program. By extracting the state with momentum q_1 from the "in" states, we now have reduced an abstract S matrix element into the matrix element of a field ϕ. Our goal, obviously, is to continue this process until all the asymptotic fields are extracted from the asymptotic states.

A small complication emerges when we extract out a second field from the "out" state with momentum p_1. We find that we must adopt the time-ordered product of two fields (otherwise, we cannot make the transition from time derivatives to the Lorentz covariant derivatives). Repeating the identical steps as before, and including this important feature of time ordering, we find:

$$
\begin{aligned}
&{}_{\text{out}}\langle p_1, \cdots |\phi(x_1)|q_2, \cdots\rangle_{\text{in}} = {}_{\text{out}}\langle p_2, \cdots |\phi(x_1) a_{\text{in}}(p_1)|q_2, \cdots\rangle_{\text{in}} \\
&\quad + iZ^{-1/2} \int d^4y_1\, e^*_{p_1}(y_1)\, (\partial_\mu^2 + m^2)_{y_1}\, {}_{\text{out}}\langle p_2, \cdots |T\phi(y_1)\phi(x_1)|q_2 \cdots\rangle_{\text{in}}
\end{aligned} \qquad (5.71)
$$

As before, the a_{in} operator now acts to the right, where it is an annihilation operator that destroys a state with exactly momentum p_1, generating a disconnected graph. Completing all steps, we now have the twice-reduced matrix element:

$$
{}_{\text{out}}\langle p_1, \cdots | \; q_1 \cdots\rangle_{\text{in}} = \text{disconnected graphs}
$$

$$+ \left(iZ^{-1/2}\right)^2 \int d^4x_1\, d^4y_1\, e^*_{p_1}(y_1) e_{q_1}(x_1) (\partial^2_\mu + m^2)_{y_1} (\partial^2_\mu + m^2)_{x_1}$$

$$\times \ {}_{\text{out}}\langle p_2, \cdots | T\phi(y_1)\phi(x_1) | q_2, \cdots \rangle_{\text{in}} \tag{5.72}$$

It is now straightforward to apply this reduction process to all fields contained within the asymptotic state vectors. After each reduction, the only complication is that we generate disconnected graphs and we must be careful with time ordering. The final result is:

$$\begin{aligned} {}_{\text{out}}\langle p_1, p_2, \cdots, p_n | q_1, q_2, \cdots, q_m \rangle_{\text{in}} &= \text{disconnected graphs} \\ &+ \left(iZ^{-1/2}\right)^{n+m} \int d^4y_1 \cdots d^4x_m \prod_{i=1}^{n} \prod_{j=1}^{m} e^*_{p_i}(y_i) e_{q_j}(x_j) \\ &\times (\partial^2_\mu + m^2)_{y_1} \cdots (\partial^2_\mu + m^2)_{x_m} \langle 0 | T\phi(y_1) \cdots \phi(x_m) | 0 \rangle \end{aligned} \tag{5.73}$$

(In general, we can choose momenta so that the disconnected graphs are zero. For values of the momenta where the disconnected graphs are non-zero, we can use this formalism to reduce them out as well.)

5.4 Reduction of Dirac Spinors

It is now straightforward to apply this formalism to the reduction of fermionic S matrices. Following the steps of the bosonic case, we write the creation and annihilation operators in terms of the original Dirac field. Then we write the asymptotic condition as:

$$x^0 \to -\infty; \qquad \langle f | \psi(x) | i \rangle \to Z_2{}^{-1/2} \langle f | \psi_{\text{in}}(x) | i \rangle \tag{5.74}$$

Next, we write the S matrix and reduce out one creation operator using Eq. (3.116):

$$\begin{aligned} {}_{\text{out}}\langle f | k, i \rangle_{\text{in}} &= {}_{\text{out}}\langle f | b^\dagger_{\text{in}}(k, \epsilon) | i \rangle_{\text{in}} \\ &= -i \lim_{t \to -\infty} \int d^3x \ {}_{\text{out}}\langle f | \bar{\psi}(x) \gamma^0 | i \rangle_{\text{in}} U_k(x, \epsilon) \\ &= {}_{\text{out}}\langle f | b^\dagger_{\text{out}}(k, \epsilon) | i \rangle_{\text{in}} \\ &\quad - iZ_2{}^{-1/2} \int d^3x \left[{}_{\text{out}}\langle f | \bar{\psi}(x) | i \rangle_{\text{in}} (-i) \overleftarrow{\partial}_0 \gamma^0 U_k(x, \epsilon) \right. \end{aligned}$$

$$+ \ {}_{\text{out}}\langle f|\bar{\psi}(x)|i\rangle_{\text{in}}(-i)\overrightarrow{\partial}_0\gamma^0 U_k(x,\epsilon)\Big] \tag{5.75}$$

As before, we must convert the $\gamma^0\partial_0$ into the covariant $\gamma^\mu\partial_\mu$. This is accomplished because we can replace the time derivative with a space derivative (because the spinor $u(k,\epsilon)e^{-ik\cdot x}$ satisfies the Dirac equation). In this way, we can now write down four different types of reduction formulas, depending on which creation or annihilation process we are analyzing.

The reduction formulas after a single reduction now read as follows:

$$\begin{aligned}
{}_{\text{out}}\langle f|b^\dagger_{\text{in}}(k,\epsilon)|i\rangle_{\text{in}} &= -iZ_2^{-1/2}\int d^4x\ {}_{\text{out}}\langle f|\bar{\psi}(x)|i\rangle_{\text{in}}(-i\overleftarrow{\not\partial}-m)U_k(x,\epsilon)\\
{}_{\text{out}}\langle f|d^\dagger_{\text{in}}(k,\epsilon)|i\rangle_{\text{in}} &= +iZ_2^{-1/2}\int d^4x\,\bar{V}_k(x,\epsilon)(i\overrightarrow{\not\partial}-m)\ {}_{\text{out}}\langle f|\psi(x)|i\rangle_{\text{in}}\\
{}_{\text{out}}\langle f|b_{\text{out}}(k,\epsilon)|i\rangle_{\text{in}} &= -iZ_2^{-1/2}\int d^4x\,\bar{U}_k(x,\epsilon)(i\overrightarrow{\not\partial}-m)\ {}_{\text{out}}\langle f|\psi(x)|i\rangle_{\text{in}}\\
{}_{\text{out}}\langle f|d_{\text{out}}(k,\epsilon)|i\rangle_{\text{in}} &= +iZ_2^{-1/2}\int d^4x\ {}_{\text{out}}\langle f|\bar{\psi}(x)|i\rangle_{\text{in}}(-i\overleftarrow{\not\partial}-m)V_k(x,\epsilon)
\end{aligned} \tag{5.76}$$

where we have dropped the disconnected graph.

Making successive reductions, until all creation and annihilation operators are reduced out, is also straightforward. As before, we find that we must take time-ordered matrix elements of the various fields. Let us take the matrix element between incoming particles and outgoing particles. The incoming particles are labeled by $p_1, p_2, \ldots$, while the incoming antiparticles are labeled by $p'_1, p'_2, \ldots$. The outgoing particles are labeled by $q_1, q_2, \ldots$, while the outgoing antiparticles are labeled by $q'_1, q'_2, \ldots$.

The matrix element, after reduction, yields the following:

$$\begin{aligned}
&{}_{\text{out}}\langle 0|b_{\text{out}}(q_1)\cdots d_{\text{out}}(q'_1)\cdots b^\dagger_{\text{in}}(p_1)\cdots d^\dagger_{\text{in}}(p'_1)\cdots|0\rangle_{\text{in}}\\
&= (-iZ_2^{-1/2})^{n/2}(iZ_2^{-1/2})^{-n'/2}\int d^4x_1\cdots d^4x'_1\cdots d^4y_1\cdots d^4y'_1\cdots\\
&\quad\times\bar{U}_{q_1}(y_1)(i\overrightarrow{\not\partial}-m)_{y_1}\cdots\bar{V}_{p'_1}(x'_1)(i\gamma^\mu\overrightarrow{\not\partial}-m)_{x'_1}\cdots\\
&\quad\times\langle 0|T\left[\bar{\psi}(y'_1)\cdots\psi(y_1)\cdots\bar{\psi}(x_1)\cdots\psi(x'_1)\cdots\right]|0\rangle\\
&\quad\times(-i\overleftarrow{\not\partial}-m)_{x_1}U_{p_1}(x_1)\cdots(-i\overleftarrow{\not\partial}-m)_{y'_1}V_{q'_1}(y'_1)\cdots
\end{aligned} \tag{5.77}$$

where we have discarded the disconnected graphs. In this way, we can reduce out even the most complex scattering processes in terms of the reduction formulas.

5.5 Time Evolution Operator

The LSZ reduction formulas have been able to convert the abstract S matrix element into the product of vacuum expectation values of the fully interacting fields. No approximations have been made. However, we still do not know how to take matrix elements of the interacting fields. Hence, we cannot yet extract out numbers out of these matrix elements. The problem is that everything is written in terms of the fully interacting fields, of which we know almost nothing. The key is now to make an approximation to the theory by power expanding in the coupling constant, which is of the order of 1/137 for QED. We begin by splitting the Hamiltonian into two distinct pieces:

$$H = H_0 + H_I \tag{5.78}$$

where H_0 is the free Hamiltonian and H_I is the interacting part. For the ϕ^4 theory, for example, the interacting part would be:

$$\begin{aligned} H_I &= \int d^3x \, \mathscr{H}_I \\ \mathscr{H}_I &= \frac{\lambda}{4!}\phi^4 \end{aligned} \tag{5.79}$$

At this point, it is useful to remind ourselves from ordinary quantum mechanics that there are several "pictures" in which to describe this time evolution. In the *Schrödinger picture*, we recall, the wave function $\psi(\mathbf{x}, t)$ and state vector are functions of time t, but the operators of the theory are constants in time. In the *Heisenberg picture*, the reverse is true; that is, the wave function and state vectors are constants in time, but the time evolution of the operators and dynamical variables of the theory are governed by the Hamiltonian:

$$\phi(\mathbf{x}, t) = e^{iHt}\phi(\mathbf{x}, 0)e^{-iHt} \tag{5.80}$$

In the LSZ formalism, we will find it convenient to define yet another picture, which resembles the interaction picture. In this new picture, we need to find a unitary operator $U(t)$ that takes us from the fully interacting field $\phi(x)$ to the free, asymptotic "in" states:

$$\phi(t, \mathbf{x}) = U^{-1}(t)\phi_{\text{in}}(t, \mathbf{x})U(t) \tag{5.81}$$

where $U(t) \equiv U(t, -\infty)$ is a time evolution operator, which obeys:

$$\begin{aligned} U(t_1, t_2)U(t_2, t_3) &= U(t_1, t_3) \\ U^{-1}(t_1, t_2) &= U(t_2, t_1) \\ U(t, t) &= 1 \end{aligned} \tag{5.82}$$

Because we now have two totally different types of scalar fields, one free and the other interacting, we must also be careful to distinguish the Hamiltonian written in terms of the free or the interacting fields. Let $H(t)$ be the fully interacting Hamiltonian written in terms of the interacting field, and let $H_0(\phi_{\text{in}})$ represent the free Hamiltonian written in terms of the free asymptotic states. Then the free field ϕ_{in} and the interacting field satisfy two different equations of motion:

$$\begin{aligned} \frac{\partial}{\partial t}\phi(t, \mathbf{x}) &= i[H(t), \phi(t, \mathbf{x})] \\ \frac{\partial}{\partial t}\phi_{\text{in}}(t, \mathbf{x}) &= i[H_0^{\text{in}}, \phi_{\text{in}}(t, \mathbf{x})] \end{aligned} \tag{5.83}$$

To solve for $U(t)$, we need to extract a few more identities. If we differentiate the expression $UU^{-1} = 1$, we find:

$$\left[\frac{d}{dt}U(t)\right]U^{-1}(t) + U(t)\frac{d}{dt}U^{-1}(t) = 0 \tag{5.84}$$

Now let us take the derivative of ϕ_{in} and use the identities that we have written down:

$$\begin{aligned} \frac{\partial}{\partial t}\phi_{\text{in}}(t, \mathbf{x}) &= \frac{\partial}{\partial t}\left[U(t)\phi(t, \mathbf{x})U^{-1}\right] \\ &= \dot{U}(t)\phi(t, \mathbf{x})U^{-1} + U(t)\dot{\phi}(t, \mathbf{x})U^{-1}(t) + U(t)\phi(t, \mathbf{x})\dot{U}^{-1}(t) \\ &= \dot{U}(t)\left(U^{-1}\phi_{\text{in}}U\right)U^{-1} + U(t)\left[iH(\phi, \pi), \phi\right]U^{-1} \\ &\quad + UU^{-1}\phi_{\text{in}}U\dot{U}^{-1} \\ &= \dot{U}U^{-1}\phi_{\text{in}} + iU[H(\phi, \pi), \phi]U^{-1} + \phi_{\text{in}}U\dot{U}^{-1} \\ &= \left[\dot{U}U^{-1} + iH(\phi_{\text{in}}, \pi_{\text{in}}), \phi_{\text{in}}\right] \end{aligned} \tag{5.85}$$

The last expression, in turn, must equal $i[H_0^{\text{in}}, \phi_{\text{in}}]$. This means that the following expression commutes with every "in" operator, and hence must be a c number:

$$\dot{U}U^{-1} + i\left[H(\phi_{\text{in}}, \pi_{\text{in}}) - H_0^{\text{in}}\right] = c \text{ number} \tag{5.86}$$

(This c number, one can show, does not contribute to the S matrix.) Thus, the $U(t)$ operator satisfies the following:

$$i\frac{\partial}{\partial t}U(t,t_0) = H_I(t)U(t,t_0) \tag{5.87}$$

where:

$$H_I(t) \equiv H(\phi_{\text{in}}, \pi_{\text{in}}) - H_0^{\text{in}} \tag{5.88}$$

that is, $H_I(t)$ is defined to be the interaction Hamiltonian defined only with free, asymptotic fields. Since $H_I(t)$ does not necessarily commute with $H_I(t')$ at different times, the integration of the previous equation is a bit delicate. However, one can show that:

$$\begin{aligned} U(t) &= U(t,-\infty) \\ &= T\ \exp\left(-i\int_{-\infty}^{t} dt_1 H_I(t_1)\right) \\ &= T\ \exp\left(-i\int_{-\infty}^{t} dt_1 \int d^3x_1\, \mathscr{H}_I(\mathbf{x}_1,t_1)\right) \end{aligned} \tag{5.89}$$

where the operator T means that, as we integrate over t_1, we place the exponentials sequentially in time order. To prove this expression, we simply insert it into Eq. (5.87). Written in this form, however, this expression is not very useful. We will find it much more convenient to power expand the exponential in a Taylor series, so we have:

$$\begin{aligned} U(t) &= 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{t} d^4x_1 \int_{-\infty}^{t} d^4x_2 \cdots \\ &\quad \times \int_{-\infty}^{t} d^4x_n\, T\left[\mathscr{H}_I(x_1)\mathscr{H}_I(x_2)\cdots\mathscr{H}_I(x_n)\right] \end{aligned} \tag{5.90}$$

(See Exercise 10 concerning the change in the upper limits of integration.) Now that we have an explicit solution for $U(t)$, let us decompose the interacting Green's function. If we take the matrix element of a series of interacting ϕ fields, we will use the $U(t)$ operator to convert the entire expression to free, asymptotic fields. To do this, let us first choose a sequence of space–time points x_i^μ, time ordered such that $x_1^0 > x_2^0 > \cdots x_n^0$, so we can drop the time ordering operator T.

Then we can replace all interacting fields with free fields by making the conversion $\phi = U^{-1}\phi_{\text{in}}U$ everywhere:

$$\langle 0|T\left(\phi(x_1)\phi(x_2)\cdots\phi(x_n)\right)|0\rangle = \langle 0|\phi(x_1)\cdots\phi(x_n)|0\rangle$$

$$
\begin{aligned}
&= \langle 0|U^{-1}(t_1)\phi_{\text{in}}(x_1)U(t_1)U^{-1}(t_2)\phi_{\text{in}}(x_2)U(t_2) \\
&\quad \cdots U(t_{n-1})U^{-1}(t_n)\phi_{\text{in}}(x_n)U(t_n)|0\rangle \\
&= \langle 0|U^{-1}(t_1)\phi_{\text{in}}(x_1)U(t_1,t_2)\phi_{\text{in}}(x_2)\cdots U(t_{n-1},t_n)\phi_{\text{in}}(x_n)U(t_n)|0\rangle \\
&= \langle 0|U^{-1}(t)U(t,t_1)\phi_{\text{in}}(x_1)\cdots\phi_{\text{in}}(x_n)U(t_n,-t)U(-t)|0\rangle \\
&= \langle 0|U^{-1}(t)T\Big[\phi_{\text{in}}(x_1)\phi_{\text{in}}(x_2)\cdots\phi_{\text{in}}(x_n)U(t,-t)\Big]U(-t)|0\rangle \\
&= \left\langle 0\middle|U^{-1}(t)T\left[\phi_{\text{in}}(x_1)\cdots\phi_{\text{in}}(x_n)\exp\left(-i\int_{-t}^{t}dt'H_I(t')\right)\right]U(-t)\middle|0\right\rangle
\end{aligned}
\tag{5.91}
$$

where t is an arbitrarily long time, much greater than t_1, and $-t$ is much less than t_n. We will later set $t = \infty$. In the last line, since we have restored the time ordering operator T, we are allowed to move all U's around inside the matrix element. We can thus combine them all into one large $U(t, -t)$, which in turn can be expanded as a function of the interacting Lagrangian (defined strictly in terms of free, asymptotic fields).

There is now one last step that must be performed. We still have the term $U(-t)|0\rangle$ and $\langle 0|U^{-1}(t)$ to eliminate in the limit as $t \to \infty$. In general, since we assume that the vacuum is stable for this theory, we know that the vacuum is an eigenstate of the U operator, up to some phase.

Since $U(t) \equiv U(t, -\infty)$, we can set:

$$\lim_{t\to\infty} U(-t)|0\rangle = |0\rangle \tag{5.92}$$

where we have taken the limit so that $U(-\infty, -\infty) = 1$. However, for the other state $\langle 0|U^{-1}(t)$, we must be a bit more careful, since the limit gives us $U^{-1}(\infty) = U(-\infty, +\infty)$. Since the vacuum is stable, this means that the vacuum at $t = -\infty$ remains the vacuum at $t = +\infty$, modulo a possible phase λ. We thus find:

$$\lim_{t\to\infty} \langle 0|U^{-1}(t) = \lambda\langle 0| \tag{5.93}$$

By hitting the equation with $|0\rangle$, the phase λ is equal to:

$$
\begin{aligned}
\lambda &= \lim_{t\to\infty} \langle 0|U^{-1}(t)|0\rangle \\
&= \lim_{t\to\infty} \langle 0|U(t)|0\rangle^{-1}
\end{aligned}
$$

$$= \left\langle 0 \left| T \exp\left(i \int d^4x \mathscr{L}_I(\phi_{\text{in}}) \right) \right| 0 \right\rangle^{-1} \tag{5.94}$$

The phase λ thus gives us the contribution of vacuum-to-vacuum graphs (i.e., graphs without any external legs). Putting everything together, we now have an expression for the Green's function defined with interacting fields:

$$G(x_1, x_2, \cdots, x_n) = \langle 0 | T \left[\phi(x_1)\phi(x_2) \cdots \phi(x_n) \right] | 0 \rangle \tag{5.95}$$

This interacting Green's function, written in terms of free fields, becomes:

$$\begin{aligned} & G(x_1, x_2, \cdots, x_n) \\ &= \frac{\langle 0 | T \phi_{\text{in}}(x_1) \cdots \phi_{\text{in}}(x_n) \exp\left\{ i \int d^4x \mathscr{L}_I(\phi_{\text{in}}) \right\} | 0 \rangle}{\langle 0 | T \exp\left\{ i \int d^4x \mathscr{L}_I \left[\phi_{\text{in}}(x) \right] \right\} | 0 \rangle} \end{aligned} \tag{5.96}$$

Using the formalism that we have constructed, we can rewrite the previous matrix element entirely in terms of the asymptotic "in" fields by a power expansion of the exponential:

$$\begin{aligned} G(x_1, x_2, \cdots, x_n) \;=\; & \sum_{m=0}^{\infty} \frac{(-i)^m}{m!} \int_{-\infty}^{\infty} d^4y_1 \cdots d^4y_m \langle 0 | T \Big[\phi(x_1)\phi(x_2) \\ & \cdots \phi(x_n) \mathscr{H}_I(y_1) \mathscr{H}_I(y_2) \cdots \mathscr{H}_I(y_m) \Big] | 0 \rangle_c \end{aligned} \tag{5.97}$$

where the subscript c refers to connected diagrams only. (From now on, we will drop the "in" subscript on all fields. However, we must remind the reader that we have made the transition from the Heisenberg picture to this new picture where all fields are free.)

The next step is to actually evaluate the time ordered product of an arbitrary number of free fields. To do this, we appeal to *Wick's theorem.*

5.6 Wick's Theorem

We begin our discussion by defining the "normal ordering" of operators. In general, unlike the classical situation, the product of two quantum fields taken at the same point is singular:

$$\lim_{x \to y} \phi(x)\phi(y) = \infty \tag{5.98}$$

(In fact, in Chapter 14, we will investigate precisely how divergent this expression is.) Unfortunately, our action consists of fields multiplied at the same point; so the transition to the quantum theory is actually slightly ambiguous. To render our expressions rigorous, we recall the normal ordered product of two fields introduced in Eq. (3.27). Let us decompose the scalar field into its creation operators [labeled by a $-$ sign, and annihilation operators, labeled by a $+$ sign: $\phi(x) = \phi^+(x)+\phi(x)^-$]. Then the normal ordered product of these fields simply rearranges the creation and annihilation parts such that the creation operators always appear on the left, and annihilation operators always appear on the right. The normal ordered product is defined as:

$$\begin{aligned} :\phi(x)\phi(y): \;\equiv\; & \phi(x)^+\phi^+(y) + \phi(x)^-\phi^+(y) + \phi(y)^-\phi^+(x) \\ + \;& \phi(x)^-\phi^-(y) \end{aligned} \tag{5.99}$$

(The normal ordered product of two fields is no longer a local object, since we have split up the components within the fields and reshuffled them.)

One consequence of normal ordering is that the vacuum expectation value of any normal ordered product vanishes, since annihilation operators always appear on the right. It vanishes because:

$$\phi^+|0\rangle = \langle 0|\phi^- = 0 \tag{5.100}$$

Similarly, one can define normal ordered products for more complicated products of fields. Now we can proceed to find the relationship between normal ordered products and time ordered products. It is easy to prove the following identity, or Wick's theorem for two fields:

$$T\,[\phi(x_1)\phi(x_2)] =: \phi(x_1)\phi(x_2): + c \text{ number} \tag{5.101}$$

The only difference between the time ordered and normal ordered products is that we have reshuffled the various annihilation and creation parts of the fields. Each time we commute parts of the fields past each other, we pick up c-number expressions. To find what this c-number expression is, we now simply take the vacuum expectation value of both sides. Since the vacuum expectation value of normal ordered products is zero, we find:

$$T\,[\phi(x_1)\phi(x_2)] =: \phi(x_1)\phi(x_2): + \langle 0|T\,[\phi(x_1)\phi(x_2)]\,|0\rangle \tag{5.102}$$

If we have three fields, then Wick's theorem[6] reads:

$$T\,(\phi(x_1)\phi(x_2)\phi(x_3)) \;=\; :\phi(x_1)\phi(x_2)\phi(x_3):$$

$$+ \langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle\phi(x_3)$$
$$+ \langle 0|T\left[\phi(x_2)\phi(x_3)\right]|0\rangle\phi(x_1)$$
$$+ \langle 0|T\left[\phi(x_3)\phi(x_1)\right]|0\rangle\phi(x_2) \qquad (5.103)$$

The last three terms can be succinctly summarized by the sum $\sum_{\text{perm}}$, which means summing over all permutations of x_1, x_2, x_3. To prove this expression with three scalar fields, we take the original identity with just two scalar fields, and multiply both sides of the equation on the right by $\phi(x_3)$, which has the earliest time. Then we merge $\phi(x_3)$ into the normal ordered product. Merging $\phi^+(x_3)$ into the normal ordered product is easy, since the annihilation operators are on the right, anyway. However, merging $\phi^-(x_3)$ into the normal ordered product is more difficult, since $\phi^-(x_3)$ must move past $\phi^+(x_1)$ and $\phi^+(x_2)$. Each time it moves past one of these terms, we pick up a c-number expression, which is equal to the time ordered product; that is:

$$\begin{aligned}\langle 0|\phi^+(x_2)\phi^-(x_3)|0\rangle &= \langle 0|\phi(x_2)\phi(x_3)|0\rangle \\ &= \langle 0|T\left[\phi(x_2)\phi(x_3)\right]|0\rangle \qquad (5.104)\end{aligned}$$

In this way, we pick up all the terms in the Wick identity. Likewise, for four fields, we have:

$$\begin{aligned}\langle 0|T\left[\phi(x_1)\cdots\phi(x_4)\right]|0\rangle &=: \phi(x_1)\cdots\phi(x_4): \\ &+ \sum_{\text{perm}}\langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle\ :\phi(x_3)\phi(x_4): \\ &+ \sum_{\text{perm}}\langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle\langle 0|T\left[\phi(x_3)\phi(x_4)\right]|0\rangle \qquad (5.105)\end{aligned}$$

By now, it should be obvious that the time ordered product of n fields can be written in terms of sum of normal ordered products. For the general n-point case (n even), Wick's theorem reads:

$$\begin{aligned}T\left[\phi(x_1)\phi(x_2)\cdots\phi(x_n)\right] &=: \phi(x_1)\phi(x_2)\cdots\phi(x_n): \\ &+ \sum_{\text{perm}}\langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle\ :\phi(x_3)\cdots\phi(x_n): \\ &+ \sum_{\text{perm}}\langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle\langle 0|T\left[\phi(x_3)\phi(x_4)\right]|0\rangle\ :\phi(x_5)\cdots\phi(x_n): \\ &\ \vdots \\ &+ \sum_{\text{perm}}\langle 0|T\left(\phi(x_1)\phi(x_2)\right)|0\rangle\cdots\langle 0|T\left(\phi(x_{n-1})\phi(x_n)\right)|0\rangle \qquad (5.106)\end{aligned}$$

For n odd, the last line reads:

$$\sum_{\text{perm}} \langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle \cdots \langle 0|T\left[\phi(x_{n-2})\phi(x_{n-1})\right]|0\rangle \phi(x_n) \tag{5.107}$$

These formulas are proved by induction on n. If we assume that they hold for $n-1$, then we multiply the entire formula on the right by $\phi(x_n)$, which has the earliest time. Then, by merging $\phi(x_n)$ with the rest of the products, we find that the formula now holds for n. The proof is straightforward, and runs the same as for the case described earlier. Merging $\phi^+(x_n)$ poses no problem, since annihilation operators are on the right anyway. However, each time $\phi^-(x_n)$ moves past $\phi^+(x_j)$, we pick up a commutator, which is equal to the vacuum expectation of the time ordered product of the two fields. In this way, it is easy to show that we pick up the nth Wick's theorem.

If we take the vacuum expectation value of both sides of the equation, then Wick's theorem for vacuum expectation values reads:

$$\begin{aligned}\langle 0|T\left[\phi(x_1)\phi(x_2) \quad \cdots \phi(x_n)\right]|0\rangle = \sum_{\text{perm}} &\langle 0|T\left[\phi(x_1)\phi(x_2)\right]|0\rangle \\ &\cdots \langle 0|T\left[\phi(x_{n-1})\phi(x_n)\right]|0\rangle\end{aligned} \tag{5.108}$$

The generalization to fermionic fields is also straightforward. [The only complication is that we pick up extra minus signs because of the anti-commutation properties of fermionic fields. For more complicated products, we must always insert (-1) whenever fermion fields move past each other.] We find:

$$T\left[\psi(x)\bar{\psi}(y)\right] =: \psi(x)\bar{\psi}(y) : + \langle 0|T\left[\psi(x)\bar{\psi}(y)\right]|0\rangle \tag{5.109}$$

Now insert Eqs. (5.97) and (5.108) into Eq. (5.73). This gives us a complete reduction of the S matrix (written as a function of interacting fields) in terms of Green's functions of free fields, as desired. In the last step, we can eliminate the $(\partial_\mu^2 + m^2)$ factors appearing in Eq. (5.73), because they act on two-point Green's functions and become delta functions:

$$\begin{aligned}(\partial_\mu^2 + m^2)_x \langle 0|T\phi(x)\phi(y)|0\rangle &= -i\delta^4(x-y) \\ (i\,\partial\!\!\!/ - m)_x \langle 0|T\psi(x)\bar{\psi}(y)|0\rangle &= i\delta^4(x-y)\end{aligned} \tag{5.110}$$

In this fashion, we have how completely reduced the S matrix element into sums of products of two-point Green's functions and certain vertex elements. Although Wick's theorem seems a bit tedious, in actual practice the decomposition proceeds

rapidly. To see this, we will analyze the four-point function taken to first order in with an interaction given by $-\lambda\phi^4/4!$. We want to expand:

$$\begin{aligned} G(x_1,\cdots x_4) &= -\frac{i\lambda}{4!}\int d^4y\langle 0|T\left[\phi(x_1)\cdots\phi(x_4)\phi(y)^4\right]|0\rangle+\cdots \\ &= (-i\lambda)\int d^4y\prod_{i=1}^{4}\langle 0|T\left[\phi(x_i)\phi(y)\right]|0\rangle+\cdots \\ &= (-i\lambda)\int d^4y\prod_{i=1}^{4}\left[i\Delta_F(x_i-y)\right]+\cdots \end{aligned} \tag{5.111}$$

where we have used Wick's theorem. The 4! term has disappeared, because there are 4! ways in which four external legs at x_i can be connected to the four fields contained within ϕ^4.

Another example of this decomposition is given by the four-point function taken to second order:

$$\begin{aligned} G(x_1,\cdots,x_4) &= \left(\frac{-i\lambda}{4!}\right)^2\frac{1}{2!}\int d^4y_1\int d^4y_2\langle 0|T\Big\{\phi(x_1)\cdots\phi(x_4) \\ &\quad\times[\phi(y_1)^4][\phi(y_2)^4]\Big\}|0\rangle \end{aligned} \tag{5.112}$$

The expansion, via Wick's theorem, is straightforward:

$$\begin{aligned} G(x_1,\cdots x_4) &= \frac{(-i\lambda)^2}{2!}\int d^4y_1\int d^4y_2\Big\{[i\Delta_F(y_1-y_2)]^2\Delta_A \\ &\quad+[i\Delta_F(y_1-y_1)][i\Delta_F(y_1-y_2)]\Delta_B\Big\} \end{aligned} \tag{5.113}$$

where:

$$\begin{aligned} \Delta_A &= \Delta_F(x_1-y_1)\Delta_F(x_2-y_1)\Delta_F(x_3-y_2)\Delta_F(x_4-y_2) \\ &\quad+\Delta_F(x_1-y_1)\Delta_F(x_3-y_1)\Delta_F(x_2-y_2)\Delta_F(x_4-y_2) \\ &\quad+\Delta_F(x_1-y_1)\Delta_F(x_4-y_1)\Delta_F(x_2-y_2)\Delta_F(x_3-y_2) \end{aligned} \tag{5.114}$$

$$\begin{aligned} \Delta_B &= \Delta_F(x_1-y_1)\Delta_F(x_2-y_2)\Delta_F(x_3-y_2)\Delta_F(x_4-y_2) \\ &\quad+\Delta_F(x_1-y_2)\Delta_F(x_2-y_1)\Delta_F(x_3-y_2)\Delta_F(x_4-y_2) \\ &\quad+\Delta_F(x_1-y_2)\Delta_F(x_2-y_2)\Delta_F(x_3-y_1)\Delta_F(x_4-y_2) \\ &\quad+\Delta_F(x_1-y_2)\Delta_F(x_2-y_2)\Delta_F(x_3-y_1)\Delta_F(x_4-y_2) \end{aligned} \tag{5.115}$$

These are shown graphically in Figure 5.4.

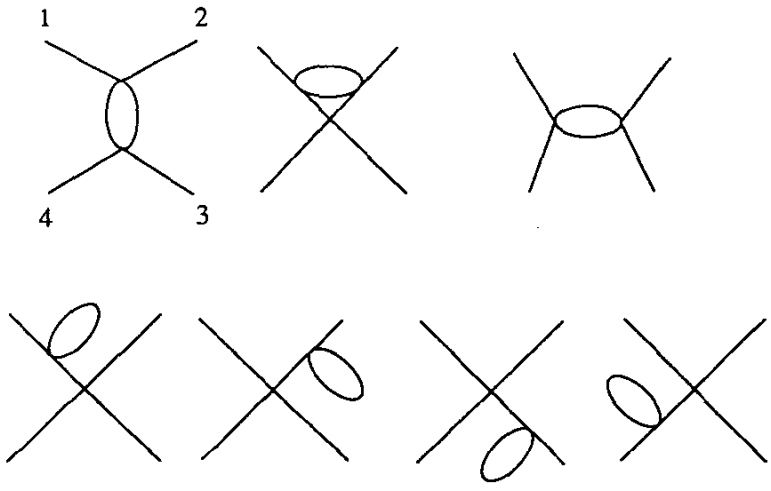


Figure 5.4. The Feynman diagrams corresponding to the Wick decomposition of ϕ^4 theory to second order.

Finally, we will convert from x-space to p-space by using Eqs. (3.50) and (3.135). When we perform the x integrations, we obtain a delta function at each vertex, which represents the conservation of momentum. When all x-space integrations are performed, we are left with one delta function representing overall momentum conservation. We are also left with a momentum integration for each internal loop. Then all Feynman's rules can be represented entirely in momentum space.

5.7 Feynman's Rules

From this, we can extract graphical rules by Feynman rules, by which we can almost by inspection construct Green's functions of arbitrary complexity. With an interaction Lagrangian given by $-\lambda\phi^4/4!$, $\mathscr{M}_{fi}$ appearing in Eq. (5.26) can be calculated as follows:

1. Draw all possible connected, topologically distinct diagrams, including loops, with n external legs. Ignore vacuum-to-vacuum graphs.
2. For each internal line, associate a propagator given by:

$$\underset{p}{- - -\!\!\succ - - -} \qquad i\Delta_F(p) = \frac{i}{p^2 - m^2 + i\epsilon} \tag{5.116}$$

3. For each vertex, associate the factor $-i\lambda$.

4. For each internal momentum corresponding to an internal loop, as sociate an integration factor:

$$\int \frac{d^4p}{(2\pi)^4} \tag{5.117}$$

5. Divide each graph by an overall symmetry factor S corresponding to the number of ways one can permute the internal lines and vertices, leaving the external lines fixed.

6. Momentum is conserved at each vertex.

The symmetry factor S is easily calculated. For the four-point function given above, the 1/4! coming from the interaction Lagrangian cancels the 4! ways in which the four external lines can be paired off with the four scalar fields appearing in ϕ^4, so $S = 1$. Now consider the connected two-point diagram at second order, which is a double-loop diagram (which has the topology of the symbol ϕ). There are 4 ways in which each external leg can be connected to each vertex. There are 3×2 ways in which the internal vertices can be paired off. So this gives us a factor of $1/S = (1/4!)(1/4!) \times (4 \times 4) \times (3 \times 2) = 1/3!$, so $S = 6$.

For QED, the Feynman's rules are only a bit more complicated. The interaction Hamiltonian becomes:

$$\mathscr{H}_I = -ie\bar{\psi}\gamma^\mu\psi A_\mu \tag{5.118}$$

As before, the power expansion of the interacting Lagrangian will pull down various factors of $\mathscr{H}_I$. Then we use Wick's theorem to pair off the various fermion and vector meson lines to form propagators and vertices.

There are only a few differences that we must note. First, when contracting over an internal fermion loop, we must flip one spinor past the others to perform the trace and Wick decomposition. This means that there must be an extra -1 factor inserted into all fermion loop integrations.

Second, various vector meson propagators in different gauges may be used, but all the terms proportional to p_μ or p_ν vanish because of gauge invariance (which will be discussed more in detail later).

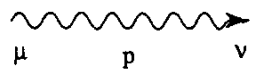
Thus, the Feynman's rules for QED become:

1. For each internal fermion line, associate a propagator given by:

$$iS_F(p) = \frac{i}{\not{p} - m + i\epsilon} = \frac{i(\not{p} + m)}{p^2 - m^2 + i\epsilon} \qquad \xrightarrow[\text{p}]{} \tag{5.119}$$

2. For each internal photon line, associate a propagator:

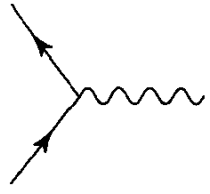
$$iD_F(p)_{\mu\nu} = -\frac{ig_{\mu\nu}}{p^2 + i\epsilon}$$



(5.120)

3. At each vertex, place a factor of:

$$-ie\gamma_\mu$$



(5.121)

4. Insert an additional factor of -1 for each closed fermion loop.

5. For each internal loop, integrate over:

$$\int \frac{d^4q}{(2\pi)^4} \tag{5.122}$$

6. A relative factor -1 appears between graphs that differ from each other by an interchange of two identical external fermion lines.

7. Internal fermion lines appear with arrows in both clockwise and counter-clockwise directions. However, diagrams that are topologically equivalent are counted only once.

8. External electron and positron lines entering a graph appear with factors $u(p, s)$ and $\bar{v}(p, s)$, respectively. External electron and positron lines leaving a graph appear with factors $\bar{u}(p, s)$ and $v(p, s)$, respectively. The direction of the positron lines is taken to be opposite of the electron lines, so that incoming positrons have momenta leaving the diagram.

Likewise, we can calculate Feynman's rules for any of the actions that we have investigated earlier.

For example, for charged scalar electrodynamics, with the additional term in the Lagrangian:

$$\mathscr{L} = D_\mu \phi^\dagger D^\mu \phi - m^2 \phi^\dagger \phi \tag{5.123}$$

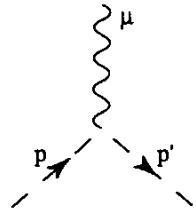
one has the following interaction Hamiltonian:

$$\mathscr{H}_I = -ie\phi^\dagger \left(\overleftarrow{\partial}_\mu - \overrightarrow{\partial}_\mu \right) \phi A^\mu - e^2 A_\mu^2 \phi^\dagger \phi \tag{5.124}$$

The Feynman rules are as follows:

1. For each scalar-scalar-vector vertex, insert the factor:

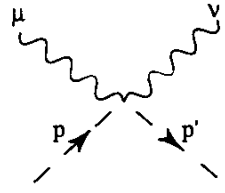
$$-ie(p+p')_\mu \qquad (5.125)$$



where p and p' are the momenta for the scalar line.

2. Insert a factor of:

$$2ie^2 g_{\mu\nu} \qquad (5.126)$$



for each "seagull" graph.

3. Insert an additional factor of $1/2$ for each closed loop with only two photon lines.

In summary, we have seen that, historically, there were two ways in which to quantize QED. The first method, pioneered by Feynman, was the propagator approach, which was simple, pictorial, but not very rigorous. The second was the more conventional operator approach of Schwinger and Tomonaga. In this chapter, we have presented the LSZ approach, which is perhaps the most convenient method for deriving the Feynman rules for any quantum field theory.

With Feynman rules, one can almost, by inspection, write down the perturbation expansion for any quantum field theory. In the next chapter, we will use these rules to calculate higher-order interactions in QED.

5.8 Exercises

1. Set up the reduction formulas for a massless and massive vector meson. Derive the counterpart of Eqs. (5.73) and (5.77).

2. Write down the Feynman rules for a massive pseudoscalar field interacting with a Dirac electron via the interaction $\bar{\psi}\gamma_5\psi\phi$.

3. Write down the Feynman rules for a massive pseudovector field interacting with the Dirac field via the term $\bar{\psi}\gamma_5\gamma^\mu\psi A_\mu$.

4. In the "old-fashioned" noncovariant canonical approach to QED in the Coulomb gauge, one derived the scattering matrix by solving the Schrödinger

equation with a second-quantized Hamiltonian, in which the interactions were ordered in time. Draw the complete set of noncovariant diagrams that are necessary to describe electron–electron scattering in QED to the one-loop level. Sketch how several noncovariant diagrams can be summed to produce a single covariant Feynman diagram.

5. Prove:

$$\frac{(-1)^{N-1}}{(N-M)!}\epsilon^{\mu_1\mu_2\cdots\mu_M\sigma_1\sigma_2\cdots\sigma_{N-M}}\epsilon_{\nu_1\nu_2\cdots\nu_M\sigma_1\sigma_2\cdots\sigma_{N-M}} = \delta^{\mu_1\mu_2\cdots\mu_M}_{\nu_1\nu_2\cdots\nu_M} \tag{5.127}$$

where:

$$\delta^{\mu_1\cdots\mu_M}_{\mu_1\cdots\mu_M} = \delta^{\mu_1}_{[\mu_1}\delta^{\mu_2}_{\nu_2}\cdots\delta^{\mu_M}_{\nu_M]} \tag{5.128}$$

6. Show the equivalence of the two expressions for the flux J in Eqs. (5.24) and (5.25) and show the equality only holds in collinear frames. [Hint: square Eq. (5.25), and then expand the product of two $\epsilon^{\mu\nu\alpha\beta}$ tensors in terms of delta functions.]

7. In Compton scattering, a photon scatters off an electron. Show that the relative velocity $|v_1 - v_2|$ appearing in the cross-section formula can exceed the speed of light. Is this a violation of relativity? Why or why not?

8. Prove Furry's theorem[9], which states that a Feynman loop diagram containing an odd number of external photon lines vanishes. (Hint: Show that a fermion loop with an odd number of legs cancels against another fermion loop with the arrows reversed, or use the fact that QED is invariant under charge conjugation. The fields transform as: $\psi \to \psi^c = C\bar{\psi}^T$ and $A_\mu \to -A_\mu$. Show that these diagrams are odd under C and hence not allowed.)

9. The factor λ in Eq. (5.94) contains Feynman graphs with no external legs. For QED, draw all such diagrams up to the second-loop level. Do not solve.

10. If we power expand Eq. (5.89) in a Taylor series, we find:

$$\begin{aligned} U(t,t') &= 1 + \sum_{n=1}^{\infty}(-i)^n \int_{t'}^{t} dt_1 \int_{t'}^{t_1} dt_2 \cdots \int_{t'}^{t_{n-1}} dt_n \\ &\quad \times T\left(\mathscr{H}_I(t_1)\cdots\mathscr{H}_I(t_n)\right) \end{aligned} \tag{5.129}$$

Prove this, and then show that it equals Eq. (5.90). (Hint: take the lowest order, and on a graph, draw the integration regions for t_1 and t_2. Show that the identity holds for this case, and then generalize to the arbitrary case.)

11. Prove that the time-ordered product in Eq. (5.73) always emerges when we make the LSZ reduction of more than one field from the asymptotic states.

12. Let us define the following Green's function:

$$\Delta'(x, x') \equiv -i\langle 0|[\phi(x), \phi(x')]|0\rangle \tag{5.130}$$

Insert a complete set of intermediate states within the commutator:

$$1 = \sum_n |n\rangle\langle n| = \int d^4q \delta^4(p_n - q) \tag{5.131}$$

Prove that the commutator becomes:

$$\Delta'(x - x') = \int_0^\infty d\sigma^2 \, \rho(\sigma^2)\Delta(x - x', \sigma) \tag{5.132}$$

where:

$$\begin{aligned} \rho(q) &= \rho(q^2)\theta(q_0) \\ &= (2\pi)^3 \sum_n \delta^4(p_n - q)\left|\langle 0|\phi(0)|n\rangle\right|^2 \end{aligned} \tag{5.133}$$

This is the Källén–Lehmann spectral representation[8,9]. [Since σ appears in this formula as a mass, this formula states that the complete, interacting value of Δ' is equal to the integral over all possible free Δ' with arbitrary mass σ, weighted by the unknown function $\rho(\sigma^2)$.]

13. Let us separate out the one-particle contribution to the smeared average. Then we find:

$$\Delta'(x - x') = Z\Delta(x - x'; m) + \int_{m^2}^\infty d\sigma^2 \, \rho(\sigma^2)\Delta(x - x'; \sigma) \tag{5.134}$$

Also, m^2 is the lowest mass squared that contributes to the continuum above the one-particle contribution. For example, for pions, $m^2 = 4m_\pi^2$. Take the time derivative of both sides, reducing the Green's functions to delta functions. Then show:

$$1 = Z + \int_{m^2}^\infty \rho(\sigma^2) \, d\sigma^2 \tag{5.135}$$

which implies:

$$0 \leq Z < 1 \tag{5.136}$$

When $Z = 1$, this means that the ρ function is zero, so the theory has collapsed into a free theory. For an interacting theory, we must have $Z < 1$.

14. Renormalization constants are usually thought to be infinite quantities, yet we have just shown that Z is less than one. Is there a contradiction?

15. For ϕ^4 theory, prove that the symmetry factor S equals:

$$S = c \prod_{n=2,3,\ldots} 2^b (n!)^{a_n} \tag{5.137}$$

where a_n equals the number of pairs of vertices that are connected by n identical lines, b is the number of lines that connect a vertex with itself, and c is the number of permutations of vertices that leave the diagram invariant when the external lines are fixed. What is S for QED?

Chapter 6
Scattering Processes and the S Matrix

I was sort of half-dreaming, like a kid would . . . that it would be funny if these funny pictures turned out to be useful, because the damned Physical Review would be full of these odd-looking things. And that turned out to be true.

—R. Feynman

6.1 Compton Effect

Now that we have derived the Feynman rules for various quantum field theories, the next step is to calculate cross sections for elementary processes involving photons, electrons, and antielectrons. At the lowest order, these cross sections reproduce classical results found with earlier methods. However, the full power of the quantum field theory will be seen at higher orders, where we calculate radiative corrections to the hydrogen atom that have been verified to great accuracy. In the process, we will solve the problem of the electron self-energy, which completely eluded earlier, classical attempts by Lorentz and others.

At the end of this chapter, we will also investigate the S matrix itself. Rather than appeal to perturbation theory and summing Feynman diagrams, we will impose mathematical constraints directly on the S matrix, like unitarity and analyticity, to obtain nontrivial constraints on π-nucleon scattering. These results hold without ever appealing to any perturbative power expansion.

The material covered in this chapter is fairly standard, and the reader is urged to consult other excellent texts for other details, such as Bjorken and Drell, Itzykson and Zuber, and Mandl and Shaw.

To begin our discussion, we will divide Feynman diagrams into two types, "trees" and "loops," on the basis of their topology. Loop diagrams, as their name suggests, have closed loops in them. Tree diagrams have no loops; that is, they only have branches. In a scattering process, we will see that the sum over tree diagrams is finite and reproduces the classical result. The loop diagrams, by contrast, are usually divergent and are purely quantum-mechanical effects.

We start by analyzing the lowest-order terms in the scattering matrix for four particles or fields. To this order, we find only tree diagrams and no loops. Thus, we should be able to reproduce generalizations of classical and nonrelativistic physics. (Interestingly enough, we will find that negative energy states cannot be omitted in these calculations, even in the nonrelativistic limit. Although these negative energy states are purely a byproduct of relativity, if we drop them, then we will fail to reproduce the classical and nonrelativistic results.)

In this chart, we will summarize the scattering processes that we will analyze in the first part of this chapter:

Compton scattering:	$e^- + \gamma \rightarrow e^- + \gamma$
Pair annihilation:	$e^- + e^- \rightarrow \gamma + \gamma$
Møller scattering:	$e^- + e^- \rightarrow e^- + e^-$
Bhabha scattering:	$e^- + e^+ \rightarrow e^- + e^+$
Bremsstrahlung:	$e^- + N \rightarrow e^- + N + \gamma$
Pair creation:	$\gamma + \gamma \rightarrow e^- + e^+$

There is a reason for writing these scattering processes in this particular order. If we take the Feynman diagrams for Compton scattering and rotate them by 90 degrees, we find that they turn into the Feynman diagrams for pair annihilation. This is called the *substitution rule*, where we take the process:

$$1 + 2 \rightarrow 3 + 4 \tag{6.1}$$

and convert it into:

$$1 + \bar{3} \rightarrow \bar{2} + 4 \tag{6.2}$$

Using the substitution rule, we can group these scattering processes into pairs:

Compton effect	$\leftrightarrow$	Pair annihilation
Møller scattering	$\leftrightarrow$	Bhabha scattering
Bremsstrahlung	$\leftrightarrow$	Pair creation

There are several advantages to using this symmetry. At the superficial level, this means that we can, almost by inspection, convert the scattering amplitude of one process into that of the other, thereby saving a considerable amount of time. At a deeper level, it signals the fact that the S matrix obeys a new kind of symmetry, called *crossing symmetry*. If we treat the S matrix as an analytic

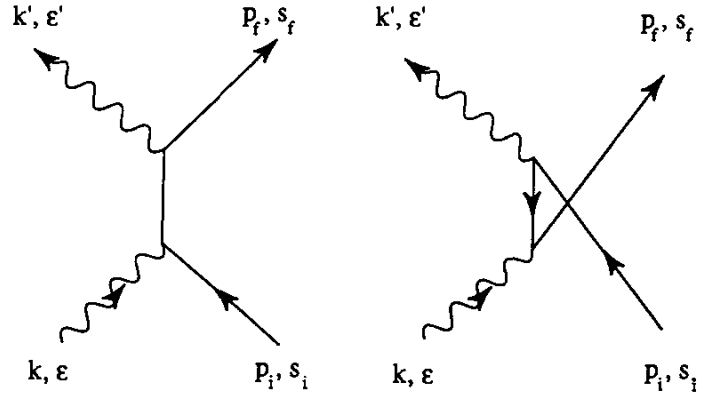


Figure 6.1. Compton scattering: a photon of momentum k scatters elastically off an electron of momentum p_i.

function of the energy variables, then crossing symmetry relates different analytic regions of the S matrix to each other in a non-trivial way.

To begin, the first process we will examine is Compton scattering, which occurs when an electron and a photon collide and scatter elastically. Historically, this process was crucial in confirming that electromagnetic radiation had particlelike properties, that is, that the photon was acting like a particle in colliding with the electron. (We will then, using the substitution rule, derive the Feynman amplitude for pair annihilation.)

We will assume that the electron has momentum p_i before the collision and p_f afterwards. The photon has momentum k before and k' afterwards. The reaction can be represented symbolically as:

$$\gamma(k) + e(p_i) \rightarrow \gamma(k') + e(p_f) \tag{6.3}$$

By energy–momentum conservation, we also have:

$$k + p_i = k' + p_f \tag{6.4}$$

Compton scattering, to lowest order, is shown in Figure 6.1.

We normalize the wave function of the photon by:

$$A_\mu(x) = \frac{1}{\sqrt{2kV}}\epsilon_\mu(e^{-ik\cdot x} + e^{ik\cdot x}) \tag{6.5}$$

To lowest order, the S matrix is:

$$S_{fi} = \frac{e^2}{V^2}(2\pi)^4\delta^4(p_f + k' - p_i - k)$$

$$\times \sqrt{\frac{m}{E_f}}\bar{u}(p_f, s_f)\left(\frac{(-i\not{\epsilon}')}{\sqrt{2k'}}\frac{i}{\not{p}_i + \not{k} - m}\frac{(-i\not{\epsilon})}{\sqrt{2k}}\right.$$

$$\left. + \frac{(-i\not{\epsilon})}{\sqrt{2k}}\frac{i}{\not{p}_i - \not{k}' - m}\frac{(-i\not{\epsilon}')}{\sqrt{2k'}}\right)\sqrt{\frac{m}{E_i}}u(p_i, s_i) \tag{6.6}$$

The differential cross section σ is found in several steps. First, we square the S matrix, which gives us a divergent result. We divide by the singular quantity $(2\pi)^4\delta(0)$ and obtain the rate of transitions. We divide by the flux $|\mathbf{v}|/V$, divide by the number of particles per unit volume $1/V$, multiply by the phase factor for outgoing particles $[V^2/(2\pi)^6]\, d^3p_f\, d^3k'$. This give us the differential cross section:

$$\begin{aligned} d\sigma &= \frac{|S_{fi}|^2}{(2\pi)^4\delta(0)}\frac{1}{|\mathbf{v}|}\frac{V^2 d^3p_f d^3k'}{(2\pi)^6} \\ &= \frac{e^4}{(2\pi)^2|\mathbf{v}|}\frac{m}{E_i}\frac{1}{2k}\int\sum_{\text{spins}}|\bar{u}_f\Gamma u_i|^2 \\ &\quad \times\delta^4(p_i + k - p_f - k')\frac{m\, d^3p_f}{E_f}\frac{d^3k'}{2k'} \end{aligned} \tag{6.7}$$

where:

$$\Gamma = \frac{\not{\epsilon}'\not{\epsilon}\not{k}}{2p_i\cdot k} + \frac{\not{\epsilon}\not{\epsilon}'\not{k}'}{2p_i\cdot k'} \tag{6.8}$$

To reduce out the spins, we will once again use the convenient formula given in Eq. (5.54):

$$\sum_{\text{spins}}|\bar{u}(p_f, s_f)\Gamma u(p_i, s_i)|^2 = \text{Tr}\left(\Gamma\frac{\not{p}_i + m}{2m}\gamma^0\Gamma^\dagger\gamma^0\frac{\not{p}_f + m}{2m}\right) \tag{6.9}$$

Although this calculation looks formidable, we can perform the trace of up to eight Dirac matrices by reducing it to the trace of six, and then four Dirac matrices, etc. We will use the formula:

$$\begin{aligned} \text{Tr}(\not{k}_1\not{k}_2\not{k}_3\cdots\not{k}_{2n}) &= k_1\cdot k_2\text{Tr}(\not{k}_3\cdots\not{k}_{2n}) - k_1\cdot k_3\text{Tr}(\not{k}_2\not{k}_4\cdots\not{k}_{2n}) \\ &\quad + \cdots + k_1\cdot k_{2n}\text{Tr}(\not{k}_2\not{k}_3\cdots\not{k}_{2n-1}) \end{aligned} \tag{6.10}$$

The problem simplifies enormously because we can eliminate entire groups of terms every time certain dot products appear, since:

$$k^2 = k'^2 = \epsilon\cdot k = \epsilon'\cdot k' = 0 \tag{6.11}$$

We can also simplify the calculation by using:

$$\epsilon^2 = \epsilon'^2 = -1; \quad p_i^2 = p_f^2 = m^2 \tag{6.12}$$

In short, each trace consists of collecting the complete set of all possible pairs of dot products of vectors, most of which vanish. Dividing the factors into smaller pieces, we now find:

$$\sum_{\text{spins}} |\bar{u}_f \Gamma u_i|^2 = \sum_{i=1}^{4} T_i \tag{6.13}$$

where:

$$\begin{aligned} T_1 &= \mathrm{Tr}\left[\not\epsilon' \not\epsilon \not k(\not p_i + m) \not k \not\epsilon \not\epsilon'(\not p_f + m)\right] \\ &= 2p_i \cdot k \,\mathrm{Tr}\left(\not\epsilon' \not\epsilon \not k \not\epsilon \not\epsilon' \not p_f\right) \\ &= 2p_i \cdot k \,\mathrm{Tr}\left(\not\epsilon' \not k \not\epsilon' \not p_f\right) \\ &= 8p_i \cdot k[2(\epsilon' \cdot k)^2 + k' \cdot p_i] \end{aligned} \tag{6.14}$$

and:

$$\begin{aligned} T_2 &= \mathrm{Tr}\left(\not\epsilon' \not\epsilon \not k(\not p_i + m) \not k' \not\epsilon' \not\epsilon(\not p_f + m)\right) \\ &= 2k \cdot p_i \,\mathrm{Tr}\left(\not k' \not\epsilon' \not\epsilon \not\epsilon' \not\epsilon \not p_i\right) + 8(k' \cdot \epsilon)^2 k \cdot p_i - 8(k \cdot \epsilon')^2 p_i \cdot k' \\ &= 8k \cdot p_i \, k' \cdot p_f[2(\epsilon' \cdot \epsilon)^2 - 1] + 8(k' \cdot \epsilon)^2 k \cdot p_i - 8(k \cdot \epsilon')^2 k' \cdot p_i \end{aligned} \tag{6.15}$$

We also have T_3 equal to T_2, and T_4 can be obtained from T_3 if we make the substitution: $(\epsilon, k) \rightarrow (\epsilon', -k')$. Since the calculation is Lorentz invariant, we can always take a specific Lorentz frame. We lose no generality by letting the electron be at rest, and let the incoming photon lie along the z axis. Let the outgoing photon scatter within the $y - z$ plane, making an angle θ with the z axis (Fig. 6.2).

Then the specific parametrization is given by:

$$\begin{aligned} p_{i,\mu} &= (m, 0, 0, 0) \\ k_\mu &= k(1, 0, 0, 1) \\ k'_\mu &= k'(1, 0, \sin\theta, \cos\theta) \\ p_{f,\mu} &= (E, 0, -k' \sin\theta, k - k' \cos\theta) \end{aligned} \tag{6.16}$$

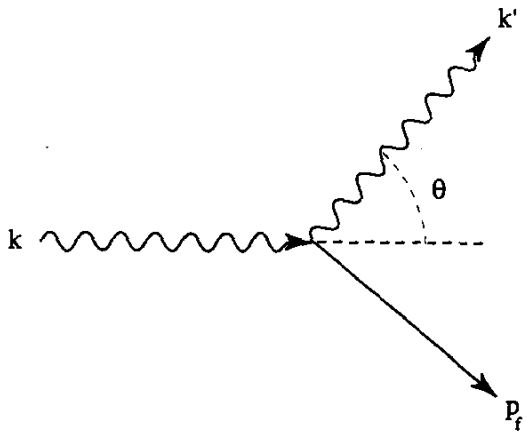


Figure 6.2. Compton scattering in the laboratory frame, where the electron is at rest.

It is important to notice that the only independent variables in this scattering are k and θ. All other variables can be expressed in terms of these two variables. For example, we can solve for k' and E in terms of the independent variables k and θ:

$$\begin{aligned} k' &= \frac{k}{1+(k/m)(1-\cos\theta)} \\ E &= m+k-k' \end{aligned} \tag{6.17}$$

Adding all four contributions, we now have:

$$\sum_{\text{spins}} |\bar{u}(p_f, s_f)\Gamma u(p_i, s_i)|^2 = \frac{1}{2m^2}\left(\frac{k'}{k} + \frac{k}{k'} + 4(\epsilon' \cdot \epsilon)^2 - 2\right) \tag{6.18}$$

We now must integrate over the momenta of the outgoing photon k' and electron p_f. Since the only independent variables in the problem are given by k and θ, all integrations are easy, except for the Jacobian, which arises when we change variables and integrate over Dirac delta functions. Thus, the integration over $d^3 p_f$ is trivial because of momentum conservation; it simply sets the momenta to be the values given above. That leaves one complication, the integration over the time components dp_{f0}. However, this integration can be rewritten in a simple fashion:

$$\frac{1}{2E_f} = \int dp_{f0}\, \delta(p_f^2 - m^2)\theta(p_{f,0}) \tag{6.19}$$

The integration over $p_{f,0}$ in the integral just sets its value to be the on-shell value. Finally, this last delta function can be removed because of the integration over k'. The only tricky part is to extract from this last integration the measure when we integrate over k'.

This last delta function can be written as:

$$\begin{aligned}\delta\left([k+p_i-k']^2-m^2\right) &= \delta\left(2m(k-k')-2kk'(1-\cos\theta)\right)\\ &= \frac{\delta\left(k'-k'(k)\right)}{(2mk/k')}\end{aligned} \tag{6.20}$$

where $k'(k)$ is the value given in Eq. (6.17). Putting all integration factors together, we now have:

$$\delta^4(p_i+k-k'-p_f)\frac{d^3p_f}{2E_f}\,d^3k' = \frac{k'^2\,d\Omega}{(2mk/k')} \tag{6.21}$$

Inserting all expressions into the cross section, we obtain the Klein–Nishina formula[1]:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4m^2}\left(\frac{k'}{k}\right)^2\left(\frac{k'}{k}+\frac{k}{k'}+4(\epsilon\cdot\epsilon')^2-2\right) \tag{6.22}$$

If we take the low-energy limit $k \to 0$, then the Klein–Nishina formula reduces to the Thompson scattering formula:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{m^2}(\epsilon\cdot\epsilon')^2 \tag{6.23}$$

If the initial and final photon are unpolarized, we can average over the initial and final polarizations ϵ and ϵ'. In the particular parametrization that we have chosen for our momenta, we can choose our polarizations ϵ and ϵ', such that they are purely transverse and perpendicular to the momenta p_i and p_f, respectively:

$$\begin{aligned}\epsilon^{(1)} &= (0,1,0,0)\\ \epsilon^{(2)} &= (0,0,1,0)\\ \epsilon'^{(1)} &= (0,1,0,0)\\ \epsilon'^{(2)} &= (0,0,\cos\theta,-\sin\theta)\end{aligned} \tag{6.24}$$

It is easy to check that these polarization vectors satisfy all the required properties. Then the sum over these polarization vectors is easy to perform:

$$\sum_{\text{spins}}(\epsilon\cdot\epsilon')^2 = \sum_{i,j}(\epsilon^{(i)}\cdot\epsilon'^{(j)})^2 = 1+\cos^2\theta \tag{6.25}$$

The averaged cross section is given by:

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{av}} = \frac{\alpha^2}{2m^2}\left(\frac{k'}{k}\right)^2\left(\frac{k'}{k}+\frac{k}{k'}-\sin^2\theta\right) \tag{6.26}$$

The integral over θ is straightforward. Let us define $z = \cos\theta$. The integration yields:

$$\begin{aligned}\sigma &= \frac{\pi\alpha^2}{m^2}\int_{-1}^{1} dz\left(\frac{1}{[1+a(1-z)]^3}\right.\\ &\quad\left.+\frac{1}{1+a(1-z)}-\frac{1-z^2}{[1+a(1-z)]^2}\right)\\ &= \left(\frac{8\pi\alpha^2}{3m^2}\right)(3/4)\left[\frac{1+a}{a^3}\left(\frac{2a(1+a)}{1+2a}-\log(1+2a)\right)\right.\\ &\quad\left.+\frac{\log(1+2a)}{2a}-\frac{1+3a}{(1+2a)^2}\right]\end{aligned} \tag{6.27}$$

where $a = k/m$. For small energies, this reduces to the usual Thompson total cross section:

$$\lim_{k\to 0}\sigma = \sigma_{\text{Thompson}} = \frac{8\pi\alpha^2}{3m^2} = 0.665\times 10^{-24}\ \text{cm}^2 \tag{6.28}$$

For high energies, the logarithm starts to dominate the cross section:

$$\lim_{k\to\infty}\sigma = \frac{\pi\alpha^2}{km}\left[\log\frac{2k}{m}+\frac{1}{2}+O\left(\frac{m}{k}\ln\frac{k}{m}\right)\right] \tag{6.29}$$

6.2 Pair Annihilation

The Feynman graphs for pair annihilation of an electron and position into two gamma rays is shown in Figure 6.3. Pair annihilation is represented by the process:

$$e^-(p_1)+e^+(p_2)\to\gamma(k_1)+\gamma(k_2) \tag{6.30}$$

However, notice that we can obtain this diagram if we simply rotate the diagram for Compton scattering in Figure 6.1. Thus, by a subtle redefinition of the various momenta, we should be able to convert the Compton scattering amplitude, which we have just calculated, into the amplitude for pair annihilation. This is the

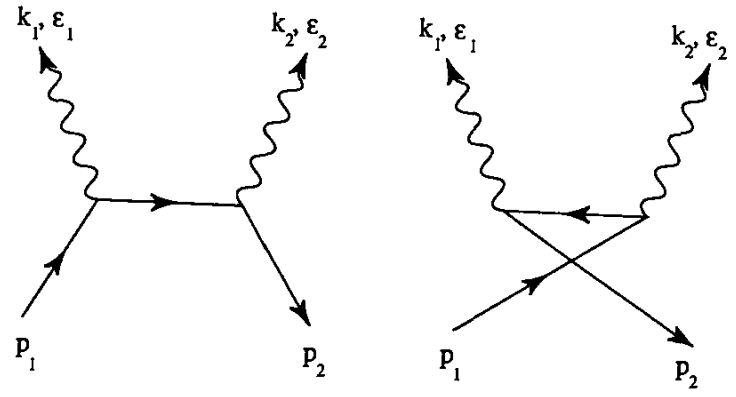


Figure 6.3. Pair annihilation: an electron of momentum p_1 annihilates with a positron of momentum p_2 into two photons.

substitution rule to which we referred earlier. For example, the S matrix now yields, to lowest order:

$$S_{fi} = -e^2(2\pi)^4\delta^4(k_1+k_2-p_1-p_2)\frac{1}{\sqrt{2k_1 2k_2}} \times \sqrt{\frac{m}{E_2}}\bar{v}(p_2,s_2)\left(\not{\epsilon}_2\frac{1}{\not{p}_1-\not{k}_1-m}\not{\epsilon}_1+\not{\epsilon}_1\frac{1}{\not{p}_1-\not{k}_2-m}\not{\epsilon}_2\right)\sqrt{\frac{m}{E_1}}u(p_1,s \tag{6.3}$$

where we have made the substitutions:

$$\begin{aligned}(k,\epsilon) &\rightarrow (-k_1,\epsilon_1)\\ (k',\epsilon') &\rightarrow (k_2,\epsilon_2)\\ u(p_i,s_i) &\rightarrow u(p_1,s_1)\\ \bar{u}(p_f,s_f) &\rightarrow \bar{v}(p_2,s_2)\end{aligned} \tag{6.32}$$

We will, as usual, take the Lorentz frame where the electron is at rest. Then our momenta become, as in Figure 6.4:

$$\begin{aligned}p_{1\mu} &= (m,0,0,0)\\ p_{2\mu} &= (E,0,0,|\mathbf{p}|)\\ k_{1\mu} &= k_1(1,0,\sin\theta,\cos\theta)\\ k_{2\mu} &= (k_2,0,-k_1\sin\theta,|\mathbf{p}|-k_1\cos\theta)\end{aligned} \tag{6.33}$$

There are only two independent variables in this process, $|\mathbf{p}|$ and θ. All other variables can be expressed in terms of them. For example, we can easily show

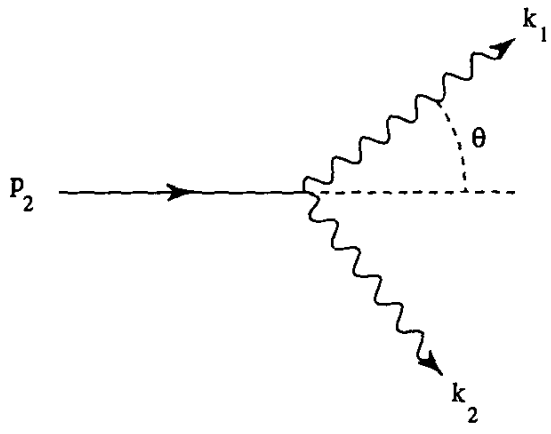


Figure 6.4. Pair annihilation in the laboratory frame, where the electron is at rest.

that:

$$\begin{aligned} k_1 &= \frac{m(m+E)}{m+E-|\mathbf{p}|\cos\theta} \\ k_2 &= \frac{(m+E)(E-|\mathbf{p}|\cos\theta)}{m+E-|\mathbf{p}|\cos\theta} \end{aligned} \tag{6.34}$$

We also have:

$$\begin{aligned} m+E &= k_1+k_2 \\ k_1\cdot k_2 &= m(m+E) = k_1(m+E-|\mathbf{p}|\cos\theta) \end{aligned} \tag{6.35}$$

When we contract over the Dirac matrices, the calculation proceeds just as before, except that we want to evaluate $|\bar{v}\Gamma u|^2$. We have to use Eq. (3.111):

$$\sum_{\text{spins}} v(p,s)\bar{v}(p,s) = \frac{\not{p}-m}{2m} \tag{6.36}$$

The trace becomes:

$$\text{Trace} = -\frac{1}{2m^2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}-4(\epsilon_1\cdot\epsilon_2)^2+2\right) \tag{6.37}$$

The integration over d^3k_1 and d^3k_2 also proceeds as before. The integrations over the delta functions are straightforward, except that we must be careful when picking up a measure term when we make a transformation on a Dirac delta function. When this additional measure term is inserted, the differential cross section becomes:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2(m+E)}{8|\mathbf{p}|(m+E-|\mathbf{p}|\cos\theta)^2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}-4(\epsilon_1\cdot\epsilon_2)^2+2\right) \tag{6.38}$$

The total cross section is obtained by summing over photon polarizations. As before, we can take a specific set of polarizations which are transverse to p_1 and p_2:

$$\begin{aligned}
\epsilon_1^{(1)} &= 0, 1, 0, 0) \\
\epsilon_1^{(2)} &= (0, 0, \cos\theta, -\sin\theta) \\
\epsilon_2^{(1)} &= (0, 1, 0, 0) \\
\epsilon_2^{(2)} &= (0, 0, |\mathbf{p}| - k_1\cos\theta, k_1\sin\theta)
\end{aligned} \tag{6.39}$$

Then we can sum over all polarizations. The only difficult sum involves:

$$\begin{aligned}
\epsilon_1^{(2)} \cdot \epsilon_2^{(2)} &= \mathbf{k}_1 \cdot \mathbf{k}_2 \\
&= 1 - \frac{(k_1 + k_2)^2}{2k_1k_2} \\
&= 1 - \frac{2m(m+E)}{2k_1k_2} \\
&= 1 - m\left(\frac{1}{k_1} + \frac{1}{k_2}\right)
\end{aligned} \tag{6.40}$$

Then the sum over spins can be written as:

$$\sum_{i,j}(\epsilon_1^{(i)} \cdot \epsilon_2^{(j)})^2 = 1 + \left[1 - \left(\frac{m}{k_1} + \frac{m}{k_2}\right)\right]^2 \tag{6.41}$$

The only integration left is the one over the solid angle, which leaves us with ($\gamma = E_2/m$):

$$\sigma = \frac{\pi\alpha^2}{m^2(1+\gamma)}\left(\frac{\gamma^2+4\gamma+1}{\gamma^2-1}\log\left(\gamma + \sqrt{\gamma^2-1}\right) - \frac{\gamma+3}{\sqrt{\gamma^2-1}}\right) \tag{6.42}$$

which is a result first obtained by Dirac.[2]

6.3 Møller Scattering

Next, we investigate electron–electron scattering. To lowest order, this scattering amplitude contains two graphs, as in Figure 6.5. This scattering is represented by:

$$e(p_1) + e(p_2) \rightarrow e(p_1') + e(p_2') \tag{6.43}$$

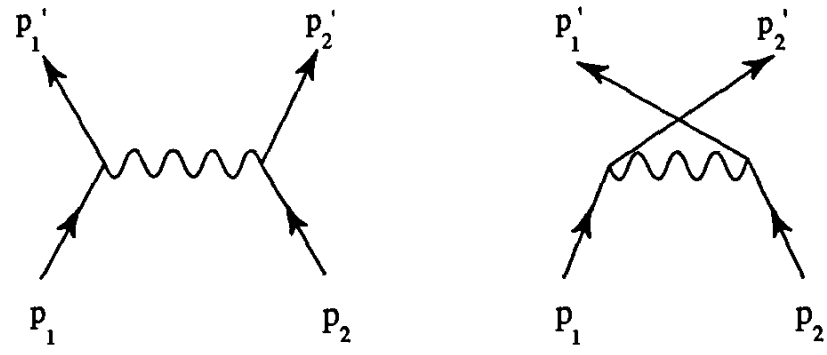


Figure 6.5. Møller scattering of two electrons with momenta p_1 and p_2.

By a straightforward application of the formulas for differential cross sections, we find:

$$d\sigma = \frac{e^4 m^4}{[(p_1 \cdot p_2)^2 - m^4]^{1/2}} (2\pi)^4 \delta^4(p_1' + p_2' - p_1 - p_2) \times \int \frac{d^3 p_1'}{(2\pi)^3 E_1'} \frac{d^3 p_2'}{(2\pi)^3 E_2'} |\mathscr{M}_{fi}|^2 \tag{6.44}$$

where, using a straightforward application of Feynman's rules for these two diagrams, we can compute $|\mathscr{M}_{fi}|^2$:

$$\begin{aligned} |\mathscr{M}_{fi}|^2 = \frac{1}{4} \Bigg[& \mathrm{Tr}\left(\gamma_\nu \frac{\not{p}_1 + m}{2m} \gamma_\sigma \frac{\not{p}_1' + m}{2m}\right) \\ & \times \mathrm{Tr}\left(\gamma^\nu \frac{\not{p}_2 + m}{2m} \gamma^\sigma \frac{\not{p}_2' + m}{2m}\right) \frac{1}{[(p_1' - p_1)^2]^2} \\ & - \mathrm{Tr}\left(\gamma_\nu \frac{\not{p}_1 + m}{2m} \gamma_\sigma \frac{\not{p}_2' + m}{2m} \gamma^\nu \frac{\not{p}_2 + m}{2m} \gamma^\sigma \frac{\not{p}_1' + m}{2m}\right) \\ & \times \frac{1}{(p_1' - p_1)^2 (p_2' - p_2)^2} + (p_1' \leftrightarrow p_2') \Bigg] \end{aligned} \tag{6.45}$$

Since the trace is over only four Dirac matrices, taking the trace is not hard to do:

$$\begin{aligned} \frac{1}{2m^4} \Bigg(& \frac{(p_1 \cdot p_2)^2 + (p_1 \cdot p_2')^2 + 2m^2 (p_1 \cdot p_2' - p_1 \cdot p_2)}{[(p_1' - p_1)^2]^2} \\ & + \frac{(p_1 \cdot p_2)^2 + (p_1 \cdot p_1')^2 + 2m^2 (p_1 \cdot p_1' - p_1 \cdot p_2)}{[(p_2' - p_1)^2]^2} \\ & + 2 \frac{(p_1 \cdot p_2)^2 - 2m^2 p_1 \cdot p_2}{(p_1' - p_1)^2 (p_2' - p_1)^2} \Bigg) \end{aligned} \tag{6.46}$$

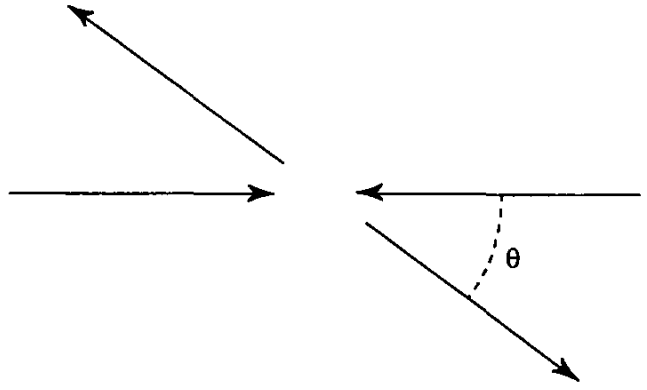


Figure 6.6. Møller scattering in the center-of-mass frame.

The kinematics is illustrated in Figure 6.6. Without any loss of generality, we can choose the center-of-mass frame, where the electron momenta p_1 and p_2 lie along the z axis:

$$\begin{aligned} p_{1,\mu} &= (E, 0, 0, |\mathbf{p}|) \\ p_{2,\mu} &= (E, 0, 0, -|\mathbf{p}|) \\ p'_{1,\mu} &= (E, 0, |\mathbf{p}| \sin\theta, |\mathbf{p}| \cos\theta) \\ p'_{2,\mu} &= (E, 0, -|\mathbf{p}| \sin\theta, -|\mathbf{p}| \cos\theta) \end{aligned} \tag{6.47}$$

The only independent variables are θ and $|\mathbf{p}|$. In terms of this parametrization, we easily find:

$$\begin{aligned} p_1 \cdot p_2 &= 2E^2 - m^2 \\ p_1 \cdot p'_1 &= E^2(1 - \cos\theta) + m^2 \cos\theta \\ p_1 \cdot p'_2 &= E^2(1 + \cos\theta) - m^2 \cos\theta \end{aligned} \tag{6.48}$$

Then the entire cross section can be written in terms of these independent variables. We finally obtain the Møller formula[3] in the center-of-mass frame:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2(2E^2 - m^2)^2}{4E^2(E^2 - m^2)^2} \left[\frac{4}{\sin^4\theta} - \frac{3}{\sin^2\theta} + \frac{(E^2 - m^2)^2}{(2E^2 - m^2)^2} \left(1 + \frac{4}{\sin^2\theta} \right) \right] \tag{6.49}$$

In the relativistic limit, as $E \to \infty$, this formula reduces to:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{E^2} \left(\frac{4}{\sin^4\theta} - \frac{2}{\sin^2\theta} + \frac{1}{4} \right) \tag{6.50}$$

For the low-energy, nonrelativistic result, we find:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{m^2} \frac{1}{4v^2} \left(\frac{4}{\sin^4\theta} - \frac{3}{\sin^2\theta} \right) \tag{6.51}$$

6.4 Bhabha Scattering

To calculate the cross section for electron–positron scattering (Fig. 6.7), we can use the substitution rule. By rotating the diagram for Møller scattering, we find the Feynman diagrams for Bhabha scattering.[4] The only substitutions we must make are:

$$\begin{aligned} u(p_1) &\rightarrow u(p_1) \\ u(p_1') &\rightarrow u(p_1') \\ u(p_2) &\rightarrow v(-p_2) \\ u(p_2') &\rightarrow v(-p_2') \end{aligned} \tag{6.52}$$

for the process:

$$e^-(p_1) + e^+(-p_2) \rightarrow e^-(p_1') + e^+(-p_2') \tag{6.53}$$

The calculation and the traces are performed exactly as before with these simple substitutions. We merely quote the final result, due to Bhabha (1935):

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{\alpha}{2E^2}\left(\frac{5}{4} - \frac{8E^4 - m^4}{E^2(E^2-m^2)(1-\cos\theta)} + \frac{(2E^2-m^2)^2}{2(E^2-m^2)^2(1-\cos\theta)^2}\right. \\ &\quad + [16E^4]^{-1}\Big[2E^4(-1+2\cos\theta+\cos^2\theta) \\ &\quad + 4E^2m^2(1-\cos\theta)(2+\cos\theta) + 2m^4\cos^2\theta\Big]\Big) \end{aligned} \tag{6.54}$$

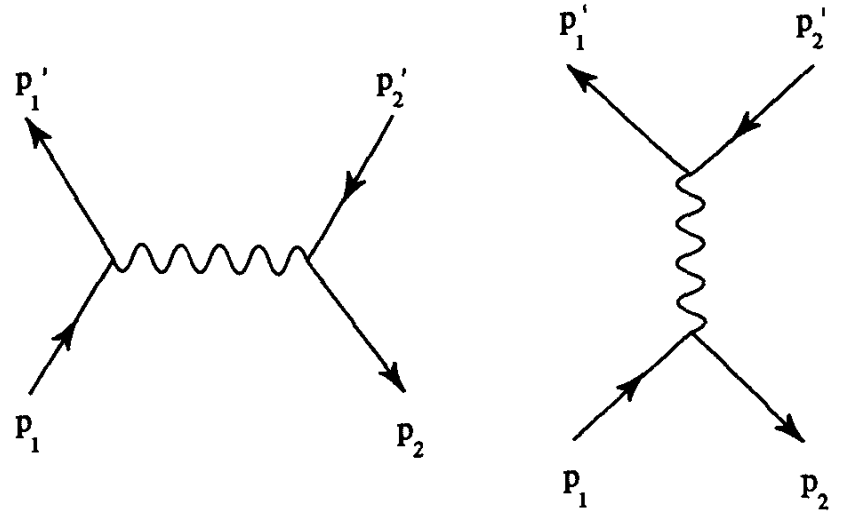


Figure 6.7. Bhabha scattering of an electron off a positron.

In the relativistic limit, we have:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{8E^2}\left(\frac{1+\cos^4\theta/2}{\sin^4\theta/2} + \frac{1}{2}(1+\cos^2\theta) - 2\frac{\cos^4\theta/2}{\sin^2\theta/2}\right) \tag{6.55}$$

In the nonrelativistic limit, we find:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{m^2}\frac{1}{16v^4\sin^4\theta/2} \tag{6.56}$$

6.5 Bremsstrahlung

Bremsstrahlung is the process by which radiation is emitted from an electron as it moves past a nucleus (Fig. 6.8). Momentum conservation gives us:

$$p_i + q = k + p_f \tag{6.57}$$

Classically, one can calculate the radiation emitted by a moving charge as it accelerates past a proton. (*Bremsstrahlung* means "braking radiation.") However, unlike the previous scattering processes, which agree to first order with the experimental data, we find a severe problem with this amplitude, which is the *infrared divergence*. The quantum field theory calculation, to lowest order, reproduces the classical result, including the unwanted infrared divergence, which has its roots in the classical theory.

Although the infrared divergence first arose (in another form) in the classical theory, the final resolution of this problem comes when we take into account higher

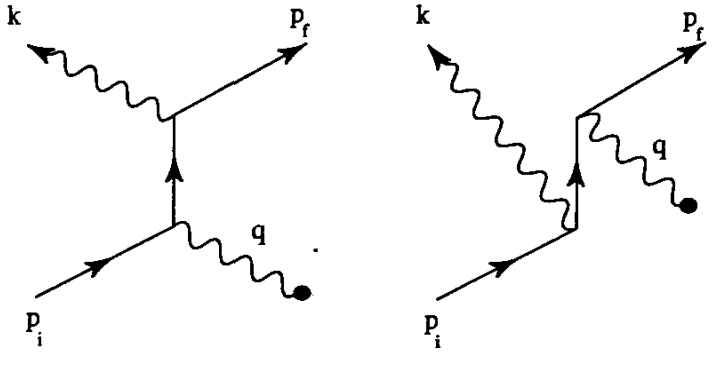


Figure 6.8. Bremsstrahlung, or the radiation emitted by an electron scattering in the presence of a nucleus.

quantum loop corrections to the scattering amplitude. The scattering matrix, using Feynman's rules, is:

$$S_{fi} = -\frac{Ze^3}{V^{3/2}} 2\pi\delta(E_f + k - E_i)\sqrt{\frac{m^2}{E_f E_i}}\bar{u}(p_f, s_f)\left(\frac{(-i\not{\epsilon})}{\sqrt{2k}}\frac{i}{\not{p}_f + \not{k} - m}\frac{(-i\gamma_0)}{|\mathbf{q}|^2}\right.$$

$$\left. + \frac{(-i\gamma_0)}{|\mathbf{q}|^2}\frac{i}{\not{p}_i - \not{k} - m}\frac{(-i\not{\epsilon})}{\sqrt{2k}}\right) u(p_i, s_i) \tag{6.58}$$

The differential cross section now becomes:

$$d\sigma = \frac{mZ^2e^6}{E_i|\mathbf{v}_i|}\int 2\pi\delta(E_f + \omega - E_i)|\bar{u}_f\Gamma u_i|^2$$

$$\times \frac{1}{|\mathbf{q}|^4}\frac{m\,d^3p_f}{E_f}\frac{d^3k}{2\omega(2\pi)^6} \tag{6.59}$$

where $\omega = k_0$ and where:

$$\Gamma = \not{\epsilon}\frac{1}{\not{p}_f + \not{k} - m}\gamma^0 + \gamma^0\frac{1}{\not{p}_i - \not{k} - m}\not{\epsilon} \tag{6.60}$$

The trace we wish to calculate is:

$$T = (2^{-5}m^{-2})(T_1 + T_2 + T_3)$$

$$= \frac{1}{2}\sum_{\epsilon}\mathrm{Tr}\left[\left(\not{\epsilon}\frac{\not{p}_f + \not{k} + m}{2p_f\cdot k}\gamma^0 - \gamma^0\frac{\not{p}_i - \not{k} + m}{2p_i\cdot k}\not{\epsilon}\right)\left(\frac{\not{p}_i + m}{2m}\right)\right.$$

$$\left.\times\left(\gamma^0\frac{\not{p}_f + \not{k} + m}{2p_f\cdot k}\not{\epsilon} - \not{\epsilon}\frac{\not{p}_i - \not{k} + m}{2p_i\cdot k}\gamma^0\right)\left(\frac{\not{p}_f + m}{2m}\right)\right] \tag{6.61}$$

The traces involved in the calculation yield:

$$T_1 = (p_f\cdot k)^{-2}\sum_{\epsilon}\mathrm{Tr}\left[\not{\epsilon}(\not{p}_f + \not{k} + m)\gamma^0(\not{p}_f + \not{k} + m)\not{\epsilon}(\not{p}_f + m)\right]$$

$$= 8(p_f\cdot k)^{-2}\sum_{\epsilon}[2(\epsilon\cdot p_f)^2(m^2 + 2p_i^0p_f^0 + 2p_i^0\omega - p_i\cdot p_f - p_i\cdot k)$$

$$+ 2\epsilon\cdot p_f\,\epsilon\cdot p_i\,k\cdot p_f + 2p_i^0\omega\,k\cdot p_f - p_i\cdot k\,p_j\cdot k] \tag{6.62}$$

and:

$$T_2 = T_1(p_i \leftrightarrow -p_f) \tag{6.63}$$

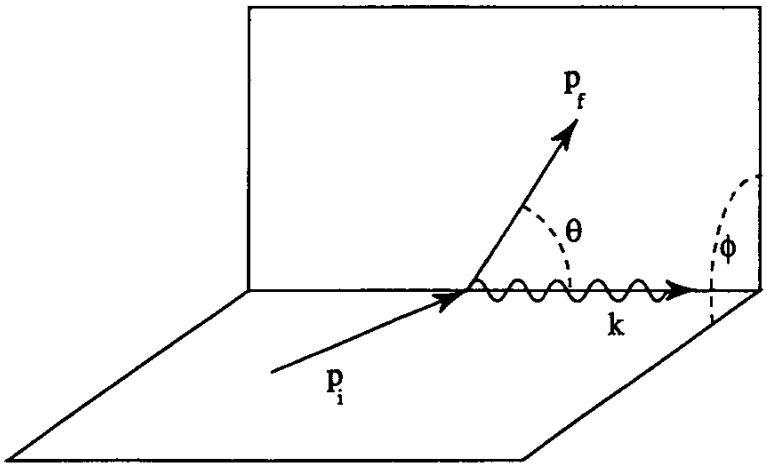


Figure 6.9. Bremsstrahlung, where the emitted photon is in the z direction and the emitted electron is in the $y-z$ plane. The incoming electron is in a plane rotated by angle ϕ from the $y-z$ plane.

and lastly:

$$\begin{aligned} T_3 &= -(p_f\cdot k)^{-1}(p_i\cdot k)^{-1}\mathrm{Tr}\left\{\gamma^0(\not{p}_i-\not{k}+m)\not{\epsilon}(\not{p}_i+m)\gamma^0\right. \\ &\quad \left.\times(\not{p}_f+\not{k}+m)\not{\epsilon}(\not{p}_f+m)+(p_i\leftrightarrow -p_f)\right\} \\ &= 16(p_f\cdot k)^{-1}(p_i\cdot k)^{-1}\sum_{\epsilon}\Big[\epsilon\cdot p_i\epsilon\cdot p_f(p_i\cdot k-p_f\cdot k \\ &\quad +2p_i\cdot p_f-4p_i^0\cdot p_f^0-2m^2) \\ &\quad +(\epsilon\cdot p_f)^2k\cdot p_i-(\epsilon\cdot p_i)^2k\cdot p_f+p_i\cdot kp_f\cdot k-m^2\omega^2 \\ &\quad +\omega(\omega p_i\cdot p_f-p_i^0p_f\cdot k-p_f^0p_i\cdot k)\Big] \end{aligned} \tag{6.64}$$

The parametrization of the momenta is a bit complicated, since the reaction does not take place in a plane. In Figure 6.9, we place the emitted photon momenta in the z direction, the emitted electron momenta p_f in the $y-z$ plane, and the incoming electron momenta p_i in a plane that is rotated by an angle ϕ from the $y-z$ plane.

The specific parametrization is equal to:

$$\begin{aligned} k_\mu &= \omega(1,0,0,1) \\ p_{f,\mu} &= (E_f,0,p_f\sin\theta_f,p_f\cos\theta_f) \\ p_{i,\mu} &= (E_i,p_i\sin\theta_i\sin\phi,p_i\sin\theta_i\cos\phi,p_i\cos\theta_i) \end{aligned} \tag{6.65}$$

where $p_i = |\mathbf{p}_i|$ and $p_f = |\mathbf{p}_f|$. Now we must calculate the sum over transverse photon polarizations. Since k_μ points in the z direction, we can choose:

$$\begin{aligned} \epsilon^{(1)} &= (0,1,0,0) \\ \epsilon^{(2)} &= (0,0,1,0) \end{aligned} \tag{6.66}$$

which satisfies all the desired properties of the polarization tensor. With this choice of parametrization for the momenta and polarizations, we easily find:

$$\begin{aligned} \sum_i (\epsilon^{(i)} \cdot p_f)^2 &= p_f^2 \sin^2\theta_f; \quad \sum_i (\epsilon^{(i)} \cdot p_i)^2 = p_i^2 \sin^2\theta_i \\ \sum_i (\epsilon^{(i)} \cdot p_f)(\epsilon^{(i)} \cdot p_i) &= p_i p_f \sin\theta_f \sin\theta_i \cos\phi \end{aligned} \tag{6.67}$$

It is now a simple matter to collect everything together, and we now have the Bethe–Heitler formula[5] (1934), which was first computed without using Feynman's rules:

$$\begin{aligned} d\sigma &= \frac{Z^2\alpha^3}{(2\pi)^2}\frac{p_f}{p_i q^4}\frac{d\omega}{\omega} d\Omega_\gamma d\Omega_e \\ &\times \left(\frac{p_f^2 \sin^2\theta_f}{(E_f - p_f\cos\theta_f)^2}(4E_i^2 - q^2) + \frac{p_i^2 \sin\theta_i}{(E_i - p_i\cos\theta_i)^2}(4E_f^2 - q^2) \right. \\ &+ 2\omega^2 \frac{p_i^2 \sin^2\theta_i + p_f^2 \sin^2\theta_f}{(E_f - p_f\cos\theta_f)(E_i - p_i\sin\theta_i)} \\ &- 2\frac{p_f p_i \sin\theta_i \sin\theta_f \cos\phi}{(E_f - p_f\cos\theta_f)(E_i - p_i\cos\theta_i)} \\ &\left. \times (4E_i E_f - q^2 + 2\omega^2) \right) \end{aligned} \tag{6.68}$$

Now let us make the approximation that $\omega \to 0$. In the soft bremsstrahlung limit, we find a great simplification, and the differential cross section becomes the one found by classical methods:

$$\frac{d\sigma}{d\Omega} \sim \left(\frac{d\sigma}{d\Omega}\right)_{\text{elastic}} \frac{e^2\, d^3k}{2\omega(2\pi)^3} \sum_\epsilon \left(\frac{\epsilon \cdot p_f}{k \cdot p_{-f}} - \frac{\epsilon \cdot p_i}{k \cdot p_i} \right)^2 \tag{6.69}$$

Here the infrared divergence[6–8] appears for the first time. This problem was first correctly analyzed by Bloch and Nordsieck.[8] The integral d^3k/ω is divergent for small ω, and therefore the amplitude for soft photon emission makes no sense. This is rather discouraging, and revealed the necessity of properly adding all quantum corrections. The resolution of this question only comes when the

one-loop vertex corrections are added in properly, which we will discuss later in this chapter. For now, it is important to understand where the divergence comes from and its general form.

The infrared divergence always emerges whenever we have massless particles in a theory that can be emitted from an initial or final leg that is on the mass shell. For example, whenever we emit a soft photon of momentum k from an on-shell electron with momentum p, we find that the propagator just before the emission is given by:

$$\frac{1}{(p-k)^2 - m^2} \sim \frac{1}{-2p \cdot k} \sim \infty \tag{6.70}$$

Because $p^2 = m^2$ and k is small, we find that an integration over momentum k inevitably produces an infrared divergence. In order to quantify this infrared divergence, let us perform the integration over the momentum k, separating out the angular part $d\Omega$ from d^3k. To parametrize the divergence, we will regulate the integral by allowing the photon to have a small but finite mass μ. (This is, of course, a bit delicate since we are breaking gauge invariance by having massive photons, but one can show that, at this order of approximation, there are no problems.) We will integrate k from μ to some energy E given by the sensitivity of the detector. Expanding out the expression in the square, the amplitude now becomes:

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{\alpha}{4\pi^2} \int_\mu^E k\, dk \int d\Omega \left(\frac{2p_f \cdot p_i}{(k \cdot p_f)(k \cdot p_i)} - \frac{m^2}{(k \cdot p_f)^2} - \frac{m^2}{(k \cdot p_i)^2} \right) \tag{6.71}$$

In our approximation, we can perform the angular integral over the last three terms in the large parentheses.

Let us calculate the last two terms appearing on the right-hand side of the equation. We use the fact that:

$$k \cdot p_f \sim E(1 - \cos\theta_f) \tag{6.72}$$

Because $d\Omega = 2\pi d(\cos\theta)$, we can trivially integrate the last two terms:

$$\int \frac{d\Omega}{4\pi} \frac{m^2}{(k \cdot p_f)^2} = \frac{m^2}{2E^2} \int_{-1}^{+1} d(\cos\theta) \frac{1}{(1 - \beta\cos\theta)^2} = 1 \tag{6.73}$$

The integral over the first term is also easy to perform. We have to use the fact that:

$$\frac{2p_f \cdot p_i}{(k \cdot p_f)(k \cdot p_i)} \sim \frac{2E^2(1 - \beta^2\cos\theta)}{E(1 - \beta\cos\theta_i)E(1 - \beta\cos\theta_f)} \tag{6.74}$$

We will now introduce the Feynman parameter trick, which is often used to evaluate Feynman integrals:

$$\frac{1}{ab} = \int_0^1 \frac{dx}{[ax + b(1-x)]^2} \tag{6.75}$$

By introducing a new variable x, we are able to perform the angular integration. We find:

$$\begin{aligned}
\int \frac{d\Omega}{4\pi} \frac{2p_f \cdot p_f}{(k \cdot p_f)(k \cdot p_i)} &\sim 2(1 - \beta^2 \cos^2\theta) \int_0^1 dx \\
&\times \int d(\cos\theta)(1/2) \frac{1}{[1 - \beta \cos\theta_f x - \beta \cos\theta_i (1-x)]^2} \\
&= 2(1 - \beta^2 \cos\theta) \int_0^1 \frac{dx}{1 - \beta^2 + 4\beta^2 \sin^2(\theta/2) x(1-x)} \\
&= \begin{cases} 2[1 + \frac{4}{3}\beta^2 \sin^2(\theta/2] + O(\beta^4) & \text{if} \quad \beta \ll 1 \\ 2\log(-q^2/m^2) + O(m^2/q^2) & \text{if} \quad m^2/q^2 \ll 1 \end{cases}
\end{aligned} \tag{6.76}$$

Inserting this value back into the previous expression, we find that the final soft bremsstrahlung cross section is then given by:

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{\alpha}{\pi} \ln \frac{E^2}{\mu^2} \begin{cases} \frac{4}{3}\beta^2 \sin^2 \frac{\theta}{2} + O(\beta^4) & \beta \ll 1 \\ \log(-q^2/m^2) - 1 + O(m^2/q^2) & \beta \sim 1 \end{cases} \tag{6.77}$$

Although this formula agrees well with experiment at large photon momenta, this amplitude is clearly divergent if we let the fictitious mass of the photon μ go to zero. Thus, the infrared divergence occurs because we have massless photons present in the theory.

We should mention that the infrared problem arose (in another form) in classical physics, before the advent of quantum mechanics. The essential point is that, even at the classical level, we have the effects due to the long-range Coulomb field. If one were to calculate the radiation field created by a particle being accelerated by a stationary charge, one would find a similar divergence using only classical equations. If one tries to divide the energy by k_0 to calculate the number of photons emitted by bremsstrahlung, it turns out to be proportional to the result presented above. Thus, as the momentum of the emitted photon goes to zero, the number of emitted photons becomes infinite. (Classically, the infrared divergence appears in the *number* of emitted photons, not the emitted *energy*.)

Quantum field theory gives us a novel, but rigorous, solution to the infrared problem, which goes to the heart of the measurement process and the quantum

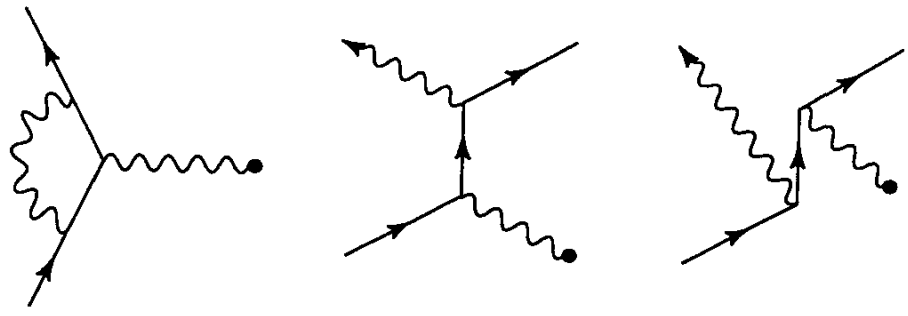


Figure 6.10. The infrared divergence cancels if we add the contributions of two different physical processes. These diagrams can be added together because the resolution of any detector is not sensitive enough to select out just one process.

theory. To this order of approximation, we have to add the contribution of *two* different physical processes to find the cross section of electrons scattering off protons or other charged particles (Fig. 6.10).

The first diagram describes the bremsstrahlung amplitude for the emission of an electron and a photon. The divergence of this amplitude is classical. Second, we have to sum over a purely quantum-mechanical effect, the radiative one-loop corrections of the electron elastically colliding off the charged proton. This may seem strange, because we are adding the cross sections of two different physical processes together, one elastic and one inelastic, to cancel the infrared divergence. However, this makes perfect sense from the point of view of the measuring process.

The essential point is to observe that our detectors cannot differentiate the presence of pure electrons from the presence of electrons accompanied by sufficiently soft photons. This is not just a problem of having crude measuring devices. No matter how precise our measuring apparatus may become, it can never be perfect; there will always be photons with momenta sufficiently close to zero that will sneak past them. Therefore, from an experimental point of view, our measuring apparatus cannot distinguish between these two types of processes and we must necessarily add these two diagrams together. Fortunately, we get an exact cancellation of the infrared divergences when these two scattering amplitudes are added together.

A full discussion of this cancellation, however, cannot be described until we discuss one-loop corrections to scattering amplitudes. Therefore, in section 6.7, we will prove that the bremsstrahlung amplitude, given by:

$$\frac{d\sigma}{d\Omega} \sim \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{\alpha}{\pi} \log \frac{E^2}{\mu^2} \log \frac{-q^2}{m^2} \tag{6.78}$$

must be added to the one-loop vertex correction in order to yield a convergent integral. We will return to this problem at that time.

Finally, we note that, by the substitution rule, we can show the relationship between bremsstrahlung and pair annihilation. Once again, by rotating the diagram around, we can convert bremsstrahlung into pair creation.

This ends our discussion of the tree-level, lowest-order scattering matrix. Although we have had great success in reproducing and extending known classical results, there are immense difficulties involved in extending quantum field theory beyond the tree level. When loop corrections are calculated, we find that the integrals diverge in the ultraviolet region of momentum space. In fact, it has taken over a half century, involving the combined efforts of several generations of physicists, to resolve many of the difficulties of renormalization theory.

We now turn to a detailed calculation of single-loop radiative corrections. Although the calculations are often long and tedious, involving formally divergent quantities, the final conclusions are simple and show that the various infinities can be consistently absorbed into a redefinition of the physical constants of the theory, such as the electric charge and electron mass. Most important, the agreement with experiment is astonishing.

We will begin our discussion of radiative corrections by first examining the self-energy correction to the photon propagator, called the vacuum polarization graph. We will show that the divergence of this graph can be absorbed into a renormalization of the electric charge.

Then, we will calculate the single-loop correction to the electron–photon vertex and show that this leads to corrections to the magnetic moment of the electron. The theoretical value of the anomalous magnetic moment will agree with experiment to one part in 10^8. After that, we will show that the radiative correction to the vertex function is also infrared divergent. Fortunately, the sign of this infrared divergence is opposite the sign found in the bremsstrahlung amplitude. When added together, we will find that the two cancel exactly, giving us a quantum mechanical resolution of the infrared problem.

And finally, we will close this chapter by analyzing the Lamb shift between the energy levels of the $2S_{1/2}$ and $2P_{1/2}$ orbitals of the hydrogen atom. The calculation is rather intricate, because the hydrogen atom is a bound state, and also there are various contributions coming from the vertex correction, the anomalous magnetic moment, the self-energy of the electron, the vacuum polarization graph, etc. However, when all these contributions are added, we find agreement with experiment to within one part in 10^6.

6.6 Radiative Corrections

The simplest higher-order radiative correction is the vacuum polarization graph, shown in Figure 6.11. This graph is clearly divergent. For large momenta,

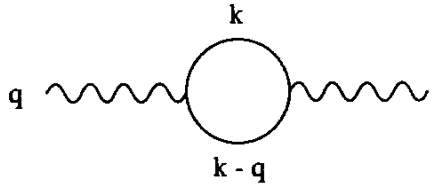


Figure 6.11. First-loop correction to the photon propagator, the vacuum polarization graph, which gives us a correction to the coupling constant and contributes to the Lamb shift.

the Feynman propagators of the two electrons give us two powers of p in the denominator, while the overall integration over d^4p gives us four powers of p in the numerator. So this graph diverges quadratically in the ultraviolet region of momentum space:

$$\Pi_{\mu\nu,m} = -e^2 \,\mathrm{Tr} \int \frac{d^4k}{(2\pi)^4} \gamma_\mu \frac{1}{\not{k} - m + i\epsilon} \gamma_\nu \frac{1}{\not{k} - \not{q} - m + i\epsilon} \tag{6.79}$$

We will perform this integration via the Pauli–Villars method,[9] although the dimensional regularization method, which we will present in the next chapter, is significantly simpler. The Pauli–Villars method replaces this divergent integral with a convergent one by assuming that there are fictitious fermions with mass M in the theory with ghost couplings. At the end of the calculation, these fictitious particles will decouple if we take the limit as their masses tend to infinity. Therefore, M gives us a convenient way of cutting off the divergences of the self-energy correction.

The graph then becomes modified as follows:

$$\tilde{\Pi}_{\mu\nu} = \Pi_{\mu\nu,m} - \Pi_{\mu\nu,M} \tag{6.80}$$

The most convenient way in which to perform the integration is to add additional auxiliary variables. This allows us to reverse the order of integration. We can then perform the integration over the momenta, and save the integration over the auxiliary variables to the very end. We will use:

$$\frac{i}{k^2 - m^2 + i\epsilon} = \int_0^\infty d\alpha \, e^{i\alpha(k^2 - m^2 + i\epsilon)} \tag{6.81}$$

Inserting this expression for the electron propagators and performing the trace, we find:

$$\begin{aligned} \Pi_{\mu\nu,m} \;=\; & 4e^2 \int_0^\infty d\alpha_1 \int_0^\infty d\alpha_2 \int \frac{d^4k}{(2\pi)^4} \\ & \times \exp\left\{ i\alpha_1(k^2 - m^2 + i\epsilon) + i\alpha_2[(k-q)^2 - m^2 + i\epsilon] \right\} \end{aligned}$$

$$\times \left[k_\mu(k-q)_\nu + k_\nu(k-q)_\mu - g_{\mu\nu}(k^2 - k\cdot q - m^2)\right] \quad (6.82)$$

With the insertion of these auxiliary variables, we can now perform the integration over d^4k. First, we shift momenta and complete the square:

$$p = k - \frac{q\alpha_2}{\alpha_1+\alpha_2} \quad (6.83)$$

Then we use the fact that:

$$\begin{aligned}\int \frac{d^4p}{(2\pi)^4}\, e^{ip^2(\alpha_1+\alpha_2)} &= \frac{1}{16\pi^2 i(\alpha_1+\alpha_2)^2} \\ \int \frac{d^4p}{(2\pi)^4}\, p_\mu p_\nu e^{ip^2(\alpha_1+\alpha_2)} &= \frac{ig_{\mu\nu}}{32\pi^2 i(\alpha_1+\alpha_2)^3}\end{aligned} \quad (6.84)$$

Putting everything back into $\Pi_{\mu\nu}$, we have:

$$\Pi_{\mu\nu,m} = (g_{\mu\nu}q^2 - q_\mu q_\nu)\Pi_1 + g_{\mu\nu}\Pi_2 \quad (6.85)$$

where:

$$\begin{aligned}\Pi_1 &= -i\frac{2\alpha}{\pi}\int_0^\infty\int_0^\infty \frac{\alpha_1\alpha_2\, d\alpha_1\, d\alpha_2}{(\alpha_1+\alpha_2)^4}\, e^{f(\alpha_1,\alpha_2)} \\ \Pi_2 &= -i\frac{\alpha}{\pi}\int_0^\infty\int_0^\infty i\frac{d\alpha_1\, d\alpha_2}{(\alpha_1+\alpha_2)^3}[f(\alpha_1,\alpha_2)-1]\, e^{f(\alpha_1,\alpha_2)}\end{aligned} \quad (6.86)$$

and where:

$$f(\alpha_1,\alpha_2) = iq^2\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2} - i(m^2 - i\epsilon)(\alpha_1+\alpha_2) \quad (6.87)$$

There is a similar expression for $\Pi_{\mu\nu,M}$. We notice that Π_2 diverges quadratically, which is bad. However, since we have carefully regularized this integral using the Pauli–Villars method, the integral is finite for fixed M, and we are free to manipulate this expression. We can then show that Π_2 vanishes. This rather remarkable fact can be proved using simple scaling arguments. Consider integrals of the following type:

$$\int_0^\infty \frac{dx}{x}\, f(x)e^{f(x)} = \frac{\partial}{\partial\rho}\int_0^\infty \frac{dx}{x}\, e^{\rho f(x)}\Big|_{\rho=1} \quad (6.88)$$

Assume that $f(x/\rho) = f(x)/\rho$. After this simple rescaling, this integral equals:

$$\frac{\partial}{\partial\rho}\int_0^\infty \frac{dx}{x}e^{f(x)}\Big|_{\rho=1} = 0 \quad (6.89)$$

since all dependence on ρ has vanished. With a few modifications, this argument can be used to show that $\Pi_2 = 0$. Thus, the vacuum polarization graph is only logarithmically divergent.

To perform the integrations in Π_1, we use one last identity:

$$1 = \int_0^\infty \frac{d\rho}{\rho}\,\delta\left(1 - \frac{\alpha_1+\alpha_2}{\rho}\right) \tag{6.90}$$

Inserting this into the expression for Π_1 and rescaling α_i, we find:

$$\begin{aligned}\Pi_1 &= -i\frac{2\alpha}{\pi}\int_0^\infty\int_0^\infty \frac{\alpha_1\alpha_2\,d\alpha_1\,d\alpha_2}{(\alpha_1+\alpha_2)^4}\int_0^\infty\frac{d\rho}{\rho} \\ &\quad\times\delta\left(1-\frac{\alpha_1+\alpha_2}{\rho}\right)e^{f(\alpha_1,\alpha_2)} \\ &= -\frac{2i\alpha}{\pi}\int_0^\infty\int_0^\infty d\alpha_1\,d\alpha_2\;\alpha_1\alpha_2\,\delta(1-\alpha_1-\alpha_2) \\ &\quad\times\int_0^\infty\frac{d\rho}{\rho}\exp\left[i\rho(q^2\alpha_1\alpha_2 - m^2 + i\epsilon)\right]\end{aligned} \tag{6.91}$$

As expected, this integral is logarithmically divergent. At this point, we now use the Pauli–Villars regulator, which lowers the divergence of the theory. To perform the tricky ρ integration, we use the fact that $m^2 - \alpha_1\alpha_2 q^2$ is positive, so we can rotate the contour integral of ρ in the complex plane by -90 degrees. Using integration by parts and rescaling, we have the following identity:

$$\int_\epsilon^\infty\frac{d\rho}{\rho}e^{-a\rho} = \log\rho e^{-a\rho}\Big|_\epsilon^\infty - \int_{a\epsilon}^\infty \log a e^{-\rho}\,d\rho + \int_{a\epsilon}^\infty \log\rho e^{-a\rho}\,d\rho \tag{6.92}$$

where $a(m) = m^2 - \alpha_1\alpha_2 q^2$. The dangerous divergence comes from the last term. However, since the last term is independent of a, it cancels against the same term, with a minus sign, coming from the Pauli–Villars contribution. Thus, we have the identity:

$$\begin{aligned}\lim_{\epsilon\to\infty}\int_\epsilon^\infty\frac{d\rho}{\rho}\left(e^{-a(m)\rho} - e^{-a(M)\rho}\right) &= -\log a(m) + \log a(M) \\ &= -\log\left(1-\frac{\alpha_1\alpha_2 q^2}{m^2}\right) + \log\frac{M^2}{m^2}\end{aligned} \tag{6.93}$$

for large but finite M. We can now take the limit at $\epsilon\to 0$ and M becomes large. Then:

$$\tilde{\Pi}_1 = -\frac{i\alpha}{3\pi}\log\frac{M^2}{m^2} + \frac{2i\alpha}{\pi}\int_0^\infty d\alpha_1\,\alpha_1(1-\alpha_1)\log\left(1-\alpha_1(1-\alpha_1)\frac{q^2}{m^2}\right) \tag{6.94}$$

If we perform the last and final integration over α_1, we arrive at:

$$\tilde{\Pi}_1 = \frac{i\alpha}{3\pi}\left\{-\log\frac{M^2}{m^2}+\frac{1}{3}+2\left(1+\frac{2m^2}{q^2}\right)\left[\left(\frac{4m^2}{q^2}-1\right)^{1/2}\right.\right.$$
$$\left.\left.\times\operatorname{arccot}\left(\frac{4m^2}{q^2}-1\right)^{1/2}-1\right]\right\} \tag{6.95}$$

This is our final result. After a long calculation, we find a surprisingly simple result that has a physical interpretation. We claim that the logarithmic divergence can be cancelled against another logarithmic divergence coming from the bare electric charge e_0. In fact, we will simply *define* the divergence of the electric charge so that it precisely cancels against the logarithmic divergence of Π_1.

To lowest order, we find that we can add the usual photon propagator $D_{\mu\nu}$ to the one-loop correction, leaving us with a revised propagator:

$$-i\frac{g_{\mu\nu}}{q^2}\left(1-\frac{\alpha}{3\pi}\log\frac{M^2}{m^2}-\frac{\alpha}{15\pi}\frac{q^2}{m^2}\right) \tag{6.96}$$

in the limit as $q^2 \to 0$. This leaves us with the usual theory, except that the photon propagator is multiplied by a factor:

$$-i\frac{g_{\mu\nu}}{q^2} \to Z_3(-i)\frac{g_{\mu\nu}}{q^2} \tag{6.97}$$

where:

$$Z_3 \sim 1-\frac{\alpha}{3\pi}\log\frac{M^2}{m^2} \tag{6.98}$$

Now let us absorb this divergence into the coupling constant e_0. We are then left with the usual theory with an extra (infinite) factor Z_3 multiplying each propagator. Since the photon propagator is connected to two electron vertices, with coupling e_0, we can absorb Z_3 into the coupling constant, so we have, to lowest order:

$$e = \sqrt{Z_3}e_0 \tag{6.99}$$

where e is called the renormalized electric charge. Since the infinity coming from Z_3 cancels (by construction) against the infinity coming from the bare electric charge, the renormalized electric charge e is finite. (Other renormalization constants will be discussed in Chapter 7.)

Although this works at the lowest level, it remains to be seen whether we can extend this procedure to all orders in perturbation theory. This will be further

discussed in the next chapter. Next, we turn to the calculation of the single-loop correction to the vertex function, which gives us the anomalous magnetic moment of the electron.

6.7 Anomalous Magnetic Moment

We recall that in Chapter 4 we derived the magnetic moment of the electron by analyzing the coupling of the electron to the vector potential. At the tree level, we know that the coupling of an electron to the photon is given by $A^\mu \bar{u}\gamma_\mu u$, which in turn gives us a gyromagnetic ratio of $g = 2$. However, the experimentally observed value differed from this predicted value by a small but important amount. Schwinger's original calculation[10] of the anomalous magnetic moment of the electron helped to establish QED as the correct theory of electrons and photons.

To calculate the higher-order corrections to the magnetic moment of the electron, we will use the Gordon identity:

$$\bar{u}(p')\gamma_\mu u(p) = \frac{1}{2m}\bar{u}(p')\left[(p+p')_\mu + i\sigma_{\mu\nu}q^\nu\right]u(p) \tag{6.100}$$

(To prove this, we simply use the Dirac equation repeatedly on the left and right spinors, which are on-shell.) The magnetic moment of the electron comes from the second term $\bar{u}\sigma_{\mu\nu}q^\nu u A^\mu$. To see this, we take the Fourier transform, so q^ν becomes ∂^ν, and the coupling becomes $\bar{u}\sigma_{\mu\nu}F^{\mu\nu}u$. The magnetic field B_i is proportional to $\epsilon_{ijk}F^{jk}$, so this coupling term in the rest frame now becomes $\bar{u}\sigma_i B_i u$, where we use the Dirac representation of the Dirac matrices. Since σ_i is proportional to the spin of the electron, which in turn is proportional to the magnetic moment of the electron, the coupling becomes $\boldsymbol{\mu}\cdot\mathbf{B}$. This is the energy of a magnet with moment $\boldsymbol{\mu}$ in a magnetic field $\mathbf{B}$.

In this section, we will calculate the one-loop vertex correction, which gives us a correction to the electron–photon coupling given by (to lowest order in α):

$$\bar{u}(p')\left[\frac{(p+p')_\mu}{2m} + \left(1+\frac{\alpha}{2\pi}\right)\frac{i\sigma_{\mu\nu}q^\nu}{2m}\right]u(p) \tag{6.101}$$

for the process given in Figure 6.12.

Notice that the $\bar{u}\sigma_{\mu\nu}q^\nu u$ term is modified by the one-loop correction, so that the g of the electron becomes:

$$\frac{g}{2} = 1 + \frac{\alpha}{2\pi} \tag{6.102}$$

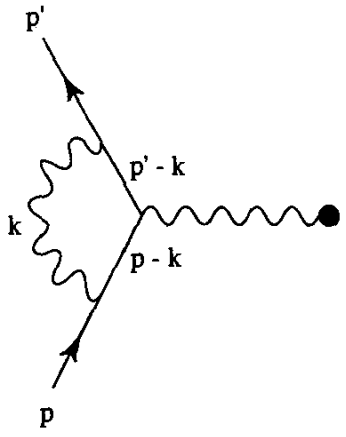


Figure 6.12. First-loop correction to the electron vertex function, which contributes to the anomalous electron magnetic moment.

Thus, QED predicts a correction to the moment of the electron. To show this, we will begin our calculation with the one-loop vertex correction:

$$\Lambda_\mu(p', p) = (-ie)^2 \int \frac{d^4k}{(2\pi)^4} \frac{(-i)}{k^2 - \mu^2 + i\epsilon} \gamma_\nu \frac{i}{\not{p}' - \not{k} - m + i\epsilon} \times \gamma_\mu \frac{i}{\not{p} - \not{k} - m + i\epsilon} \gamma^\nu \tag{6.103}$$

Anticipating that the integral is infrared divergent, we have added μ, the fictitious mass of the photon. The integral is also divergent in the ultraviolet region, so we will use the Pauli–Villars cutoff method later to isolate the divergence.

Throughout this calculation, we tacitly assume that we have sandwiched this vertex between two on-shell spinors, so we can use the Gordon decomposition and the mass-shell condition. Our goal is to write this expression in the form:

$$\Lambda_\mu \sim \gamma_\mu F_1(q^2) + \frac{i\sigma_{\mu\nu}q^\nu}{2m} F_2(q^2) \tag{6.104}$$

sandwiched between $\bar{u}(p')$ and $u(p)$, where F_1 and F_2 are the form factors that measure the deviation from the simple γ_μ vertex. We will calculate explicit forms for these two form factors. (We will find that F_1 cancels against the infrared divergence found in the bremsstrahlung calculation, giving us a finite result. We will show that F_2 gives us a correction to the magnetic moment of the electron.)

We begin with the Feynman parameter trick, generalizing Eq. (6.75). The following can be proved by induction:

$$\frac{1}{a_1 \cdots a_n} = (n-1)! \int_0^\infty \int_0^\infty \int_0^\infty \prod_{i=1}^{n} dz_i \frac{\delta\left(1 - \sum_i z_i\right)}{\Delta^n} \tag{6.105}$$

where: $\Delta = \sum_i a_i z_i$. For our purposes, we want:

$$\begin{aligned} a_1 &= k^2 - \mu^2 + i\epsilon \\ a_2 &= (p' - k)^2 - m^2 + i\epsilon \\ a_3 &= (p - k)^2 - m^2 + i\epsilon \end{aligned} \tag{6.106}$$

Therefore:

$$\Lambda_\mu(p', p) = 2ie^2 \int \frac{d^4k}{(2\pi)^4} \int_0^\infty dz_1\, dz_2\, dz_3\, \delta(1 - z_1 - z_2 - z_3) \left(\frac{N_\mu(k)}{\Delta^3} \right) \tag{6.107}$$

where:

$$N_\mu(k) = \gamma_\nu \left(\not{p}' - \not{k} + m \right) \gamma_\mu \left(\not{p} - \not{k} + m \right) \gamma^\nu \tag{6.108}$$

Next, we would like to perform the integration by completing the square:

$$\begin{aligned} \Delta &= \sum_{i=1}^{3} a_i z_i = (k - p' z_2 - p z_3)^2 - \Delta_0 \\ \Delta_0 &= m^2(1 - z_1)^2 + \mu^2 z_1 - q^2 z_2 z_3 - i\epsilon \\ &\quad -[(p')^2 - m^2]^2 z_2(1 - z_2) - (p^2 - m^2) z_3(1 - z_3) \end{aligned} \tag{6.109}$$

This allows us to make the shift in integration:

$$k \to k + p' z_2 + p z_3 \tag{6.110}$$

Therefore, after the shift, we have:

$$\begin{aligned} \Lambda_\mu(p', p) &= \frac{2ie^2}{(2\pi)^4} \int_0^1 dz_1\, dz_2\, dz_3\, \delta(1 - z_1 - z_2 - z_3) \\ &\quad \times \int_{-\infty}^{\infty} d^4k\, \frac{N_\mu(k + p' z_2 + p z_3)}{(k^2 - \Delta_0 + i\epsilon)^3} \end{aligned} \tag{6.111}$$

By power counting, the integral diverges. This is why we must subtract off the contribution of the Pauli–Villars field, which has mass Λ. Let us expand:

$$N_\mu(k + p' z_2 + p z_3) = -k^2 \gamma_\mu + 2k_\mu \not{k} + A_\mu(k) + N_\mu(p' z_2 + p z_3) \tag{6.112}$$

where $A_\mu(k)$ is linear in the k_μ variable. (This term can be dropped, since its integral over d^4k vanishes.)

Therefore, the leading divergence behaves like:

$$\int d^4k \left(\frac{k_\mu k_\nu}{\left(k^2 - \Delta_0\right)^3} - \frac{k_\mu k_\nu}{\left(k^2 - \Delta_\Lambda\right)^3} \right) \tag{6.113}$$

where Δ_Λ represents the Pauli–Villars contribution, where μ is replaced by Λ. With this insertion, the integral converges for finite but large Λ. To perform this integration, we must do an analytic continuation of the previous equation. We know that (see Appendix):

$$\int d^4k \frac{k_\mu k_\nu}{\left(k^2 - \Delta\right)^\alpha} = \frac{i\pi^2}{2\Gamma(\alpha)(-\Delta)^{\alpha-3}} \Gamma(\alpha - 3) g_{\mu\nu} \tag{6.114}$$

Now let us take the limit at $\alpha \to 3$. This expression, of course, diverges, but the Pauli–Villars term subtracts off the divergence. If we let $\alpha - 3 = \epsilon$, then we have:

$$\begin{aligned}
&\lim_{\alpha \to 3} \int d^4k \left(\frac{k^2}{\left(k^2 - \Delta_0\right)^\alpha} - \frac{k^2}{\left(k^2 - \Delta_\Lambda\right)^\alpha} \right) \\
&= \lim_{\epsilon \to 0} i\pi^2 \Gamma(\epsilon) \left(\frac{1}{(-\Delta_0)^\epsilon} - \frac{1}{(-\Delta_\Lambda)^\epsilon} \right) \\
&= \lim_{\epsilon \to 0} i\pi^2 \frac{1}{\epsilon} \left(e^{\epsilon \log(-\Delta_0)} - e^{\epsilon \log(-\Delta_\Lambda)} \right) \\
&= -i\pi^2 \log \left(\frac{\Delta_\Lambda}{\Delta_0} \right) \\
&= -i\pi^2 \log \left(\frac{z_1 \Lambda^2}{\Delta_0} \right)
\end{aligned} \tag{6.115}$$

where we drop terms like Λ^{-n}. Because this expression is sandwiched between two on-shell spinors, we can also reduce the term:

$$\begin{aligned}
N_\mu(p' z_2 + p z_3) &= -\gamma_\mu \left[2m^2(1 - 4z_1 + z_1^2) + 2q^2(1 - z_2)(1 - z_3) \right] \\
&\quad -2mz_1(1 - z_1) \left[\not{q}, \gamma_\mu \right]
\end{aligned} \tag{6.116}$$

At this point, all integrals can be evaluated. Putting everything together, and dropping all terms of order Λ^{-1} or less, we now have:

$$\Lambda_\mu(p', p) = \frac{\alpha}{2\pi} \int_0^\infty dz_1\, dz_2\, dz_3\, \delta(1 - z_1 - z_2 - z_3) \left[\log \left(\frac{z_1 \Lambda^2}{\Delta_0} \right) \gamma_\mu \right.$$

$$+ \gamma_\mu \frac{m^2(1-4z_1+z_1^2)+q^2(1-z_2)(1-z_3)}{\Delta_0} + \frac{im\sigma_{\mu\nu}q^\nu z_1(1-z_1)}{\Delta_0}\Big] \quad (6.117)$$

Now let us compare this expression with the form factors F_1 and F_2 appearing in Eq. (6.104). It is easy to read off:

$$F_1(q^2) \rightarrow 1 + \frac{\alpha}{2\pi}\int_0^\infty dz_1\, dz_2\, dz_3\, \delta(1-z_1-z_2-z_3) \times \left(-\log\Delta_0 + \frac{m^2(1-4z_1+z_1^2)+q^2(1-z_2)(1-z_3)}{\Delta_0} - (q^2=0)\right) \quad (6.118)$$

[We have deliberately subtracted off the integrand defined at $q^2 = 0$ in order to maintain the constraint $F_1(q^2 = 0) = 1$, which preserves the correct normalization of the vertex function. This extra subtraction term comes from the fact that we have separated out the divergent logarithmic part, which is absorbed in an infinite constant called Z_1 which renormalizes the vertex function. This point will be discussed in more detail in Chapter 7.]

The value of F_2 can similarly be read off:

$$F_2(q^2) = \frac{\alpha}{2\pi}\int_0^\infty dz_1\, dz_2\, dz_3\, \delta(1-z_1-z_2-z_3)\left(\frac{2m^2 z_1(1-z_1)}{\Delta_0}\right) \quad (6.119)$$

The calculation for $F_2(q^2)$ is a bit easier, since there are no ultraviolet or infrared divergences. Because of this, we can set $\mu = 0$. Let us choose new variables:

$$p\cdot p' = m^2\cosh\theta; \quad q^2 = -4m^2\sinh^2(\theta/2) = 2m^2(1-\cosh\theta) \quad (6.120)$$

where $q^2 = (p-p')^2$. Then we have:

$$\begin{aligned} F_2(q^2) &= \frac{\alpha}{\pi}\int_0^1 dz_2\, dz_3\, \theta(1-z_2-z_3)\frac{z_1+z_3-(z_2+z_3)^2}{z_2^2+z_3^2+2z_2z_3\cosh\theta} \\ &= \frac{\alpha}{2\pi}\int_0^1 d\beta \frac{1}{\beta^2+(1-\beta)^2+2\beta(1-\beta)\cosh\theta} \end{aligned} \quad (6.121)$$

This leaves us with the exact result:

$$F_2(q^2) = \frac{\alpha}{2\pi}\frac{\theta}{\sinh\theta} \quad (6.122)$$

We are especially interested in taking the limit as $|q^2| \rightarrow 0$ and $\theta \rightarrow 0$:

$$F_2(0) = \frac{\alpha}{2\pi} \quad (6.123)$$

This gives us the correction to the magnetic moment of the electron, as in Eq. (6.101):

$$\frac{g-2}{2} = \frac{\alpha}{2\pi} \sim 0.0011614 \tag{6.124}$$

This is only a first-order calculation, yet already we are very close to the experimental value. Since the calculation was originally performed by Schwinger, the calculation since has been taken to α^3 order (where there are 72 Feynman diagrams). The theoretical value to this order is given by:

$$\begin{aligned} a_{\text{th.}} &\equiv \frac{1}{2}(g-2) \\ &= 0.5(\frac{\alpha}{\pi}) - 0.32848\left(\frac{\alpha}{\pi}\right)^2 + 1.49\left(\frac{\alpha}{\pi}\right)^3 + \cdots \end{aligned} \tag{6.125}$$

The final results for both the theoretical and experimental values are[11]:

$$\begin{aligned} a_{\text{th}} &= 0.001159652411(166) \\ a_{\text{expt}} &= 0.001159652209(31) \end{aligned} \tag{6.126}$$

where the estimated errors are in parentheses.

The calculation agrees to within one part in 10^8 for a and to one part in 10^9 for g, which is graphic vindication of QED. (To push the calculation to the fourth order involves calculating 891 diagrams and 12,672 diagrams at the fifth order.)

6.8 Infrared Divergence

The calculation for F_1 is much more difficult. However, it will be very important in resolving the question of the infrared divergence, which we found in the earlier discussion of bremsstrahlung. We will find that the infrared divergence coming from the bremsstrahlung graph and F_1 cancel exactly. Although the calculation of F_1 is difficult, one can can extract useful information from the integral by taking the limit as μ becomes small. Then F_1 integration in Eq. (6.118) splits up into four pieces:

$$F_1 = \sum_{i=1}^{4} P_i + \cdots \tag{6.127}$$

where the ellipsis represents constant terms.

After changing variables, each of the P_i pieces can be exactly evaluated:

$$\begin{aligned} P_1 &= -\frac{\alpha}{\pi}\int_0^1 dz_2 \int_0^{1-z_2} dz_3 \frac{\cosh\theta}{z_2^2+z_3^2+2z_2z_3\cosh\theta+(\mu^2/m^2)(1-z_2-z_3)} \\ &= \frac{\alpha}{\pi}\theta\coth\theta\log\frac{\mu}{m} - \frac{2\alpha}{\pi}\coth\theta\int_0^{\theta/2} d\phi\,\phi\tanh\phi \end{aligned} \tag{6.128}$$

where:

$$1-2z_2 = \frac{\tanh\phi}{\tanh(\theta/2)} \tag{6.129}$$

The others are given by:

$$\begin{aligned} P_2 &= \frac{\alpha}{\pi}\cosh\theta\int_0^1 dz_2\int_0^{1-z_2} dz_3 \frac{z_2+z_3}{z_2^2+z_3^2+2z_2z_3\cosh\theta} \\ &= \frac{\alpha}{\pi}\theta\coth\theta \\ P_3 &= \frac{2\alpha}{\pi}(1-\cosh\theta)\int_0^1 dz_2\int_0^{1-\alpha_2} dz_3 \frac{z_2z_3}{z_2^2+z_3^2+2z_2z_3\cosh\theta} \\ &= \frac{\alpha}{2\pi}\left(\frac{\theta}{\sinh\theta}-1\right) \\ P_4 &= -\frac{\alpha}{2\pi}\int_0^1 dz_2\int_0^{1-z_2} dz_3 \log\left[(z_2+z_3)^2+4z_2z_3\sinh^2(\theta/2)\right] \\ &= -\frac{\alpha}{2\pi}\frac{\theta/2}{\tanh(\theta/2)} + \text{const.} \end{aligned} \tag{6.130}$$

Thus, adding the pieces together, we find:

$$\begin{aligned} F_1(q^2) &= \frac{\alpha}{\pi}\Big[\left(\log\frac{\mu}{m}+1\right)(\theta\coth\theta-1) \\ &\quad - 2\coth\theta\int_0^{\theta/2} d\phi\,\phi\tanh\phi - \frac{\theta}{4}\tanh\frac{\theta}{2}\Big] \end{aligned} \tag{6.131}$$

For $|q^2| \ll m^2$, we find using Eqs. (6.104), (6.120), (6.122), and (6.131):

$$\gamma_\mu + \Lambda_\mu^c(p',p) \sim \gamma_\mu\left[1+\frac{\alpha}{3\pi}\frac{q^2}{m^2}\left(\log\frac{m}{\mu}-\frac{3}{8}\right)\right] + \frac{\alpha}{8\pi m}[\not{q},\gamma_\mu] \tag{6.132}$$

For $|q^2| \gg m^2$, we find:

$$\gamma_\mu + \Lambda^c_\mu(p', p) \sim \gamma_\mu \left\{ 1 - \frac{\alpha}{\pi} \log \frac{m}{\mu} \left[\log \left(\frac{-q^2}{m^2} \right) - 1 + O\left(m^2/q^2\right) \right] \right\} \quad (6.133)$$

Plugging all this into the cross-section formula, we now find our final result:

$$\left(\frac{d\sigma}{d\Omega} \right)_\mu = \left(\frac{d\sigma}{d\Omega} \right)_0 \left[1 - \frac{2\alpha}{\pi} \log \frac{m}{\mu} \chi(q^2) \right] \quad (6.134)$$

where:

$$\chi(q^2) = \begin{cases} -\frac{1}{3} q^2/m^2 & \text{if } -q^2 \ll m^2 \\ \log -q^2/m^2 - 1 & \text{if } -q^2 \gg m^2 \end{cases} \quad (6.135)$$

Now we come to the final step, the comparison of the bremsstrahlung amplitude in Eq. (6.77) and the vertex correction for electron scattering. Although they represent different physical scattering processes, they must be added because there is always an uncertainty in our measuring equipment in measuring soft photons. Comparing the two amplitudes, we recall that the bremsstrahlung amplitude in Eq. (6.78) was given by:

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega} \right)_0 \frac{\alpha}{\pi} \log \frac{-q^2}{m^2} \log \frac{E^2}{\mu^2} \quad (6.136)$$

while the vertex correction graph yields:

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega} \right)_0 \left(1 - \frac{\alpha}{\pi} \log \frac{-q^2}{m^2} \log \frac{-q^2}{\mu^2} \right) \quad (6.137)$$

Clearly, when these two amplitudes are added together, we find a finite, convergent result independent of μ^2, as desired. The cancellation of infrared divergences to all orders in perturbation theory is a much more involved process. However, there are surprising simplifications that give a very simple result for this calculation (see Appendix).

6.9 Lamb Shift

Two of the great accomplishments of QED were the determination of the anomalous magnetic moment of the electron, which we discussed earlier, and the Lamb shift.[12] The fact that these two effects could not be explained by ordinary quantum

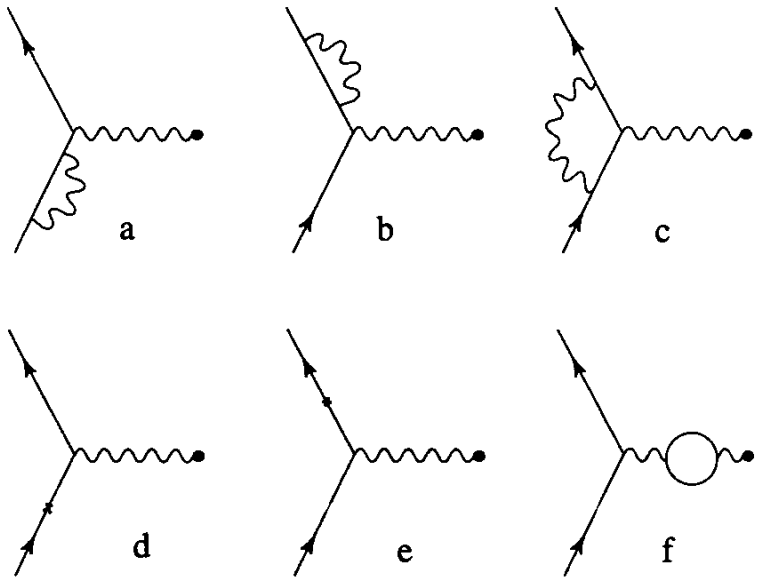


Figure 6.13. The various higher-order graphs that contribute to the Lamb shift: (a) and (b) the electron self-energy graphs; (c) the vertex correction; (d) and (e) the electron mass counterterm; (f) the photon self-energy correction.

mechanics, and the fact that the QED result was so accurate, helped to convince the skeptics that QED was the correct theory of the electron–photon system.

In 1947, Lamb and Retherford demonstrated that the $2S_{1/2}$ and the $2P_{1/2}$ energy levels of the hydrogen atom were split; the $2P_{1/2}$ energy level was depressed more than 1000 MHz below the $2S_{1/2}$ energy level. (The original Dirac electron in a classical Coulomb potential, as we saw earlier in Section 4.2, predicted that the energy levels of the hydrogen atom should depend only on the principal quantum number n and the total spin j, so these two levels should be degenerate.)

The calculation of the Lamb shift is rather intricate, because we are dealing with the hydrogen atom as a bound-state problem, and also because we must sum over all radiative corrections to the electron interacting with a Coulomb potential that modify the naive $\bar{u}\gamma_0 u A^0$ vertex. These corrections include the vertex correction, the anomalous magnetic moment, the self-energy of the electron, the vacuum polarization graph, and even infrared divergences (Fig. 6.13).

The original nonrelativistic bound-state calculation of Bethe,[13] which ignored many of these subtle higher-order corrections, could account for about 1000 MHz of the Lamb shift, but only a fully relativistic quantum treatment could calculate the rest of the difference. Because of the intricate nature of the calculation, we will only sketch the highlights of the calculation. To begin the discussion, we first see that the vacuum polarization graph can be attached to the photon line, changing the photon propagator to:

$$D_{\mu\nu} = -\frac{g_{\mu\nu}}{k^2}\left(1 - \frac{e^2}{60\pi^2}\frac{k^2}{m^2} + O(k^4)\right) \tag{6.138}$$

This, of course, translates into a shift in the effective coupling of an electron to the Coulomb potential.[14,15] Analyzing the zeroth component of this propagator, we see that the coupling of the electron to the Coulomb potential changes as follows:

$$ie_0^2\frac{\bar{u}\gamma_0 u}{q^2} \rightarrow ie^2\frac{\bar{u}\gamma_0 u}{q^2}\left(1-\frac{\alpha q^2}{15\pi m^2}+O(\alpha^2)\right) \tag{6.139}$$

To convert this back into x space, let us take the Fourier transform. We know that the Fourier transform of $1/q^2$ is proportional to $1/r$. This means that the static Coulomb potential that the electron sees is given by:

$$\left(1-\frac{\alpha}{15\pi m^2}\nabla^2\right)\frac{e^2}{4\pi r}=\frac{e^2}{4\pi r}+\frac{e^2}{60\pi^2 m^2}\delta^3(\mathbf{x}) \tag{6.140}$$

meaning that there is a correction to Coulomb's law given by QED. This correction, in turn, shifts the energy levels of the hydrogen atom. We know from ordinary nonrelativistic quantum mechanics that, by taking matrix elements of this modified potential between hydrogen wave functions, we can calculate the first-order correction to the energy levels of the hydrogen atom due to the vacuum polarization graph.

Now let us generalize this discussion to include the other corrections to the calculation of the Lamb shift. Our method is the same: calculate the corrections to the vertex function $\bar{u}\gamma_\mu u$, take the zeroth component, and then take the low-energy limit. In Eqs. (6.104), (6.132), and (6.133), we saw how radiative corrections modified the vertex function with additional form factors $F_1(q^2)$ and $F_2(q^2)$. If we add the various contributions to the vertex correction, we find:

$$\begin{aligned}\bar{u}\gamma_\mu u \rightarrow \bar{u}\Bigg\{&\gamma_\mu\left[1-\frac{\alpha q^2}{3\pi m^2}\left(\log\frac{m}{\mu}-\frac{3}{8}-\frac{1}{5}\right)\right]\\&+\frac{i\alpha}{4\pi m}\sigma_{\mu\nu}q^\nu\Bigg\}u\end{aligned} \tag{6.141}$$

(For example, the vacuum polarization graph contributes the factor $-1/5$ to the vertex correction. The logarithm term comes from the vertex correction, and the μ term is eventually cancelled by the infrared correction.) Now take the low-energy limit of this expression. The $\sigma_{\nu\mu}q^\nu$ term reduces down to a spin-orbit correction, and we find the effective potential given by:

$$\Delta V_{\text{eff}}\sim\frac{4\alpha^2}{3m^2}\left(\log\frac{m}{\mu}-\frac{3}{8}-\frac{1}{5}+\frac{3}{8}\right)\delta^3(\mathbf{x})+\frac{\alpha^2}{4\pi m^2 r^3}\boldsymbol{\sigma}\cdot\mathbf{L} \tag{6.142}$$

By taking the matrix element of this potential between two hydrogen wave functions, we can calculate the energy split due to this modified potential. The

vertex correction, for example, gives us a correction of 1010 MHz. The anomalous magnetic moment of the electron contributes 68 MHz. And the vacuum polarization graph, calculated earlier, contributes -27.1 MHz. Adding these corrections together, we find, to the lowest loop level, that we arrive at the Lamb shift to within 6 MHz accuracy.

Since then, higher-order corrections have been calculated, so the difference between experiment and theory has been reduced to 0.01 MHz. Theoretically, the $2S_{1/2}$ level is above the $2P_{1/2}$ energy level by 1057.864 ± 0.014 MHz. The experimental result is 1057.862 ± 0.020 MHz. This is an excellent indicator of the basic correctness of QED.

6.10 Dispersion Relations

So far, we have been discussing the scattering matrix from the point of view of perturbation and renormalization theory, that is, as a sum of increasingly complicated divergent Feynman diagrams. However, the calculations and the renormalization procedures rapidly become extraordinarily tedious and complicated, involving hundreds of Feynman diagrams at the fourth order.

There is another approach one may take to the scattering matrix that completely avoids the complicated and often counterintuitive operations used in renormalization theory. Following Heisenberg and Chew,[16] one may also take the approach that the S matrix by itself satisfies so many stringent physical requirements that perhaps the S matrix is uniquely determined. This approach avoids the formidable apparatus that one must introduce in order to renormalize even the simplest quantum field theory. Although this approach does not give a systematic method to construct the S matrix for various physical processes, it does give us rigid constraints that are sometimes strong enough to solve for certain properties of the S matrix, such as sum rules and dispersion relations.

The S matrix, for example, should satisfy the following properties:

1. Unitarity or conservation of probability.
2. Analyticity in the various energy variables.
3. Lorentz invariance.
4. Crossing symmetry.

These, in turn, impose an enormous number of constraints on the S matrix, independent of perturbation theory, that may give clues to the final answer. One of the great successes of this approach was the use of *dispersion relations*[17] to calculate constraints on the pion–nucleon cross sections. It gave dramatic verification of the power of analyticity in deriving nontrivial relations among

the cross sections. These relations were used extensively in strong interaction physics where a successful, renormalizable field theory eluded physicists for many decades.

In classical optics, we know that the imposition of causality on the propagation of a wave front is sufficient for us to write down a dispersion relation. Classically, we find that causality implies that the S matrix found in optics is analytic. Following this analogy to classical optics, we can show that the microscopic causality of the Green's functions is sufficient to prove that the S matrix is analytic and hence satisfies certain dispersion relations.

We know that functions $f(z)$ analytic in the upper half plane obey very stringent constraints; they can be written as Cauchy contour integrals:

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(u)}{u-z}du = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(u)}{u-z}dz \tag{6.143}$$

where the contour C is a large closed semicircle sitting on the real axis in the upper half complex plane. In the limit that f vanishes sufficiently rapidly at infinity, we can drop the integral around the outer semicircle. Notice that z sits in the upper half plane. If we slowly let z approach the real axis, then we must take the principal part of the integral:

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\left(\int_{-\infty}^{z-\epsilon} + \int_{z+\epsilon}^{\infty}\right)\frac{f(u)}{u-z}du + \frac{1}{2}\frac{1}{2\pi i}\int_{C'}\frac{f(u)}{u-z}du \\ &= \frac{1}{\pi i}P\int_{-\infty}^{\infty}\frac{f(u)}{u-z}du \end{aligned} \tag{6.144}$$

where C' is a circular contour integral taken infinitesimally around the point z.

Taking the real part of both sides, we then have:

$$\mathrm{Re}\, f(z) = \frac{1}{\pi}P\int_{-\infty}^{\infty}du\,\frac{\mathrm{Im}\, f(u)}{u-z} \tag{6.145}$$

We can therefore convert this formula to an integral from 0 to ∞:

$$\begin{aligned} \mathrm{Re}\, f(z) &= \frac{1}{\pi}P\int_{-\infty}^{0}du\frac{\mathrm{Im}\, f(u)}{u-z} + \frac{1}{\pi}\int_0^{\infty}du\frac{\mathrm{Im}\, f(u)}{u-z} \\ &= \frac{1}{\pi}P\int_0^{\infty}du\frac{\mathrm{Im}\, f(u)}{u+z} + \frac{1}{\pi}P\int_0^{\infty}du\frac{\mathrm{Im}\, f(u)}{z-u} \\ &= \frac{1}{\pi}P\int_0^{\infty}2u\,du\frac{\mathrm{Im}\, f(u)}{u^2-z^2} \end{aligned} \tag{6.146}$$

where we assume $\mathrm{Im} f(z) = -\mathrm{Im} f(z)$.

In ordinary nonrelativistic quantum mechanics, we know that the scattering amplitude of a wave scattering off a stationary, hard sphere is given by a function $f(p)$, which in turn is given as a function of the cross section via the optical theorem:

$$\text{Im } f(p) = \frac{p}{4\pi}\sigma_{\text{tot}}(p) \tag{6.147}$$

Then we have the forward dispersion relation:

$$\text{Re } f(p) = \frac{1}{2\pi^2} P \int_0^\infty dq \frac{q^2\sigma_{\text{tot}}(q)}{q^2 - p^2} \tag{6.148}$$

If often turns out that f does not vanish fast enough to make the above relation valid. For example, for photon–proton scattering (see Section 6.1), we have, for zero frequency:

$$f(0) = -\frac{\alpha}{M} \tag{6.149}$$

which is the Thompson amplitude. Inserting this into the dispersion relation, we have:

$$-\frac{\alpha}{M} = \frac{1}{2\pi^2}\int_0^\infty \sigma_{\text{tot}}(q)dq \tag{6.150}$$

But this is obviously incorrect, because the right-hand side is positive, while the left-hand side is not. The error we have made is assuming that f vanishes at infinity. Instead, we will write a dispersion relation for $f(p)/p$, which does vanish at infinity. Repeating the same steps as before, we find the following dispersion relation:

$$\frac{f(p)}{p} = -\frac{\alpha}{pM} + \frac{1}{2\pi i}\lim_{\epsilon\to 0}\int dq \frac{f(q)}{q(q - p - i\epsilon)} \tag{6.151}$$

Taking the real part, we obtain our final result:

$$\begin{aligned}\text{Re } f(p) &= -\frac{\alpha}{M} + \frac{p^2}{\pi} P \int_0^\infty 2q\, dq \frac{\text{Im } f(q)}{q^2(q^2 - p^2)} \\ &= -\frac{\alpha}{M} + \frac{p^2}{2\pi^2} P \int_0^\infty dq \frac{\sigma_{\text{tot}}(q)}{q^2 - p^2}\end{aligned} \tag{6.152}$$

Now that we have shown how to use the power of analyticity to write dispersion relations, let us combine this with the added condition of unitarity to obtain

conditions on the S matrix. We know that the scattering matrix satisfies:

$$\sum_n S_{nf}^* S_{ni} = \delta_{fi} \tag{6.153}$$

By subtracting off the delta function that describes when the particles do not interact with each other, the S matrix can be written in terms of a $\mathscr{T}$ matrix:

$$S_{fi} = \delta_{fi} - i(2\pi)\delta^4(\sum p_f - \sum p_i)\mathscr{T}_{fi} \tag{6.154}$$

Inserting the expression for S in terms of $\mathscr{T}$, and taking the case of forward scattering, $f = i$, we have:

$$\mathscr{T}_{ii} - \mathscr{T}_{ii}^* = -i(2\pi)^4 \sum_n \delta^4(P_i - P_n)\mathscr{T}_{ni}^* \mathscr{T}_{ni} \tag{6.155}$$

Earlier, in Eqs. (5.21) to (5.26) we derived:

$$\sigma_{\text{tot}} = \frac{(2\pi)^6}{v_{\text{lab}}} \sum_n (2\pi)^4 \delta^4(P_i - P_n)|\mathscr{T}_{ni}|^2 \tag{6.156}$$

where the sum over n includes integrating over final states. Inserting this into the unitarity condition, we find:

$$\text{Im}\,\mathscr{T}_{ii} = -\frac{1}{2} v_{\text{lab}} \frac{\sigma_{\text{tot}}(\omega)}{(2\pi)^6} \tag{6.157}$$

We now wish to apply this formalism to calculate dispersion relations for pion–nucleon scattering.[18] We first need to define the kinematics of the collision. We begin with the scattering of a pion of momentum q_1 off a proton of momentum p_1, producing a pion of momentum q_2 and a proton of momentum p_2.

Let us define the following *Mandelstam variables*:

$$\begin{aligned} s &= (q_1 + p_1)^2 \\ t &= (q_2 - q_1)^2 \\ u &= (q_1 - p_2)^2 \end{aligned} \tag{6.158}$$

For the case of forward scattering, the most convenient variable is the laboratory energy ω of the pion:

$$\omega = \frac{p_1 \cdot q_1}{M} = \frac{s - M^2 - m_\pi^2}{2M} \tag{6.159}$$

By Lorentz invariance, the matrix $\mathscr{M}$ appearing in the transition matrix is given by:

$$\mathscr{M}(q_i, p_i) = \bar{u}(p_2)\left(A(s,t) + \frac{1}{2}(\not{q}_1 + \not{q}_2)B(s,t)\right)u(p_1) \tag{6.160}$$

where nothing is known about the form factors A and B.

In the limit of forward scattering, these two form factors A and B merge into one unknown function:

$$\begin{aligned} \mathscr{T}_{ii} = \frac{1}{(2\pi)^6 2\omega}\mathscr{M}(q_1, p_1) &= \frac{1}{(2\pi)^6 2\omega}4\pi\bar{u}(p_1)T(\omega)u(p_1) \\ 4\pi T(\omega) &= A(s,0) + \omega B(s,0) \end{aligned} \tag{6.161}$$

It can be shown that $T(\omega)$ is an analytic function of ω. Then we can use crossing relations to relate the π^+–nucleon scattering amplitude to the π^-—nucleon scattering amplitude:

$$T(\omega) = T^{\pi^+ p}(\omega) = T^{\pi^- p}(-\omega) \tag{6.162}$$

Written in this fashion, we can write the unitarity condition as:

$$\begin{aligned} \text{Im } T^{\pi^+ p}(-\omega) &= \frac{1}{2i}\left[T^{\pi^+ p}(-\omega + i\epsilon) - T^{\pi^+ p}(-\omega - i\epsilon)\right] \\ &= \frac{k}{4\pi}\sigma_{\text{tot}}^{\pi^- p}(\omega) \end{aligned} \tag{6.163}$$

where $k = \sqrt{\omega^2 - m_\pi^2}$.

Let us introduce the functions $T^\pm$:

$$T^\pm = \frac{1}{2}\left[T(\omega) \pm T(-\omega)\right] = \frac{1}{2}\left[T^{\pi^+ p}(\omega) \pm T^{\pi^- p}(\omega)\right] \tag{6.164}$$

The point of this discussion is to use dispersion relations to write down nontrivial constraints for T that can then be compared with experiment. To write down the dispersion relations, we must deduce the analytic structure of T, which can be done by an analysis of Feynman diagrams. First, if we analyze four-point Feynman diagrams, there is a pole corresponding to the exchange of a single particle. This pole occurs at $\omega_B = m_\pi^2/2M$, which is due to the propagator. The intermediate particle corresponding to this pole is off-shell, and the value of ω is below the energy at which physical scattering takes place. Second, the intermediate states become on-shell in the region $[m_\pi, \infty]$ or $[-\infty, -m_\pi]$. For these values of the energy, real scattering takes place with real intermediate states. Analytically, this means that there is a Riemann cut in the real ω axis.

Putting these two facts together, we can now write down the dispersion relation for T, consisting of the contour integral and also the pole term:

$$T(\omega) = \frac{1}{\pi}\int_{-\infty}^{-m_\pi} d\omega \frac{\mathrm{Im}\, T(\omega')}{\omega' - \omega - i\epsilon} + \frac{1}{\pi}\int_{m_\pi}^{\infty} d\omega \frac{\mathrm{Im}\, T(\omega')}{\omega' - \omega - i\epsilon} + \frac{2f^2}{\omega - \omega_B} \quad (6.165)$$

where the f term comes from the pole contribution, and:

$$f = \left(\frac{g^2}{4\pi}\right)\frac{m_\pi^2}{4M^2} \quad (6.166)$$

Inserting the unitarity condition, we can rewrite the dispersion relation as:

$$\begin{aligned} \mathrm{Re}\, T^-(\omega)/\omega &= \frac{1}{4\pi^2} P\int_{m_\pi}^{\infty} d\omega' \frac{\sqrt{\omega'^2 - m_\pi^2}}{\omega'^2 - \omega^2}\left[\sigma_{\mathrm{tot}}^{\pi^+ p}(\omega') - \sigma_{\mathrm{tot}}^{\pi^- p}(\omega')\right] \\ &\quad - \frac{2f^2}{\omega^2 - \omega_B^2} \end{aligned} \quad (6.167)$$

This is our final result, which agrees well with experiment.

In summary, we have seen that a straightforward application of Feynman's rules allows us to calculate, to lowest order, the interactions of electrons and photons. In this approximation, we find good agreement with earlier, classical results and also experimental data. At higher orders, we find disastrous divergences that require careful renormalizations of the coupling constant and the masses. However, at the one-loop level, QED predicts results for the Lamb shift and the anomalous magnetic moment of the electron that are in excellent agreement with experimental data. This success, in fact, convinced the scientific community of the correctness of QED.

We will now turn to the conclusion of Part I, which is a proof that QED is finite to any order in perturbation theory.

6.11 Exercises

1. Prove the Feynman parameter formula in Eqs. (6.75) and (6.105). (Hint: use induction.)

2. Assume that electrons interact with massless neutrinos via the interaction Lagrangian $gA_\mu \bar{\psi}_e \gamma^\mu (1 - \gamma_5)\psi_\nu$, where A_μ is a massive vector field and ψ_ν is the neutrino field. To second order in the coupling constant, draw all Feynman diagrams for the four-point function $e^- + \nu \to e^- + \nu$ and write down the corresponding expression for the scattering formula. Do not solve.

3. Draw all the Feynman diagrams that appear in the electron–photon vertex function as well as the photon propagator to the third-loop order. Do not solve.

4. Explicitly reduce out the traces in Eq. (6.45).

5. Fill in the missing steps necessary to prove Eqs. (6.49) and (6.68).

6. Consider the scattering of an electron off a stationary spin-0 particle, using the Yukawa interaction $\bar{\psi}\psi\phi$ to lowest order. Compute the matrix element for this scattering. How will the formula differ from the standard Coulomb formula?

7. Compute the scattering matrix to lowest order of an electron with a stationary neutron if the interaction is given by $g\bar{\psi}\sigma^{\mu\nu}\psi F_{\mu\nu}$. Reduce out all gamma matrices. How does this result compare with the Coulomb formula?

8. Show by symmetry arguments that the most general coupling between the proton and the electromagnetic current is given by:

$$\bar{u}_{p'}\left(\gamma_\mu F_1(q^2) + i\frac{\kappa}{2m_p}\sigma_{\mu\nu}q^\nu F_2(q^2)\right)u_p \tag{6.168}$$

where κ is the anomalous magnetic moment of the proton in units of $e/2m_p$, and $F_1(q^2)$ and $F_2(q^2)$ are form factors and functions of the momentum transferred squared.

9. From this, derive the Rosenbluth formula for electrons scattering off a stationary proton target:

$$\begin{aligned}\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab}} &= \frac{\alpha^2\cos^2(\theta/2)}{4E^2\sin^4(\theta/2)}\left[1+2E\sin^2(\theta/2)/m_p\right]^{-1} \\ &\times\left\{|F_1(q^2)|^2 + \frac{q^2}{4m_p^2}\left[2|F_1(q^2)+\kappa F_2(q^2)|^2\tan^2(\theta/2)\right.\right. \\ &\left.\left.+\kappa^2|F_2(q^2)|^2\right]\right\}\end{aligned} \tag{6.169}$$

where:

$$q^2 = \frac{4E^2\sin^2(\theta/2)}{1+(2E/m_p)\sin^2(\theta/2)} \tag{6.170}$$

and where we have assumed that the laboratory energy of the electrons E is much larger than m_e.

10. Prove that $\Pi_2 = 0$ using the identities in Eqs. (6.88) and (6.89).

11. Prove Eqs. (6.62) and (6.64). Fill in the missing steps.

12. Prove Eqs. (6.17) and (6.34).

13. Show that the Mandelstam variables obey the following relationship:

$$s + t + u = \sum_{i=1}^{4} m_i^2 \tag{6.171}$$

where m_i are the masses of the external particles. Show that a four-point scattering amplitude has poles for various values of s, t, and u. Show how the scattering amplitude changes under crossing symmetry as a function of the Mandelstam variables. For example, prove Eq. (6.162).

14. Consider a higher derivative theory with a Lagrangian given by $\phi[\partial_\mu^2 - (\partial_\mu^2)^2/m^2]\phi$. Calculate its propagator and show that it corresponds to a Pauli–Villars-type propagator; that is, it propagates a negative norm ghost. Analyze its ultraviolet divergences, if it has any, if we add the interaction term ϕ^4. Show that the Lagrangian can be written in canonical form; that is, $\mathscr{L} = p\dot{q} - \mathscr{H}$. (Hint: add in auxiliary fields in order to absorb the large number of time derivatives in the action.)

15. Take an arbitrary Feynman graph. Using the Feynman parameter formula, prove[19,20]:

$$\begin{aligned} I &= \int \frac{d^4q_1 \cdots d^4q_k}{\prod_{i=1}^{n}(p_i^2 - m_i^2)} \\ &= \int d^4q_1 \cdots d^4q_k \int_0^1 d\alpha_1 \cdots d\alpha_n \delta\left(1 - \sum_{j=1}^{n} \alpha_j\right) \\ &\quad \times \left(\sum_j (k_j^2 - m_j^2)\alpha_j + 2\sum_{j,s} k_j \alpha_j \eta_{js} \cdot q_s + \sum_{j,s,t} \alpha_j \eta_{js} \eta_{jt} q_s \cdot q_t\right)^{-n} \end{aligned} \tag{6.172}$$

where:

$$p_j = k_j + \sum_{s=1}^{k} \eta_{js} q_s \tag{6.173}$$

and η_{js} equals ± 1 or zero, depending on the placement of the jth and sth lines. If the j line lies in the sth loop, then $\eta = +1$ (-1) if p_j and q_s are parallel (antiparallel). Otherwise, η equals zero.

16. Prove that the previous integral can be written as:

$$\int_0^\infty d\alpha_1 \cdots d\alpha_n \frac{\delta\left(1 - \sum_{j=1}^n \alpha_j\right)}{\Delta^2 \left[\sum_{j=1}^n (k_j^2 - m_j^2)\alpha_j\right]^{n-2k}} \tag{6.174}$$

where:

$$\Delta = \det \left(\sum_{j=1}^n \eta_{js}\eta_{jt}\alpha_j\right)_{st} \tag{6.175}$$

(Hint: choose the k_i in order to set the cross-terms to zero:

$$\sum_{j=1}^n k_j \alpha_j \eta_{js} = 0 \tag{6.176}$$

Then diagonalize the integration over q_i.)

17. In the previous problem, we can make an analogy between a Feynman integral and Kirchhoff's laws found in electrical circuits. Show that we can make the analogy:

$$\begin{aligned} k &\rightarrow \text{current} \\ \alpha &\rightarrow \text{resistance} \\ k\alpha &\rightarrow \text{voltage} \\ k^2\alpha &\rightarrow \text{power} \end{aligned} \tag{6.177}$$

Show that the statement:

$$\sum_{\text{loop}} k_j \alpha_j = 0 \tag{6.178}$$

for momenta taken around a closed loop corresponds to the statement that the voltage around a closed loop is zero. Second, we also have the equation:

$$\sum_{\text{vertex}} k_j = 0 \tag{6.179}$$

which is the statement of current conservation. Because the Feynman parameters α are strictly real numbers, show that capacitors and inductors are not allowed in the circuit. In this way, we can intuitively analyze the analytic structure of a Feynman graph.

Chapter 7
Renormalization of QED

The war against infinities was ended. There was no longer any reason to fear the higher approximations. The renormalization took care of all infinities and provided an unambiguous way to calculate with any desired accuracy any phenomenon resulting from the coupling of electrons with the electromagnetic field .... It is like Hercules' fight against Hydra, the many-headed sea monster that grows a new head for every one cut off. But Hercules won the fight, and so did the physicists.

—V. Weisskopf

7.1 The Renormalization Program

One of the serious complications found in quantum field theory is the fact that the theory is naively divergent. When higher-order corrections are calculated for QED or ϕ^4 theory, one finds that the integrals diverge in the ultraviolet region, for large momentum p.

Since the birth of quantum field theory, several generations of physicists have struggled to renormalize it. Some physicists, despairing of ever extracting meaningful information from quantum field theory, even speculated that the theory was fundamentally sick and must be discarded. In hindsight, we can see that the divergences found in quantum field theory were, in some sense, inevitable. In the transition from quantum mechanics to quantum field theory, we made the transition from a finite number of degrees of freedom to an infinite number. Because of this, we must continually sum over an infinite number of internal modes in loop integrations, leading to divergences. The divergent nature of quantum field theory then reflects the fact that the ultraviolet region is sensitive to the infinite number of degrees of freedom of the theory. Another way to see this is that the divergent graphs probe the extremely small distance region of space-time, or, equivalently, the high-momentum region. Because almost nothing is known about the nature of physics at extremely small distances or momenta, we are disguising our ignorance of this region by cutting off the integrals at small distances.

Since that time, there have been two important developments in renormalization theory. First was the renormalization of QED via the covariant formulation developed by Schwinger and Tomonaga and by Feynman (which were shown to be equivalent by Dyson[1]). This finally showed that, at least for the electromagnetic interactions, quantum field theory was the correct formalism. Subsequently, physicists attacked the problem of the strong and weak interactions via quantum field theory, only to face even more formidable problems with renormalization that stalled progress for several decades.

The second revolution was the proof by 't Hooft that spontaneously broken Yang–Mills theory was renormalizable, which led to the successful application of quantum field theory to the weak interactions and opened the door to the gauge revolution.

There have been many renormalization proposals made in the literature, but all of them share the same basic physical features. Although details vary from scheme to scheme, the essential idea is that there is a set of "bare" physical parameters that are divergent, such as the coupling constants and masses. However, these bare parameters are unmeasurable. The divergences of these parameters are chosen so that they cancel against the ultraviolet infinities coming from infinite classes of Feynman diagrams, which probe the small-distance behavior of the theory. After these divergences have been absorbed by the bare parameters, we are left with the physical, renormalized, or "dressed" parameters that are indeed measurable.

Since there are a finite number of such physical parameters, we are only allowed to make a finite number of such redefinitions. Renormalization theory, then, is a set of rules or prescriptions where, after a finite number of redefinitions, we can render the theory finite to any order.

(If, however, an infinite number of redefinitions were required to render all orders finite, then *any* quantum field theory could be "renormalized." We could, say, find a different rule or prescription to cancel the divergences for each of the infinite classes of divergent graphs. Unless there is a well-defined rule that determines how this subtraction is to be carried out to all orders, the theory is not well defined; it is infinitely ambiguous.)

We should stress that, although the broad features of the renormalization program are easy to grasp, the details may be quite complicated. For example, solving the problem of "overlapping divergences" requires detailed graphical and combinatorial analysis and is perhaps the most important complication of renormalization theory. Due to these tedious details, there have been several errors made in the literature concerning renormalization theory. (For example, the original Dyson/Ward proof of the renormalization of QED actually breaks down at the 14th order[2,3] because of overlapping diagrams. The original claims of renormalization were thus incomplete. However, the proof can presumably be patched up.)

Mindful of these obscure complications, which tend to conceal the relatively simple essence of renormalization theory, in this chapter we will first try to approach the problem of renormalization from a schematic point of view, and then present the details later. Instead of presenting the complications first, which may be quite involved, we will discuss the basic components of renormalization theory, which occur in four essential steps:

1. *Power counting*
   By simply counting the powers of p in any Feynman graph, we can, for large p, tell whether the integral diverges by calculating the degree of divergence of that graph: each boson propagator contributes p^{-2}, each fermion propagator contributes p^{-1}, each loop contributes a loop integration with p^4, and each vertex with n derivatives contributes at most n powers of p. If the overall power of p; that is, the degree of divergence D, is 0 or positive, then the graph diverges. By simple power counting arguments, we can then calculate rather quickly whether certain theories are hopelessly nonrenormalizable, or whether they can be potentially renormalized.

2. *Regularization*
   Manipulating divergent integrals is not well defined, so we need to cutoff the integration over d^4p. This formally renders each graph finite, order by order, and allows us to reshuffle the perturbation theory. At the end of the calculation, after we have rearranged the graphs to put all divergent terms into the physical parameters, we let the cutoff tend to infinity. We must also show that the resulting theory is independent of the regularization method.

3. *Counterterms or multiplicative renormalization*
   Given a divergent theory that has been regularized, we can perform formal manipulations on the Feynman graphs to any order. Then there are at least two equivalent ways in which to renormalize the theory:

   First, there is the method of *multiplicative renormalization*, pioneered by Dyson and Ward for QED, where we formally sum over an infinite series of Feynman graphs with a fixed number of external lines. The divergent sum is then absorbed into a redefinition of the coupling constants and masses in the theory. Since the bare masses and bare coupling constants are unmeasurable, we can assume they are divergent and that they cancel against the divergences of corresponding Feynman graphs, and hence the theory has absorbed all divergences at that level.

   Second, there is the method of *counterterms*, pioneered by Bogoliubov, Parasiuk, Hepp, and Zimmerman (BPHZ), where we add new terms directly to the action to subtract off the divergent graphs. The coefficients of these counterterms are chosen so that they precisely kill the divergent graphs. In a

renormalizable theory, there are only a finite number of counterterms needed to render the theory finite to any order. Furthermore, these counterterms are proportional to terms in the original action. Adding the original action with the counterterms gives us a renormalization of the masses and coupling constants in the action. These two methods are therefore equivalent; that is, by adding counterterms to the action, they sum up, at the end of the calculation, to give a multiplicative rescaling of the physical parameters appearing in the action.

These methods then give us simple criteria that are necessary (but not sufficient) to prove that a theory is renormalizable:

a. The degree of divergence D of any graph must be a function only of the number of external legs; that is, it must remain constant if we add more internal loops. This allows us to collect all N-point loop graphs into one term. (For super-renormalizable theories, the degree of divergence actually decreases if we add more internal loops).

b. The number of classes of divergent N-point graphs must be finite. These divergences must cancel against the divergences contained within the bare parameters.

4. *Induction*
The last step in the proof of renormalizability is to use an induction argument. We assume the theory is renormalizable at the nth order in perturbation theory. Then we write down a recursion relation that allows us to generate the n + 1st-order graphs in terms of the nth-order graphs. By proving the n + 1st-order graphs are all finite, we can prove, using either multiplicative or counterterm renormalization, that the entire perturbation theory, order by order, is finite. Since there are various recursion relations satisfied by field theory (e.g., Schwinger–Dyson equations, renormalization group equations, etc.), there are also a variety of renormalization programs. However, all induction proofs ultimately rely on *Weinberg's theorem* (which states that a Feynman graph converges if the degree of divergence of the graph and all its subgraphs is negative).

7.2 Renormalization Types

Based on simple power counting of graphs, we can begin to catalog a wide variety of different field theories on the basis of their renormalizability. We will group quantum field theories into four distinct categories:

1. Nonrenormalizable theories.

2. Renormalizable theories.
3. Super-renormalizable theories.
4. Finite theories.

7.2.1 *Nonrenormalizable Theories*

To determine the degree of divergence of any graph, we need to know the dimension of the various fields and coupling constants. We can determine the dimension of a field by analyzing the behavior of the propagator at large momenta, or by analyzing the free action. We demand that the action has the dimension of $\hbar$, (i.e., has zero dimension). Since the volume element d^4x has dimension cm^4, this means that the Lagrangian must have dimension cm^{-4}.

For example, the Klein–Gordon action contains the term $(\partial_\mu \phi)^2$. Since the derivative has dimension cm^{-1}, it means that ϕ also has dimension cm^{-1}. The mass m has dimensions cm^{-1} in our units, so that $m^2\phi^2$ has the required dimension $(\text{cm})^{-4}$. By the same reasoning, the massless Maxwell field also has dimension $(\text{cm})^{-1}$. The Dirac field, however, has dimension $\text{cm}^{-3/2}$, so that the term $\bar{\psi} i \not{\partial} \psi$ has dimension cm^{-4}.

It is customary to define the dimension of a field in terms of inverse centimeters (or, equivalently, grams). If $[\phi]$ represents the dimension of the field in inverse centimeters, then the dimensions of the fields in d space–time dimensions can be easily computed by analyzing the free action, which must be dimensionless:

$$\begin{aligned} [\phi] &= \frac{d-2}{2} \\ [\psi] &= \frac{d-1}{2} \end{aligned} \tag{7.1}$$

The simplest example of a nonrenormalizable theory is one that has a coupling constant with negative dimension, like ϕ^5 theory in four dimensions. To keep the action dimensionless, the coupling constant g must have dimension -1. Now let us analyze the behavior of an N-point function. If we insert a $g\phi^5$ vertex into the N-point function, this increases the number of g's by one, decreasing the dimension of the graph. This must be compensated by an increase in the overall power of k by 1, which increases the dimension of the graph, such that the total dimension of the graph remains the same. Inserting a vertex into the N-point graph has thus made it more divergent by one factor of k. By inserting an arbitrary number of vertices into the N point function, we can arbitrarily increase the overall power of k and hence make the graph arbitrarily divergent. The same remarks apply for ϕ^n for $n > 4$ in four dimensions.

(Since the presence or absence of dimensional coupling constants depends so crucially on the dimension of space–time, we will find that in different space–times the set of renormalizable and nonrenormalizable theories are quite different.)

Some examples of non-renormalizable theories include:

1. *Nonpolynomial actions*
   These actions have an infinite number of terms in them, and typically look like $\sum_{n=3}^{\infty} \phi^n$. They necessarily have coupling constants with negative dimension, and hence are not renormalizable.

2. *Gravity*
   Quantum gravity has a coupling constant κ with negative dimension. ($\kappa^2 \sim G_N$, where G_N is Newton's constant, which has dimension -2.) This means that we cannot perform the standard renormalization program. Also, a power expansion in the coupling constant yields a nonpolynomial theory. Thus, quantum gravity is not renormalizable and is infinitely ambiguous.

3. *Supergravity*
   By the same arguments, supergravity is also nonrenormalizable. Even though it possesses highly nontrivial Ward identities that kill large classes of divergences, the gauge group is not large enough to kill all the divergences.

4. *Four-fermion interactions*
   These actions, like the original Fermi action or the Nambu–Jona–Lasinio action, contain terms like $(\bar{\psi}\psi)^2$. By power counting, we know that ψ has dimensions $\mathrm{cm}^{-3/2}$, so the four-fermion action has dimension 6. This requires a coupling constant with dimension -2, so the theory is nonrenormalizable.

5. $\bar{\psi}\sigma^{\mu\nu}\psi F_{\mu\nu}$
   This coupling, which seems to be perfectly well defined and gauge invariant, is not renormalizable because it has dimension 5, so its coupling constant has dimension -1.

6. *Massive vector theory with non-Abelian group*
   A propagator like:

$$\delta_{ab} \frac{g_{\mu\nu} - \frac{k_\mu k_\nu}{M^2}}{k^2 - M^2 + i\epsilon} \tag{7.2}$$

   goes like $O(1)$ for large k, and hence does not damp fast enough to give us a renormalizable theory. So the theory fails to be renormalizable by power counting arguments.

7. *Theories with anomalies*
   Ward–Takahashi identities are required to prove the renormalizability of gauge theories. However, sometimes a classical symmetry of an action does not survive the rigorous process of quantization, and the symmetry is broken. We say that there are anomalies in the theory that destroy renormalizability. Anomalies will be studied in greater length in Chapter 12.

7.2.2 *Renormalizable Theories*

The renormalizable theories only form a tiny subset of possible quantum field theories. They have only a finite number of counterterms. They also have no dimensional coupling constants; so the dimension of each term in the Lagrangian is cm^{-4}.

Some well-known renormalizable theories include:

1. ϕ^4
   This is the simplest renormalizable theory one can write in four dimensions. Because ϕ has dimension 1, this interaction has dimension 4, and hence the coupling constant is dimensionless. This theory is a prototype of much more complicated actions. (However, it should be pointed out that this theory, when summed to all orders, is probably a free theory.)

2. *Yukawa theory*
   The Yukawa theory has a coupling between fermions and scalars given by:

$$\mathscr{L} = g\bar{\psi}\tau^a\psi\phi^a \tag{7.3}$$

   where τ^a is the generator of some Lie group and g is dimensionless.

3. *Massive vector Abelian theory*
   Although this theory has a propagator similar to the massive vector non-Abelian gauge theory, the troublesome $k_\mu k_\nu$ term drops out of any Feynman graph by $U(1)$ gauge invariance. (This term cannot be dropped in a gauge theory with a non-Abelian group.)

4. *QED*
   There are no dimensional coupling constants, and, by power counting arguments, we need only a finite number of counterterms.

5. *Massless non-Abelian gauge theory*
   By power counting, this theory is renormalizable. A more detailed proof of the renormalizability of this theory will be shown in Chapter 13.

6. *Spontaneously broken non-Abelian gauge theory*
   Although massive non-Abelian gauge theories are, in general, nonrenormalizable, there is one important exception. If the gauge symmetry of a massless non-Abelian theory is spontaneously broken, then renormalizability is not destroyed. The fact that renormalizability persists even after the gauge group is spontaneously broken helped to spark the gauge revolution.
   Since the renormalizability of a theory is dependent on the dimension of space–time, we also have the following renormalizable theories:

7. $\bar{\psi}\psi\phi^2$
   This is renormalizable in three dimensions.

8. ϕ^3
   This is renormalizable in six dimensions.

9. ϕ^5, ϕ^6
   These are renormalizable in three dimensions.

10. $(\bar{\psi}\psi)^2$
    This is renormalizable in two dimensions. Although this interaction is nonrenormalizable in four dimensions, in two dimensions it requires no dimensional coupling constant.
    Finally, we can write down the complete set of interactions for spin 0, $\frac{1}{2}$, and 1 fields that are potentially renormalizable. In four dimensions, they are given symbolically by (if we omit isospin and Lorentz indices):

$$\phi^4, \quad \bar{\psi}\psi\phi, \quad (A^2)^2, \quad \bar{\psi}\not{A}\psi, \quad \phi^\dagger\partial_\mu\phi A^\mu, \quad \phi^\dagger\phi A^2 \tag{7.4}$$

    (The Yang–Mills theory has an additional ∂A^3 interaction, which we will discuss separately.)

7.2.3 *Super-renormalizable Theories*

Super-renormalizable theories converge so rapidly that there are only a finite number of graphs that diverge in the entire perturbation theory. The degree of divergence actually goes down as we add more internal loops. The simplest super-renormalizable theories have coupling constants with positive dimension, such as ϕ^3 in four dimensions. Repeating the argument used earlier, this means that increasing the order g of an N-point function must necessarily decrease the number of momenta k appearing in the integrand, such that the overall dimension of the graph remains the same. Thus, as the order of the graph increases, sooner or later the graph becomes convergent. Thus, there are only a finite number of divergent graphs in the theory.

Some examples of super-renormalizable theories include:

1. ϕ^3
   This is super-renormalizable in four dimensions (but the theory is sick because it is not bounded from below and the vacuum is not stable).

2. ϕ^4
   This is super-renormalizable in three dimensions because it has three superficially divergent graphs, which contribute to the two-point function.

3. $P(\phi)$
   ϕ has zero dimension in two dimensions; so we can have an arbitrary polynomial in the action yet still maintain renormalizability. The interaction $P(\phi)$ produces only a finite number of divergences, all of them due to self-contractions of the lines within the various vertices.

4. $P(\phi)\bar{\psi}\psi$
   This is also super-renormalizable in two dimensions.

7.2.4 Finite Theories

Although Dirac was one of the creators of quantum field theory, he was dissatisfied with the renormalization approach, considering it artificial and contrived. Dirac, in his later years, sought a theory in which renormalization was not necessary at all. Dirac's verdict about renormalization theory was, "This is just not sensible mathematics. Sensible mathematics involves neglecting a quantity when it turns out to be small—not neglecting it because it is infinitely great and you do not want it!"

Instead, Dirac believed that a new theory was needed in which renormalizations were inherently unnecessary. Until recently, it was thought that Dirac's program was a dead end, that renormalization was inherent within any quantum field theory. However, because of the introduction of supersymmetry, we have two possible types of theories that are finite to any order in perturbation theory:

1. *Super Yang–Mills theory*
   Supersymmetry gives us new constraints among the renormalization constants Z that are not found in ordinary quantum theories. In fact, for the $SO(4)$ super Yang–Mills theory, one can show that these constraints are enough to guarantee that $Z = 1$ for all renormalization constants to all orders in perturbation theory. In addition, the $SO(2)$ super Yang–Mills theory, coupled to certain classes of supersymmetric matter, is also finite to all orders in perturbation theory. Although these super Yang–Mills theories are uninteresting from the point of view of phenomenology, the fact that supersymmetry is powerful enough to render certain classes of quantum field theories finite to all orders is reason enough to study them.

2. *Superstrings*
   Supersymmetry also allows us to construct actions much more powerful than the super Yang-Mills theory, such as the superstring theory. Superstring theory has two important properties. First, it is finite to all orders in perturbation theory and is free of all anomalies. Second, it contains quantum gravity, as well as all known forces found in nature, as a subset. The fact that superstring theory is the only candidate for a finite theory of quantum gravity is remarkable. (Because there is no experimental evidence at all to support the existence of supersymmetry, we will discuss these supersymmetric theories later in Chapters 20 and 21.)

7.3 Overview of Renormalization in ϕ^4 Theory

Since renormalization theory is rather intricate, we will begin our discussion by giving a broad overview of the renormalization program for the simplest quantum field theory in four dimensions, the ϕ^4 theory, and then for QED. We will stress only the highlights of how renormalization is carried out. After sketching the overall renormalization program for these two theories, we will then present the details, such as the regularization program and the induction argument.

Our goal is to present the arguments that show that (1) the degree of divergence D of any Feynman graph in ϕ^4 theory is dependent only on the number of external lines, and that (2) these divergent classes can be absorbed into the physical parameters.

Given any graph for ϕ^4, we can analyze its divergent structure by power counting as follows:

$$\begin{aligned} E &= \text{number of external legs} \\ I &= \text{number of internal lines} \\ V &= \text{number of vertices} \\ L &= \text{number of loops} \end{aligned} \tag{7.5}$$

The degree of divergence of an arbitrary Feynman graph is easily computed. Each internal propagator contributes $1/p^2$, while each loop contributes d^4p, so the degree of divergence is given by:

$$D = 4L - 2I \tag{7.6}$$

Now we use some simple counting arguments about graphs to reduce this expression. Each vertex has four lines connecting to it. Each of these lines, in

turn, either ends on an external leg, or on one end of an internal leg, which has two ends. Thus, we must have:

$$4V = 2I + E \tag{7.7}$$

Also, the loop number L can be calculated by analyzing the independent momenta in any Feynman graph. The number of independent momenta is equal to the number of internal lines I minus the constraints coming from momentum conservation. There are V such momentum constraints, minus the overall momentum conservation from the entire graph. Since the number of independent momenta in a Feynman graph is also equal to the number of loop momenta, we have, therefore:

$$L = I - V + 1 \tag{7.8}$$

Inserting these graphical rules into our expression for the divergence of a Feynman graph, we now have:

$$D = 4 - E \tag{7.9}$$

This means that the degree of divergence of any graph in four dimensions is strictly dependent on the number of external lines, which is a necessary condition for renormalizability. The degree of divergence is hence independent of the number of internal loops in the graph.

This also a gratifying result because it means that only the two-point and the four-point graphs are divergent. This, in turn, gives us the renormalization of the two physical quantities in the theory: the mass and the coupling constant. Thus, by using only power counting arguments, in principle we can renormalize the entire theory with only two redefinitions corresponding to two physical parameters. [In d dimensions, however, the degree of divergence is $D = d+(1-d/2)E+(d-4)V$. Because D increases with the number of internal vertices, there are problems with renormalizing the theory in higher dimensions.]

Next, we want to sketch the two methods by which renormalization is carried out: multiplicative renormalization and counterterms. For ϕ^4 theory, we first begin with an elementary discussion of multiplicative renormalization.

We start with an action defined with "unrenormalized" or "bare" coupling constants and masses:

$$\mathscr{L}_0 = \frac{1}{2}\partial_\mu\phi_0\partial^\mu\phi_0 - \frac{1}{2}m_0^2\phi_0^2 - \frac{\lambda_0}{4!}\phi_0^4 \tag{7.10}$$

(We remind the reader that m_0 and λ_0 are formally infinite, but are not measurable.) We define $\Sigma(p^2)$ to be the sum of all *proper* (or one-particle irreducible) two-point graphs. (A graph is called improper or one-particle reducible if we can, by cutting

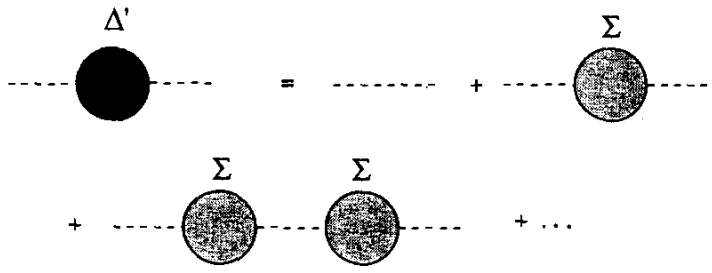


Figure 7.1. The complete propagator Δ' is the sum of an infinite chain of one-particle irreducible graphs Σ. Δ' itself is improper.

just one line, break the graph into two distinct parts. A proper diagram cannot be split into two parts by cutting one line.) For example, the first-loop contribution to $\Sigma(p^2)$ is given by the following:

$$-i\Sigma(p^2) = -i\frac{\lambda_0}{2}\int \frac{d^4p}{(2\pi)^4}\frac{1}{p^2 - m_0^2 + i\epsilon} + \cdots \tag{7.11}$$

which is obtained by taking a four-point vertex and joining two legs together into a loop. It is a proper graph because if we cut the loop, it becomes a four-point vertex and hence does not split apart into two distinct pieces.

Now let $\Delta'(p)$ represent the sum over all possible two-point graphs. We call it the complete or full propagator. $\Delta'(p)$ is obviously the sum of two parts, a proper and improper part. By definition, the proper part cannot be split any further by cutting an intermediate line, but the improper part can. If we split the improper part in half, then each piece, in turn, is the sum of a proper and improper part. Then the smaller improper part can be further split into a proper and improper part. By successively cutting all improper pieces into smaller pieces, we can eliminate all improper parts and write the complete propagator entirely in terms of proper parts. This successive cutting process obviously creates a sequence of proper parts strung together along a string, as in Figure 7.1. $\Delta'(p)$ itself is improper, since it can be split into two pieces by cutting one line.

We can iterate the complete self-energy graph $\Delta'(p^2)$ an infinite number of times with respect to Σ, so that we can formally sum the series, using the fact that:

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \tag{7.12}$$

The sum equals:

$$\begin{aligned} i\Delta'(p) &= i\Delta_F(p) + i\Delta_F(p)[-i\Sigma(p^2)]i\Delta_F(p)\cdots \\ &= i\Delta_F(p)\left(\frac{1}{1 + i\Sigma(p^2)i\Delta_F(p)}\right) \end{aligned}$$

$$= \frac{i}{p^2 - m_0^2 - \Sigma(p^2) + i\epsilon} \tag{7.13}$$

Now power expand this self-energy correction $\Sigma(p^2)$ around $p^2 = m^2$, where m is, at this point, finite but arbitrary:

$$\Sigma(p^2) = \Sigma(m^2) + (p^2 - m^2)\Sigma'(m^2) + \tilde{\Sigma}(p^2) \tag{7.14}$$

where $\tilde{\Sigma}(p^2) \sim O(p^2 - m^2)^2$, and where $\Sigma(m^2)$ and $\Sigma'(m^2)$ are divergent. The net effect of summing this infinite series of graphs is that the complete propagator is now modified to:

$$\begin{aligned} i\Delta'(p) &= \frac{i}{p^2 - m_0^2 - \Sigma(m^2) - (p^2 - m^2)\Sigma'(m^2) - \tilde{\Sigma}(p^2) + i\epsilon} \\ &= \frac{i}{p^2 - [m_0^2 + \Sigma(m^2)] - (p^2 - m^2)\Sigma'(m^2) - \tilde{\Sigma}(p^2) + i\epsilon} \\ &= \frac{i}{[1 - \Sigma'(m^2)](p^2 - m^2) - \tilde{\tilde{\Sigma}}(p^2) + i\epsilon} \end{aligned} \tag{7.15}$$

At this point, m_0 is infinite but arbitrary. Since $\Sigma(m^2)$ is also divergent, we will define m_0 and m such that m_0 cancels against the divergent part coming from $\Sigma(m^2)$, giving us the finite piece m. We will choose:

$$m_0^2 + \Sigma(m^2) = m^2 \tag{7.16}$$

[There is a certain arbitrariness in how two infinite terms cancel, because an infinite term plus a finite term is still infinite. $\Sigma(m^2)$ may have a finite piece in addition to an infinite piece, so the value of m^2 is, at this point, arbitrary. We will comment on this important ambiguity later.]

With this choice, we have:

$$i\Delta'(p) = \frac{iZ_\phi}{p^2 - m^2 - \tilde{\tilde{\Sigma}}(p^2) + i\epsilon} \tag{7.17}$$

where we have made the following redefinitions:

$$\begin{aligned} m^2 &= m_0^2 + \Sigma(m^2) = m_0^2 + \delta m^2 \\ Z_\phi &= \frac{1}{1 - \Sigma'(m^2)} \\ \tilde{\tilde{\Sigma}}(p^2) &= \tilde{\Sigma}(p^2)[1 - \tilde{\Sigma}(m^2)]^{-1} = Z_\phi \tilde{\Sigma}(p^2) \end{aligned} \tag{7.18}$$

Since the divergence of the propagator is contained within Z_ϕ, we can extract this divergent piece. From the unrenormalized propagator Δ', we will remove this divergence and hence obtain the *renormalized propagator* $\tilde{\Delta}'(p)$:

$$\Delta'(p) = Z_\phi \tilde{\Delta}'(p) \tag{7.19}$$

This simple example demonstrates some of the main features of renormalization theory. First, the pole structure of the bare propagator has changed by summing up all graphs. Although we started with a bare propagator $\Delta_F(p)$ with a simple pole at the (infinite) bare mass m_0, the effect of summing all possible graphs is to shift the bare mass to the renormalized or "dressed" mass m. The p dependence of the complete propagator Δ' has changed in a nontrivial fashion. It no longer consists of just a simple pole. However, the complete propagator Δ' still has a pole at the shifted mass squared $p^2 = m^2$ because $\tilde{\Sigma}(m^2) = 0$,

Second, the divergent unrenormalized propagator $\Delta'(p)$ has been converted to the convergent, renormalized propagator $\tilde{\Delta}'(p)$ by a multiplicative rescaling by Z_ϕ. This is crucial in our discussion of renormalization, because it means that we can extract the divergence of self-energy graphs by a simple rescaling (which will eventually be absorbed by redefining the physical coupling constants and wave functions of the theory).

We could also have written this in the language of proper vertices. Let $\Gamma^{(n)}$ represent the one-particle irreducible n-point vertex function, with the n propagators on the external lines removed. (The proper vertex will be defined more rigorously in Chapter 8.) In the free theory, the proper vertex for the two-point function is defined as the inverse of the propagator:

$$i\Gamma_0^{(2)}(p) = p^2 - m_0^2 \tag{7.20}$$

Once we sum to all orders in perturbation theory, then the proper vertex function becomes infinite. If we divide out by this infinite factor Z_ϕ, then we can write the renormalized vertex as:

$$i\Gamma^{(2)}(0) = -m^2 \tag{7.21}$$

where we have taken $p = 0$.

Now consider the effect of renormalization on the coupling constant λ_0. Let $\Gamma^{(4)}$ represent the four point proper vertex, summed over all possible graphs, with propagators on the external lines removed. To lowest order, this four-point graph equals:

$$i\Gamma_0^{(4)}(p_i) = \lambda_0 \tag{7.22}$$

The one-loop correction to this is given by:

$$i\Gamma_0^{(4)} = \lambda_0 - \frac{i}{2}\lambda_0^2 \int \frac{d^4p}{[(p-q)^2 - m_0^2 + i\epsilon][p^2 - m_0^2 + i\epsilon]} \tag{7.23}$$

Although we cannot evaluate $\Gamma^{(4)}$ to all orders in perturbation theory, we know that it is Lorentz invariant and hence can be written in terms of the three Mandelstam variables:

$$s = (p_1 + p_2)^2; \quad t = (p_1 + p_3)^2; \quad u = (p_1 + p_4)^2 \tag{7.24}$$

Therefore:

$$i\Gamma_0^{(4)}(p_i) = \lambda_0 + f(s) + f(t) + f(u) \tag{7.25}$$

for some divergent function f. For the value $p = 0$, let Z_λ^{-1} be this overall infinite factor contained within f. If we divide out the infinite factor from the vertex function, then we have:

$$i\Gamma^{(4)}(0) = \lambda \tag{7.26}$$

where λ is the physical, renormalized coupling constant. In fact, we can take Eqs. (7.21) and (7.26) to be the *definition* of the mass and coupling constant, measured at the point $p = 0$. Because of the ambiguity in manipulating finite and infinite quantities, this definition is not unique. We could have defined the physical mass and coupling constant at some arbitrary momentum scale μ as well as $p = 0$. In other words, these quantities are actually functions of μ. (Of course, when we perform an experiment and actually measure the physical masses and coupling constants, we do so at a fixed momentum, so there is no ambiguity. However, if these experiments could be performed at different momenta, then the effective physical constants may change.)

For example, we could have defined the vertex function at the point $p^2 = m^2$ and $s = u = t = 4m^2/3$. In general, we can define the masses and coupling constants at a different momentum $p = \mu$ via:

$$\begin{aligned} i\Gamma^{(2)}(\mu) &= p^2 - m^2(\mu) \\ i\Gamma^{(4)}(\mu) &= \lambda(\mu) \end{aligned} \tag{7.27}$$

This point μ is called the *renormalization point* or the *subtraction point*. The ambiguity introduced by this subtraction point μ will appear repeatedly throughout this book, and will be studied in more detail in Chapter 14.

Now that we have isolated the divergent multiplicative quantities, we can easily pull them out of any divergent Feynman diagram by redefining the coupling constants and masses. To see how this is done, let us split the Z_ϕ factor occurring with every renormalized propagator in a Feynman diagram into two pieces $(\sqrt{Z_\phi})^2$. In this diagram, move each factor of $\sqrt{Z_\phi}$ into the nearest vertex function. Since each vertex function has four legs, it means that the renormalized vertex function will receive the contribution of four of these factors, or $\sqrt{Z_\phi}^4 = Z_\phi^2$. Since the renormalization of the vertex function contributes an additional factor of Z_λ^{-1}, then the λ_0 sitting in front of the vertex function picks up a factor of Z_ϕ^2/Z_λ. But this means that the original bare coupling constant λ_0 is now modified by this multiplicative renormalization as follows:

$$\lambda_0 \to \lambda = \lambda_0 Z_\lambda^{-1} Z_\phi^2 \tag{7.28}$$

which we define to be renormalized coupling constant.

In this way, we can move all factors of Z_ϕ into the various vertex functions, renormalizing the coupling constant, except for the propagators that are connected to the external legs. We have a left over factor of $\sqrt{Z_\phi}$ for each external leg. This last factor can be eliminated by wave function renormalization. (As we saw in Chapter 5, it was necessary to include a wave function renormalization factor $Z^{-1/2}$ in the definition of "in" and "out" states.) Since the wave function is not a measurable quantity, we can always eliminate the last factors of Z_ϕ by renormalizing the wave function.

In summary, we first began with the ϕ^4 theory defined totally in terms of the unrenormalized, bare coupling constants and masses, m_0^2 and λ_0, which are formally infinite. Then, by summing over infinite classes of diagrams, we found that the modified propagator had a pole at the shifted renormalized mass: $p^2 = m^2$. Furthermore, the propagator and vertex function were multiplied in front by infinite renormalization constants Z_ϕ and Z_λ^{-1}, respectively. By splitting up the Z_ϕ at each propagator into two pieces, we could redistribute these constants such that the new coupling constant became $\lambda_0 Z_\phi^2/Z_\lambda$. The reshuffling of renormalization constants can be summarized as follows:

$$\begin{aligned} \phi &= Z_\phi^{-1/2}\phi_0 \\ \lambda &= Z_\lambda^{-1} Z_\phi^2 \lambda_0 \\ m^2 &= m_0^2 + \delta m^2 \end{aligned} \tag{7.29}$$

where $\delta m^2 = \Sigma(m^2)$.

This can also be generalized to the n-point function as well. If we take an unrenormalized vertex $\Gamma_0^{(n)}(p_i, \lambda_0, m_0)$, we can begin summing the graphs within the unrenormalized vertex to convert it to the renormalized one. In doing so, we

pick up extra multiplicative factors of Z_ϕ and Z_λ, which allow us to renormalize the coupling constant and shift the bare mass. There are n external legs within $\Gamma_0^{(n)}$ that have no propagators, and hence do not contribute their share of the $Z_\phi^{1/2}$ factors. This leaves us with an overall factor of $Z_\phi^{-n/2}$. Then we are left with:

$$\Gamma_0^{(n)}(p_i, \lambda_0, m_0) = Z_\phi^{-n/2}\Gamma^{(n)}(p_i, \lambda, m, \mu) \tag{7.30}$$

Although renormalization of ϕ^4 seems straightforward, unfortunately, there are two technical questions that remain unanswered.

First, as we mentioned earlier, a more rigorous proof must grapple with the problem of overlapping divergences, which is the primary source of complication in any renormalization program. Salam[4] was perhaps the first to fully appreciate the importance of these overlapping divergences. Because of these divergences, the final steps needed to renormalize ϕ^4 are, in some sense, more difficult than the renormalization of QED. As a result, the problem of overlapping divergences for ϕ^4 theory will be discussed at length later in this chapter. We will then solve the problem of overlapping divergences in Chapter 13, when we study the BPHZ renormalization method. We will then complete the renormalizability of ϕ^4 theory in Chapter 14, where we develop the theory of the renormalization group.

Besides the multiplicative renormalization method, there is yet another way to perform renormalization, and this is to proceed *backwards*, that is, start with the usual ϕ^4 action defined with the physical coupling constants and masses, which, of course, are finite. Then, as we calculate Feynman diagrams to each order, we find the usual divergences. The key point is that we can cancel these divergences by adding counterterms to the original action (which are proportional to terms in the original action). This second method is called the counterterm method.

These counterterms contribute new terms to the Feynman series that cancel the original divergences to that order. At the next order, we then find new divergences, so we add new counterterms (again, proportional to terms in the original action), which cancel the divergences to that order. The final action is then the original renormalized action plus an infinite sequence of counterterms, to all orders. Because all the counterterms are proportional to terms in the original action, we wind up with the unrenormalized action defined in terms of unrenormalized parameters (which was the starting point of the previous procedure).

To see the close link between the multiplicative renormalization approach and the counterterm approach, let us start with the renormalized action:

$$\mathscr{L} = \frac{1}{2}[(\partial_\mu\phi)^2 - m^2\phi^2] - \frac{\lambda}{4!}\phi^4 \tag{7.31}$$

where λ and m are the renormalized, finite quantities. We then find divergences with this action, so we add counterterms to the action, such that these counterterms

cancel the divergences that appear to that order. Since the infinities that arise are similar to the infinities we encountered with the multiplicative renormalization program, we find that the coefficients of the counterterms can be summed to yield the quantities Z_ϕ, Z_λ, and δm^2 found earlier.

To see this, let us add the counterterm $\Delta\mathscr{L}$ to the Lagrangian:

$$\mathscr{L} \to \mathscr{L} + \Delta\mathscr{L} \tag{7.32}$$

where the counterterm $\mathscr{L}$ must cancel the divergences coming from the two-point and four-point graphs, which are contained within $\Sigma(p^2)$ and $i\Gamma^{(4)}(p^2)$. Using the subtraction point $\mu = 0$, we find that the counterterm that cancels these divergences can be written as (to lowest loop order):

$$\Delta\mathscr{L} = \frac{\Sigma(0)}{2}\phi^2 + \frac{\Sigma'(0)}{2}(\partial_\mu\phi)^2 + \frac{i\Gamma^{(4)}(0)}{4!}\phi^4 \tag{7.33}$$

Now make the definitions:

$$\begin{aligned} \Sigma'(0) &= Z_\phi - 1 \\ \Sigma(0) &= -(Z_\phi - 1)m^2 + \delta m^2 \\ \Gamma^{(4)}(0) &= -i\lambda(1 - Z_\lambda) \end{aligned} \tag{7.34}$$

which are equivalent (to lowest loop order) to the definitions in Eq. (7.18).

With these definitions, we now have:

$$\Delta\mathscr{L} = \frac{1}{2}[Z_\phi - 1][(\partial_\mu\phi)^2 - m^2\phi^2] + \frac{\delta m^2}{2}Z_\phi\phi^2 - \frac{\lambda(Z_\lambda - 1)}{4!}\phi^4 \tag{7.35}$$

If we add $\mathscr{L}$ and $\Delta\mathscr{L}$ together, we find that we retrieve the original unrenormalized action $\mathscr{L}_0$:

$$\mathscr{L}_0 = \mathscr{L} + \Delta\mathscr{L} \tag{7.36}$$

This intuitively shows the equivalence of the multiplicative renormalization and the counterterm methods. Thus, it is a matter of taste which method we use. In practice, however, this second method of adding counterterms is perhaps more widely used. The point is, however, that both the multiplicative approach and the counterterm approach are equivalent. The fact that the counterterms were proportional to terms in the original action made this equivalence possible.

7.4 Overview of Renormalization in QED

Let us now begin an overview of the renormalization of QED. As in any quantum field theory, the first step involves power counting. Once again, our goal is to show that the degree of divergence of any graph is independent of the number of internal loops.

Let us count the superficial degree of divergence of each graph in QED. We define:

$$\begin{aligned} L &= \text{number of loops} \\ V &= \text{number of vertices} \\ E_\psi &= \text{number of external electron legs} \\ I_\psi &= \text{number of internal electron legs} \\ E_A &= \text{number of external photon legs} \\ I_A &= \text{number of internal photon legs} \end{aligned} \tag{7.37}$$

Then the superficial degree of divergence is:

$$D = 4L - 2I_A - I_\psi \tag{7.38}$$

We can rewrite this equation so that it is only a function of the external legs of the graph, no matter how many internal legs or loops it may have. Each vertex, for example, connects to one end of an internal electron leg. For external electron legs, only one end connects onto a vertex. Thus:

$$V = I_\psi + \frac{1}{2}E_\psi \tag{7.39}$$

Likewise, each vertex connects to one end of an internal photon line, unless it is external. Thus:

$$V = 2I_A + E_A \tag{7.40}$$

Also, we know that the total number of independent momenta is equal to L, which in turn equals the total number of internal lines in the graph minus the number of vertices (since we have momentum conservation at each vertex) plus one (since we also have overall momentum conservation). Thus:

$$L = I_\psi + I_A - V + 1 \tag{7.41}$$

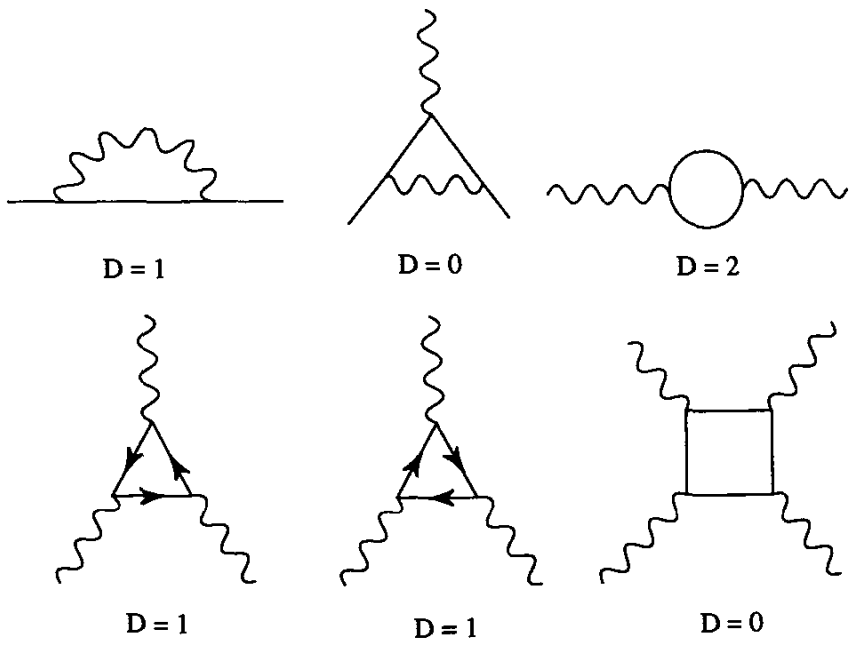


Figure 7.2. The list of divergent classes of graphs in QED. Only the first three graphs, however, are truly divergent if we use gauge invariance and Furry's theorem.

Putting everything together, we find the final formula:

$$D = 4 - \frac{3E_\psi}{2} - E_A \tag{7.42}$$

This is very fortunate, because once again it shows that the total divergence of any Feynman graph, no matter how complicated, is only dependent on the number of external electron and photon legs. In fact, we find that the only graphs that diverge are the following (Fig. 7.2):

1. Electron self-energy graph ($D = 1$).
2. Electron–photon vertex graph ($D = 0$).
3. Photon vacuum polarization graph ($D = 2$) (by gauge invariance, one can reduce the divergence of this graph by two).
4. Three-photon graph ($D = 1$)—these graphs cancel because the internal electron line travels in both clockwise and counterclockwise direction. (If the internal fermion line is reversed, then this can be viewed as reversing the charge of the electron at the vertex. Since there are an odd number of vertices, the overall sign of the graph flips.) Then the two graphs cancel against each other by Furry's theorem.
5. Four-photon graph ($D = 0$)—this graph is actually convergent if we use gauge invariance.

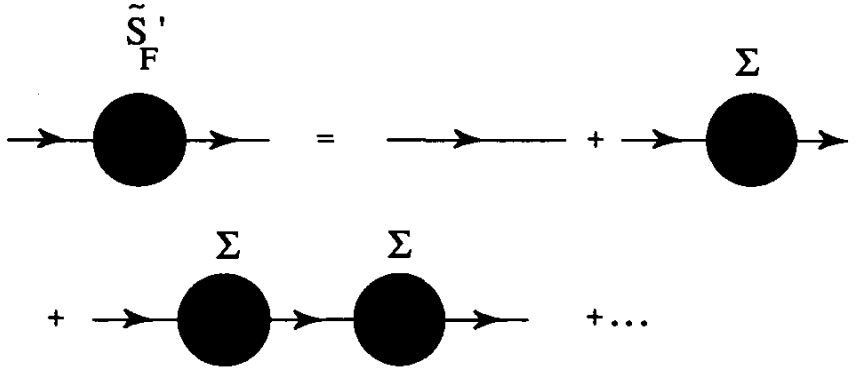


Figure 7.3. The complete propagator S'_F is the sum over one-particle irreducible graphs Σ arranged along a chain.

Only the first three classes of graphs are actually divergent. Fortunately, this is also the set of divergent graphs that can be absorbed into the redefinition of physical parameters, the coupling constant, and the electron mass. (The photon mass, we shall see, is not renormalized because of gauge invariance.) The one-loop graphs that we want particularly to analyze are therefore the electron self-energy correction $\Sigma(\not p)$, the photon propagator $D_{\mu\nu}$, and the vertex correction Γ_μ:

$$-i\Sigma(\not p) = (-ie)^2 \int \frac{d^4k}{(2\pi)^4} \gamma^\mu \frac{i}{\not p - \not k - m} \frac{-ig_{\mu\nu}}{k^2} \gamma^\nu \tag{7.43}$$

$$i\Pi^{\mu\nu}(k) = -(-ie)^2 \int \frac{d^4p}{(2\pi)^4} \mathrm{Tr}\left(\gamma^\mu \frac{i}{\not p - m} \gamma^\nu \frac{i}{\not p - \not k - m}\right)$$

$$iD_{\mu\nu}(k) = -i\frac{g_{\mu\nu}}{k^2} + \left(\frac{-ig_{\mu\alpha}}{k^2}\right) i\Pi^{\alpha\beta}(k) \left(\frac{-ig_{\beta\nu}}{k^2}\right)$$

$$-ie\Lambda_\mu(p,q) = (-ie)^3 \int \frac{d^4k}{(2\pi)^4} \frac{-ig_{\rho\sigma}}{(k+p)^2} \gamma^\rho \frac{i}{\not k - \not q - m} \gamma_\mu \frac{i}{\not k - m} \gamma^\sigma \tag{7.44}$$

To begin renormalization theory, we will, as in ϕ^4 theory, sum all possible graphs appearing in the complete propagator (Fig. 7.3):

$$\begin{aligned} iS'_F(\not p) &= iS_F(\not p) + iS_F(\not p)[-i\Sigma(\not p)]iS_F(\not p) + \cdots \\ &= \frac{i}{\not p - m_0 - \Sigma(\not p) + i\epsilon} \end{aligned} \tag{7.45}$$

Then we make a Taylor expansion of the mass correction $\Sigma(\not{p})$ around $\not{p} = m$, where m is the finite renormalized mass, which is arbitrary:

$$\Sigma(\not{p}) = \Sigma(m) + (\not{p} - m)\Sigma'(m) + \tilde{\Sigma}(\not{p}) \tag{7.46}$$

where $\tilde{\Sigma}(\not{p}) \sim O(\not{p} - m)^2$ and vanishes for $\not{p} = m$. Since m_0 is divergent and arbitrary, we will choose m_0 and m so that m_0 cancels the divergence coming from $\Sigma(m)$. We will choose:

$$m_0 + \Sigma(m) = m \tag{7.47}$$

(Again, there is a certain ambiguity in how this cancellation takes place.)

Inserting the value of $\Sigma(\not{p})$ into the renormalized electron propagator, we now can rearrange terms, just as in the ϕ^4 case, to find:

$$\begin{aligned} iS_F'(\not{p}) &= \frac{i}{\not{p} - m_0 - \Sigma(m) - (\not{p} - m)\Sigma'(m) - \tilde{\Sigma}(\not{p})} \\ &= \frac{i}{[1 - \Sigma'(m)](\not{p} - m) - \tilde{\tilde{\Sigma}}(\not{p}) + i\epsilon} \\ &= \frac{iZ_2}{\not{p} - m - \tilde{\tilde{\Sigma}}(\not{p}) + i\epsilon} \end{aligned} \tag{7.48}$$

where we have defined:

$$\begin{aligned} m &= m_0 + \Sigma(m) \\ Z_2 &= \frac{1}{1 - \Sigma'(m)} \\ \tilde{\tilde{\Sigma}}(\not{p}) &= \tilde{\Sigma}(\not{p})[1 - \Sigma'(m)]^{-1} = Z_2\tilde{\Sigma}(\not{p}) \end{aligned} \tag{7.49}$$

Since the divergence of the complete propagator S_F' is contained within Z_2, we can remove this term and define the renormalized propagator $\tilde{S}_F'$:

$$S_F'(\not{p}) = Z_2\tilde{S}_F'(\not{p}) \tag{7.50}$$

(Throughout this chapter, we will label the divergent, complete propagators like S_F' with the prime, and the finite, renormalized propagators like $\tilde{S}_F'$ with a tilde and a prime.) As before, by summing this subclass of diagrams for the electron self-energy, we have been able to redefine the mass of the electron, such that the physical mass m is actually finite, and also show that the electron propagator is multiplicatively renormalized by the factor Z_2:

$$\psi = \sqrt{Z_2}\psi_0 \tag{7.51}$$

Next, we analyze the photon propagator $D_{\mu\nu}$ in the same way, summing over infinite classes of one-particle irreducible graphs:

$$D'_{\mu\nu} = D_{\mu\nu} - D_{\mu\alpha}\Pi^{\alpha\beta}D_{\beta\nu} + \cdots \tag{7.52}$$

However, a naive application of the previous methods would generate a renormalized mass for the photon, which would be disastrous, as we insist that the physical photon be massless. This is the first example of how gauge invariance helps preserve certain properties of the theory to all orders in perturbation theory. The photon propagator must be gauge invariant, meaning that $k_\mu\Pi^{\mu\nu} = 0$. This constraint, in turn, allows us to make the following decomposition of the second-rank tensor into a scalar quantity $\Pi(k^2)$:

$$\Pi^{\mu\nu}(k^2) = (k^\mu k^\nu - g^{\mu\nu}k^2)\Pi(k^2) \tag{7.53}$$

As before, we now make the infinite summation of corrections to the photon propagator:

$$D'_{\mu\nu} = \frac{1}{k^2[1+\Pi(k^2)]}\left(-g_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\Pi(k^2)\right) \tag{7.54}$$

We then power expand $\Pi(k^2)$ around $k^2 = 0$:

$$\Pi(k^2) = \Pi(0) + \tilde{\Pi}(k^2) \tag{7.55}$$

Inserting this back into the propagator, we find:

$$D'_{\mu\nu} = -\frac{g_{\mu\nu}Z_3}{k^2[1+\tilde{\tilde{\Pi}}(k^2)]} + \text{gauge terms} \tag{7.56}$$

where the last term is proportional to $k_\mu k_\nu$ and hence will be dropped. We have also defined:

$$\begin{aligned} Z_3 &= \frac{1}{1+\Pi(0)} \\ \tilde{\tilde{\Pi}}(k^2) &= \tilde{\Pi}(k^2)[1+\Pi(0)]^{-1} = Z_3\tilde{\Pi}(k^2) \end{aligned} \tag{7.57}$$

Then the finite, renormalized propagator $\tilde{D}'$ can be defined by extracting out Z_3:

$$D'_{\mu\nu} = Z_3\tilde{D}'_{\mu\nu} \tag{7.58}$$

As before, this allows us to make the following wave function renormalization:

$$A_\mu = \sqrt{Z_3} A_{\mu 0} \tag{7.59}$$

Finally, we wish to study the effect of renormalization on the electron–photon vertex function. We first define the vertex function:

$$\Gamma_\mu(p, p') = \gamma_\mu + \Lambda_\mu(p, p') \tag{7.60}$$

After summing all graphs, we find that the vertex graph is infinite, and that we can parametrize this divergence by introducing a third infinite quantity, Z_1, such that:

$$\Gamma_\mu(p, p') = \frac{1}{Z_1} \tilde{\Gamma}_\mu(p, p') \tag{7.61}$$

or:

$$\gamma_\mu + \Lambda_\mu(p, p') = \frac{1}{Z_1} \left[\gamma_\mu + \tilde{\Lambda}_\mu(p, p')\right] \tag{7.62}$$

As in the case of the electron mass, which was renormalized by the infinite quantity δm, we will also use Z_1 to renormalize the coupling constant. The renormalized coupling constant e, as shown above, receives a multiplicative correction $1/Z_1$ from the divergent part of the vertex graph. However, we also know that we have to renormalize the fermion and photon lines. Since there are two fermion lines and one photon line attached to each vertex function, we must multiply the coupling constant by another factor $\sqrt{Z_2}^2 \sqrt{Z_3}$.

The renormalized coupling constant is therefore:

$$e = \frac{Z_2 \sqrt{Z_3}}{Z_1} e_0 \tag{7.63}$$

(Later, using what are called Ward–Takahashi identities, we will show that some of these renormalization constants are equal, i.e. $Z_1 = Z_2$; so that the condition on the renormalized coupling constant reduces to: $e = \sqrt{Z_3} e_0$.)

As we mentioned earlier, instead of using multiplicative renormalization, we could have alternatively used the counterterm approach. This means starting with the theory defined with the physical parameters and then computing divergent self-energy and vertex corrections. Then we add counterterms to the action which, order by order, cancel these divergences. If we then add the renormalized action to the counterterms, we will reproduce the bare action.

If we start with the action defined with renormalized parameters:

$$\mathscr{L} = -\frac{1}{4}(F_{\mu\nu})^2 + \bar{\psi}(i\not{\partial} - m)\psi - e\bar{\psi}\not{A}\psi \tag{7.64}$$

then the theory produces divergent amplitudes, which can be cancelled order by order by adding counterterms to the action. Since the infinities that we encounter are identical to the ones we found with the multiplicative renormalization program, it is not surprising that the coefficients of these counterterms can be written in terms of the same Z's. A careful analysis yields:

$$\mathscr{L} \to \mathscr{L} + \Delta\mathscr{L} \tag{7.65}$$

where:

$$\Delta\mathscr{L} = -\frac{1}{4}(Z_3 - 1)(F_{\mu\nu})^2 + (Z_2 - 1)\bar{\psi}(i\not{\partial} - m)\psi + Z_2\delta m\bar{\psi}\psi - e(Z_1 - 1)\bar{\psi}\not{A}\psi \tag{7.66}$$

Adding these two terms together, we find:

$$\mathscr{L}_0 = -\frac{1}{4}Z_3(F_{\mu\nu})^2 + Z_2\bar{\psi}(i\not{\partial} - m)\psi - Z_1 e\bar{\psi}\not{A}\psi \tag{7.67}$$

If we change variables to the unrenormalized quantities:

$$\begin{aligned} \psi_0 &= \sqrt{Z_2}\psi \\ A_{\mu 0} &= \sqrt{Z_3}A_\mu \\ e_0 &= Z_1 Z_2^{-1} Z_3^{-1/2} e \\ m_0 &= m - \delta m \end{aligned} \tag{7.68}$$

then our action becomes the unrenormalized one:

$$\mathscr{L} + \Delta\mathscr{L} = \mathscr{L}_0 \tag{7.69}$$

Finally, we would like to clarify the arbitrariness introduced into the theory by the subtraction point μ, the point at which we define the masses and coupling constants in Eq. (7.27). This ambiguity arose because of the way that finite parts were handled when canceling infinities. For example, we recall that the one-loop calculation of the photon vacuum polarization graph in Section 6.6, using a cutoff Λ, yielded:

$$\Pi(k^2) = \frac{\alpha_0}{3\pi}\log\frac{\Lambda^2}{k^2} \tag{7.70}$$

In order to extract out the renormalization constant Z_3, we must split $\Pi(k^2)$ into two parts, a constant part and a momentum-dependent part, as in Eq. (7.55). However, *there are an infinite number of ways that we can perform this split.* We could equally well have split the momentum-dependent and momentum-independent parts as:

$$\Pi(k^2) = \frac{\alpha_0}{3\pi}\left(\log\frac{\Lambda^2}{\mu^2} + \log\frac{\mu^2}{k^2}\right) = \Pi(\Lambda, \mu) + \tilde{\Pi}(k^2, \Lambda, \mu) \tag{7.71}$$

where μ is arbitrary. Of course, we have done nothing. We have added and subtracted the same term, $\log\mu^2$. However, by adding finite terms to infinite quantities, we have conceptually made a significant change by altering the nature of the split. With this new split, the photon propagator in Eq. (7.56) can now be rewritten as:

$$D'_{\mu\nu} = \frac{-g_{\mu\nu}Z_3(\Lambda, \mu)}{k^2\left[1 + \tilde{\tilde{\Pi}}(k^2, \Lambda, \mu)\right]} \tag{7.72}$$

It is essential to notice that Z_3 now has an explicit dependence on μ:

$$\begin{aligned} Z_3(\Lambda, \mu) &= \frac{1}{1 + \frac{\alpha_0}{3\pi}\log\frac{\Lambda^2}{\mu^2}} \\ \tilde{\tilde{\Pi}}(k^2, \Lambda, \mu) &= Z_3(\Lambda, \mu)\tilde{\Pi}(k^2, \Lambda, \mu) \end{aligned} \tag{7.73}$$

However, since the renormalized coupling constant α is a function of Z_3, we can isolate the dependence of α on μ. To do this, we note that the unrenormalized α_0 is, by definition, independent of μ. Now choose two different renormalization points, μ_1 and μ_2. Since α can be written in terms of μ_1 or μ_2, we have:

$$\alpha_0 = \frac{\alpha(\mu_1)}{Z_3(\Lambda, \mu_1)} = \frac{\alpha(\mu_2)}{Z_3(\Lambda, \mu_2)} \tag{7.74}$$

where we have set $Z_1 = Z_2$. Substituting in the value of $Z_3(\Lambda, \mu)$, we have:

$$\frac{1}{\alpha(\mu_1)} = \frac{1}{\alpha(\mu_2)} - \frac{2}{3\pi}\log\frac{\mu_1}{\mu_2} \tag{7.75}$$

To lowest order, we have therefore derived a nontrivial relation, following from the *renormalization group* analysis,[5,6] which expresses how the effective coupling constants depend on the point μ, where we define the masses and coupling constants via Eq. (7.27). These renormalization group equations will prove crucial in our discussion of the asymptotic behavior of gauge theory in Chapter 14, where we give a more precise derivation of these relations. They will prove essential in demonstrating that QCD is the leading theory of the strong interactions.

7.5 Types of Regularization

This completes a brief sketch of how the renormalization program works for ϕ^4 and QED. Now comes the difficult part of filling in the essential details of the renormalization program. We will now concentrate on the regularization schemes, certain complications such as the Ward–Takahashi identity and overlapping graphs, and then we will present the proof that QED is renormalizable.

Over the decades, a wide variety of regularization schemes have been developed, each with their own distinct advantages and disadvantages. Each regularization scheme necessarily breaks some feature of the original action:

1. *Pauli–Villars regularization*
   Until recently, this was one of the most widely used regularization scheme. We cutoff the integrals by assuming the existence of a fictitious particle of mass M. The propagator becomes modified by:

$$\frac{1}{p^2 - m^2} - \frac{1}{p^2 - M^2} = \frac{m^2 - M^2}{(p^2 - m^2)(p^2 - M^2)} \tag{7.76}$$

   The relative minus sign in the propagator means that the new particle is a ghost; that is, it has negative norm. This means that we have explicitly broken the unitarity of the theory. The propagator now behaves as $1/p^4$, which is usually enough to render all graphs finite. Then, we take the limit as $M^2 \to \infty$ so that the unphysical fermion decouples from the theory. The advantage of the Pauli–Villars technique is that it preserves local gauge invariance in QED; hence the Ward identities are preserved (although they are broken for higher groups). There have been a large number of variations proposed to the Pauli–Villars technique, such as higher covariant derivatives in gauge theory and higher R^2 terms in quantum gravity.

2. *Dimensional regularization*
   This is perhaps the most versatile and simplest of the recent regularizations. Dimensional regularization involves generalizing the action to arbitrary dimension d, where there are regions in complex d space in which the Feynman integrals are all finite. Then, as we analytically continue d to four, the Feynman graphs pick up poles in d space, allowing us to absorb the divergences of the theory into the physical parameters. Dimensional regularization obviously preserves all properties of the theory that are independent of the dimension of space-time, such as the Ward–Takahashi identities.

3. *Lattice regularization*
   This is the most widely used regularization scheme in QCD for nonperturbative calculations. Here, we assume that space–time is actually a set of discrete points arranged in some form of hypercubical array. The lattice spacing then serves as the cutoff for the space–time integrals. For QCD, the lattice is gauge invariant, but Lorentz invariance is manifestly broken. The great advantage of this approach is that, with Monte Carlo techniques, one can extract qualitative and some even some quantitative information from QCD. One disadvantage with this approach is that it is defined in Euclidean space; so we are at present limited to calculating only the static properties of QCD; the lattice has difficulty describing Minkowski space quantities, such as scattering amplitudes.

In this chapter, we will mainly stress the dimensional regularization method.[7–17] We will now show explicitly, to lowest order only, how dimensional regularization can regulate the divergences of the theory so that we are only manipulating finite quantities (albeit in an unphysical dimension). Then later we will show how to put this all together and renormalize field theory to all orders in perturbation theory.

Our starting point is the action for scalar mesons and for QED in d dimensions:

$$\begin{aligned}\mathscr{L} &= \frac{1}{2}(\partial_\mu\phi_0)^2 - \frac{m_0^2}{2}\phi_0^2 - \frac{\mu^{4-d}}{4!}\lambda_0\phi_0^4 \\ \mathscr{L} &= \bar{\psi}_0(i\not{\partial} - m_0)\psi_0 - e_0\mu^{2-d/2}A_0^\mu\bar{\psi}_0\gamma_\mu\psi_0 - \frac{1}{4}F_{0\mu\nu}^2 \end{aligned} \tag{7.77}$$

where μ is an arbitrary parameter with the dimension of mass. It is necessary to insert this dimensional parameter because the dimension of the fermion field is $[\psi] = (d-1)/2$ while the dimension of the boson field is $[A_\mu] = (d/2) - 1$. (Our final result must be independent of the choice of μ. In this formalism, μ takes the place of the subtraction point introduced earlier in the renormalization-group equations.)

Generalizing space–time to d dimensions, we are interested in evaluating the integral in Eq. (7.11):

$$-i\Sigma(p^2) = -i\frac{\lambda_0\mu^{4-d}}{2}\int\frac{d^dp}{(2\pi)^d}\frac{1}{p^2 - m_0^2 + i\epsilon} \tag{7.78}$$

Our goal is to find a region in complex d space where this integral is finite, and then analytically continue to $d = 4$, where we expect to pick up poles in divergent quantities. Then the renormalization scheme consists of absorbing all these poles into the coupling constants and masses of the theory.

To begin the dimensional regularization process, we first must express the integration over d^dp for arbitrary dimension. To calculate the volume integral over

d space, we remind ourselves that the change from Cartesian to polar coordinates in two dimensions is given by:

$$\begin{aligned} x_1 &= r\cos\theta_1 \\ x_2 &= r\sin\theta_1 \end{aligned} \tag{7.79}$$

In three dimensions, we use the transformation to spherical coordinates given by:

$$\begin{aligned} x_1 &= r\cos\theta_1 \\ x_2 &= r\sin\theta_1\cos\theta_2 \\ x_3 &= r\sin\theta_1\sin\theta_2 \end{aligned} \tag{7.80}$$

Given these two examples, it is not hard to write down the transformation from Cartesian to d-dimensional spherical coordinates, which spans all of d space. (We will make a change of variables in the time parameter, converting real time to imaginary time. This is called a Wick rotation, which takes us from Minkowski space to Euclidean space, where we can perform all d-dimensional integrals all at once. We then Wick rotate back to Minkowski coordinates at the end of the calculation. Alternatively, we could have done the calculation completely in Minkowski space, where the dx^0 integration is handled differently from the other integrations.) The d-dimensional transformation is given by:

$$\begin{aligned} x_1 &= r\cos\theta_1 \\ x_2 &= r\sin\theta_1\cos\theta_2 \\ x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3 \\ x_4 &= r\sin\theta_1\sin\theta_2\sin\theta_3\cos\theta_4 \\ &\vdots \\ x_k &= r\left(\prod_{i=1}^{k-1}\sin\theta_i\right)\cos\theta_k \\ &\vdots \\ x_d &= r\prod_{i=1}^{d-1}\sin\theta_i \end{aligned} \tag{7.81}$$

By induction, one can prove that the Jacobian from Cartesian to spherical coordinates is given by:

$$\begin{aligned} J &= \det\left(\frac{\partial(x_1, x_2, \cdots, x_d)}{\partial(r, \theta_1, \theta_2, \cdots, \theta_{d-1})}\right) \\ &= r^{d-1}\prod_{i=1}^{d-1}\sin^{i-1}\theta_i \end{aligned} \tag{7.82}$$

The volume element in d space is therefore given by:

$$\begin{aligned} \prod_{i=1}^{d} dx_i &= J dr \prod_{i=1}^{d-1} d\theta_i \\ &= r^{d-1} dr d\Omega_{d-1} \\ &= r^{d-1} dr \prod_{i=1}^{d-1}\sin^{i-1}\theta_i d\theta_i \end{aligned} \tag{7.83}$$

To evaluate this integral, we use the fact that:

$$\int_0^\pi \sin^m\theta d\theta = \frac{\sqrt{\pi}\Gamma\left(\frac{1}{2}(m+1)\right)}{\Gamma\left(\frac{1}{2}(m+2)\right)} \tag{7.84}$$

Then in d dimensions, we have:

$$\begin{aligned} \int d^d p &= \int_0^\infty r^{d-1} dr \int_0^{2\pi} d\theta_1 \int_0^\pi \sin\theta_2 \, d\theta_2 \int_0^\pi \sin^2\theta_3 \, d\theta_3 \cdots \\ &\quad \times \int_0^\pi \sin^{d-2}\theta_{d-1} \, d\theta_{d-1} \\ &= \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty r^{d-1} \, dr \end{aligned} \tag{7.85}$$

The integration that we want to perform can therefore be written in the form:

$$I = \int \frac{d^d p}{(p^2 + 2p\cdot q - m^2)^\alpha} = \int d\Omega_{n-1} \int dr \frac{r^{n-1}}{(r^2 - q^2 - m^2)^\alpha} \tag{7.86}$$

(In the last expression, we have changed coordinates $p \to p + q$ in order to eliminate the cross term between p and q. We are also working in Euclidean space and will Wick rotate back to Minkowski space later.) The integral over the

solid angle is given by:

$$\int d\Omega_{d-1} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{7.87}$$

We are left, therefore, with an integral over r. This integral over r is of the form of an Euler Beta function:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = 2\int_0^\infty dt\, t^{2x-1}(1+t^2)^{-x-y} \tag{7.88}$$

With this formula, we can prove:

$$\int_0^\infty \frac{r^\beta dr}{(r^2+C^2)^\alpha} = \frac{\Gamma\left(\frac{1}{2}(1+\beta)\right)\Gamma\left(\alpha - \frac{1}{2}(1+\beta)\right)}{2(C^2)^{\alpha-(1+\beta)/2}\Gamma(\alpha)} \tag{7.89}$$

We are left with the integral:

$$I = \left(\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)\left(\frac{\Gamma(d/2)\Gamma(\alpha - d/2)}{2(-q^2-m^2)^{\alpha-d/2}\Gamma(\alpha)}\right) \tag{7.90}$$

The final result is therefore:

$$\int \frac{d^d p}{(p^2+2p\cdot q - m^2)^\alpha} = \frac{i\pi^{d/2}\Gamma(\alpha - d/2)}{\Gamma(\alpha)(-q^2-m^2)^{\alpha-d/2}} \tag{7.91}$$

(The extra i appears when we rotate back from Euclidean space to Minkowski space.)

Given these formulas, let us first analyze the one-loop correction to the ϕ^4 theory, which is given by Σ, and then generalize this discussion for the Dirac theory. The one-loop correction in ϕ^4 theory in Eq. (7.78) is given by:

$$\begin{aligned} -i\Sigma(p^2) &= \frac{-i\lambda_0 m_0^2\Gamma\left(1-(d/2)\right)}{32\pi^2}\left(\frac{4\pi\mu^2}{-m_0^2}\right)^{2-d/2} \\ &\to \frac{i\lambda_0 m_0^2}{16\pi^2}\left(\frac{1}{4-d}\right) + \cdots \end{aligned} \tag{7.92}$$

where we have extracted the pole term in the Γ function as $d \to 4$. These poles in d space, as we pointed out earlier, can all be absorbed into the renormalization of physical coupling constants and masses. This is our final result for the one-loop, two-point divergence to this order.

To generalize this discussion to the Dirac propagator, we must perform the dimensional integral with two Feynman propagators, so there are two terms in the

denominator. It is easier if we combine the two factors in the denominator together; so we will once again use Feynman's parameter trick. First, let us calculate the electron self-energy correction, which can be written in d-dimensional space as:

$$\begin{aligned}
\Sigma(\not p) &= -ie_0^2\mu^{4-d}\int\frac{d^dk}{(2\pi)^d}\gamma_\mu\left(\frac{\not p-\not k+m_0}{[(p-k)^2-m_0^2]k^2}\right)\gamma^\mu \\
&= -ie_0^2\mu^\epsilon\int_0^1 dx\int\frac{d^dk}{(2\pi)^d}\gamma_\mu\frac{\not p-\not k+m_0}{[(p-k)^2x-m_0^2x+k^2(1-x)]^2}\gamma^\mu \\
&= -ie_0^2\mu^\epsilon\int_0^1 dx\;\gamma_\mu(\not p-\not p x-\not q+m_0)\gamma^\mu \\
&\quad\times\int\frac{d^dq}{(2\pi)^d}\frac{1}{[q^2-m_0^2x+p^2x(1-x)]^2} \\
&= \mu^\epsilon e_0^2\frac{\Gamma(2-d/2)}{(4\pi)^{d/2}}\int_0^1 dx\;\gamma_\mu[\not p(1-x)+m_0]\gamma^\mu \\
&\quad\times[m_0^2x-p^2x(1-x)]^{d/2-2} \qquad (7.93)
\end{aligned}$$

where we have made a sequence of steps: (1) we used the Feynman parameter trick; (2) we made the substitution $q = k - px$; (3) integrated over q, dropping terms linear in q. Finally, we look for the pole in $\Gamma(2-d/2)$ as $d\to 4$. Let us perform the integration over x:

$$\begin{aligned}
\Sigma(\not p) &= -\frac{e_0^2}{16\pi^2}\Gamma(\epsilon/2)\int_0^1 dx\,\{2\not p(1-x)-4m_0-\epsilon[\not p(1-x)+m_0]\} \\
&\quad\times\left(\frac{m_0^2x-p^2x(1-x)}{4\pi\mu^2}\right)^{-\epsilon/2} \\
&= \frac{e_0^2}{8\pi^2\epsilon}(-\not p+4m_0)+\frac{e_0^2}{16\pi^2}\left[\not p(1+\gamma)-2m_0(1+2\gamma)\right. \\
&\quad\left.+2\int_0^1 dx\,[\not p(1-x)-2m_0]\log\left(\frac{m_0^2x-p^2x(1-x)}{4\pi\mu^2}\right)\right] \\
&= \frac{e_0^2}{8\pi^2\epsilon}(-\not p+4m_0)+\text{finite} \qquad (7.94)
\end{aligned}$$

(where γ is the Euler–Mascheroni constant).

Comparing this with the expansion for $\Sigma(\not{p})$ in Eq. (7.49), we now have a result for:

$$Z_2 = 1 - \frac{e_0^2}{8\pi^2\epsilon} \tag{7.95}$$

Next, we evaluate the vacuum polarization contribution to the photon propagator. Once again, we will use the Feynman parameter trick:

$$\begin{aligned}
\Pi_{\mu\nu} &= ie_0^2\mu^{4-d}\int\frac{d^dp}{(2\pi)^d}\left(\frac{\mathrm{Tr}\,[\gamma_\mu(\not{p}+m_0)\gamma_\nu(\not{p}-\not{k}+m_0)]}{(p^2-m_0^2)[(p^2-k^2)^2-m_0^2]}\right)\\
&= ie_0^2\mu^\epsilon\int_0^1 dx\int\frac{d^dq}{(2\pi)^d}\\
&\times\left(\frac{\mathrm{Tr}\,\{\gamma_\mu(\not{q}+\not{k}x+m_0)\gamma_\nu[\not{q}-\not{k}(1-x)+m_0]\}}{[q^2-m_0^2+k^2x(1-x)]^2}\right)\\
&= ie_0^2\mu^\epsilon f(d)\int_0^1 dx\int\frac{d^dq}{(2\pi)^d}\left(\frac{2q_\mu q_\nu}{[q^2-m_0^2+k^2x(1-x)]^2}\right.\\
&\left.-\frac{2x(1-x)(k_\mu k_\nu-g_{\mu\nu}k^2)}{[q^2-m_0^2+k^2x(1-x)]^2}-\frac{g_{\mu\nu}}{[q^2-m_0^2+k^2x(1-x)]}\right)
\end{aligned} \tag{7.96}$$

where $q \to q - kx$ and $f(d) = \mathrm{Tr}\, I$. The first and third terms on the right-hand side of the equation cancel, leaving us with only a logarithmic divergence. (Gauge invariance has thus reduced a quadratically divergent graph to only a logarithmically divergent one.) Extracting the pole as $d \to 4$, we find:

$$\begin{aligned}
\Pi_{\mu\nu} &= \frac{e_0^2}{2\pi^2}(k_\mu k_\nu - g_{\mu\nu}k^2)\\
&\times\left[\frac{1}{3\epsilon}-\frac{\gamma}{6}-\int_0^1 dx\,x(1-x)\log\left(\frac{m_0^2-k^2x(1-x)}{4\pi\mu^2}\right)+\cdots\right]\\
&= \frac{e_0^2}{6\pi^2\epsilon}(k_\mu k_\nu - g_{\mu\nu}k^2)+\cdots
\end{aligned} \tag{7.97}$$

Using Eq. (7.57), this means that we can write:

$$Z_3 = 1 - \frac{e_0^2}{6\pi^2\epsilon} \tag{7.98}$$

Finally, we must write down the dimensionally regularized vertex correction:

$$\begin{aligned}
\Lambda_\mu(p,q,p') &= -i(e\mu^{2-d/2})^2 \int \frac{d^dk}{(2\pi)^d} \frac{\gamma_\nu(\not{p}'-\not{k}+m_0)\gamma_\mu(\not{p}-\not{k}+m_0)\gamma^\nu}{k^2[(p-k)^2-m_0^2][(p'-k)^2-m_0^2]} \\
&= \frac{2ie_0^2\mu^\epsilon}{(2\pi)^d}\int_0^1 dx \int_0^{1-x} dy \int d^dk \\
&\quad \times \frac{\gamma_\nu(\not{p}'-\not{k}+m_0)\gamma_\mu(\not{p}-\not{k}+m_0)\gamma^\nu}{[k^2-m_0^2(x+y)-2k(px+p'y)+p^2x+p'^2y]^3} \\
&= \frac{2ie_0^2\mu^\epsilon}{(2\pi)^d}\int_0^1 dx \int_0^{1-x} dy \int d^dk \\
&\quad \times \frac{\gamma_\nu[\not{p}'(1-y)-\not{p}x-\not{k}+m_0]\gamma_\mu[\not{p}(1-x)-\not{p}'y-\not{k}+m_0]\gamma^\nu}{[k^2-m_0^2(x+y)+p^2x(1-x)+p'^2y(1-y)-2p\cdot p'xy]^3}
\end{aligned} \tag{7.99}$$

where we have made the substitution $k \to k - px - p'y$.

This expression, in turn, contains a divergent and a convergent part. The convergent part, as before, gives us the contribution to the anomalous magnetic moment of the electron.

From Eq. (7.61), the divergent part can be isolated as:

$$\begin{aligned}
\Lambda_\mu^{(1)} &= \frac{e_0^2\mu^\epsilon}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 dx \int_0^{1-x} dy \\
&\quad \times \frac{\gamma_\nu\gamma_\rho\gamma_\mu\gamma^\rho\gamma^\nu}{[m_0^2(x+y)+p^2x(1-x)+p'^2y(1-y)+2p\cdot p'xy]^{2-d/2}} \\
&= \frac{e_0^2}{8\pi^2\epsilon}\gamma_\mu + \cdots
\end{aligned} \tag{7.100}$$

This, in turn, means that we can write:

$$Z_1 = 1 - \frac{e_0^2}{8\pi^2\epsilon} \tag{7.101}$$

Notice that $Z_1 = Z_2$, as we expected from the Ward–Takahashi identity.

In summary, we have, to lowest order, the correction to the electron self-energy, the photon propagator, and the vertex function:

$$\begin{aligned}
\Sigma(\not{p}) &= \frac{e_0^2}{8\pi^2\epsilon}(-\not{p}+4m_0) + \text{finite} \\
\Pi_{\mu\nu} &= \frac{e_0^2}{6\pi^2\epsilon}(k_\mu k_\nu - k^2 g_{\mu\nu}) + \text{finite}
\end{aligned}$$

$$\Lambda_\mu^{(1)} = \frac{e_0^2}{8\pi^2\epsilon}\gamma_\mu + \text{finite} \tag{7.102}$$

These, in turn, give us the expression for the renormalization constants:

$$\begin{aligned} Z_1 &= Z_2 = 1 - \frac{e^2}{8\pi^2\epsilon} \\ Z_3 &= 1 - \frac{e^2}{6\pi^2\epsilon} \\ e_0 &= \frac{Z_1}{Z_2\sqrt{Z_3}}e \end{aligned} \tag{7.103}$$

Comparing this result with the Pauli–Villars regularization method as in Eq. (6.98), we find that, to lowest order, the divergences are identical if we make the following correspondence between the pole $1/\epsilon$ and the cutoff Λ:

$$\frac{1}{\epsilon} \leftrightarrow \frac{1}{2}\log\frac{\Lambda^2}{m^2} \tag{7.104}$$

Thus, to this order, our results are independent of the regularization procedure.

7.6 Ward–Takahashi Identities

When we generalize our discussion to all orders in perturbation theory, our work will be vastly simplified by a set of *Ward–Takahashi* identities,[18,19] which reduce the number of independent renormalization constants Z_i. When we generalize our discussion to include gauge theories in later chapters, we will see that Ward–Takahashi identities are an essential ingredient in proving that a theory is renormalizable.

Specifically, we will use the fact that:

$$\Lambda_\mu(p, p) = -\frac{\partial}{\partial p_\mu}\Sigma(\not{p}) \tag{7.105}$$

where we have written the vertex correction in terms of the electron self-energy correction.

To prove this and more complicated identities arising from gauge theories, the most convenient formalism is the path integral formalism. However, for our purposes, we can prove the simplest Ward–Takahashi identities from graphical methods. To prove this identity, we observe that:

$$-\frac{\partial}{\partial p_\mu}\frac{1}{\not{p} - m + i\epsilon} = \frac{1}{\not{p} - m + i\epsilon}\gamma_\mu\frac{1}{\not{p} - m + i\epsilon} \tag{7.106}$$

(To prove this matrix equation, we set up a difference equation and then take the limit for small δp_μ.) Now let us use the above identity on an arbitrary graph that appears within $\Sigma(\not{p})$. In general, we can write $\Sigma(\not{p})$ as follows to bring out its p dependence:

$$\Sigma(\not{p}) \sim \cdots \frac{1}{(\not{p}+\not{q}_i)-m} \cdots \tag{7.107}$$

where the ellipses may represent very complicated integrals that are not important to our discussion. Also, there may be a large number of propagators with p dependence within $\Sigma(\not{p})$, which we suppress for the moment. Now take the derivative with respect to p_μ, and we find:

$$-\frac{\partial}{\partial p_\mu}\Sigma(\not{p}) = \sum_i \cdots \frac{1}{\not{p}+\not{q}_i-m}\gamma_\mu \frac{1}{\not{p}+\not{q}_i-m} \cdots \tag{7.108}$$

where we must differentiate over all propagators that have a p dependence. The term on the right is therefore the sum over all terms that have electron propagators containing p dependence.

The right-hand side of the previous equation is now precisely the graph corresponding to a vertex correction connected to an external photon line with zero momentum. This in turn means that the right-hand side equals $\Gamma_\mu(p, p)$, as promised. The purpose of this rather simple exercise is to prove a relation between renormalization constants. Now let us insert all this into the Ward–Takahashi identity. Differentiating the full $\Sigma(\not{p})$ to all orders, we find:

$$\begin{aligned} -\frac{\partial}{\partial p_\mu}\Sigma(\not{p}) &= -\frac{\partial}{\partial p_\mu}[\Sigma(m)+(\not{p}-m)\Sigma'(m)+\tilde{\Sigma}(\not{p})] \\ &= -\gamma_\mu\Sigma'(m)-\partial_\mu\tilde{\Sigma}(\not{p}) \\ &= \Lambda_\mu(p, p) \\ &= \gamma_\mu(Z_1^{-1}-1)+Z_1^{-1}\tilde{\Lambda}_\mu(p, p) \end{aligned} \tag{7.109}$$

where we have used Eq. (7.61).

Comparing terms proportional to the Dirac matrix, we see that the term $(Z_1^{-1}-1)\gamma_\mu$ must be proportional to $-\Sigma'(m)$, which we calculated earlier to be equal to $-1+(1/Z_2)$ in Eq. (7.49).

Comparing these two coefficients of γ_μ, we easily find:

$$Z_1 = Z_2 \tag{7.110}$$

which is our desired result. For a complete proof of renormalization, however, we need a more powerful version of the Ward–Takahashi identity, rather than the

one defined for zero-momentum external photons. We will use the version:

$$(p' - p)_\mu \Gamma^\mu(p', p) = \left[S'^{-1}_F(p') - S'^{-1}_F(p) \right] \tag{7.111}$$

This is easily shown to lowest order, where $S_F^{-1} = \not{p} - m$. Then the Ward–Takahashi identity reduces to the simple expression $(p'-p)_\mu \gamma^\mu = (\not{p}' - m - \not{p} + m)$. We can also show that this expression, in the limit that $p \to p'$, reduces to the previous version of the Ward–Takahashi identity.

This can also be rewritten in terms of the one-particle irreducible graph Σ. Because $S'^{-1}_F = \not{p} - m - \Sigma(\not{p})$, we have:

$$(p' - p)_\mu \Lambda^\mu = -\left[\Sigma(p') - \Sigma(p) \right] \tag{7.112}$$

The proof of this identity is also performed by graphical methods. We use the fact that:

$$\frac{1}{\not{p}+\not{q} - m} \not{q} \frac{1}{\not{p} - m} = \frac{1}{\not{p} - m} - \frac{1}{\not{p}+\not{q} - m} \tag{7.113}$$

which uses the identity:

$$\frac{1}{A}(A - B)\frac{1}{B} = \frac{1}{B} - \frac{1}{A} \tag{7.114}$$

To prove the general form of the Ward–Takahashi identity, we simply use this identity everywhere along a fermion line or along a fermion loop. This procedure is schematically represented in Figure 7.4.

Consider a specific member of $\Sigma(p)_{\alpha\beta}$ and follow the fermion line that connects the index α with β. Then schematically, we have:

$$\Sigma(p_0) = \sum \int \cdots \int \cdots \left(\prod_{i=0}^{n} \not{a}_i \frac{1}{\not{p}_i - m} \right) \not{a}_n \tag{7.115}$$

where $p_n = p_0$, where we have omitted all momentum integrations, and where a_i denote the momentum-dependent photon lines that are connected to other fermion loops (which are temporarily dropped). The matrices are ordered sequentially. Also, the total momenta contained within the various a_i must sum to zero.

Let us do the same for Λ_μ. If we follow the fermion line connecting the two fermion ends of Λ_μ, we find:

$$\Lambda_\mu(p', p) = \sum \int \cdots \int \cdots \sum_{r=0}^{n-1} \left(\prod_{i=0}^{r} \not{a}_i \frac{1}{\not{p}_i - m} \right) \gamma_\mu \frac{1}{\not{p}_r + \not{q} - m}$$

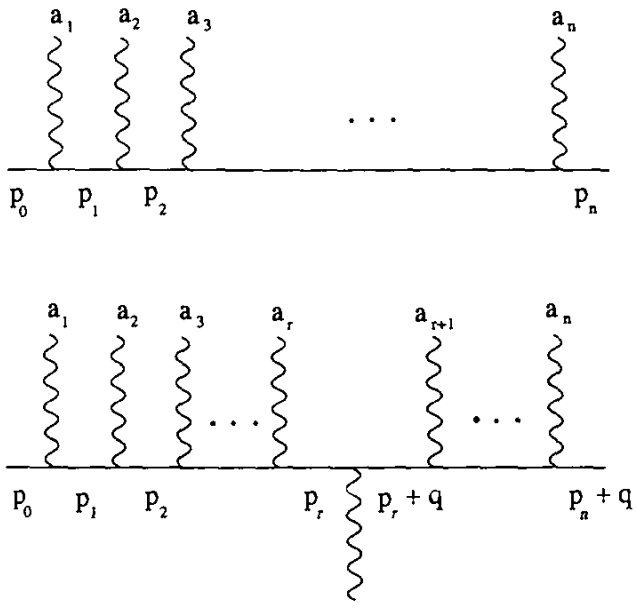


Figure 7.4. The Ward–Takahashi identity is proved by examining the insertion of a photon line along a fermion line contained within Σ and showing that we retrieve Λ_μ.

$$\times \left(\prod_{s=r+1}^{n-1} \not{a}_s \frac{1}{\not{p}_s + \not{q} - m} \right) \not{a}_n \tag{7.116}$$

The γ_μ is inserted at the rth point along the fermion line, and we sum over all possible values of r. In other words, we sum over all possible insertion points of γ_μ along the fermion line. The essential point in the proof is the observation that there is a graphical similarity between the vertex correction and the self-energy correction. Graphically speaking, if we take all the graphs within Σ and attach an extra photon line along all possible fermion propagators, then the resulting graphs are identical to the graphs within Λ_μ.

To show this rigorously, let us now contract Λ_μ with q^μ and use the identity shown previously. Then each term in the rth sum in Λ_μ splits into two parts. We now have $2r$ possible terms, which cancel pairwise because of the minus sign. The only terms that do not cancel are the first and last terms:

$$\begin{aligned} q^\mu \Lambda_\mu &= \sum \int \cdots \int \cdots \sum_{r=0}^{n-1} \left(\prod_{i=0}^{r-1} \not{a}_i \frac{1}{\not{p}_i - m} \right) \not{a}_r \left(\frac{1}{\not{p}_r - m} - \frac{1}{\not{p}_r + \not{q} - m} \right) \\ &\quad \times \left(\prod_{s=r+1}^{n-1} \not{a}_s \frac{1}{\not{p}_s + \not{q} - m} \right) \not{a}_n \\ &= \Sigma(p_0) - \Sigma(p_0 + q) \end{aligned} \tag{7.117}$$

where $p_0 = p$ and $p' = p_0 + q$.

The last step in the proof is to notice that we could have attached the photon leg of momentum q along any closed fermion loop as well. However, then the pairwise cancellation is exact, and the contribution of all closed loops is zero. This completes the proof of the Ward–Takahashi identity to all orders.

7.7 Overlapping Divergences

As we mentioned earlier, the renormalization program either cuts off the divergences with counterterms or absorbs them by multiplicative renormalization. In order to study how this is done systematically, we can use the method of *skeletons*. Draw a box around each divergent electron and photon self-energy graph and each vertex graph. Then we can replace the self-energy insertions with a line and the vertex insertions with a point. In this way, we obtain a *reduced graph*. Then we repeat this process, drawing boxes around the self-energy and vertex insertions in the reduced graph, and reduce it once more. Eventually, we obtain a graph that can no longer be reduced; that is, it is *irreducible*. The final graph after all these reductions is called a skeleton. An irreducible graph is one that is its own skeleton.

The advantage of introducing this concept is that the renormalization program is reduced to canceling the divergences within each box. For example, consider the complete vertex function Γ_μ summed to all orders. If we make the reduction of Γ_μ, we wind up with a sum over skeleton graphs, such that each line in the skeleton corresponds to the proper self-energy graph S'_F and $D'_{\mu\nu}$ and each vertex is the proper vertex. This can be summarized as:

$$\Gamma_\mu(p, p') = \gamma_\mu + \Lambda^S_\mu(S', D', \Gamma, e_0^2, p, p') \tag{7.118}$$

In this way, all complete vertex functions can be written entirely in terms of skeleton graphs over proper self-energy graphs and proper vertices.

This process of drawing boxes around the self-energy insertions and vertex insertions is unambiguous as long as the boxes are disjoint or nested; that is, the smaller boxes lie wholly within the larger one. Then the skeleton graph is unique. However, the skeleton reduction that we have described has an ambiguity that infests self-energy diagrams. For example, in Figure 7.5, we see examples of *overlapping diagrams* where the boxes overlap and the skeleton is not unique. One can show that these overlapping divergences occur for vertex insertions within self-energy parts. Unless these overlapping divergences appearing in self-energy graphs are handled correctly, we will overcount the number of graphs. These overlapping divergences are the only real difficulty in proving the renormalization of ϕ^4 and QED.

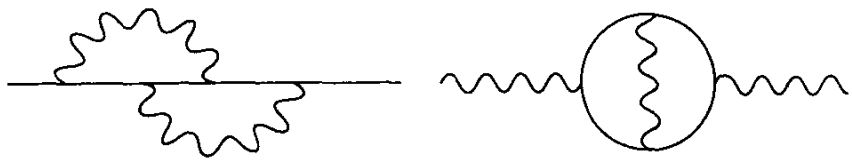


Figure 7.5. Examples of overlapping divergences found in QED.

There are several solutions to this delicate problem of overlapping divergences. The cleanest and most powerful is the BPHZ program, which we will present in Chapter 13. For QED, the most direct solution was originally given by Ward, who used the fact that, although $\Sigma(\not{p})$ has overlapping divergences, the vertex function Λ_μ does not. Therefore, by using the Ward–Takahashi identity:

$$\frac{\partial \Sigma(p)}{\partial p^\mu} = -\Lambda_\mu \tag{7.119}$$

we can reduce all calculations to functions in which overlapping divergences are absent.

Mathematically, taking the derivative with respect to p^μ is identical to inserting a zero photon line at every electron propagator, as we saw in Eq. (7.106). This means that every time an electron propagator appears in Σ, we replace it with two electron propagators sandwiching a zero momentum photon insertion matrix γ^μ. (However, we stress that there are still some serious, unresolved questions concerning overlapping divergences that emerge in 14th order diagrams, which we will discuss later.)

To see how this actually works, let us take the derivative of an overlapping divergence appearing in Σ. In Figure 7.6, we see the effect that $\partial/\partial p_\mu$ has on the overlapping divergence.

A single overlapping divergence has now split up into three pieces, each of which has a well-defined skeleton decomposition. Therefore, the Ward–Takahashi identity has helped to reduce the overlapping divergences within Σ to the more

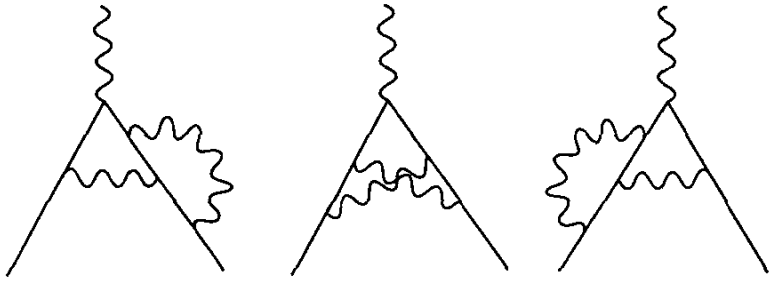


Figure 7.6. The effect that taking the partial derivative with respect to p_μ has on the overlapping divergence.

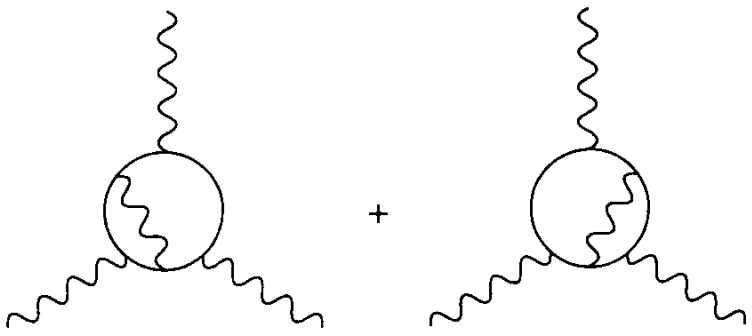


Figure 7.7. Taking the derivative of the photon self-energy graph with respect to the momentum yields these diagrams.

manageable problem of calculating the skeleton decomposition of the vertex Λ_μ. As long as we reduce all electron self-energy ambiguities to vertex insertions via the Ward–Takahashi identity, the overlapping divergence problem is apparently solved.

For photon self-energy graphs, there is also an overlapping divergence problem, which is also treated in much the same way. Although there is no Ward–Takahashi identity for the photon self-energy graph, we can solve the problem by introducing a new function W_μ, which is defined by:

$$W_\mu(k) \equiv \frac{\partial}{\partial k^\mu}\left[iD'(k^2)\right]^{-1} \tag{7.120}$$

where:

$$D'(k^2) = -\frac{1}{k^2\left(1 - e_0^2\Pi(k^2)\right)} \tag{7.121}$$

More explicitly, we can write:

$$\begin{aligned} W_\mu(k) &= 2ik_\mu + ik_\mu T(k^2) \\ T(k^2) &= \frac{e_0^2}{k^2}k^\mu \frac{\partial}{\partial k^\mu}\left[k^2\Pi(k^2)\right] \end{aligned} \tag{7.122}$$

The whole point of introducing this function $W_\mu(k)$ is that it has no overlapping divergences. Once again, the net effect of the partial derivative is to convert the photon self-energy graph into a vertex graph for T, which has no overlapping divergences. To see how this works, in Figure 7.7 we have performed the differentiation of an overlapping photon self-energy graph and created a vertex graph that has no ambiguous skeleton decomposition.

In summary, as long as we properly replace all electron and photon self-energy graphs with their vertex counterparts, there is usually no need to worry about overlapping divergences.

7.8 Renormalization of QED

Over the years, a large number of renormalization programs have been developed, with various degrees of rigor, and each with their own advantages and disadvantages. Since the reader will encounter one or more of these renormalization programs in his or her research, it is important that the reader be familiar with a few of these approaches. In this book, we will present three renormalization proofs:

1. The original Dyson/Ward proof, (which apparently breaks down at the fourteenth order, due to ovelapping divergences).
2. The BPHZ proof.
3. Proof based on the Callan–Symanzik equations.

We will first present the Dyson/Ward proof, which uses the auxiliary function W_μ to handle the overlapping divergence problem. In a later chapter, we will renormalize quantum field theory using more modern and sophisticated techniques, such as the renormalization group and the BPHZ program, which are more general than the proof that we will present here.

The Dyson/Ward proof of renormalization can be described in four steps:

1. First write down the complete set of coupled equations containing all the divergences of the theory for all self-energies and vertices. Everything is written in terms of skeleton graphs defined over the (divergent) complete propagators S'_F and $D'_{\mu\nu}$.
2. Subtract off the infinite divergences only in the vertex functions T and Λ_μ, which are free of overlapping divergences. Then define the new renormalized self-energy functions $\tilde{S}'_F$ and $\tilde{D}'_{\mu\nu}$ in terms of these subtracted vertex parts. This will enable us to write down an equivalent set of coupled equations for the finite set of self-energy functions free of overlapping divergences.
3. Rewrite the subtraction process as a multiplicative rescaling of the vertex and self-energy parts.
4. Absorb all multiplicative renormalizations into the coupling constant, masses, and wave functions.

7.8.1 Step One

The proof begins by writing down the expression for the unrenormalized self-energy graphs and vertices, summed to all orders, in terms of their skeletons. It

will prove useful to summarize the notation that we will use:

$$\begin{aligned}
\Sigma(p) &\rightarrow \text{proper } e \text{ self-energy graph} \\
\Pi_{\mu\nu} &\rightarrow \text{proper } \gamma \text{ self-energy graph} \\
\Gamma_\mu &\rightarrow \text{proper vertex (unrenormalized)} \\
S'_F &\rightarrow \text{complete } e \text{ self-energy graph} \\
D'_{\mu\nu} &\rightarrow \text{complete } \gamma \text{ self-energy graph} \\
\tilde{S}'_F &\rightarrow \text{renormalized } e \text{ self-energy graph} = Z_1^{-1} S'_F \\
\tilde{D}'_{\mu\nu} &\rightarrow \text{renormalized } \gamma \text{ self-energy graph} = Z_3^{-1} D'_{\mu\nu} \\
\tilde{\Gamma}_\mu &\rightarrow \text{renormalized vertex} = Z_1 \Gamma_\mu
\end{aligned} \tag{7.123}$$

Then the unrenormalized self-energy graphs and vertices satisfy:

$$\begin{aligned}
\Gamma_\mu(p, p') &= \gamma_\mu + \Lambda_\mu^S(S', D', \Gamma, e_0^2, p, p') \\
W_\mu(k^2) &= 2ik_\mu + ik_\mu \Gamma^S(S', D', \Gamma, W, e_0^2, k^2) \\
S'(p)^{-1} &= S'(p_0)^{-1} + (p - p_0)^\mu \Gamma_\mu(p, p_0) \\
D'(k^2)^{-1} &= \int_0^1 dx\, k^\mu W_\mu(xk)
\end{aligned} \tag{7.124}$$

where the superscript S denotes skeleton graphs, the prime denotes the complete propagator, which is the sum over divergent one-particle irreducible graphs.

The first two equations are really definitions, telling us that the vertex graphs Γ_μ and W_μ can be written in terms of skeleton graphs. (Because they have no overlapping divergences, this can always be done.) The third equation is the Ward–Takahashi identity, and the last equation is an integrated version of the definition of W_μ. Since these equations are all divergent, we must make the transition to the renormalized quantities.

7.8.2 *Step Two*

To find the convergent set of equations, we want to perform a subtraction only on those functions that have no overlapping divergences; that is, we perform the subtraction on T and Λ. In this way, we remove the possibility of overcounting. The subtracted functions are denoted by a tilde:

$$\begin{aligned}
\tilde{\Lambda}_\mu(p, p') &\equiv \Lambda_\mu^S(p, p') - \Lambda_\mu^S(p_0, p_0)\big|_{\not{p}_0 = \mu} \\
\tilde{T}(k^2) &\equiv T^S(k^2) - T^S(\mu^2)
\end{aligned} \tag{7.125}$$

Only one subtraction is necessary since the diagrams are only logarithmically divergent. (As before, we stress that there is an infinite degree of freedom in choosing μ, the subtraction point. Whatever value of μ we choose, we demand that the physics be independent of the choice.)

We note that $\Lambda_\mu^S(p_0, p_0)$ can be further reduced. Since it is defined for $\not{p}_0 = \mu$, it can only be a function of γ_μ. Thus, we can also write:

$$\Lambda_\mu^S(p_0, p_0)\big|_{\not{p}_0=\mu} = L\gamma_\mu \tag{7.126}$$

where L is divergent.

The important point is that we have only performed the subtractions on the quantities that have no overlapping divergences and no overcounting ambiguities, that is, the vertex functions T and Λ. This, in turn, allows us to *define* the finite self-energy parts (with a tilde) via the following equations:

$$\begin{aligned}
\tilde{\Gamma}_\mu(p, p') &\equiv \gamma_\mu + \tilde{\Lambda}_\mu(\tilde{S}, \tilde{D}, \tilde{\Gamma}, e^2, p, p') \\
\tilde{W}_\mu(k) &\equiv 2ik_\mu + ik_\mu \tilde{T}(\tilde{S}, \tilde{D}, \tilde{\Gamma}, \tilde{W}, e^2, k^2) \\
\tilde{S}(p)^{-1} &\equiv \tilde{S}(p_0)^{-1} + (p - p_0)^\mu \tilde{\Gamma}_\mu(p, p_0) \\
\tilde{D}(k^2) &\equiv \int_0^1 dx\, k^\mu W_\mu(xk)
\end{aligned} \tag{7.127}$$

The advantage of these definitions is that everything is now defined in terms of $\tilde{T}$ and $\tilde{\Lambda}$, which have no overlapping divergences. However, the quantities $\tilde{\Gamma}_\mu$, $\tilde{W}_\mu$ were simply defined by the previous equations. We still have no indication that these subtracted functions have any relationship to the actual renormalized self-energy and vertex parts.

7.8.3 *Step Three*

Now comes the important step. Up to now, we have made many rather arbitrary definitions that, as yet, have no physical content. We must now show that these quantities with the tilde are, indeed, the renormalized quantities that we want.

To see how this emerges, let us focus on the vertex graph. In terms of the subtraction, we can write:

$$\begin{aligned}
\tilde{\Gamma}_\mu &\equiv \gamma_\mu + \tilde{\Lambda}_\mu \\
&= \gamma_\mu + \Lambda_\mu^S - L\gamma_\mu \\
&= (1 - L)\left(\gamma_\mu + \frac{1}{1 - L}\Lambda_\mu^S\right)
\end{aligned}$$

$$= Z_1 \left(\gamma_\mu + \frac{1}{Z_1} \Lambda^S_\mu \right) \tag{7.128}$$

where we have defined:

$$Z_1 = 1 - L \tag{7.129}$$

Although we have factored out the renormalization constant Z_1 from this equation, the right-hand side is still not in the correct form. We want the right-hand side to be written in terms of the unrenormalized quantities, not the renormalized ones.

To find the scaling properties, it is useful to write the vertex functions symbolically in terms of propagators S and D and vertices γ_μ. Symbolically, by deleting integrals, traces, etc., the vertices can be written as products of propagators and vertices:

$$\begin{aligned} \Lambda^S_{2n} &\sim e_0^{2n} S^{2n} (D)^n (\gamma_\mu)^{2n+1} \\ T^S_{2n} &\sim e_0^{2n} S^{2n+\sigma} (D)^{n-\sigma} (\gamma_\mu)^{2n+\sigma} (2ik_\mu)^{1-\sigma} \end{aligned} \tag{7.130}$$

where σ is an integer that increases by one for every differentiation of an electron line.

Given this symbolic decomposition, we want to study their behavior under a scaling given by:

$$\begin{aligned} \gamma_\mu &\rightarrow a\gamma_\mu \\ D_{\mu\nu} &\rightarrow bD_{\mu\nu} \\ S &\rightarrow a^{-1} S \end{aligned} \tag{7.131}$$

(where a and b are proportional to renormalization constants).

Then it is easy to show that the vertex functions scale as:

$$\begin{aligned} a\Lambda^S_\mu(S, D, \gamma_\mu, e_0^2, p, p') &= \Lambda^S_\mu(a^{-1}S, bD, a\gamma_\mu, b^{-1}e_0^2, p, p) \\ b^{-1}T^S(S, D, \gamma_\mu, 2ik_\mu, e_0^2, k^2) &= T^S(a^{-1}S, bD, a\Gamma_\mu, b^{-1}2ik_\mu, b^{-1}e_0^2, k^2) \end{aligned} \tag{7.132}$$

With these rescaling relations, we can now absorb the factor Z_1 back into the renormalized vertex function Λ^S_μ and convert it into an unrenormalized one. Let $a = Z_1^{-1}$ and $b = Z_3$. We then have:

$$\tilde{\Gamma}_\mu = Z_1 \left(\gamma_\mu + \frac{1}{Z_1} \Gamma^S_\mu(\tilde{S}, \tilde{D}, \tilde{\Gamma}, e^2, p, p') \right)$$

$$\begin{aligned} &= Z_1\left[\gamma_\mu + \Lambda^S_\mu(Z_1\tilde{S}, Z_3\tilde{D}, Z_1^{-1}\tilde{\Gamma}, Z_3^{-1}e^2, p, p')\right] \\ &= Z_1\left[\gamma_\mu + \Lambda^S_\mu(S', D', \Gamma, e_0^2, p, p')\right] \\ &= Z_1\Gamma_\mu \end{aligned} \tag{7.133}$$

where we have used $Z_1 = Z_2$. This is the result that we wanted. We have now shown that the subtracted quantity $\tilde{\Gamma}_\mu$, after a rescaling by Z_1 and Z_3, can be written multiplicatively in terms of the unrenormalized quantity Γ_μ. This justified the original definition of $\tilde{\Gamma}_\mu$ that we introduced earlier.

7.8.4 *Step Four*

Now that we have renormalized the vertex, the rest is now easy. The vertex $\tilde{W}_\mu$ can now be renormalized in the same way. With $a = Z_1^{-1}$ and $b = Z_3$, we have:

$$\begin{aligned} \tilde{W}_\mu(k) &= 2ik_\mu + ik_\mu\left(T^S(k^2) - T^S(\mu^2)\right) \\ &= \left(1 - \frac{1}{2}T^S(\mu^2)\right)\left(2ik_\mu + \frac{1}{1 - \frac{1}{2}T^S(\mu^2)}ik_\mu T^S(k^2)\right) \\ &= Z_3\left(2ik_\mu + \frac{1}{Z_3}ik_\mu T^S(k^2)\right) \\ &= Z_3\left[2ik_\mu + ik_\mu T^S(Z_2\tilde{S}, Z_3\tilde{D}, Z^{-1}\tilde{\Gamma}_\mu, Z^{-1}\tilde{W}, Z_3^{-1}e^2, k^2)\right] \\ &= Z_3\left[2ik_\mu + ik_\mu T^S(S', D', \Gamma, W, e_0^2, k^2)\right] \end{aligned} \tag{7.134}$$

Thus, we have the other renormalized relation:

$$\tilde{W}_\mu = Z_3 W_\mu \tag{7.135}$$

From the renormalization of these vertex functions, it is easy to renormalize the self-energy terms as well, since everything is multiplicative. We now have the following relations that show the link between renormalized and unrenormalized quantities:

$$\begin{aligned} \Gamma_\mu &= Z_1^{-1}\tilde{\Gamma}_\mu & W_\mu &= Z_3^{-1}\tilde{W}_\mu \\ S' &= Z_2\tilde{S} & D'_{\mu\nu} &= Z_3 D_{\mu\nu} \\ e_0^2 &= Z_3^{-1}e^2 & Z_1 &= Z_2 \end{aligned} \tag{7.136}$$

In summary, we have proved that a subtraction of the divergences can be re-expressed in terms of a multiplicative rescaling of the vertex and self-energy

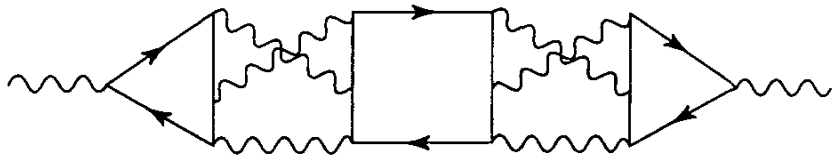


Figure 7.8. This fourteenth-order photon self-energy Feynman diagram was shown by Yang and Mills to suffer from a serious overlapping divergence problem, because of ambiguities introduced when we take the momentum derivative. This graph apparently invalidates the original Dyson/Ward renormalization proof.

parts. This means that all divergences are multiplicative and can be absorbed into a renormalization of the coupling constants, masses, and wave functions.

Although this was thought to be the first complete proof of the renormalizability of QED, this proof may be questioned in terms of its rigor. For example, it was shown by Yang and Mills[20,22] that there is an overlapping ambiguity at the fourteenth order in QED, thereby ruining this proof. For the electron self-energy, the Ward–Takahashi identity solves the problem of overlapping divergences, but for the photon self-energy, they showed that the operation of taking a momentum derivative is ambiguous at that level, thereby invalidating the proof (Fig. 7.8). (They also showed how it might be possible to remedy this problem, but did not complete this step.)

Second, another criticism of this proof is that we necessarily had to manipulate functions that were sums of an infinite number of graphs. Although each graph may be finite, the sum certainly is not, because QED certainly diverges when we sum over all orders; that is, it is an asymptotic theory, not a convergent one. More specifically, what we want is a theory based on an induction process, such that at any finite order, all functions are manifestly finite. We need a functional equation that allows one to calculate all self-energy and vertex parts at the n + 1st level when we are given these functions at the nth level.

Since there are many functional equations that link the nth-order functions to the n + 1st-order functions, there are also many inductive schemes that can renormalize field theory.[23,24] One of the most useful of these schemes is the inductive process using the BPHZ and renormalization group equations (which will be discussed in further detail in Chapters 13 and 14).

In summary, we have seen that renormalization theory gives us a solution to the ultraviolet divergence problem in quantum field theory. The renormalization program proceeded in several steps. First, by power counting, we isolated the divergences of all graphs, which must be a simple function of the number of external lines. Second, we regulated these divergences via cutoff or dimensional regularization. Third, we showed that these divergences can be canceled either by adding counterterms to the action, or by absorbing them into multiplicative renormalizations of the physical parameters. Finally, we showed that all divergences can be absorbed in this way.

The weakness of our proof, however, is that it handles overlapping divergences in an awkward (possibly incorrect) way, and that it is not general enough to handle different kinds of field theories. It remains to be seen if the overlapping divergence problem can be truly solved in this formalism. In Chapters 13 and 14, we will present the BPHZ and renormalization group proofs of renormalization, which are not plagued by overlapping divergences and are much more versatile.

This completes our discussion of QED. Next, in Part II we will discuss the Standard Model.

7.9 Exercises

1. Show that $Z_1 = Z_2$ to one-loop order for the electromagnetic field coupled to a triplet of π mesons, where Z_1 is the vertex renormalization constant, and Z_2 is for the π self-energy. (Hint: use the Ward–Takahashi identity.)

2. Do a power counting analysis of ϕ^p in q dimensions; isolate the graphs that are divergent. Confirm the statements made in the text concerning the renormalizability or super-renormalizability of the theory in various dimensions.

3. Draw all possible graphs necessary to prove the Ward–Takahashi identity for QED to order α^3.

4. Do a power counting of the massive Yukawa theory, with interaction term $\bar{\psi}\psi\phi$. Isolate all divergent graphs. Show that all divergences can be, in principle, moved into the physical parameters of the system. Outline the renormalization program.

5. Analyze the renormalization properties of a derivative coupling theory:

$$g\bar{\psi}\gamma^{\mu}\psi\partial_{\mu}\phi \tag{7.137}$$

   Is the S matrix equal to one? Is the theory trivial? Consider making a field redefinition on the ψ and $\bar{\psi}$. Calculate the self-energy correction to the fermion propagator and verify your conjecture.

6. Consider the one-loop electron self-energy diagram in QED. Let this electron also interact with an external scalar field via a derivative coupling, as in the previous problem. To first order in g, attach this derivative coupling term in all possible places along each electron propagator. There are three such graphs. Show that the sum of these three graphs is zero.

7. Repeat the previous problem, except consider an electron line with all possible photon lines attached to it to all orders in e. To first order in g, attach the

derivative coupling interaction along all electron propagators and show that this also sums to zero.

8. Consider the one-loop correction to the propagator in a massive ϕ^3 theory. Calculate the one-loop correction using both the Pauli–Villars method and the dimensional regularization method in order to find a relationship between ϵ and Λ/m.

9. Prove Eq. (7.91). Fill in the missing steps in its derivation that were omitted in the text.

10. Fill in the missing steps in Eqs. (7.93), (7.94), (7.96), and (7.97).

11. Let $K_{\alpha\beta\gamma\delta}$ equal the complete four-electron Green's function, where the Greek letters label the Dirac spinor indices. From this, we can construct what are called the Schwinger–Dyson equations for the electron vertex, with electron momenta p and p':

$$\begin{aligned}\tilde{\Gamma}_\mu(p', p)_{\gamma\delta} &= Z_1(\gamma_\mu)_{\gamma\delta} - \int \frac{d^4q}{(2\pi)^4}\big[\tilde{S}'_F(p'+q)\tilde{\Gamma}_\mu(p'+q, p+q) \\ &\quad \times \tilde{S}'_F(p+q)\big]_{\beta\alpha}\tilde{K}_{\alpha\beta,\delta\gamma}(p+q, p'+q, q) \qquad (7.138)\end{aligned}$$

and for the photon propagator:

$$\Pi_{\mu\nu}(q) = iZ_1 \int \frac{d^4k}{(2\pi)^4}\mathrm{Tr}\left[\gamma_\mu \tilde{S}'_F(k)\tilde{\Gamma}_\nu(k, k+q)\tilde{S}'_F(k+q)\right] \qquad (7.139)$$

Graphically, write down what these recursion relations look like. Then show that they are graphically correct to two-loop order.

12. Show that K does not suffer from overlapping divergences, which means that the Schwinger–Dyson equations (instead of the Ward–Takahashi identities) may be used to renormalize QED to all orders.[23,24]

Part II

Gauge Theory and the Standard Model

Chapter 8
Path Integrals

One feels as Cavalieri must have felt calculating the volume of a pyramid before the invention of calculus.

—R. Feynman

8.1 Postulates of Quantum Mechanics

Previously, we outlined how to quantize field theories with various spins using the canonical quantization approach. However, for increasingly complex systems, such as gauge theory, quantum gravity, and superstring theory, canonical quantization proves to be a very clumsy formalism since manifest Lorentz invariance is broken. Instead, we will explore a new method in this chapter.

Perhaps the most powerful quantization method is the *path integral* approach, which was developed by Feynman,[1,2] based on an idea of Dirac.[3] The path integral method is versatile enough to handle a variety of different types of gauge theories. The path integral approach has many advantages over the other techniques:

1. The path integral formalism yields a simple, covariant quantization of complicated systems with constraints, such as gauge theories. While calculations with the canonical approach are often prohibitively tedious, the path integral approach yields the results rather simply, vastly reducing the amount of work.

2. The path integral formalism allows one to go easily back and forth between the other formalisms, such as the canonical or the various covariant approaches. In the path integral approach, these various formalisms are nothing but different choices of gauge.

3. The path integral formalism is based intuitively on the fundamental principles of quantum mechanics. Quantization prescriptions, which may seem rather arbitrary in the operator formalism, have a simple physical interpretation in the path integral formalism.

4. The path integral formalism can be used to calculate nonperturbative as well as perturbative results.
5. The path integral formalism is based on c-number fields, rather than q-number operators. Hence, the formalism is much easier to manipulate.
6. At present, there are a few complex systems with constraints that can only be quantized in the path integral formalism.
7. Renormalization theory is much easier to express in terms of path integrals.

Our discussion of the path integral formalism begins with two deceptively simple principles:

8.1.1 *Postulate I*

The probability $P(b, a)$ of a particle moving from point a to point b is the square of the absolute value of a complex number, the transition function $K(b, a)$:

$$P(b, a) = |K(b, a)|^2 \tag{8.1}$$

8.1.2 *Postulate II*

The transition function $K(b, a)$ is given by the sum of a phase factor $e^{iS/\hbar}$, where S is the action, taken over all possible paths from a to b:

$$K(b, a) = \sum_{\text{paths}} k e^{iS/\hbar} \tag{8.2}$$

where the constant k can be determined by:

$$K(c, a) = \sum_{\text{paths}} K(c, b) K(b, a) \tag{8.3}$$

where we sum over all intermediate points b connecting a and c.

These postulates incorporate the essence of the celebrated double slit experiment, where a beam of electrons passes through a barrier with two small holes. A screen is placed behind the barrier to detect the presence of the electrons. As a point particle, an electron cannot, of course, go through both holes simultaneously. Classically, therefore, we expect that the electrons will go through one slit or the other, leaving two distinct marks on the screen just behind the two holes.

However, experiments show that the pattern created on the screen by repeated passages of the electrons through these holes is an interference pattern, associated with wave-like, not particle-like, behavior. Classically, we are therefore left with

a paradox. A point particle cannot go through both holes at once, yet the passage of a large number of electrons successively going past the barrier clearly leaves an interference pattern, with minima and maxima, as if the electron somehow went through both holes.

In the path integral approach, as in quantum mechanics, this puzzle can be resolved. The postulates of the path integral approach and quantum mechanics do not allow us to calculate the precise motion of a single point particle. They only allow us to calculate probability amplitudes. The probability that an electron will go from the source past the slits to the screen is given by summing over *all* possible paths. These probabilities, in turn, may have wave-like behavior, even if the electron itself is a point particle.

The sum over paths reproduces the interference pattern that is experimentally seen on the screen. Thus, the path integral approach incorporates the philosophy behind the double-slit experiment, which, in turn, embodies the essence of the quantum principle.

As in quantum mechanics, we make the transition to classical mechanics by taking the limit $\hbar \to 0$. For large values of S, the exponential of $iS/\hbar$ undergoes large fluctuations, and hence cancels out to zero. Hence, the contribution of the paths that maximize the action S do not contribute much to the sum over paths:

$$\delta S \gg \hbar : \quad \sum_{\text{paths}} e^{iS/\hbar} \sim 0 \tag{8.4}$$

In the classical limit, the paths that dominate the sum are the ones where $\delta S/\hbar$ is as small as possible. However, the path for which δS is minimized is just the classical path:

$$\delta S = 0 \quad \to \quad \text{classical mechanics} \tag{8.5}$$

Thus, we recover classical mechanics in the limit as $\hbar \to 0$. The picture that emerges from the path integral approach is therefore intuitively identical to the principles of quantum mechanics. To calculate the probability that a particle at point a goes to a point b, one must sum over all possible paths connecting these two points, including the classical one. The path preferred by classical mechanics is the one that minimizes the action for $\delta S \ll \hbar$ (Fig. 8.1).

Although the path integral method gives us an elegant formalism in which to reformulate all of quantum field theory, one should also point out the potential drawbacks of the formalism. One problem is that the path integral is not well defined in Minkowski space. In this chapter, we will assume that all path integrals are computed with the Euclidean metric. Then, the functional integral is taken over e^{-S}, which has much better convergence properties than integrals over e^{iS}. At the end of the calculation, we assume that we can analytically continue back

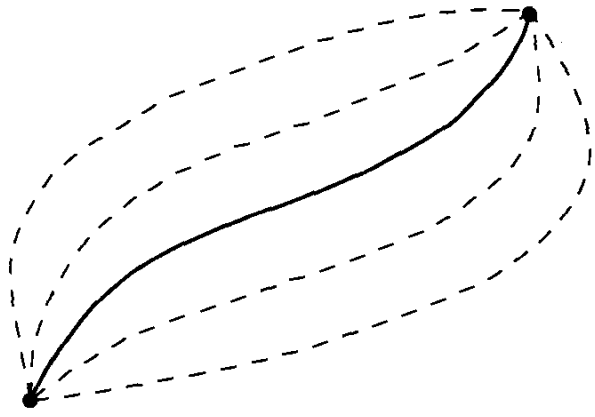


Figure 8.1. The path integral sums over all possible paths connecting two points, including the one favored by classical mechanics. In this way, the path integral sums over quantum corrections to classical mechanics.

to Minkowski space. (The question of whether this analytic continuation from Minkowski space to Euclidean space and back again is rigorously defined is a highly nontrivial question. This is a delicate matter, the subject of a field called axiomatic field theory, which is beyond the scope of this book.)

Another problem is that the transition between c numbers and operators becomes illdefined when the Hamiltonian has ordering problems. The path integral over a system with the Hamiltonian of the form $p^2 f(q)$, for example, becomes ambiguous when making the transition to the operator language, since p and q do not commute. For systems more complex than the harmonic oscillator, the integrals may not be Gaussian, and ordering problems may creep into the path integral. For complicated systems, one must often use "point splitting" methods, that is, separating two fields by a small infinitesimal amount in space-time in order to regularize the integrals. Unfortunately, a detailed elaboration of these delicate points is also beyond the scope of this book.

With these problems in mind, let us now compute with the path integral approach. We first divide up a path by discretizing space–time. Let us divide up each path in three-space into N points (Fig. 8.2). Then the "sum over all paths" can be transformed into a functional integral:

$$\sum_{\text{paths}} = \lim_{N\to\infty} \prod_{i=1}^{3} \prod_{n=1}^{N} \delta x_n^i \to \int Dx \tag{8.6}$$

The integral $\int Dx$ is not an ordinary integral. It is actually an infinite product of integrals, taken over all possible $dx(t)$. Whenever we use the differential symbol D, we should remember that it is actually an infinite product of differentials taken

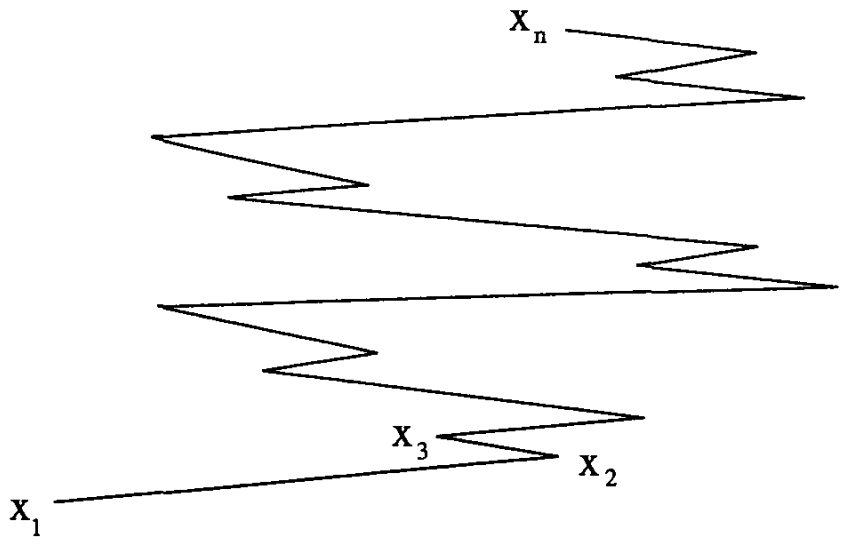


Figure 8.2. To calculate with path integrals, we break the path into a discrete number of intermediate points, and then integrate over the position of these intermediate points.

over all points. In this functional language, the transition function becomes:

$$K(b,a) = k \int_a^b Dx\, e^{iS/\hbar} \tag{8.7}$$

where k can be determined as follows:

$$K(c,a) = \int K(c,b)K(b,a)\, Dx_b \tag{8.8}$$

where we integrate over all possible intermediate points x_b which link points a and c.

To give this approach some substance, let us begin with the simplest of all possible classical systems, the free nonrelativistic point particle in the first quantized formalism. Our discussion begins with the classical action:

$$S = \int dt\, \frac{1}{2} m\dot{x}_i^2 \tag{8.9}$$

Let us now discretize the paths. We take dt to be a small interval ϵ and then discretize the Lagrangian:

$$\begin{aligned} dt &\rightarrow \epsilon \\ \frac{1}{2} m\dot{x}_i^2 dt &\rightarrow \frac{1}{2} m(x_n - x_{n+1})_i^2 \epsilon^{-1} \end{aligned} \tag{8.10}$$

The transition function $K(b, a)$ can be written as the path integral over e^{iS}:

$$\begin{aligned} K(b,a) &= \lim_{\epsilon \to 0} \int \int \cdots \int dx_2\, dx_3 \cdots dx_{N-1} \\ &\times k \exp\left(\frac{im}{2\epsilon} \sum_{n=1}^{N-1} (x_n - x_{n+1})^2\right) \end{aligned} \tag{8.11}$$

Unfortunately, one of the drawbacks of the path integral formalism is that embarrassingly few functional integrals can actually be performed. However, we will find that the simplest Gaussian path integral is also the one most frequently found for free systems. Specifically, we will repeatedly use the Gaussian integral:

$$\int_{-\infty}^{\infty} dx\, x^{2n} e^{-r^2 x^2} = \frac{\Gamma(n + \frac{1}{2})}{r^{2n+1}} \tag{8.12}$$

To evaluate the expression for $K(b, a)$, we now perform one of these Gaussian integrations:

$$\begin{aligned} &\int_{-\infty}^{\infty} dx_2 \exp\left[-a(x_1 - x_2)^2 - a(x_2 - x_3)^2\right] \\ &= \sqrt{\frac{\pi}{2a}} \exp\left[-\frac{1}{2} a (x_1 - x_3)^2\right] \end{aligned} \tag{8.13}$$

The key point is that the Gaussian integral over x_2 has left us with another Gaussian integral over the remaining variables. This process can be repeated an arbitrarily large number of times: Each time we perform a Gaussian integral on an intermediate point, we find a Gaussian integral among the remaining variables.

After repeated integrations, we find:

$$\begin{aligned} &\int_{-\infty}^{\infty} dx_2 \quad \cdots dx_{N-1} \exp\left[-a(x_1 - x_2)^2 - \cdots - a(x_{N-1} - x_N)^2\right] \\ &= \sqrt{\frac{\pi^{N-2}}{(N-1)a^{N-2}}} \exp\left[-\frac{a}{N-1}(x_1 - x_N)^2\right] \end{aligned} \tag{8.14}$$

This, in turn, allows us to calculate the constant k:

$$k = \left(\frac{2\pi i \epsilon}{m}\right)^{-(1/2)N} \tag{8.15}$$

If we take the limit as the number of intermediate points goes to infinity, then we are left with the final result for the transition function:

$$K(b,a) = \left|\frac{m}{2\pi(t_b - t_a)}\right|^{1/2} \exp\frac{(1/2)im(x_b - x_a)^2}{t_b - t_a} \tag{8.16}$$

This is a pleasant result. This is exactly the Green's function we derived in Eq. (3.58) for nonrelativistic quantum mechanics. Beginning with only the postulates of the path integral approach and the simplest possible classical action, we have derived the Green's function found in quantum mechanics that propagates Schrödinger waves. It obeys the equation:

$$-\frac{1}{2m}\frac{\partial^2}{\partial x_b^2}K(b,a) = i\frac{\partial}{\partial t_b}K(b,a) \tag{8.17}$$

Our first exercise in the path integral formalism gave us encouraging results. Now let us tackle more general and more difficult problems, such as (1) the transition between the Lagrangian and the Hamiltonian approaches and (2) the transition from c-number expressions to operator q-number expressions. In the usual canonical approach, these two transitions appear rather ad hoc and counterintuitive.

In the path integral approach, the transition between the Lagrangian and Hamiltonian systems is easily performed by adding an infinite sequence of Gaussian integrations for the momentum p_i. For each infinitesimal integration, we use the fact that:

$$\int_{-\infty}^{\infty} dp\, e^{iap^2+ibp} = \sqrt{\frac{i\pi}{a}}e^{-ib^2/4a} \tag{8.18}$$

which can be proved by completing the square. If we let $a = -1/2m$ and $b = \dot{x}$ and integrate over an infinite number of these momenta, then we have:

$$\begin{aligned} K(b,a) &= \int_{x_a}^{x_b} Dx \exp i\int_{t_a}^{t_b} dt\left[\frac{1}{2}m(\dot{x}_i)^2 - V(x)\right] \\ &= \int_{x_a}^{x_b} Dx\, Dp \exp i\int_{t_a}^{t_b} dt\left(p\dot{x}_i - \frac{p_i^2}{2m} - V(x)\right) \end{aligned} \tag{8.19}$$

The Lagrangian appears on the first line, but the Hamiltonian, defined by $H(p,x) = p^2/2m + V(x)$, appears on the second line. By performing the functional integral over Dp, we can go back and forth between the Lagrangian and Hamiltonian formalisms. In the path integral formalism, the relationship between the Lagrangian and the Hamiltonian formalism is no mystery, but simply the byproduct of performing an additional functional integration over momentum.

(For clarity, because normalization factors, such as $1/2\pi$, appear repeatedly throughout our discussion, we have absorbed them into the definition of Dx and Dy. We will henceforth drop these trivial normalization factors, since they can can always be explicitly written out later.) Thus, in the path integral formalism, the difference between the two formalisms only lies in a Gaussian integration over momentum. The path integral formalism allows us to go between these formalisms with ease:

$$L = \frac{1}{2}m(\dot{x}_i)^2 - V(x) \leftrightarrow H = \frac{p_i^2}{2m} + V(x) \tag{8.20}$$

So far, everything has been defined in terms of c-number expressions. Operators, which are the basis of the canonical approach, do not enter into the picture at all. Now, let us make the second transition, this time from the path integral formalism to the operator formalism, to show that the operator formalism that we have patiently developed in Chapters 3 and 4 is nothing but a specific representation of the path integral.

We recall that in the canonical formalism, the starting point was the canonical equal-time commutation relation between fields $\phi(x)$ and their conjugates $\pi(x)$. Only later could we calculate the propagators and finally the S matrix. In the path integral formalism, the sequence is roughly the reverse. We begin with the S matrix as the starting point, and we later derive the operator formalism as a consequence.

To see how operators naturally emerge in a formalism defined entirely without operators, let us write the transition function between point x_1 at time t_1 to point x_N at time t_N in the Heisenberg representation. We will carefully divide the path into N intermediate points. In this formalism, the transition probability of a particle at point x_1 and time t_1 going to x_N and time t_N is given by the matrix element between eigenstates $|x, t\rangle$

The Heisenberg representation, we recall, is based on a complete set of position eigenstates $|x\rangle$ of the position operator $\hat{x}$, which is now treated as an operator with eigenvalue x:

$$\hat{x}|x\rangle = x|x\rangle \tag{8.21}$$

We also introduce eigenstates of the momentum operator p:

$$\begin{aligned} 1 &= \int |x\rangle\, dx\, \langle x| \\ 1 &= \int |p\rangle\, dp\, \langle p| \end{aligned} \tag{8.22}$$

such that they are normalized as follows:

$$\begin{aligned} \langle x|y\rangle &= \delta(x-y) \\ \langle p|x\rangle &= \frac{e^{ipx}}{\sqrt{2\pi}}; \quad \langle x|p\rangle = \frac{e^{-ipx}}{\sqrt{2\pi}} \end{aligned} \tag{8.23}$$

To check the consistency of this normalization, we perform the following manipulations:

$$\begin{aligned} \langle x|y\rangle &= \langle x|p\rangle \int dp \langle p|y\rangle \\ &= \int dp \frac{e^{-ipx}}{\sqrt{2\pi}} \frac{e^{ipy}}{\sqrt{2\pi}} \\ &= \int \frac{dp}{2\pi} e^{-ip(x-y)} \\ &= \delta(x-y) \end{aligned} \tag{8.24}$$

Our normalizations are thus consistent.

Our task is now to rewrite the functional integration over p at an intermediate point along the path in terms of an operator expression defined in the Heisenberg picture. We will use the fact that the transition element between two neighboring points can be written as:

$$\langle x_2|e^{-iH(t_2-t_1)}|x_1\rangle = \langle x_2, t_2|x_1, t_1\rangle \tag{8.25}$$

Let us take a specific value of $\dot{x} \sim (x_1 - x_2)\delta t$ and dp that appears within the functional integral and carefully rewrite the integral over dp and its integrand as follows:

$$\begin{aligned} \int \frac{dp}{2\pi} e^{i(p\dot{x}-H(x,p))\delta t} &= \int \frac{dp}{2\pi} e^{-iH(x,p)\delta t} e^{ip(x_1-x_2)} \\ &= e^{-iH(x,\partial_x)\delta t} e^{-ix_2 p} \int \frac{dp}{\sqrt{2\pi}} e^{ipx_1} \\ &= e^{-iH(x,\partial_x)\delta t} \langle x_2|p\rangle \int dp \, \langle p|x_1\rangle \\ &= e^{-iH(x,\partial_x)\delta t} \langle x_2|x_1\rangle \\ &= \langle x_2|e^{-iH(x,\partial_x)\delta t}|x_1\rangle \\ &= \langle x_2, t_2|x_1, t_1\rangle \end{aligned} \tag{8.26}$$

We have now made the transition between a Lagrangian defined in terms of x and $\dot{x}$ and a Hamiltonian defined in terms of x and its derivative ∂_x. The transition was made possible because the derivative of the exponential brings down a p:

$$\begin{aligned} \partial_x\, e^{ipx} &= ip\, e^{ipx} \\ e^{-iH(x,\partial_x)\delta t}\, e^{ipx} &= e^{-iH(x,p)\delta t}\, e^{ipx} \end{aligned} \tag{8.27}$$

In the path integral formalism, this is the origin of the transition between c numbers and q-number operators; that is, the insertion of intermediate states defined in p space allows us to replace the p variable with a ∂_x operator. Thus, we have made the transition between:

$$\begin{aligned} H(x,p) &\leftrightarrow H(x,\partial_x) \\ p &\leftrightarrow -i\frac{\delta}{\delta x} \end{aligned} \tag{8.28}$$

In summary, we have now shown that the path integral formalism can express the propagator $K(b,a)$ in three different ways, in the Lagrangian or Hamiltonian formalism, or in the operator formalism in the Heisenberg picture. This can be summarized by the following identity:

$$\begin{aligned} K(N,1) &= \langle x_N, t_N | x_1 t_1 \rangle \\ &= \langle x_N, t_N | x_{N-1}, t_{N-1} \rangle \int dx_{N-1}\, \langle x_{N-1}, t_{N-1}| \\ &\quad \ldots |x_2, t_2\rangle \int dx_2 \langle x_2, t_2 | x_1, t_1 \rangle \\ &= \int Dx\, \exp\left(i \int_{t_1}^{t_N} dt\, L(x,\dot{x}) \right) \\ &= \int Dx\, Dp\, \exp\left(i \int_{t_1}^{t_N} dt\, (p\dot{x} - H(x,p)) \right) \end{aligned} \tag{8.29}$$

Finally, let us reanalyze, from the point of view of path integrals, how the time ordering operator T enters into the propagator. Let us analyze the matrix element of an initial state $|x_i, t_i\rangle$ with a final state $\langle x_n, t_n|$, with the operators $x_j(t_j)$ and $x_k(t_k)$ sandwiched between them. We will assume that $t_j > t_k$. As before, we will take time slices and insert a series of complete intermediate states between the states at each slice:

$$\langle x_n, t_n | x(t_j) x(t_k) | x_1, t_1 \rangle = \langle x_n, t_n | x_{n-1}, t_{n-1} \rangle \int dx(t_{n-1}) \langle x_{n-1}, t_{n-1}|$$

$$\cdots |x_{j+1}, t_{j+1}\rangle \int dx(t_{j+1}) \langle x_{j+1}, t_{j+1}|x_j, t_j\rangle x(t_j)$$

$$\cdots |x_{k+1}, t_{k+1}\rangle \int dx(t_{k+1}) \langle x_{k+1}, t_{k+1}|x_k, t_k\rangle x(t_k)$$

$$\cdots |x_2, t_2\rangle \int dx(t_2) \langle x_2, t_2|x_1, t_1\rangle \tag{8.30}$$

Taking the limit as the number of time slices goes to infinity, we have:

$$\begin{aligned} \langle x_n, t_n|x(t_j)x(t_k)|x_1, t_1\rangle &= \int Dx\,Dp\; x(t_j)x(t_k) \\ &\quad \times \exp\left(i \int_{t_1}^{t_n} (p\dot{x} - H(p,x))\, dt\right) \\ &= \int Dx\; x(t_j)x(t_k) \exp i \left(\int_{t_1}^{t_n} L(x,\dot{x}) dt\right) \end{aligned} \tag{8.31}$$

Now, let us reverse the order of the times, such that $t_j < t_k$. In this case, the previous formula must be modified because we can no longer take time slices. Thus, whenever $t_j < t_k$, we cannot make the transition from operators to path integrals unless we reverse the ordering of the operators. In order for this formalism to make sense, we will always reverse the order of the operators, such that the later times always appear to the left, so that we can proceed with taking time slices. To enforce this condition, we must use the time ordered product in this case:

$$\begin{aligned} &\langle x_n, t_n|T\left[x(t_j)x(t_k)\right]|x_1, t_1\rangle \\ &\quad = \int Dx\,Dp\; x(t_j)x(t_k) \exp\left(i \int_{t_1}^{t_n} dt\, (p\dot{x} - H(p,x))\right) \end{aligned} \tag{8.32}$$

For a large number of insertions, we have obviously:

$$\begin{aligned} &\langle x_n, t_n|T\left[x(t_j)x(t_k)\cdots x(t_m)\right]|x_1, t_1\rangle \\ &\quad = \int Dx\,Dp\; x(t_j)x(t_k)\cdots x(t_m) \exp\left(i \int_{t_1}^{t_n} [p\cdot x - H(p,x)]\right) \end{aligned} \tag{8.33}$$

We emphasize that the left-hand side consists of operators, so the ordering of the times is important. However, on the right-hand side we have a c-number expression, where the ordering of the $x(t_i)$ makes no difference. The correspondence between operators and these c-number expressions in the path integral only holds

when we can make time slices, that is, when the operators are time ordered. From the path integral point of view, this is the origin of the time ordering in the matrix elements.

8.2 Derivation of the Schrödinger Equation

In a first quantized formalism, where the action is a function of x^i and not fields, the path integral formalism gives us an added bonus: It gives us a derivation of the Schrödinger equation. Usually, introductory courses in quantum mechanics begin by postulating the Schrödinger wave equation. Certain conventions, such as the quantization of x and p, seem rather arbitrary. Only later emerges the probabilistic interpretation. Here, we reverse this order: we begin with the probabilistic postulates of quantum mechanics and derive the Schrödinger wave equation as a consequence, thus giving a new physical interpretation to that equation.

In the path integral approach, the evolution of a state is given by the transition function $K(b, a)$. From a classical point of view, this can be viewed as the analogue of Huygen's principle, where the evolution of a wave can be determined by assuming that each point along a wave front emits a new wave front. The integration over all these infinitesimal wave fronts then gives us the overall evolution of the wave front. Mathematically, this is given by:

$$\psi(x_j, t_j) = \int_{-\infty}^{\infty} K(x_j, t_j; x_i, t_i)\psi(x_i, t_i)\, dx_i \tag{8.34}$$

Earlier, we derived, assuming only the Lagrangian $(1/2)mv_i^2$, an expression for the nonrelativistic transition function. Now let us calculate how wave fronts move with this transition function. The time evolution, from t to $t + \delta t$, is given by:

$$\psi(x, t + \epsilon) = \int_{-\infty}^{\infty} A^{-1} \exp\left(\frac{im(x - y)^2}{2\epsilon}\right) \psi(y, t)\, dy \tag{8.35}$$

where:

$$A = \left(\frac{2\pi i\epsilon}{m}\right)^{1/2} \tag{8.36}$$

To perform this integration, let dy be replaced by $d\eta$, where $\eta = y - x$:

$$\psi(x, t + \epsilon) = \int_{-\infty}^{\infty} A^{-1} e^{im\eta^2/2\epsilon} \psi(x + \eta, t)\, d\eta \tag{8.37}$$

Now Taylor expand the left-hand side in terms of t, and the right-hand side in terms of η:

$$\psi(x,t)+\epsilon\frac{\partial\psi}{\partial t} = \int_{-\infty}^{\infty} A^{-1}e^{im\eta^2/2\epsilon} \times\left(\psi(x,t)+\eta\frac{\partial\psi}{\partial x}+\frac{1}{2}\eta^2\frac{\partial^2\psi}{\partial x^2}+\cdots\right)d\eta \quad (8.38)$$

The integration over $d\eta$ is easily performed. The integration over the linear term in η vanishes because it is linear, and the integration over the higher terms vanish in the limit $\epsilon \to 0$. This gives us:

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m}\frac{\partial^2\psi}{\partial x^2} \quad (8.39)$$

This is the Schrödinger wave equation, as desired. It is straightforward to insert a potential into the path integral, in which case we derive the Schrödinger wave equation in a potential, which is the traditional starting point for quantum mechanics.

8.3 From First to Second Quantization

So far, we have only investigated the path integral formalism in the first-quantized formalism, reproducing known results. The reader may complain that the path integral formalism is an elaborate, powerful machinery that has only rederived simple results. However, when we make the transition to the second quantized formalism and eventually to gauge theory, we will find that the path integral approach is the preferred formalism for quantum field theory. We saw earlier that the integration over all intermediate points along a path was enforced by inserting the number "1" at each intermediate point:

$$1 = |x_i, t_i\rangle \int dx_i \langle x_i, t_i| \quad (8.40)$$

The transition to field theory is made by introducing yet another expression for the number "1," this time based on an integration over an infinite number of degrees of freedom. We will use the familiar Gaussian integration:

$$\delta_{ij} = \frac{2}{(\sqrt{\pi})^n}\int\prod_{l=1}^{n}dx_l\ (x_i\,x_j)\exp\left(-\sum_{k=1}^{n}(x_k)^2\right) \quad (8.41)$$

Now let us replace the variable x_i with a function $\psi(x)$, which is temporarily viewed as a discretized number ψ_x, where x is now seen as an infinite discrete index.

The transition from finite degrees of freedom to infinite degrees of freedom is then made by inserting the following expression for "1" into the path integral:

$$\delta_{x,y} \sim \int D\psi \; D\psi^* \; \psi_x^* \psi_y \exp - \left(\sum_z \psi_z^* \psi_z \right) \tag{8.42}$$

where:

$$D\psi = \prod_z d\psi_z \tag{8.43}$$

Written in terms of functions $\psi(x)$ rather than discretized variables ψ_z, we now have:

$$\delta(x - y) = \int D\psi \; D\psi^* \; \psi^*(x)\psi(y) \; \exp - \left(\int Dx |\psi(x)|^2 \right) \tag{8.44}$$

These expressions can also be rewritten in terms of bra and ket vectors as follows:

$$\begin{aligned} \psi(x) &= \langle x|\psi\rangle \\ \psi^*(x) &= \langle \psi|x\rangle \end{aligned} \tag{8.45}$$

This allows us to write:

$$\begin{aligned} \delta(x - y) &= \int D^2\psi \; \psi^*(x)\psi(y) \exp - \left(\int Dx \; \psi^*(z)\psi(z) \right) \\ &= \langle x|\psi\rangle \int D^2\psi \; \exp - \left(\int \langle \psi|z\rangle \; Dz \; \langle z|\psi\rangle \right) \langle \psi|y\rangle \\ &= \langle x|1|y\rangle \end{aligned} \tag{8.46}$$

Written in this language, the number "1" now becomes:

$$1 = |\psi\rangle \int d^2\psi \, e^{-\langle \psi|\psi\rangle} \langle \psi| \tag{8.47}$$

We can now repeat all the steps used in making the transition from the Lagrangian approach to the operator approach in the Heisenberg picture by inserting this new expression for the number "1" into the path integral. When we do this, we then

have an expression for the transition function written entirely in terms of $\psi(x)$. A straightforward insertion of this new set of intermediate states yields:

$$\begin{aligned} K(b,a) &= \int_{x_a}^{x_b} Dx\, e^{i\int dt\, \frac{1}{2}m\dot{x}_i^2} \\ &= \int D\psi\, D\psi^*\, \psi(x_a)^*\psi(x_b) e^{i\int dt\, \psi^*(i\partial_t - H)\psi} \end{aligned} \tag{8.48}$$

where the Lagrangian is equal to:

$$L = \psi^* \left(i\frac{\partial}{\partial t} - H \right) \psi \tag{8.49}$$

At this point, we have now derived a second quantized version of the nonrelativistic Schrödinger equation. This may seem odd, since usually quantum field theory is associated with the merger of relativity and quantum mechanics. But quantum field theory can be viewed independently from relativity; that is, the essence of quantum field theory is that it has an infinite number of quantum degrees of freedom. In this sense, the path integral formalism can accomodate a nonrelativistic Schrödinger field theory.

Next, we would like to compute the familiar expressions found in Chapter 3 and 4 in terms of the path integral approach. First, we define the average $\langle O \rangle$ of the expression O by inserting it into the integral:

$$\langle O \rangle \equiv N \int DX \exp\left(-\sum_{i,j=1}^{n} \frac{1}{2} x_i D_{ij} x_j \right) O \tag{8.50}$$

Our goal is to find an expression for $\langle O \rangle$, and later make the transition to an infinite number of degrees of freedom ($n \to \infty$). To find an expression for this average, we will find it convenient to introduce an intermediate stage in the calculation. We define the generating functional as follows:

$$I(D, J) \equiv \int \prod_{i=1}^{n} dx_i \exp\left(-\sum_{i,j=1}^{n} \frac{1}{2} x_i D_{ij} x_j + \sum_{i=1}^{n} J_i x_i \right) \tag{8.51}$$

where we fix N by setting $I(D, 0) = N^{-1}$.

To find an expression for $I(D, J)$, we first make a similarity transformation $x_i' = S_{ij} x_j$, such that S diagonalizes the matrix D. We are then left with the eigenvalues of the matrix D in the integral. The integration separates into a product of independent integrations over x_i'. We then perform each integration separately, giving us the square root of the eigenvalues of D matrix. The product

of the eigenvalues, however, is equal to the determinant of the D matrix, giving us the final result:

$$I(D, J) = (2\pi)^{n/2}(\det D_{ij})^{-1/2} \exp\left(\sum_{i,k}^{n} \frac{1}{2} J_i (D^{-1})_{ik} J_k\right) \tag{8.52}$$

Our goal is to evaluate the average of an arbitrary product $x_1 x_2 \cdots x_n$:

$$\langle x_1 x_2 \cdots x_n \rangle \equiv N \int \prod_{i=1}^{n} dx_i \, x_1 x_2 \cdots x_n \exp\left(-\sum_{i,j}^{n} \frac{1}{2} x_i D_{ij} x_j\right) \tag{8.53}$$

where the normalization constant N can be fixed via:

$$I(D, 0) = N^{-1} = (2\pi)^{n/2} \left(\det D_{ij}\right)^{-1/2} \tag{8.54}$$

We can also take repeated derivatives of $I(D, J)$ with respect to J, and then set J equal to zero. Each time we take the derivative $\partial/\partial J_i$, we bring down a factor of x_i into the integral:

$$\begin{aligned} \langle x_1 \cdots x_n \rangle &= \prod_{i=1}^{n} \frac{\partial}{\partial J_i} I(D, J)\Big|_{J=0} \\ &= \sum_{\text{pairings}} D^{-1}_{k_1 k_2} \cdots D^{-1}_{k_{n-1} k_n} \end{aligned} \tag{8.55}$$

This expression, for an odd number of x's, vanishes. However, for two x's, we have:

$$\langle x_i x_j \rangle = (D^{-1})_{ij} \tag{8.56}$$

For four x's, we have:

$$\begin{aligned} \langle x_i x_j x_k x_l \rangle &= \Big[(D^{-1})_{ij}(D^{-1})_{kl} \\ &\quad + (D^{-1})_{ik}(D^{-1})_{jl} + (D^{-1})_{il}(D^{-1})_{jk}\Big] \end{aligned} \tag{8.57}$$

Not surprisingly, if we analyze the way in which these indices are paired off, we see Wick's theorem beginning to emerge. The point here is that Wick's theorem, which was based on arguments concerning normal ordering of operators, is now emerging from entirely c-number integrations.

The transition to quantum field theory, as before, is now made by making the transition from a finite number of variables x_i to an infinite number of variables

$\phi(x)$. As before, we are interested in evaluating the transition probability between a field at point x and a field at point y:

$$\int D\phi\, \phi(x)\phi(y) \exp i \left(\int d^4x\, L(\phi) \right) \tag{8.58}$$

To evaluate this integral, we will find it convenient, as before, to introduce the generating functional:

$$Z(J) = N \int D\phi\, e^{i \int d^4x [L(x) + J(x)\phi(x)]} \tag{8.59}$$

where:

$$\begin{aligned} D\phi &\equiv \prod_x d\phi(x) \\ N^{-1} &\equiv \int D\phi\, e^{i \int d^4x\, L(x)} \end{aligned} \tag{8.60}$$

To perform this integration for a Klein–Gordon field, we will repeat the steps we used for the simpler theory based on finite number of degrees of freedom. We first make a shift of variables:

$$\phi(x) \rightarrow \phi(x) + \phi(x)_{\text{cl}} \tag{8.61}$$

where ϕ_{cl} satisfies the Klein–Gordon equation with a source term. We recall that the Feynman propagator is defined via:

$$(\partial_\mu \partial^\mu + m^2)_x \Delta_F(x - y) = -\delta^4(x - y) \tag{8.62}$$

A classical solution can then be defined via:

$$\phi_{\text{cl}} = -\int \Delta_F(x - y) J(y)\, d^4y \tag{8.63}$$

which satisfies:

$$(\partial_\mu \partial^\mu + m^2)\phi_{\text{cl}} = J(x) \tag{8.64}$$

We can now perform the integral by performing a Gaussian integration:

$$Z(J) = \exp\left(-\frac{i}{2} \int d^4x d^4y\, J(x) \Delta_F(x - y) J(y) \right) \tag{8.65}$$

where:

$$N^{-1} = \left[\det\left(\partial_\mu^2 + m^2\right)\right]^{1/2} = \int D\phi \exp i\left(\int d^4x L(\phi)\right) \tag{8.66}$$

Using the fact that:

$$\frac{\delta}{\delta J(x)} J(y) = \delta^4(x - y) \tag{8.67}$$

we find that the transition function is given by:

$$-i\Delta_F(x - y) = \frac{\delta}{\delta J(x)} \frac{\delta}{\delta J(y)} Z(J)\Big|_{J=0} \tag{8.68}$$

The average of several fields taken at points x_i is now given as follows:

$$\begin{aligned} \Delta(x_1, x_2, \cdots, x_n) &= i^n \langle 0|T\phi(x_1)\phi(x_2)\cdots\phi(x_n)|0\rangle \\ &= \frac{\delta^n Z(J)}{\delta J(x_1)\delta J(x_2)\cdots\delta J(x_n)}\Big|_{J=0} \end{aligned} \tag{8.69}$$

By explicit differentiation, we can take the derivatives for four fields and find:

$$\begin{aligned} -\left(\prod_{i=1}^{4} \frac{\delta}{\delta J(x_i)}\right) Z(J)\Big|_{J=0} &= \Delta_F(x_1 - x_2)\Delta_F(x_3 - x_4) \\ &+ \Delta_F(x_1 - x_3)\Delta_F(x_2 - x_4) + \Delta_F(x_1 - x_4)\Delta_F(x_2 - x_3) \end{aligned} \tag{8.70}$$

In this way, we have derived Wick's theorem starting with purely c-number expressions.

$Z(J)$ can also be written as a power expansion in J. If we power expand the generating functional, then we have:

$$Z(J) = \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int d^4x_1 \cdots d^4x_n J(x_1)\cdots J(x_n) Z^{(n)}(x_1 \cdots x_n) \tag{8.71}$$

where:

$$\begin{aligned} Z^{(n)}(x_1, \cdots, x_n) &= \frac{\delta^n}{\delta J(x_1)\cdots J(x_n)} Z(J)\Big|_{J=0} \\ &= i^n \langle 0|T\phi(x_1)\cdots\phi(x_n)|0\rangle \end{aligned} \tag{8.72}$$

In this way, the path integral method can derive all the expressions found earlier in the canonical formalism.

8.4 Generator of Connected Graphs

In analyzing complicated Feynman diagrams, we must distinguish between two types of graphs: connected and disconnected graphs. A graph is called disconnected when it can be separated into two or more distinct pieces without cutting any line.

The generating functional $Z(J)$ that we have been analyzing generates all types of Feynman graphs, both connected and disconnected. However, when we apply the formalism of path integrals to a variety of physical problems, including renormalization theory, it is often desirable to introduce a new functional that generates just the connected graphs. The path integral formalism is versatile enough to give us this new generating functional, which is denoted $W(J)$. We define this generator as follows:

$$\begin{aligned} Z(J) &= e^{iW(J)} \\ W(J) &= -i \log Z(J) \end{aligned} \tag{8.73}$$

If we take repeated derivatives of $W(J)$ to calculate the relationship between Z and W, we find:

$$\frac{\delta^2 W}{\delta J(x_1)\,\delta J(x_2)} = \frac{i}{Z^2}\frac{\delta Z}{\delta J(x_1)}\frac{\delta Z}{\delta J(x_2)} - \frac{i}{Z}\frac{\delta^2 Z}{\delta J(x_1)\,\delta J(x_2)} \tag{8.74}$$

and:

$$\begin{aligned} \frac{\delta^4 W}{\delta J(x_1)\,\delta J(x_2)\,\delta J(x_3)\,\delta J(x_4)} &= \left(\frac{i}{Z^2}\frac{\delta^2 Z}{\delta J(x_1)\,\delta J(x_2)}\frac{\delta^2 Z}{\delta J(x_3)\,\delta J(x_4)} + \text{perm.}\right) \\ &\quad - \frac{i}{Z}\frac{\delta^4 Z}{\delta J(x_1)\,\delta J(x_2)\,\delta J(x_3)\,\delta J(x_4)} \end{aligned} \tag{8.75}$$

To analyze the content of these equations, let us power expand $W(J)$ in powers of J:

$$W(J) = \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_1 \cdots dx_n J(x_1) \cdots J(x_n) W^{(n)}(x_1, \cdots, x_n) \tag{8.76}$$

Taking $J = 0$, we arrive at:

$$iW^{(2)}(x_1, x_2) = Z^{(2)}(x_1, x_2) \tag{8.77}$$

This is not surprising, since the propagator is connected. The expansion, however, becomes nontrivial when we consider expanding out to fourth order, where

disconnected graphs enter into the functional:

$$\begin{aligned} W^{(4)}(x_1, x_2, x_3, x_4) &= i\left[Z^{(2)}(x_1, x_2)Z^{(2)}(x_3, x_4) + \text{perm.}\right] \\ &- iZ^{(4)}(x_1, x_2, x_3, x_4) \end{aligned} \tag{8.78}$$

This equation can be checked to show that W generates only connected graphs. For example, in ϕ^4 theory to order λ, we can show that this works as indicated. In this case:

$$Z^{(2)}(x_1, x_2) = -i\Delta_F(x_1 - x_2) + \frac{\lambda}{2}\int d^4z\, \Delta_F(x_1 - z)\Delta_F(z - x_2)\Delta_F(z, z) \tag{8.79}$$

while:

$$\begin{aligned} Z^{(4)}(x_1, x_2, x_3, x_4) &= -[\Delta_F(x_1 - x_2)\Delta_F(x_3 - x_4) + 2\,\text{terms}] \\ &- \frac{i\lambda}{2}\left(\int d^4z \Delta_F(x_1 - z)\Delta_F(z, z)\Delta_F(z - x_2)\Delta_F(x_3 - x_4) + 5\ \text{terms}\right) \\ &- \frac{i\lambda}{4!}\left(\int d^4z \Delta_F(x_1 - z)\Delta_F(x_2 - z)\right. \\ &\left.\times\, \Delta_F(x_3 - z)\Delta_F(x_4 - z) + 23\ \text{terms}\right) \end{aligned} \tag{8.80}$$

Inserting these factors back into the identity for $W^{(4)}(x_1, x_2, x_3, x_4)$, we find that the disconnected pieces cancel, and the only term which survives is the connected piece, which forms the topology of a cross.

Next, we would like to find the generating functional for proper vertices Γ, which is essential in a discussion of renormalization theory. Proper vertices (or one-particle irreducible vertices), we recall, appear when we consider renormalizing coupling constants. We define $\Gamma(\phi)$, the generator of proper vertices, via a Legendre transformation as:

$$\Gamma(\phi) = W(J) - \int d^4x\, J(x)\phi(x) \tag{8.81}$$

(From now on, we will use the symbol ϕ to represent the c-number field.)

The fields ϕ and J have a nontrivial relationship between them. They are not independent of each other. In fact, by taking repeated derivatives, we can establish the relationship between them. Let us take the partial derivative of the previous equations with respect to J (keeping ϕ fixed) and with respect to ϕ, keeping J

fixed. Then by differentiating both sides, we have:

$$\frac{\delta W(J)}{\delta J(x)} = \phi(x); \quad \frac{\delta \Gamma(\phi)}{\delta \phi(x)} = -J(x) \tag{8.82}$$

Let us take repeated differentials of the above equations. Differentiating by J and by ϕ, we find:

$$\begin{aligned} G(x, y) &= -\frac{\delta^2 W}{\delta J(x)\, \delta J(y)} = -\frac{\delta \phi(x)}{\delta J(y)} \\ \Gamma(x, y) &= \frac{\delta^2 \Gamma}{\delta \phi(x)\, \delta \phi(y)} = -\frac{\delta J(x)}{\delta \phi(y)} \end{aligned} \tag{8.83}$$

If we treat $\Gamma(x, y)$ and $G(x, y)$ as matrices with continuous space–time indices, then they are inverses of each other, as can be seen as follows:

$$\begin{aligned} \int d^4y\, G(x, y)\Gamma(y, z) &= -\int d^4y \frac{\delta^2 W}{\delta J(x)\, \delta J(y)} \frac{\delta^2 \Gamma}{\delta \phi(y)\, \delta \phi(z)} \\ &= \int d^4y \frac{\delta \phi(x)}{\delta J(y)} \frac{\delta J(y)}{\delta \phi(z)} \\ &= \frac{\delta \phi(x)}{\delta \phi(z)} \\ &= \delta^4(x - z) \end{aligned} \tag{8.84}$$

We would now like to establish a relationship between third derivatives of the functionals. If we differentiate the previous equation by $J(u)$, we find that it vanishes. Thus, we find the relationship:

$$\begin{aligned} &\int d^4y \frac{\delta^3 W}{\delta J(x)\, \delta J(u)\, \delta J(y)} \frac{\delta^2 \Gamma}{\delta \phi(y)\, \delta \phi(z)} \\ &= \int d^4y \frac{\delta^2 W}{\delta J(x)\, \delta J(y)} \int d^4y'\, G(u, y') \frac{\delta^3 \Gamma}{\delta \phi(y)\, \delta \phi(z)\, \delta \phi(y')} \end{aligned} \tag{8.85}$$

where we have the fact that:

$$\frac{\delta}{\delta J(u)} = \int d^4y' \frac{\delta \phi(y')}{\delta J(u)} \frac{\delta}{\delta \phi(y')} = -\int d^4y'\, G(u, y') \frac{\delta}{\delta \phi(y')} \tag{8.86}$$

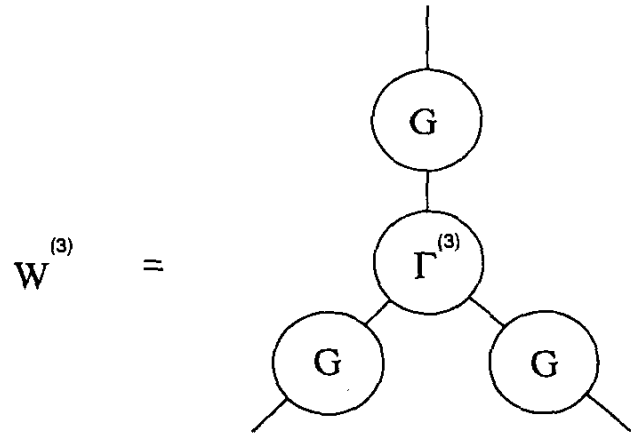


Figure 8.3. Graphic representation of the relationship between $W(J)$ and $\Gamma(\phi)$ out to third order.

We can simplify this equation a bit by inverting the matrix $\Gamma(y, z)$ that appears. Then we find:

$$\frac{\delta^3 W}{\delta J(x)\,\delta J(y)\,\delta J(z)} = -\int d^4x'\, d^4y'\, d^4z' \times G(x, x')G(y, y')G(z, z')\frac{\delta^3 \Gamma}{\delta\phi(x')\,\delta\phi(y')\,\delta\phi(z')} \tag{8.87}$$

Graphically, this is represented in Figure 8.3. In this way, we can derive relationships between the generating functionals. However, taking repeated derivatives becomes quite involved when we increase the number of legs. There is yet another, perhaps more direct way in which to see the relationship between the various generating functions by taking power expansions. As before, we can power expand as follows:

$$\Gamma(\phi) = \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_1 \cdots dx_n\, \phi(x_1) \cdots \phi(x_n)\Gamma^{(n)}(x_1, \cdots x_n) \tag{8.88}$$

We want a way to compare $\Gamma^{(n)}$ and $W^{(n)}$. To solve this problem, we would like to power expand ϕ in terms of J. By Taylor's theorem, we know that a power expansion of ϕ can be expressed in terms of powers of J, with coefficients given by the nth derivative of ϕ with respect to J. However, we already know from Eq. (8.83) that the derivative of ϕ with respect to J is given by the second derivative of W with respect to J. Thus, from Eq. (8.82), we find:

$$\phi(x) = \int d^4x_1\, W^{(2)}(x, x_1)J(x_1) + \frac{1}{2}\int d^4x_1\, d^4x_2\, W^{(3)}(x, x_1, x_2)J(x_1)J(x_2) + \cdots \tag{8.89}$$

We also know the converse, that we can power expand J in terms of ϕ. Likewise, we know from Taylor's theorem that the coefficient of each term is given by the nth derivative of J with respect to ϕ. Thus, Eqs. (8.82) and (8.87) give us:

$$\begin{aligned} J(x) &= -\int d^4x_1 \left[W^{(2)}\right]^{-1}(x, x_1)\phi(x_1) \\ &- \frac{1}{2!}\int d^4y_1 d^4y_2 d^4y_3 d^4x_1 d^4x_2 \left[W^{(2)}\right]^{-1}(x, y_3)\left[W^{(2)}\right]^{-1}(x_1, y_1) \\ &\times \left[W^{(2)}\right]^{-1}(x_2, y_2) W^{(3)}(y_1, y_2, y_3)\phi(x_1)\phi(x_2) + \cdots \end{aligned} \tag{8.90}$$

Let us introduce what is called the "amputated" functional:

$$\tilde{W}^{(n)}(x_1, \cdots, x_n) = \int \prod_{i=1}^{n} d^4y_i \, G^{-1}(x_i, y_i) W^{(n)}(y_1, \cdots y_n) \tag{8.91}$$

This has a simple meaning. Since the connected piece $W^{(n)}$ always has propagators connected to each external leg, this means that $\tilde{W}^{(n)}$ is just the connected part minus these external legs. That is the reason why it is often called the "amputated" Green's functions for connected graphs.

Now that we have solved for $\tilde{W}^{(n)}$, we can now solve for the relationship between $\tilde{W}$ and Γ. Equating terms with the same power of ϕ, we find:

$$\begin{aligned} W^{(2)}(x_1, x_2) &= -\left[\Gamma^{(2)}(x_1, x_2)\right]^{-1} \\ \tilde{W}^{(3)}(x_1, x_2, x_3) &= \Gamma^{(3)}(x_1, x_2, x_3) \\ \tilde{W}^{(4)}(x_1, x_2, x_3, x_4) &= \Gamma^{(4)}(x_1, x_2, x_3, x_4) \\ &\quad + \int d^4y d^4z \, \Gamma^{(3)}(x_1, x_2, y) W^{(2)}(y, z) \\ &\quad \times \Gamma^{(3)}(z, x_3, x_4) + 2 \text{ terms} \end{aligned} \tag{8.92}$$

Thus, to any desired level of expansion, we can find the relationship between the proper vertices and the amputated Green's function for connected graphs. The advantage of this path integral approach is that our results are independent of perturbation theory. Without having to use complicated graphical techniques, we can rapidly prove nontrivial relations between different types of vertices and propagators. This will prove useful in renormalization theory.

8.5 Loop Expansion

Up to now, we have only explored the power expansion of the S matrix in powers of the coupling constant. However, in quantum field theory it is often convenient to expand in a different power series, one based on the number of loops in a Feynman diagram. In this section, we will show that the loop expansion, in turn, corresponds to a power expansion in $\hbar$, Planck's constant. (To show this, we must reinsert all $\hbar$ factors that were eliminated when we originally set $\hbar = 1$.)

The expansion in loop number or $\hbar$ has important implications. For example, in Chapter 6 we learned that the Feynman tree diagrams for various scattering amplitudes simply reproduced the results of the classical theory. A complicated tree diagram may appear with a large number of coupling constants, but it still only corresponds to the classical theory. To see the true effects of quantization, we have to go beyond the tree diagrams and study loops. A power expansion in the loop number or $\hbar$, rather than the coupling constant, therefore measures the deviation of the quantum theory from the classical theory.

Similarly, loop effects are also important when discussing radiative corrections to a quantum field theory. If a Klein–Gordon theory, for example, has an interacting potential $V(\phi)$, then radiative corrections will modify this potential. These radiative corrections, in turn, are calculated in the loop expansion. These loop corrections are important because they shift the minimum of $V(\phi)$ and hence change the vacua of the theory. This expansion will prove useful when calculating radiative corrections to the effective potential in Chapter 10.

In this section, we will show that the path integral gives a very convenient way in which to power expand a theory in the loop order. Let us now rewrite our previous expressions for the generating functional, explicitly putting back all factors of $\hbar$ that we previously omitted. We know from dimensional arguments and the definition of the path integral that the action appears in the functional as $S/\hbar$, so the generating functional $Z(J)$ can be written as:

$$Z(J) = \int D\phi \exp\left(\frac{i}{\hbar}\int [\mathscr{L} + \hbar J(x)\phi(x)]\, d^4x\right) \tag{8.93}$$

Since we are interested in the relation between $\hbar$ and the standard perturbation theory, let us divide the Lagrangian into the free and interacting parts, $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_I$. We can extract $\mathscr{L}_I$ from the path integral by converting it into an operator:

$$Z(J) = \exp\left[\frac{i}{\hbar}\int d^4x\, \mathscr{L}_I\left(-\frac{\delta}{\delta J}\right)\right] Z_0(J) \tag{8.94}$$

where $Z_0(J)$ is just a function of the free Lagrangian $\mathscr{L}_0$. If we evaluate the derivatives, then the exponential in front of Z_0 simply reproduces $\mathscr{L}_I(\phi)$ in the exponential, giving us back the original expression for $Z(J)$.

The advantage of writing the functional in this fashion is that we can now explicitly perform the functional integral over ϕ in the free partition functions $Z_0(J)$, leaving us with the standard expression:

$$Z_0(J) = \exp\left(-\frac{i\hbar}{2}\int d^4x\, d^4y\, J(x)\Delta_F(x-y)J(y)\right) \tag{8.95}$$

Now we can begin the counting of $\hbar$ in a typical Feynman graph by inserting Z_0 back into the expression for $Z(J)$. For any Feynman diagram, each propagator, from the previous expression for Z_0, is multiplied by $\hbar$. However, each vertex, because it appears in the combination $\mathscr{L}_I/\hbar$, is multiplied by a factor of $\hbar^{-1}$.

For an arbitrary Feynman graph, the total counting of $\hbar$ is given by $\hbar$ raised to the power of $P - V$, that is, the number of propagators minus the number of vertices. However, we know from our discussion of renormalization theory in the previous chapter that:

$$L = P - V + 1 \tag{8.96}$$

where L is the number of loops in a Feynman diagram. For any Feynman diagram, we therefore pick up an overall factor of:

$$\hbar^{P-V} = \hbar^{L-1} \tag{8.97}$$

It is now easy to see that a power expansion in $\hbar$ is also a power expansion in the loop number. In Chapter 10, we will use this formalism of loop expansions to calculate the radiative corrections to several quantum field theories, showing that the loop expansion is powerful enough to shift the minimum of the potential $V(\phi)$ via radiative corrections. This will prove essential in isolating the true vacuum of a theory with a broken symmetry.

8.6 Integration over Grassmann Variables

Matter, of course, is not just bosonic. To incorporate fermions in the path integral formalism, we must define how to integrate over anticommuting variables. To do this, we must use the *Grassmann variables*, which are a set of anticommuting numbers satisfying $\theta_i\theta_j = -\theta_j\theta_i$. Integration over Grassmann variables is problematic, since $\theta^2 = 0$ for an anticommuting number, and hence the entire

foundation of calculus seems to collapse. However, a clever choice allows us to generalize the path integral formalism to fermions.

One of the features that we would like to incorporate in an integration over Grassmann variables is the fact that the integral over all space is translationally invariant:

$$\int_{-\infty}^{\infty} dx\, \phi(x) = \int_{-\infty}^{\infty} dx\, \phi(x + c) \tag{8.98}$$

Let us try to incorporate this feature in an integration over a Grassmann variable:

$$\int d\theta\, \phi(\theta) = \int d\theta\, \phi(\theta + c) \tag{8.99}$$

An arbitrary function $\phi(\theta)$ can be easily decomposed in a Taylor expansion, which terminates after only one term, since higher terms are zero:

$$\phi(\theta) = a + b\theta \tag{8.100}$$

Let us now define:

$$\begin{aligned} I_0 &= \int d\theta \\ I_1 &= \int d\theta\, \theta \end{aligned} \tag{8.101}$$

Inserting this power expansion into the integral, we find:

$$\int d\theta\, \phi(\theta) = aI_0 + bI_1 = (a + bc)I_0 + bI_1 \tag{8.102}$$

In order to maintain this identity, we choose the following normalizations:

$$\begin{aligned} I_0 &= 0 \\ I_1 &= 1 \end{aligned} \tag{8.103}$$

This, in turn, forces us to make the following unorthodox definitions:

$$\int d\theta = 0; \quad \int d\theta\, \theta = 1 \tag{8.104}$$

This is a novel choice of definitions, for it means that the integral over a Grassmann variable equals the derivative:

$$\int d\theta = \frac{\partial}{\partial \theta} \tag{8.105}$$

Now, we generalize this discussion to a nontrivial case of many Grassmann variables. With N such variables, we wish to perform the integration:

$$I(A) = \int \prod_{i=1}^{N} d\theta_i d\bar{\theta}_i \exp\left(\sum_{i,j=1}^{N} \bar{\theta}_i A_{ij} \theta_j\right) \tag{8.106}$$

where θ_i and $\bar{\theta}_i$ are two distinct sets of Grassmann variables. To evaluate this integral, we simply power expand the exponential. Because the Grassmann integral of a constant is zero, the only term that survives the integration is the Nth expansion of the exponential:

$$I(A) = \int \prod_{i=1}^{N} d\theta_i \, d\bar{\theta}_i \frac{1}{N!} \left(\sum_{i,j=1} \bar{\theta}_i A_{ij} \theta_j\right)^N \tag{8.107}$$

Most of the terms in the integral of the Nth expansion are zero. The only terms that survive are given by:

$$\begin{aligned} I(A) &= \int \left(\prod_{i=1}^{N} d\theta_i \, d\bar{\theta}_i \, \bar{\theta}_i \theta_i\right) \left(\sum_{\text{perm}} \epsilon^{i_1 i_2 \cdots i_N} A_{1 i_1} A_{2 i_2} A_{3 i_3} \cdots A_{N i_N}\right) \\ &= \det A \end{aligned} \tag{8.108}$$

An essential point is that the determinant appears in the numerator, rather than the denominator. This will have some significant implications later when we discuss ghosts and the Faddeev–Popov quantization program.

Now we make the transition from θ_i to the fermionic field $\psi(x)$. We introduce two sources η and $\bar{\eta}$ and define the generating functional:

$$Z(\eta, \bar{\eta}) = N \int D\psi \, D\bar{\psi} \, e^{i \int d^4x [\mathscr{L}(x) + \bar{\eta}\psi + \bar{\psi}\eta]} \tag{8.109}$$

where:

$$\mathscr{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{8.110}$$

As before, we can perform the functional integral by shifting variables:

$$\psi(x) \rightarrow \psi(x) + \psi_{\text{cl}}(x) \tag{8.111}$$

where:

$$\psi_{\rm cl}(x) = -\int S_F(x-y)\eta(y)\,d^4y \tag{8.112}$$

Performing the integral, we find:

$$Z(\eta,\bar{\eta}) = \exp\int d^4x\,d^4y\big(-i\bar{\eta}(x)S_F(x-y)\eta(y)\big) \tag{8.113}$$

where:

$$N = \det\,(i\not{\partial} - m) \tag{8.114}$$

The averages over the field variables can now be found by successively differentiating with respect to the η and $\bar{\eta}$ fields and then setting them equal to zero:

$$\begin{aligned} -iS_F(x-y) &= \frac{\delta}{\delta\bar{\eta}(x)}\frac{\delta}{\delta\eta(y)}Z(\eta)\Big|_{\eta=0,\bar{\eta}=0} \\ &= N\int D\psi\,D\bar{\psi}\;\bar{\psi}(y)\psi(x)e^{i\int d^4x\,L(x)} \end{aligned} \tag{8.115}$$

Successive integrations over the source fields gives us:

$$\begin{aligned} &\prod_{i=1}^{n}\frac{\delta}{\delta\eta(x_i)}\prod_{j=1}^{n}\frac{\delta}{\delta\bar{\eta}(y_j)}Z(\eta,\bar{\eta})\Big|_{\eta=\bar{\eta}=0} \\ &= (i)^{2n}\langle 0|T\bar{\psi}(x_1)\cdots\bar{\psi}(x_n)\psi(y_1)\cdots\psi(y_n)|0\rangle \end{aligned} \tag{8.116}$$

By performing the functional integration with respect to η and $\bar{\eta}$, we once again retrieve Wick's expansion for fermionic fields.

8.7 Schwinger–Dyson Equations

The functional technique allows us to formulate QED based on the Schwinger–Dyson integral equation. This equation, when power expanded, yields the standard perturbation theory. But since we do not necessarily have to power expand these equations, these integral equations also apply to bound-state and nonperturbative problems. The Schwinger–Dyson equation is based on the deceptively simple

observation that the integral of a derivative is zero:

$$\int D\phi \frac{\delta}{\delta\phi} \equiv 0 \tag{8.117}$$

Although this statement appears trivial from the point of view of functional analysis, it yields highly nontrivial relations among generating functionals in quantum field theory.

In particular, let us act on the generating functional $Z(J)$ for a scalar field theory:

$$0 = \int D\phi \frac{\delta}{\delta\phi} \exp\left(iS(\phi) + i\int d^4x J\phi\right) \tag{8.118}$$

The functional derivative of the source term simply pulls down a factor of J. We simply get:

$$0 = \int D\phi \left[iS'(\phi) + iJ\right] \exp\left(iS + i\int d^4x\, J(x)\phi(x)\right) \tag{8.119}$$

This can be rewritten as:

$$\left[S'\left(-i\frac{\delta}{\delta J}\right) + J\right] Z(J) = 0 \tag{8.120}$$

This is the Schwinger–Dyson relation, which is independent of perturbation theory. At this point, we can take any number of derivatives of this equation with respect to the fields and obtain a large number of integral equations involving various Green's functions. Or, we can power expand this equation and reproduce the known perturbation theory.

For QED, the generalization of Eq. (8.120) reads:

$$\left[\frac{\delta}{\delta A_\mu} S\left(-i\frac{\delta}{\delta J_\nu}, -i\frac{\delta}{\delta\bar{\eta}}, i\frac{\delta}{\delta\eta}\right) + J_\mu\right] Z(J,\eta,\bar{\eta}) = 0 \tag{8.121}$$

Our strategy will be to convert this expression for $Z(J)$ into an expression for $W(J)$ and to an expression for $\Gamma(\phi)$. Then we will take a derivative with respect to A_μ and set all sources to zero. We begin by using the fact that:

$$\frac{\delta S}{\delta A_\mu} = \left[\partial^2 g_{\mu\nu} - (1-\alpha^{-1})\partial_\mu\partial_\nu\right] A^\nu - e\bar{\psi}\gamma_\mu\psi \tag{8.122}$$

We can rewrite the previous equation as an equation on $W(J)$:

$$J_\mu + \left[\partial^2 g_{\mu\nu} - (1-\alpha^{-1})\partial_\mu\partial_\nu\right] \frac{\delta W}{\delta J_\nu}$$

$$-e\frac{\delta W}{\delta \eta}\gamma_\mu \frac{\delta W}{\delta \bar{\eta}} - e\frac{\delta}{\delta \eta}\left(\gamma_\mu \frac{\delta W}{\delta \bar{\eta}}\right) = 0 \tag{8.123}$$

Now let us convert this expression for $W(J)$ into an expression for $\Gamma(\phi)$, defined by the Legendre transformation:

$$W(J, \eta, \bar{\eta}) = \Gamma(A, \psi, \bar{\psi}) + \int d^4x \, (J_\mu A^\mu + \bar{\psi}\eta + \bar{\eta}\psi) \tag{8.124}$$

We must make the substitution:

$$\begin{aligned} A_\mu &= \frac{\delta W}{\delta J^\mu}; \qquad \psi = \frac{\delta W}{\delta \bar{\eta}} \\ J_\mu &= -\frac{\delta \Gamma}{\delta A^\mu}; \qquad \eta = -\frac{\delta \Gamma}{\delta \bar{\psi}} \end{aligned} \tag{8.125}$$

Then the Schwinger–Dyson equation can be written as:

$$\begin{aligned} \frac{\delta \Gamma}{\delta A^\mu(x)}\Big|_{\psi=\bar{\psi}=0} &= \left[\partial^2 g_{\mu\nu} - (1-\alpha^{-1})\partial_\mu \partial_\nu\right] A^\nu(x) \\ &\quad - ie\mathrm{Tr}\left[\gamma_\mu \left(\frac{\delta^2 \Gamma}{\delta \bar{\psi}\delta\psi}\right)^{-1}(x, x)\right] \end{aligned} \tag{8.126}$$

where the last term on the right is proportional to the electron propagator, and we have used the fact that:

$$-\delta_{\alpha\beta}\delta^4(x-y) = \int d^4z \frac{\delta^2 W}{\delta \eta_\alpha(x)\delta\bar{\eta}_\gamma(z)} \frac{\delta^2 \Gamma}{\delta\psi_\gamma(z)\delta\bar{\psi}_\beta(y)}\Big|_{\eta=\bar{\eta}=\psi=\bar{\psi}=0} \tag{8.127}$$

Our last step is to take the derivative with respect to A^ν. Then the term on the right is related to the photon propagator. The final expression becomes:

$$\begin{aligned} \frac{\delta^2 \Gamma}{\delta A^\mu(x)\delta A^\nu(y)}\Big|_{A=\psi=\bar{\psi}=0} &= \left[\partial^2 g_{\mu\nu} - (1-\alpha^{-1})\partial_\mu\partial_\nu\right]\delta^4(x-y) \\ &\quad + ie^2 \int d^4u \, d^4v \, \mathrm{Tr}\left[\gamma_\mu S'_F(x, u)\Lambda_\nu(y, u, v)S'_F(v, x)\right] \end{aligned} \tag{8.128}$$

where we have defined the vertex function as:

$$\frac{\delta^3 \Gamma}{\delta A^\mu(x)\,\delta\bar{\psi}(y)\,\delta\bar{\psi}(z)}\Big|_{A=\psi=\bar{\psi}=0} = e\Lambda_\mu(x, y, z) \tag{8.129}$$

We have also used the formula for taking the derivative of an inverse matrix:

$$\frac{\delta}{\delta A_\mu}\mathscr{M}^{-1} = -\mathscr{M}^{-1}\frac{\delta}{\delta A_\mu}\mathscr{M}\mathscr{M}^{-1} \tag{8.130}$$

where $\mathscr{M} = \delta^2\Gamma/\delta\bar{\psi}\,\delta\psi$ and:

$$\begin{aligned} S'_F(x,y) = \mathscr{M}^{-1} &= \left(\frac{\delta^2\Gamma}{\delta\bar{\psi}\,\delta\psi}\right)^{-1} \\ &= \int \frac{d^4p}{(2\pi)^4}\frac{e^{-ip\cdot(x-y)}}{\not{p} - m - \Sigma(p)} \end{aligned} \tag{8.131}$$

These functional relations, in turn, are identical to the Schwinger–Dyson equations introduced in Exercise (7.11).

In summary, we have seen that the path integral method of Feynman is not only elegant and powerful, it is also very close to the original spirit of quantum mechanics. The formalism is so versatile that we can reproduce the canonical formalism discussed earlier, as well as quantize increasingly complicated theories, such as Yang–Mills theory and quantum gravity. The path integral formalism, in fact, has become the dominant formalism for high-energy physics. In Chapter 9, we will see the power of the path integral approach when we quantize the Yang–Mills theory.

8.8 Exercises

1. Using path integrals for a free Dirac theory, calculate the expectation value of the product of six fermionic fields in terms of propagators. Show that the resulting expression is equivalent to the decomposition given by Wick's theorem.

2. Prove Eq. (8.14).

3. Prove that the disconnected pieces in Eq. (8.78) cancel to second order in λ. Sketch how the cancellation works at third order.

4. Let θ_i be a Grassmann column vector. Let us make the transformation of variables: $\phi_i = M_{ij}\theta_j$. We define:

$$\int d\phi_1\, d\phi_2 \cdots d\phi_n\ (\phi_1\phi_2\cdots\phi_n) = \int d\theta_1\, d\theta_2 \cdots d\theta_n\ (\theta_1\theta_2\cdots\theta_n) \tag{8.132}$$

Prove that this implies:

$$d\phi_1\, d\phi_2 \cdots d\phi_n = (\det M)^{-1} d\theta_1\, d\theta_2 \cdots d\theta_n \tag{8.133}$$

which is the opposite of the usual rule for differentials.

5. Derive the Schrödinger equation for an electron in the presence of a potential $V(x)$ using path integrals.

6. For invertible, square matrices A, B, C, D, prove:

$$\begin{pmatrix} A & C \\ D & B \end{pmatrix} = \begin{pmatrix} A - CB^{-1}D & CB^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ D & B \end{pmatrix}$$
$$= \begin{pmatrix} A & 0 \\ D & 1 \end{pmatrix} \begin{pmatrix} 1 & A^{-1}C \\ 0 & B - DA^{-1}C \end{pmatrix}$$

7. For matrices A, B, prove:

$$\log(AB) = \log A + \log B + \frac{1}{2}[\log A, \log B] + \cdots \tag{8.134}$$

8. Prove:

$$\det(1 + M) = 1 + \operatorname{Tr} M + \frac{1}{2}\left[(\operatorname{Tr} M)^2 - \operatorname{Tr}(M^2)\right] + \cdots \tag{8.135}$$

9. For matrices A and B, prove, by power expansion, that:

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] + \frac{1}{12}[B, [B, A]] + \cdots\right) \tag{8.136}$$

10. For matrices A and B, prove that:

$$\begin{aligned} e^A e^B &= \exp[A, B]\, e^B e^A \\ e^{A+B} &= \exp\left(-\frac{1}{2}[A, B]\right) e^A e^B \end{aligned} \tag{8.137}$$

Under what conditions are these identities valid?

11. For matrices A and B, prove:

$$(e^A e^B)^n = \exp\left(\frac{n(n+1)}{2}[A, B]\right) e^{nA} e^{nB} \tag{8.138}$$

Are there any restrictions on this formula?

12. Prove:

$$\int d\theta_n \cdots d\theta_2 \, d\theta_1 \exp \frac{1}{2} \sum_{i,j=1}^{n} \theta_i D_{ij} \theta_j = \sqrt{\det D} \tag{8.139}$$

for the case $n = 3$ by an explicit calculation.

13. Prove the previous relation for arbitrary n. Prove it in two ways: first, by diagonalizing the D matrix and then performing the integration over the eigenvalues of D; second, by power expanding the expression and using the known identities for the antisymmetric ϵ tensor.

14. In Eq. (8.92), we established a relationship between the $\tilde{W}^{(4)}$ and $\Gamma^{(4)}$. Find the relationship between the fifth orders, then graphically illustrate what it means.

15. For ϕ^4 theory, show that $\Gamma^{(4)}$ is actually one-particle irreducible. Expand it only to fourth order in the coupling.

Chapter 9
Gauge Theory

We did not know how to make the theory fit experiment. It was our judgment, however, that the beauty of the idea alone merited attention.
—C. N. Yang

9.1 Local Symmetry

An important revolution in quantum field theory took place in 1971, when the Yang–Mills theory was shown to be renormalizable, even after symmetry breaking, and therefore was a suitable candidate for a theory of particle interactions. The theoretical landscape in particle physics rapidly changed; a series of important papers emerged in which the weak and strong interactions quickly yielded their secrets. This revolution was remarkable, given the fact that in the relative confusion of the 1950s and 1960s, it appeared as if quantum field theory was an unsuitable framework for particle interactions.

Historically, gauge theory had a long but confused past. Although the Yang–Mills equation had been discovered in 1938 by O. Klein[1] (who was studying Kaluza–Klein theories), it was promptly forgotten during World War II. It was resurrected independently by Yang and Mills[2] in 1954 (and also by Shaw[3] and Utiyama[4]), but it was unsuitable for particle interactions because it only described massless vector particles. The discovery by 't Hooft[5] that the theory could be made massive while preserving renormalizability sparked the current gauge revolution.

Previously, we studied theories which were symmetric under *global* symmetries, so the group parameter ϵ was a constant. But the essence of Maxwell and Yang–Mills theory is that they are invariant under a symmetry that changes at every space–time point; that is, they are *locally* invariant. This simple principle of local gauge invariance, as we shall see, imposes highly nontrivial and nonlinear constraints on quantum field theory.

We begin with the generators of some Lie algebra:

$$[\tau^a, \tau^b] = if^{abc}\tau^c \tag{9.1}$$

Let the fermion field ψ_i transform in some representation of $SU(N)$, not necessarily the fundamental representation. It transforms as:

$$\psi_i(x) \rightarrow \Omega_{ij}(x)\psi_j(x) \tag{9.2}$$

where Ω_{ij} is an element of $SU(N)$.

The essential point is that the group element Ω is now a function of space–time; that is, it changes at every point in the universe. It can be parametrized as:

$$\Omega_{ij}(x) = \left(e^{-i\theta^a(x)\tau^a}\right)_{ij} \tag{9.3}$$

where the parameters $\theta^a(x)$ are local variables, and where τ^a is defined in whatever representation we are analyzing.

The problem with this construction is that derivatives of the fermion field are not covariant under this transformation. A naive transformation of the derivatives of these fields picks up terms like $\partial_\mu\Omega$. In order to cancel this unwanted term, we would like to introduce a new derivative operator D_μ that is truly covariant under the group. To construct such an operator D_μ, let us introduce a new field, called the *connection* A_μ:

$$D_\mu \equiv \partial_\mu - igA_\mu \tag{9.4}$$

where:

$$A_\mu(x) \equiv A^a_\mu(x)\tau^a \tag{9.5}$$

The essential point of this construction is that the covariant derivative of the ψ field is gauge covariant:

$$\begin{aligned}(D_\mu\psi)' &= \partial_\mu\psi' - igA'_\mu\psi' \\ &= \Omega\partial_\mu\psi + (\partial_\mu\Omega)\psi - igA'_\mu\Omega\psi \\ &= \Omega D_\mu\psi \end{aligned} \tag{9.6}$$

The troublesome term $\partial_\mu\Omega$ is precisely cancelled by the variation of the A_μ term if we set:

$$A'_\mu(x) = -\frac{i}{g}[\partial_\mu\Omega(x)]\Omega^{-1}(x) + \Omega(x)A_\mu(x)\Omega^{-1}(x) \tag{9.7}$$

Infinitesimally, this becomes:

$$\begin{cases} \delta A^a_\mu = -\frac{1}{g}\partial_\mu \theta^a + f^{abc}\theta^b A^c_\mu \\ \delta \psi = -ig\theta^a \tau^a \psi \end{cases} \tag{9.8}$$

[If we reduce $SU(N)$ down to the group $U(1)$, then we recover the field transformations for QED.]

It is also possible to construct the invariant action for the connection field itself. Since D_μ is covariant, then the commutator of two covariant derivatives is also covariant. We define the commutator as follows:

$$\begin{aligned} F_{\mu\nu} &= \frac{i}{g}[D_\mu, D_\nu] \\ &= \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \\ &= \left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu\right)\tau^a \end{aligned} \tag{9.9}$$

Because D_μ is genuinely covariant, this means that the $F^a_{\mu\nu}$ tensor is also covariant:

$$F_{\mu\nu} \to \Omega F_{\mu\nu}\Omega^{-1} \tag{9.10}$$

We can now construct an invariant action out of this tensor. We want an action that only has two derivatives (since actions with three or higher derivatives are not unitary, i.e., they have ghosts). The simplest invariant is given by the trace of the commutator. This is invariant because:

$$\mathrm{Tr}\left(\Omega F_{\mu\nu}\Omega^{-1}\Omega F^{\mu\nu}\Omega^{-1}\right) = \mathrm{Tr}\left(F_{\mu\nu}F^{\mu\nu}\right) \tag{9.11}$$

The unique action with only two derivatives is therefore given by:

$$S = \int d^4x \left(-\frac{1}{2}\mathrm{Tr}\, F_{\mu\nu}F^{\mu\nu}\right) = \int d^4x \left(-\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}\right) \tag{9.12}$$

This is the action for the Yang–Mills theory, which is the starting point for all discussions of gauge theory.

The field tensor $F_{\mu\nu}$, we should point out, also obeys the Bianchi identities. We know, by the Jacobi identity, that certain multiple commutators vanish identically. Therefore, we have:

$$\left[D_\mu, [D_\nu, D_\rho]\right] + \left[D_\nu, [D_\rho, D_\mu]\right] + \left[D_\rho, [D_\mu, D_\nu]\right] \equiv 0 \tag{9.13}$$

This is easily checked by explicitly writing out the terms in the commutators. Written in terms of the field tensor, this becomes:

$$[D_\mu, F_{\nu\rho}] + [D_\nu, F_{\rho\mu}] + [D_\rho, F_{\mu\nu}] = 0 \tag{9.14}$$

(It is important to stress that these are exact identities. They are not equations of motion, nor are they new constraints on the field tensor.)

Lastly, since $\bar{\psi} \to \bar{\psi}\Omega^\dagger$ and $D_\mu\psi \to \Omega D_\mu\psi$, it is easy to show that the invariant fermion action coupled to the gauge field is given by:

$$S = \int d^4x\, \bar{\psi}(i \not{D} - m)\psi \tag{9.15}$$

9.2 Faddeev–Popov Gauge Fixing

The real power of the path integral approach is that we have the freedom to choose whatever gauge we desire. This is impossible in the canonical approach, where the gauge has already been fixed. However, in the path integral approach, because gauge fixing is performed by inserting certain delta functions into the path integral, we can change the gauge by simply replacing these factors. This formalism was introduced by Faddeev and Popov.[6]

Historically, however, before the Faddeev–Popov method, the quantization of Yang–Mills theory was not clear for many years. In 1962, Feynman[7] showed that the theory suffered from a strange kind of disease: The naive quantization of the theory was not unitary. In order to cancel the nonunitary terms from the theory, Feynman was led to postulate the existence of a term that did not emerge from the standard quantization procedure. Today, we know this ghost, first revealed by Feynman using unitarity arguments, as the *Faddeev–Popov ghost.*

To begin, we first stress that the path integral of a theory like Maxwell's theory of electromagnetism is, in principle, undefined because of the gauge degree of freedom. Because the path integral DA and the action S are both gauge invariant, it means that functionally integrating over DA will eventually overcount the degrees of freedom of the theory. Because the Maxwell theory is invariant under the gauge transformation:

$$A_\mu^\Omega = A_\mu + \partial_\mu\Omega \tag{9.16}$$

this means that functionally integrating over both A_μ^Ω and A_μ will overcount the integrand repeatedly. In fact:

$$\int DA_\mu e^{iS} = \infty \tag{9.17}$$

If we begin with one field configuration A_μ and consider all possible A_μ^Ω, then we are sweeping out an "orbit" in functional space. The problem is that, as Ω changes along the orbit, we repeat ourselves an infinite number of times in the path integral. Our problem is therefore to "slice" the orbit once so that we do not have this infinite overcounting.

This is the origin of the gauge-fixing problem. To solve this problem, one is tempted to insert factors like:

$$\delta(\partial_\mu A^\mu); \quad \delta(\nabla \cdot \mathbf{A}) \tag{9.18}$$

into the path integral, forcing it to respect the gauge choice $\partial_\mu A^\mu = 0$ or $\nabla \cdot \mathbf{A} = 0$.

More generally, we would like to fix the gauge with an arbitrary function of the fields:

$$\delta\left(F(A_\mu)\right) \tag{9.19}$$

which would fix the gauge to be $F(A_\mu) = 0$.

The source of the problem is that inserting a delta function into the functional integration DA changes the measure of the integration. For example, we know that the delta function changes when we make seemingly trivial changes in it. For example, if we have a function $f(x)$ that has a zero at $x = a$, we recall that:

$$\delta\left(f(x)\right) = \frac{\delta(x-a)}{|f'(x)|} \tag{9.20}$$

The choice $\delta\left(f(x)\right)$ differs from the choice $\delta(x-a)$ by a term $f'(x)$. Thus, there is an ambiguity in the measure of integration. To solve this ambiguity, we insert the number "1" into the path integral, which we know has the correct measure:

$$1 = \Delta_{FP} \int D\Omega\, \delta\left(F(A_\mu^\Omega)\right) \tag{9.21}$$

where Δ_{FP} is the Faddeev–Popov determinant, which guarantees the correct measure, and $D\Omega = \prod_x d\Omega(x)$ is the invariant group measure. It satisfies the invariance property:

$$D\Omega = D(\Omega'\Omega) \tag{9.22}$$

if Ω' is a fixed element of $SU(N)$. Since $\Omega = 1 - i\theta^a\tau^a + \cdots$, the group measure (for small θ) is equal to:

$$D\Omega = \prod_{a,x} d\theta^a(x) \tag{9.23}$$

(More details concerning the invariant group measure will be presented in Chapter 15.)

Inserting the number "1" into the functional, we now have:

$$\int DA_\mu \left(\Delta_{FP} \int D\Omega\, \delta\left(F(A_\mu^\Omega)\right)\right) e^{i\int d^4x\, \mathscr{L}(A)} \tag{9.24}$$

Our task is now to calculate an explicit result for the Faddeev–Popov determinant. To do this, we first notice that it is gauge invariant:

$$\Delta_{FP}(A_\mu) = \Delta_{FP}(A_\mu^\Omega) \tag{9.25}$$

This is because it is equal to the integration over all gauge Ω factors, and hence is independent of the gauge. However, it will be instructive to see this more explicitly.

Let us change the gauge, replacing A_μ in the Faddeev–Popov determinant with A_μ^Ω. Then the definition becomes:

$$\begin{aligned}
\Delta_{FP}^{-1}(A_\mu^{\Omega'}) &= \int D\Omega'\, \delta\left[F(A_\mu^{\Omega'\Omega})\right] \\
&= \int D\left[\Omega'\Omega\right] \delta\left(F(A_\mu^{\Omega'\Omega})\right) \\
&= \int D\Omega''\delta\left(F(A_\mu^{\Omega''})\right) \\
&= \Delta_{FP}^{-1}(A_\mu)
\end{aligned} \tag{9.26}$$

Thus, it is gauge invariant. Now we come to the crucial part of the calculation. Let us make a gauge transformation on the entire path integral, so that $A_\mu \to A_\mu^{\Omega^{-1}}$. The measure DA, the Faddeev–Popov determinant, and the action S are all gauge invariant. The factor that is not gauge invariant is $F(A_\mu^\Omega)$, which changes into $F(A_\mu)$. The integral, after the gauge transformation, now becomes:

$$\left(\int D\Omega\right) \int DA_\mu\, \Delta_{FP}\, \delta\left(F(A_\mu)\right) e^{i\int d^4x\, \mathscr{L}(A)} \tag{9.27}$$

We have now accomplished the following:

1. We have explicitly isolated the infinite part of the matrix element, which is given by:

$$\int D\Omega = \infty \tag{9.28}$$

   By simply dividing out by the $\int D\Omega$, we remove the infinite overcounting.

2. The gauge choice is now enforced by $\delta\left(F(A_\mu)\right)$.

3. The essential point is that the factor Δ_{FP} gives the correct measure in the path integral. This is the factor that was missing for so many years in previous attempts to quantize gauge theories.

The problem of gauge fixing is therefore reduced to finding an explicit expression for Δ_{FP}. From Eq. (9.21), we know that the Faddeev–Popov determinant is written in terms of $\delta(F(A_\mu^\Omega))$, which in turn can be re-expressed via Eq. (9.20) as:

$$\delta\left[F(A_\mu^\Omega)\right] = \delta(\Omega - \Omega_0)\det\left|\frac{\delta F(A_\mu^\Omega)(x)}{\delta\Omega'(x')}\right|^{-1} \tag{9.29}$$

where the determinant is a functional generalization of the factor $|f'(x)|$.

Thus, the Faddeev–Popov term can be written as:

$$\Delta_{FP} = \det\left|\frac{\delta F(A_\mu^\Omega)(x)}{\delta\Omega'(x')}\right| \tag{9.30}$$

for $F(A_\mu) = 0$. To perform explicit calculations with this determinant, it is convenient to rewrite this expression in a form where we can extract new Feynman rules. To solve for the determinant, let us power expand the factor $F(A_\mu^\Omega)$ for a small group parameter θ:

$$F\left(A_\mu^\Omega(x)\right) = F(A_\mu(x)) + \int d^4y\, M(x, y)\theta(y) + \cdots \tag{9.31}$$

In this approximation, only the matrix M survives when we take the derivative. The Faddeev–Popov term can be converted into the determinant of the matrix M, which in turn can be converted into a Gaussian integral over two fields c and $c^\dagger$:

$$\Delta_{FP} = \det M = \int Dc\, Dc^\dagger \exp i\left(\int d^4x\, d^4y\, c^\dagger(x)M(x, y)c(y)\right) \tag{9.32}$$

The determinant $\det M$ appears in the numerator of the path integral, rather than the denominator. This means that when we exponentiate the Faddeev–Popov term

into the Lagrangian, we must integrate over *Grassmann variables*, rather than bosonic variables, as we discussed in Section 8.6. Thus, we find that c and $c^\dagger$ are actually "ghost" fields, that is, scalar fields obeying Fermi–Dirac statistics. This is the origin of the celebrated Faddeev–Popov ghosts.

Alternatively, we could have rewritten the determinant as follows:

$$\Delta_{FP} = \det M = \exp(\mathrm{Tr} \log M) \tag{9.33}$$

where the determinant is understood to be taken over discretized x and y and any isospin indices. (To prove that the determinant of a matrix M can be written as the exponential of the trace of the logarithm of M, simply diagonalize the M matrix and re-express the formula in terms of the eigenvalues of M. Then this identity is trivial.) This term can be written in more familiar language if we write the matrix as $M = 1 + L$ and expand as follows:

$$\begin{aligned} \det(1+L) &= \exp \mathrm{Tr}\,[\log(1+L)] \\ &= \exp\left(\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \mathrm{Tr}\, L^n\right) \end{aligned} \tag{9.34}$$

The trace over L, we shall see shortly, can now be interpreted as closed loops in the Feynman expansion of the perturbation theory.

To gain some familiarity with this Faddeev–Popov term, let us compute the Faddeev–Popov determinant for the simplest gauge theory, Maxwell's theory. Earlier, in the canonical and covariant approaches, we fixed the gauge without even thinking about complications due to the functional measure. We will find that we were lucky: The Faddeev–Popov measure is trivial for the gauges found in Maxwell theory, but highly nontrivial for non-Abelian gauge theory. To see this, let us choose the gauge:

$$F(A_\mu) = \partial^\mu A_\mu = 0 \tag{9.35}$$

The variation of this gauge fixing gives us:

$$F(A_\mu^\Omega) = \partial^\mu A_\mu + \partial^\mu \partial_\mu \theta \tag{9.36}$$

Then the M matrix becomes:

$$M(x, y) = \left[\partial_\mu \partial^\mu\right]_{x,y} \tag{9.37}$$

Written in terms of ghost variables c and $c^\dagger$, this means we must add the following term to the action:

$$\int d^4x\, d^4y\, c^\dagger(x)\partial^\mu\partial_\mu c(y) \tag{9.38}$$

where the integration over $c^\dagger$ and c must now be viewed as being Grassmannian; that is, these fields are scalar Grassmann ghosts. Fortunately, this determinant decouples from the rest of the theory. The determinant of the Laplacian does not couple to any of the fermions or gauge fields in QED, and hence gives only an uninteresting multiplicative factor in front of the S matrix, which we can remove at will.

For the Coulomb gauge, we find a similar argument, except that the gauge variation gives us a factor:

$$\det \nabla^2 \tag{9.39}$$

which also decouples from the theory. Thus, from the path integral point of view, we were fortunate that we could take these gauges in quantizing QED without suffering any problems. We are not so fortunate, however, for gauge theories, where the Faddeev–Popov term gives us highly nontrivial corrections to the naive theory. We begin by choosing a gauge for the theory, such as:

$$\partial^\mu A^a_\mu = 0 \tag{9.40}$$

If we place this into the path integral, then we must at the same time insert the Faddeev–Popov measure:

$$\Delta_{FP} = \det\,(M)_{x,y;a,b} \tag{9.41}$$

where x and y are discretized space–time variables, and a, b are isospin variables. The matrix M is easily found:

$$\begin{aligned} M_{ab}(x-y) &= \frac{\delta}{\delta\Omega^a(x)} A^{b,\Omega}_\mu(y) = \frac{\delta}{\delta\theta^a(x)}\left(\partial^\mu \frac{-1}{g}(\partial_\mu\theta^b + g f^{bcd} A^c_\mu \theta^d)\right)(y) \\ &= \frac{1}{g}\left(-\delta^{ab}\partial^\mu\partial_\mu + g f^{abc} A^c_\mu\right)_{x,y} \delta^4(x-y) \end{aligned} \tag{9.42}$$

If we rescale and exponentiate this into the functional integral, we find an additional term in the action given by:

$$\int d^4x\, c^{\dagger a}\left(\delta^{ab}\partial^\mu\partial_\mu - g f^{abc} A^c_\mu\right) c^b \tag{9.43}$$

The $c(x)$ are the celebrated Faddeev–Popov ghosts. These ghosts have peculiar properties:

1. They are scalar fields under the Lorentz group, but have anticommuting statistics, violating the usual spin-statistics theorem. They are therefore ghosts. However, we do not care, since their job is to cancel the ghosts coming from the quantization of the A^a_μ field.
2. These ghosts couple only to the gauge field. They do not appear in the external states of the theory. Therefore, they cannot appear in tree diagrams at all; they only make their presence in the loop diagrams, where we have an internal loop of circulating ghosts.
3. These ghosts are an artifact of quantization. They decouple from the physical spectrum of states.

9.3 Feynman Rules for Gauge Theory

Let us now put all the various ingredients together and write down the Feynman rules for the theory. The action plus the gauge-fixing contribution becomes:

$$\mathscr{L} + \mathscr{L}_{FP} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - \frac{1}{2\alpha}(\partial_\mu A^{a\mu})^2 \tag{9.44}$$

while the ghost contribution becomes:

$$\mathscr{L}_g = \int d^4x d^4y \sum_{ab} c^\dagger_a(x)[M(x, y)]_{ab}c_b(y) \tag{9.45}$$

where:

$$[M(x, y)]_{ab} = -\frac{1}{g}\partial^\mu \left(\delta^{ab}\partial_\mu - g\epsilon^{abc}A^c_\mu\right)\delta^4(x - y) \tag{9.46}$$

To extract the Feynman rules from this action, we will decompose the action into a free part and interacting part:

$$\mathscr{L}_0 = -\frac{1}{4}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)^2 - \frac{1}{2\alpha}(\partial^\mu A^a_\mu)^2 + c^\dagger_a\partial^2 c_a \tag{9.47}$$

and:

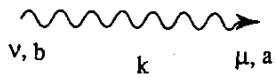
$$\mathscr{L}_I = -\frac{1}{2}g(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)\epsilon^{abc}A^{b\mu}A^{c\nu}$$

$$+ \quad \frac{1}{4}g^2\epsilon^{abc}\epsilon^{ade}A^b_\mu A^c_\nu A^{d\mu}A^{e\nu}$$

$$- \quad igc^{a\dagger}\epsilon^{abc}\partial^\mu A^c_\mu c^b \tag{9.48}$$

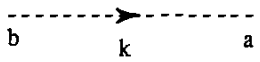
From this, we can read off the Feynman rules for Yang–Mills theory:

1. The gauge meson propagator is given by:



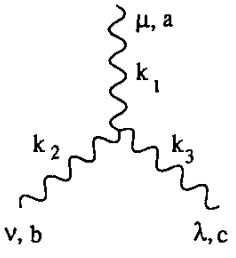
$$i\Delta^{ab}_{\mu\nu}(k) = -i\delta^{ab}\left(g_{\mu\nu} - (1-\alpha)\frac{k_\mu k_\nu}{k^2}\right)\frac{1}{k^2+i\epsilon} \tag{9.49}$$

2. The (directed) ghost propagator is given by:



$$i\Delta^{ab}(k) = -i\delta^{ab}\frac{1}{k^2+i\epsilon} \tag{9.50}$$

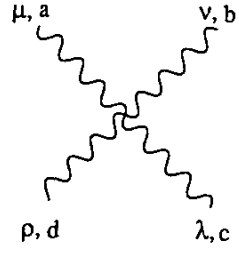
3. The three-gauge meson vertex function for mesons with momenta and quantum numbers (k_1, μ, a), (k_2, ν, b), and (k_3, λ, c) is given by:



$$i\Gamma^{abc}_{\mu\nu\lambda} = ig\epsilon^{abc}\Big[(k_1-k_2)_\lambda g_{\mu\nu} + (k_2-k_3)_\mu g_{\nu\lambda} + (k_3-k_1)_\nu g_{\mu\lambda}\Big] \tag{9.51}$$

with $\sum_{i=1}^3 k_i = 0$.

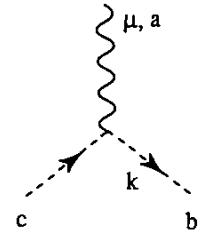
4. The four-gauge vertex is given by:



$$\begin{aligned} i\Gamma^{abcd}_{\mu\nu\lambda\rho} = ig^2\Big[&\epsilon^{abe}\epsilon^{cde}\left(g_{\mu\lambda}g_{\nu\rho} - g_{\nu\lambda}g_{\mu\rho}\right) \\ &+ \epsilon^{ace}\epsilon^{bde}\left(g_{\mu\nu}g_{\lambda\rho} - g_{\lambda\nu}g_{\mu\rho}\right) \\ &+ \epsilon^{ade}\epsilon^{cbe}\left(g_{\mu\lambda}g_{\rho\nu} - g_{\rho\lambda}g_{\mu\nu}\right)\Big] \end{aligned} \tag{9.52}$$

with $\sum_{i=1}^4 k_i = 0$

5. The two-ghost/gauge meson vertex function is given by:



$$i\Gamma^{abc}_\mu = g\epsilon^{abc}k_\mu \tag{9.53}$$

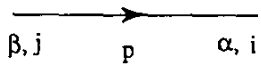
where k_μ is the incoming ghost's momentum.

Now let us derive the Feynman rules for the gauge field coupled to a fermion and a scalar field. We assume that the coupling to a scalar field is given by the matrix R^a, taken in whatever representation we desire. The coupling is given by:

$$\begin{aligned}\mathscr{L} &= \bar{\psi}\left[i\gamma^\mu(\partial_\mu - igA^a_\mu\tau^a) - m\right]\psi \\ &\quad + \left[(\partial^\mu - igA^{a\mu}R^a)\phi\right]^\dagger\left[(\partial_\mu - igA^a_\mu R^a)\phi\right] - m^2\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2\end{aligned} \tag{9.54}$$

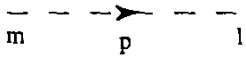
Then the Feynman rules for this theory are given by:

1. The fermion propagator is given by:



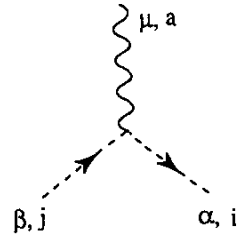
$$iS_F^{ij}(p)_{\alpha\beta} = \left(\frac{i\delta^{ij}}{\not{p} - m + i\epsilon}\right)_{\alpha\beta} \tag{9.55}$$

2. The scalar boson propagator is given by:



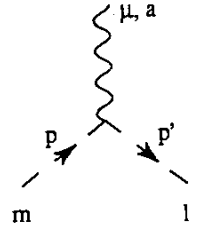
$$i\Delta_F^{lm}(p) = \frac{i\delta^{lm}}{p^2 - m^2 + i\epsilon} \tag{9.56}$$

3. The fermion–gauge-meson coupling is given by:



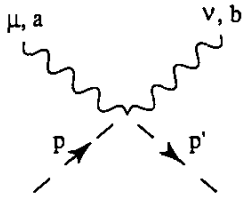
$$ig(\gamma_\mu)_{\alpha\beta}\left(\tau^a\right)_{ij} \tag{9.57}$$

4. The boson–gauge-meson coupling is given by:



$$ig(p + p')_\mu R^a_{lm} \tag{9.58}$$

5. The two-boson–gauge-meson coupling is given by:



$$-ig^2 g_{\mu\nu}\{R^a, R^b\} \tag{9.59}$$

6. The four-scalar couplings are given by $-i\lambda$.

9.4 Coulomb Gauge

To begin a discussion of gauge theory in the Coulomb gauge, it is first instructive to rewrite the original action as follows:

$$\mathscr{L} = \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - \frac{1}{2}F^{a,\mu\nu}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu\right) \tag{9.60}$$

where it is essential to notice that $F^a_{\mu\nu}$ is an *independent field*, totally unrelated to the vector field A^a_μ. (However, by eliminating the $F_{\mu\nu}$ field by its equations of motion, we can show the equivalence with the standard Yang–Mills action.) This new version of the action is invariant under:

$$\begin{aligned} \delta A^a_\mu &= -\frac{1}{g}\partial_\mu\theta^a + f^{abc}\theta^b A^c_\mu \\ \delta F^a_{\mu\nu} &= f^{abc}\theta^b F^c_{\mu\nu} \end{aligned} \tag{9.61}$$

Now let us take the variation of the action with respect to $F^a_{\mu\nu}$ as well as with respect to A^a_μ. The two equations of motion yield:

$$\begin{aligned} F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \\ 0 &= \partial_\mu F^{a\mu\nu} + gf^{abc}A^{b,\mu}F^c_{\mu\nu} \end{aligned} \tag{9.62}$$

This new action will prove to be useful when quantizing the theory in the Coulomb gauge. First, we will eliminate F^a_{ij} in terms of the A^a_i fields, keeping F^a_{0i} an independent field. Written in terms of these variables, we find that the Lagrangian becomes:

$$\mathscr{L} = -\frac{1}{4}F^a_{ij}(A)F^{aij}(A) + \frac{1}{2}(F^a_{0i})^2 - F^{a,0i}\left(\partial_i A^a_0 - \partial_0 A^a_i + gf^{abc}A^b_0 A^c_i\right) \tag{9.63}$$

Let us define:

$$\begin{aligned} E^a_i &= F^a_{0i} \\ B^a_i &= \frac{1}{2}\epsilon_{ijk}F^{a,jk}(A) \end{aligned} \tag{9.64}$$

In terms of these fields, we now have:

$$\mathscr{L} = E^a_i\dot{A}^a_i - \mathscr{H} \tag{9.65}$$

where:

$$\mathscr{H} = \frac{1}{2}\left[(E_i^a)^2 + (B_i^a)^2\right] + A_0^a\left(\partial_i E_i^a + g f^{abc} E_i^b A_i^c\right) \tag{9.66}$$

Written in this form, the similarity to the quantization of Maxwell's theory in the Coulomb gauge is apparent. Of particular importance is the Lagrange multiplier A_0^a, which multiplies the covariant divergence of E_i^a, which is the gauge generalization of Gauss's Law found earlier for QED in Eq. (4.54). The equation of motion for A_0^a yields:

$$D_i E_i^a = \partial_i E_i^a + g f^{abc} E_i^b A_i^c = 0 \tag{9.67}$$

which indicates that not all conjugate momenta are independent, a situation common to all gauge theories.

There are two ways in which to solve this important constraint. First, as in the Gupta–Bleuler approach, we can apply this constraint directly onto the state vectors of the theory:

$$D_i E_i^a |\Psi\rangle = 0 \tag{9.68}$$

However, it is easy to show that $D_i E_i^a$, acting on a field A_i^a, generates the standard gauge transformation; that is:

$$\left[\int d^3x\, \Lambda^a D_i E_i^a(x),\, A_j^b(y)\right]_{x_0=y_0} = -i\left(\partial_j \Lambda^b(y) + g f^{bcd} \Lambda^c A_j^d(y)\right) \tag{9.69}$$

Therefore the constraint equation means that the state vectors $|\Psi\rangle$ of the theory must be *singlets* under the gauge group. This is a rather surprising result, because it implies that free vector mesons A_i^a are not part of the physical spectrum. This is consistent, however, with our understanding of the quark model, where nonperturbative calculations indicate that the only allowed states are singlets under the "color" gauge group $SU(3)$. The allowed singlets under the color group include quark–antiquark and three-quark bound states, which are the only ones seen experimentally. This will be important when we discuss the phenomenon of confinement in Chapters 11 and 15.

The second approach in solving this constraint is to assume, for the moment, that perturbation theory is valid and simply drop the higher-order terms. Then the constraint equation reduces to the statement that the E_i^a field is transverse. To see this, we will decompose the E_i^a field into transverse and longitudinal modes:

$$E_i^a = E_i^{aT} + E_i^{aL} \tag{9.70}$$

where:

$$\nabla_i E_i^{aT} = 0 \tag{9.71}$$

An explicit decomposition of any field E_i^a into transverse and longitudinal parts can be given as:

$$E_i^a = \left(\delta_{ij} - \nabla_i \frac{1}{\nabla^2} \nabla_j\right) E_j^a + \nabla_i \frac{1}{\nabla^2} \nabla_j E_j^a \tag{9.72}$$

We will find it convenient to factor out the longitudinal mode explicitly by introducing the field f^a:

$$E_i^{aL} = -\nabla_i f^a \tag{9.73}$$

Now insert these definitions back into the Gauss's Law constraint, which now becomes:

$$\begin{aligned} &\nabla_i E_i^a + g f^{abc} E_i^b A_i^c \\ &= \quad -\nabla_i^2 f_i^a - g f^{abc} \nabla_i f^b A_i^c + g f^{abc} E_i^{bT} A_i^c = 0 \end{aligned} \tag{9.74}$$

For the case of Maxwell's field, this constraint was trivial to solve, since the cross term was not present. For the Yang–Mills theory, this cross term gives us a nontrivial complication.

Fortunately, we can now solve for f^a by inverting this equation as a perturbation series. Let us rewrite it as:

$$D_{ab} f^b = f^{abc} E_i^{bT} A_i^c \tag{9.75}$$

where:

$$D_{ab} = \nabla^2 \delta_{ab} - g f^{abc} A_i^c \nabla_i \tag{9.76}$$

To solve this equation, we must introduce the Green's function:

$$\left(\nabla^2 \delta^{ab} - g f^{abd} A_i^d \nabla_i\right) \Delta^{bc}(x, y; A) = \delta^{ac} \delta^3(\mathbf{x} - \mathbf{y}) \tag{9.77}$$

Let us assume that we can invert this equation. (This statement, as we shall shortly see, is actually incorrect.) Inverting this equation, we find the solution for f^a:

$$f^a(x) = g \int d^3y \, \Delta^{ab}(x, y; A) f^{bcd} A_k^c(y) E_k^d(y) \tag{9.78}$$

where:

$$\Delta^{ab}(x, y; A) = \frac{\delta^{ab}}{4\pi |\mathbf{x} - \mathbf{y}|} + g \int d^3z \frac{1}{4\pi |\mathbf{x} - \mathbf{z}|} f^{acb} A_k^c \nabla_k \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} + \cdots \quad (9.79)$$

Let us now insert this expression back into the Hamiltonian, where we have explicitly solved for the Gauss's Law constraint. The Hamiltonian becomes:

$$H = \frac{1}{2} \int d^3x \left[(E_i^{aT})^2 + (B_i^a)^2 + (\nabla_i f^a)^2 \right] \quad (9.80)$$

and the generating functional becomes:

$$\begin{aligned} Z(J) \;=\; & \int DE_i^a \, DA_i^a \, \delta\left(\nabla_i E_i^a\right) \delta\left(\nabla_i A_i^a\right) \\ & \times \exp\left(i \int d^4x \left[E_k^a \dot{A}_k^a - \frac{1}{2}(E_k^a)^2 - \frac{1}{2}(B_k^a)^2 - \frac{1}{2}(\nabla_k f^a)^2 - A_k^a J_k^a \right] \right) \end{aligned} \quad (9.81)$$

From this, we can read off the Feynman rules for the Yang–Mills theory in the Coulomb gauge. We have written the path integral entirely in terms of the physical, transverse states. The theory is hence unitary, since all longitudinal modes with negative norm have been explicitly removed.

There is, however, another form for the Coulomb path integral that is more covariant looking and whose Feynman rules are easier to work with. Both forms of the Coulomb path integral, of course, yield the same S matrix. If we start with the original action in Eq. (9.60) written in terms of the auxiliary field $F_{\mu\nu}$, then the path integral can be written as:

$$Z(J) = \int DA_\mu^a \, DF_{\mu\nu}^a \, \delta\left(\nabla_i A_i^a\right) \Delta_{FP}(A) \exp\left(iS(A, F) + i \int d^4x \, J^\mu A_\mu \right) \quad (9.82)$$

where the Faddeev–Popov factor for the Coulomb gauge $\nabla_i A_i^a = 0$ can be written as:

$$\Delta_{FP}(A) = \det\left(\frac{\delta \nabla_i A_i^\Omega(x)}{\delta \Omega(y)} \right)_{x,y} = \det\left[M^{ab}(x, y) \right] \quad (9.83)$$

where:

$$M^{ab}(x, y) = D^{ab}(A) \delta^4(x - y) \quad (9.84)$$

To calculate the Feynman rules from this action, we simply note that we can write:

$$M^{ab} = \nabla^2(\delta^{ab} + L^{ab}) \tag{9.85}$$

where:

$$L^{ab} = gf^{abc}\frac{1}{\nabla^2}A_i^c\nabla_i \tag{9.86}$$

Now we use the expression:

$$\begin{aligned}
\det M &= \exp(\mathrm{Tr}\,\log M) = \left[\det\nabla^2\right](\det(1+L)) \\
&\sim \exp\left[\mathrm{Tr}\,\log(1+L)\right] \\
&= \exp\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n} \\
&\times \int d^4x_1\, d^4x_2\cdots d^4x_n\ \mathrm{Tr}\,L(x_1,x_2)L(x_2,x_3)\cdots L(x_n,x_1)
\end{aligned} \tag{9.87}$$

where we have thrown away a factor of $\det\nabla^2$, which contributes closed loops that do not couple to anything.

In summary, we have shown that there are two equivalent ways of writing the Coulomb gauge. The first, with all redundant ghost modes removed, is explicitly unitary (but difficult to calculate with). The second form, although not manifestly unitary, has a covariant form. The only correction to this covariant form is the determinant factor, which we can see from the previous expression consists of nothing but closed loops coupled to the A_μ field. Thus, from a calculational point of view, the only correction to the Feynman rules is to insert closed loops into the theory connected to gauge fields.

9.5 The Gribov Ambiguity

There is, however, one tricky point that we have glossed over. As we mentioned earlier, the operator $D^{ab}(A)$ in Eq. (9.76) is actually not invertible, and hence the Coulomb gauge does not, in fact, completely eliminate the gauge degree of freedom of the theory.

This unexpected result is due to the fact that the Coulomb gauge $\nabla_i A_i^a = 0$ does not uniquely fix the gauge; that is, it is possible to find another A_i', which is

gauge equivalent to A_i, that satisfies:

$$\nabla_i A_i' = 0 \tag{9.88}$$

To see this, we write A_i' out in more detail:

$$A_i' = \Omega A_i \Omega^{-1} - \frac{i}{g}(\partial_i \Omega)\Omega^{-1} \tag{9.89}$$

The important point is that A_i' contains a term proportional to $1/g$. Perturbation theory, as a power expansion in g, will never pick up this factor. As an example, take the group $SU(2)$, and take A_i to be gauge equivalent to the number 0:

$$A_i = -\frac{i}{g}(\partial_i \Omega)\Omega^{-1} \tag{9.90}$$

with $\nabla_i A_i^a = 0$. If the Coulomb gauge were a good gauge, then the only solution of this equation should be $A_i = 0$. However, this is not so. For example, we can parametrize the previous gauge using radial coordinates:

$$\Omega = \cos\omega(r)/2 + i\boldsymbol{\sigma}\cdot\mathbf{n}\sin\omega(r)/2 \tag{9.91}$$

where $n^i n^i = 1$ and $n^i = x^i/r$.

Then the Coulomb gauge condition becomes:

$$\frac{d^2\omega}{dt^2} + \frac{d\omega}{dt} - \sin 2\omega = 0; \quad t = \log r \tag{9.92}$$

which is the equation of a damped pendulum in a constant gravitational field. If $\omega = 0$, then $A_i = 0$, and this is the solution we desire. The problem, however, is that there are obviously many other solutions to this equation other than $\omega = 0$, so the uniqueness of the Coulomb gauge-fixing procedure is violated. For a nonsingular solution, we want $\omega = 0, 2\pi, 4\pi, \ldots$ at $t = -\infty$. But then the pendulum can either fall clockwise or counterclockwise many times and then eventually wind up at a position of stable equilibrium at $\omega = -\pi$. For these nontrivial solutions at $t \to \infty$, this means that we have the asymptotic condition:

$$\Omega \to \pm i\frac{\sigma^i x^i}{r} \tag{9.93}$$

Because the Coulomb gauge does not uniquely fix the gauge, we will, in general, find an infinite sequence of identical copies, each related by a gauge transformation, each satisfying the Coulomb constraint. We will call these *Gribov*

copies.[8] Another way of saying this is that $M^{ab}(x, y; A)$ is a matrix that has a zero eigenvalue. Not only do we have nonzero eigenvalues λ_n:

$$M^{ab}\psi_n^b(x) = \lambda_n \psi_n^a \tag{9.94}$$

but we also have eigenfunctions with zero eigenvalue:

$$M^{ab}\phi_m^b(x) = 0 \tag{9.95}$$

Since the determinant of M can be written as the product of its eigenvalues:

$$\det M^{ab} = \prod_n \lambda_n \prod_m 0 = 0 \tag{9.96}$$

the determinant itself is zero and hence M is noninvertible. Thus, M^{ab} cannot be inverted. As a consequence, the formulas we have given for the Coulomb gauge are actually slightly incorrect. The presence of these zero eigenvalue functions $\phi_n(x)$ spoils the inversion process, and hence spoils the Coulomb gauge fixing. Thus the canonical quantization program for gauge theory, based on quantizing the physical fields, does not exist, technically speaking.

Although we cannot fully fix the gauge with the choice $\nabla_i A_i^a = 0$, we can still salvage our calculation in the Coulomb gauge. In fact, we can show, to any order in perturbation theory, that these zero eigenvalues do not affect our perturbative results. The reason we can ignore these states with zero eigenvalues is that they do not couple to the physical Hilbert space. For example, we can show:

$$\langle \nabla^2 f^a | \phi_n^a \rangle = 0 \tag{9.97}$$

To see how this helps us, let us construct a modified propagator:

$$D^{ab}(A)\tilde{G}^{bc}(x, y; A) = \delta^{ac}\delta^3(\mathbf{x} - \mathbf{y}) - \sum_n \phi_n^a(x)\phi_n^{c*}(y) \tag{9.98}$$

With these zero modes explicitly subtracted out, we can define the inverse operator $\tilde{G}^{ab}$ without any problems. The perturbation theory with $\tilde{G}^{ab}$ can be used because:

$$\tilde{G}\nabla^2 f = \Delta \nabla^2 f \tag{9.99}$$

because Δ [in Eq. (9.77)] and $\tilde{G}$ only differ by the zero eigenvalues ϕ_n^a, which vanish when contracted onto $\nabla^2 f$.

The moral of this exercise is that, although the Coulomb gauge is riddled with Gribov copies, we can safely ignore them as long as one stays within perturbation

theory. These Gribov copies carry an additional $1/g$ dependence beyond the first copy, and hence cannot be picked up in perturbation theory. Once one leaves perturbation theory to discuss nonperturbative phenomena, such as confinement, then presumably these Gribov copies become important.

The fact that these Gribov copies can be ignored in perturbation theory is fortunate, because the Faddeev–Popov quantization procedure becomes difficult to work with in the presence of these copies. To see how the Faddeev–Popov program is modified, let us enumerate the Gribov copies with an index n. Then are an infinite number of solutions to the equation:

$$\nabla_i A_i^{a(n)} = 0 \tag{9.100}$$

The Faddeev–Popov determinant presumably becomes modified as follows:

$$\sum_n \Delta_{FP}^{-1}[A_\mu^{a(n)}] = \int D\Omega \delta\left(\nabla_i A_i^\Omega\right) \tag{9.101}$$

Then the generating functional becomes:

$$\begin{aligned} Z(J) &= \int DA_\mu \left(\sum_n \Delta_{FP}^{-1}(A_\mu^{a(n)})\right)^{-1} \delta\left(\nabla_i A_i^{a(n)}\right) \\ &\quad \times \exp\left(iS + i\int d^4x\, J^\mu A_\mu\right) \end{aligned} \tag{9.102}$$

It is quite difficult to extract Feynman rules from a path integral as complicated as this. Thus, the standard ghost/loop interpretation of the Faddeev–Popov factor is now lost, and we have difficulty setting up the Feynman rules. However, as we have stressed, in perturbation theory the part analytic at $g = 0$ is only a function of the first Gribov copy, and so we can throw away all the infinite contributions from the ambiguity as long as we stick to perturbation theory.

9.6 Equivalence of the Coulomb and Landau Gauge

In general, the Green's functions of a gauge theory are dependent on the gauge. However, because the Green's functions are not directly measurable, this does not cause any harm. But the S matrix, because it is, by definition, measurable, should be independent of the gauge.

We will now present a functional proof that the S matrix is independent of the gauge. We will first establish how the generator of the Green's functions changes when we go from the Coulomb gauge to the Landau gauge, and then show that these modifications vanish when we go on-shell. We stress that our proof easily generalizes to the arbitrary case, proving that the S matrix for gauge theory is gauge independent. This will prove useful in the next chapter, where we discuss gauge theory in different gauges. In one gauge, the "unitary" gauge, the theory is unitary but not manifestly renormalizable. In the other, the "renormalizable" gauge, the theory is renormalizable but not manifestly unitary. Because the S matrix is gauge independent, this will show that the theory is both unitary and renormalizable.

Our starting point is the generator of the Green's functions in the Coulomb gauge:

$$Z_C(J) = \int DA_\mu \, \Delta_C[A_\mu] \prod_x \delta(\nabla_i A_i(x)) \exp\left(iS[A_\mu] + i\int d^4x \, J^\mu A_\mu\right) \tag{9.103}$$

where $\Delta_C[A_\mu]$ is the Faddeev–Popov measure coming from the Coulomb gauge. Using functional methods, we will now change the gauge to the Landau gauge. We will use the fact that the number "1" can be written, as usual, as:

$$1 = \Delta_L[A_\mu] \int D\Omega(x) \prod_x \delta\left(\partial^\mu A^\Omega_\mu(x)\right) \tag{9.104}$$

where this expression is written in the Landau gauge.

We will now insert the number "1" (written in the Landau gauge) into the expression for the generator of Green's functions (written in the Coulomb gauge). This insertion does not change anything, so the generator becomes:

$$\begin{aligned} Z_C(J) \;=\; & \int DA_\mu \Big(\int D\Omega(x) \prod_x \delta\left(\partial^\mu A^\Omega_\mu(x)\right) \Delta_L[A_\mu] \Big) \Delta_C[A_\mu] \\ & \times \prod_x \delta(\nabla_i A_i) \exp\left(iS(A_\mu) + i\int d^4x J^\mu A_\mu\right) \end{aligned} \tag{9.105}$$

The insertion is contained within the first set of parentheses. Next, in order to remove the Ω in the previous expression, we will make an inverse gauge transformation on A_μ:

$$A_\mu \to A^{\Omega^{-1}}_\mu \quad . \tag{9.106}$$

We will now use the fact that $S(A_\mu)$, the Faddeev–Popov determinant, and the measure DA_μ are all gauge independent. The term $\delta\left(\partial^\mu A^\Omega_\mu\right)$ loses its dependence

on Ω, as desired. Thus, all terms involving the Landau gauge lose their dependence on Ω. On the other hand, we must carefully keep track of those terms in the Coulomb gauge that are affected by this transformation. Then the expression for $Z_C(J)$ becomes:

$$\begin{aligned} Z_C(J) &= \int DA_\mu \, \Delta_L(A_\mu) \prod_x \delta\left(\partial^\mu A_\mu\right) \exp\left[iS(A_\mu)\right] \\ &\times \left(\Delta_C(A_\mu) \int D\Omega(x) \prod_x \delta\left(\nabla_i A_i^{\Omega^{-1}}\right)\right) \exp\left(i \int d^4x J^\mu A_\mu^{\Omega^{-1}}\right) \end{aligned} \tag{9.107}$$

Almost all the dependence on the Coulomb gauge is now concentrated within the first set of large parentheses. Our next step is to show that the factor appearing within these large parentheses can be set equal to one, thereby eliminating almost all trace of the Coulomb gauge. To analyze the term within these parentheses, we first observe that the integration over $D\Omega$ contains the delta function, which forces us to pick out a particular value for Ω^{-1}, which we will call Ω_0. This specific value of Ω_0, because of the delta function constraint, must satisfy:

$$\nabla_i A_i^{\Omega_0} \equiv 0 \tag{9.108}$$

In other words, we can rewrite the delta function term as:

$$\int D\Omega \prod_x \delta\left(\nabla_i A_i^{\Omega^{-1}}\right) = \Delta_0 \int D\Omega \delta\left(\Omega^{-1} - \Omega_0\right) \tag{9.109}$$

where Δ_0 and Ω_0 are the terms that we must calculate. In general, this task is not an easy one, because Ω_0 is a function of A_i itself. However, we can determine Ω_0 because it must, by construction, satisfy:

$$\nabla_i A_i^{\Omega_0} = \nabla_i \left\{\Omega_0 \left[A_i - ig^{-1}\Omega_0^{-1}\nabla_i \Omega_0\right] \Omega_0^{-1}\right\} = 0 \tag{9.110}$$

This expression cannot be solved exactly for $A_i^{\Omega_0}$. However, we can always power expand this expression to find the perturbative solution, which is given by:

$$A_i^{\Omega_0} = \left(\delta_{ij} - \nabla_i \frac{1}{\nabla^2} \nabla_j\right) A_j + O(A^2) \tag{9.111}$$

Thus, to any order of accuracy, we can solve for $A_\mu^{\Omega_0}$ and Ω_0. The advantage of introducing this expression for Ω_0 is that we can now show that the object in the

first set of large parentheses in Eq. (9.107) equals one:

$$\Delta_C(A_\mu)\int D\Omega \prod_x \delta\left(\nabla_i A_i^{\Omega^{-1}}\right) = \Delta_C(A_\mu)\det\left(\frac{\delta\left(\nabla_i A_i^{\Omega^{-1}}(x)\right)}{\delta\Omega(y)}\right)_{x,y}$$
$$= \Delta_C(A_\mu)\Delta_C^{-1}[A_\mu^{\Omega_0}] \tag{9.112}$$

where we have used the fact that Δ_{FP} is gauge invariant. Thus, the term in the large parentheses can be set equal to one, and the expression for $Z_C(J)$ can be written as:

$$Z_C(J) = \int DA_\mu\, \Delta_L(A_\mu)\prod_x \delta\left(\partial^\mu A_\mu\right)\exp\left(iS(A_\mu) + i\int d^4x\, J^\mu A_\mu^{\Omega_0}\right) \tag{9.113}$$

The important point here is that we have lost all dependence on the Coulomb gauge, except for the term $J^\mu A_\mu^{\Omega_0}$. This means that the Green's functions, as expected, are all gauge dependent. The next task is to show that the S matrix is independent of the gauge choice, even if the Green's functions are gauge dependent. To do this, we must extract out the dependence on Coulomb gauge parameters from the Landau gauge parameters. The source J_μ in the functional, because it originally came from the Coulomb gauge, satisfies:

$$J_0 = \nabla_i J_i = 0 \tag{9.114}$$

We may therefore write:

$$\int d^4x\, J^\mu A_\mu^{\Omega_0} = \int d^4x\, J^\mu F_\mu(A) \tag{9.115}$$

where F is defined, to lowest order, as:

$$F_\mu(A) = A_\mu(x) + O(A_\mu^2) \tag{9.116}$$

Now let us extract out all dependence on the Coulomb gauge out of the generator of the Green's function:

$$Z_C(J) = \exp\left[i\int d^4x\, J^\mu F_\mu\left(-i\frac{\delta}{\delta j_\nu}\right)\right] Z_L(j)\big|_{j=0} \tag{9.117}$$

We now have a functional expression relating the generating functional of the Landau gauge to the generating functional of the Coulomb gauge. Next, we must investigate the dependence of the S matrix on the gauge-dependent parameters.

In order to compare our results for the S matrix, we must go on-shell; that is, i.e., we must set the $p^2 \to 0$ on the external legs of any graph. Then, if we look at the contribution of the F term to any Feynman graph in this on-shell limit, the only terms that survive are self-energy corrections (i.e., radiative corrections to the propagator). In Chapter 7, we showed that the net effect of self-energy corrections is to give us a multiplicative renormalization of the overall diagram. This renormalization, in turn, can be absorbed in the overall renormalization of the S matrix that must always be performed when calculating radiative corrections. Thus, the S matrices calculated in the Coulomb or Landau gauges only differ by a multiplicative constant, given by the renormalization of the self-energy parts, which in turn can be absorbed into the overall renormalization of the S matrix. We therefore find that the two formalisms give the same on-shell S matrix, as desired.

In summary, we have seen how the path integral method gives us a convenient formalism in which to quantize gauge theory. The only complications are the Faddeev–Popov ghosts, which arise from the functional measure of integration. The power of the path integral method is that we can rapidly move from one gauge choice to another in gauge theory. This is crucial in order to show that gauge theory is both unitary and renormalizable.

In the next chapter, we will construct a realistic theory of the weak interactions from gauge theory. The essential ingredient will be spontaneous symmetry breaking, which allows us to have massive vector mesons without spoiling renormalization.

9.7 Exercises

1. Choose the gauge $A_1^a = A_2^a$. Is this a legitimate gauge? If so, construct the Faddeev–Popov ghost term.

2. Quantize the theory in the gauge:

$$\partial_\mu A^\mu + cA_\mu A^\mu = 0 \tag{9.118}$$

   Calculate the Faddeev–Popov ghosts.

3. Repeat the analysis for the gauge:

$$\partial_\mu A_0 \partial^\mu A_0 = 0 \tag{9.119}$$

4. Calculate the propagator in the gauge $\eta_\mu A^\mu = 0$, where η_μ is a constant and normalized to $\eta_\mu^2 = 1$.

5. Analyze whether the axial gauge $A_3 = 0$ suffers from a Gribov ambiguity or not. (Hint: see whether the gauge constraint is invertible or not.) Do other gauges suffer from a Gribov ambiguity? Discuss the delicate points involved with this problem.

6. A spin-3/2 field ψ_μ has both vector and spinor indices (we will suppress the spinor index). Its action is given by:

$$\mathscr{L} = \epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\partial_\rho\psi_\sigma \tag{9.120}$$

Show that it is invariant under a local gauge symmetry $\delta\psi_\mu = \partial_\mu\alpha$, where α is a spinor. Break the gauge by adding $-\frac{1}{2}(\bar{\psi}\cdot\gamma)\not{\partial}(\gamma\cdot\psi)$ to the action. Show that the propagator equals $\gamma_\mu \not{k}\gamma_\nu/k^2$.

7. If the spin-3/2 field also carries an isospin index, can it be coupled to the Yang–Mills field in a gauge covariant fashion? Justify your answer.

8. Consider an $SU(N)$ Yang–Mills field coupled locally to a multiplet of massive mesons in the adjoint representation of $SU(N)$. Write down the Feynman rules for the scalar–vector interaction vertices.

9. From a canonical point of view, the purpose of the Faddeev–Popov ghost is to cancel the ghost modes coming from the Yang–Mills propagator. In the Landau gauge, prove that this cancellation occurs at the one loop level for vector meson scattering.

10. Prove Eq. (9.34).

11. Prove that Eq. (9.78) solves Eq. (9.75).

12. By repeating the same steps used to show the equivalence between the Coulomb and Landau gauge, show the equivalence between *any* two gauges allowed by the Faddeev–Popov formalism.

Chapter 10
The Weinberg–Salam Model

If my view is correct, the universe may have a kind of domain structure. In one part of the universe, you may have one preferred direction of the axis; in another part, the direction of the axis may be different.

—Y. Nambu

10.1 Broken Symmetry in Nature

In nature, a variety of beautiful and elegant symmetries surrounds us. However, there are also many examples of symmetries in nature that are broken. Rather than put explicit symmetry-breaking terms into the Hamiltonian by hand, which seems artificial and unappealing, we would like to break these symmetries in a way such that the equations retain their symmetry.

Nature seems to realize this by exploiting the mechanism of *spontaneous symmetry breaking*; that is, the Hamiltonian is invariant under some symmetry, but the symmetry is broken because the vacuum state of the Hamiltonian is not invariant. The simplest examples come from solid-state physics, where the phenomenon of spontaneous symmetry breaking is quite common. Consider a ferromagnet, such that the atoms possess a spin σ_i. Although the Hamiltonian does not select out any particular direction in space, the ground state of the theory, however, can consist of atoms whose spins are all aligned in the same direction. Thus, rotational symmetry can be broken by the vacuum state, even when the Hamiltonian remains fully symmetric. To restore the symmetry, we have to heat the ferromagnet to a high temperature T, where the atoms once again become randomly aligned.

In addition, spontaneous symmetry breakdown may also be associated with the creation of massive vector fields. In the theory of superconductivity, for example, spontaneous symmetry breaking occurs at extremely low temperatures, giving us the Meissner effect, in which magnetic flux lines are expelled from the interior of a superconductor. However, the magnetic field penetrates slightly into

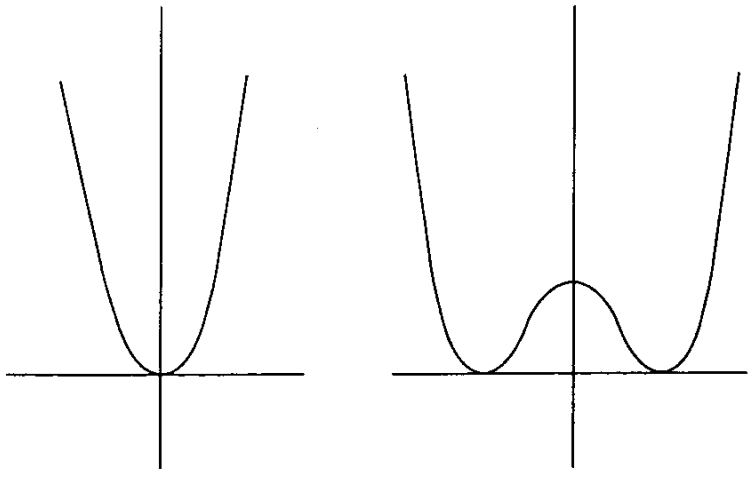


Figure 10.1. The first potential corresponds to a unique vacuum, or a Klein-Gordon field with a positive mass. The second potential, which exhibits spontaneous symmetry breaking, corresponds to a scalar field with a tachyon mass.

the medium, so there is a finite-range electromagnetic field, which corresponds to a "massive photon." We conclude that spontaneous symmetry breaking can, under certain circumstances, give mass to a massless vector field.

Similarly, spontaneous symmetry breaking gives us a solution to the problem that faced physicists trying to write down a theory of the weak interactions in 1950s and 1960s. One candidate was the massive vector meson theory. However, the massive vector meson theory was known to be nonrenormalizable by simple power counting arguments. Spontaneous symmetry breaking, however, solves this problem. It preserves the renormalizability of the original gauge theory even after symmetry breaking, giving us a renormalizable theory of massive vector mesons.

To illustrate spontaneous symmetry breaking, let us begin our discussion with a scalar field with a ϕ^4 interaction which has the symmetry $\phi \rightarrow -\phi$:

$$\mathscr{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 \tag{10.1}$$

If m^2 is negative, then m is imaginary; that is, we have tachyons in the theory. However, quantum mechanically, we can reinterpret this theory to mean that we have simply expanded around the wrong vacuum.

In Figure 10.1, we see two potentials, one described by $m^2\phi^2/2 + \lambda\phi^4/4!$ with positive m^2, which gives us a unique vacuum, and another with negative m^2, which corresponds to a tachyon mass.

For the second potential, a particle would rather not sit at the usual vacuum $\phi = 0$. Instead, it prefers to move down the potential to a lower-energy state given by the bottom of one of the wells. Thus, we prefer to power expand around the new minimum $\phi = v$.

(Classically, a rough analogy can be made with a vertical, stationary rod that is hanging from the ceiling. Normally, the rod's lowest-energy state is given by $\theta = 0$. If we displace the rod by an angle θ, then the potential resembles the first figure in Fig. 10.1. However, if the rod is sent spinning around the vertical axis, then it will reach a new equilibrium at some fixed angle $\theta = \theta_0$, and its effective potential will resemble the second figure in Fig. 10.1. For the spinning rod, $\theta = 0$ is no longer the lowest energy state.)

For the scalar particle we discussed before, if m^2 is positive, then there is only one minimum at the usual vacuum configuration $\phi = 0$ and the potential is still symmetric; that is, it has the symmetry $\phi \rightarrow -\phi$.

However, if m^2 is negative, then quantum mechanically the vacuum expectation value of the ϕ field no longer vanishes because we have chosen the wrong vacuum. For the tachyonic potential, we can easily find the location of the new minimum:

$$\phi_0 = v = \pm\sqrt{-6m^2/\lambda} \tag{10.2}$$

Normally, we demand that the vacuum expectation value of the scalar field vanishes. However, if the naive vacuum is the incorrect one, then we find instead:

$$\langle 0|\phi|0\rangle = v \tag{10.3}$$

Clearly, our troubles have emerged because we have power expanded around the wrong vacuum. To correct this situation, we must shift the value of the ϕ field as follows:

$$\tilde{\phi} = \phi - v \tag{10.4}$$

In terms of the shifted field $\tilde{\phi}$, we now have broken the original symmetry $\phi \rightarrow -\phi$ and:

$$\langle 0|\tilde{\phi}|0\rangle = 0 \tag{10.5}$$

We also have a new action given by:

$$\frac{1}{2}\partial_\mu\tilde{\phi}\partial^\mu\tilde{\phi} + m^2\tilde{\phi}^2 - \frac{1}{6}\lambda v\tilde{\phi}^3 - \frac{1}{4!}\lambda\tilde{\phi}^4 \tag{10.6}$$

Because m^2 is negative, we have an ordinary scalar particle with a positive mass squared given by $-2m^2$. The original symmetry between ϕ and $-\phi$ has now been spontaneously broken because the field has been shifted and there is a new vacuum.

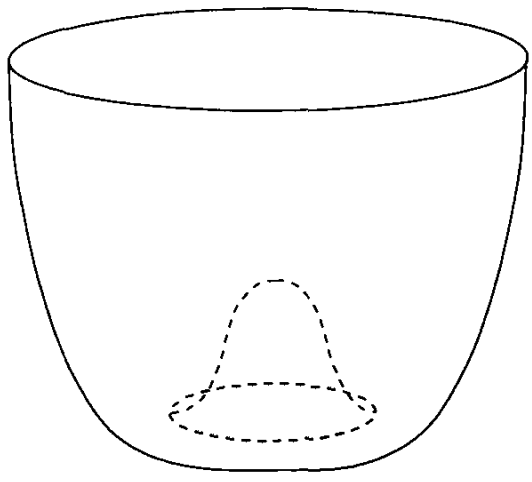


Figure 10.2. The new solution to the minimum of the potential corresponds to a ring of solutions.

Let us now examine a less trivial example given by the global $O(N)$ scalar theory, where the field ϕ^i transforms as an N-vector:

$$\mathscr{L} = \frac{1}{2}\partial_\mu \phi^i \partial^\mu \phi^{i\mu} - \frac{1}{2}m^2 \phi^i \phi^i - \frac{\lambda}{4!}(\phi^i \phi^i)^2 \tag{10.7}$$

Again, if the mass term has the wrong sign, (i.e., if m^2 is negative), then there is a new vacuum solution given by:

$$\phi^i \phi^i = v^2 = -6m^2/\lambda \tag{10.8}$$

Contrary to the previous potential that we examined, which had two minima, this theory has an infinite number of vacua. In Figure 10.2, we can see that there is actually a degenerate ring of solutions sitting in the bottom of the potential well.

Any solution of this equation yields a new vacuum. Now let us break this degeneracy by singling out a specific direction in isospin space, the last entry in the column vector:

$$\langle \phi \rangle_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ v \end{pmatrix} = (v_i) \tag{10.9}$$

where $v_i = 0$ if $i = 1, \cdots, N-1$ and $v_N = v$.

This new vacuum is still invariant under the group that leaves the last entry $v_N = v$ invariant, that is, the subgroup $O(N-1)$. We are thus breaking the original symmetry group $O(N)$ down to $O(N-1)$ with this choice of vacuum.

Let τ^i be the generators of $O(N)$. These generators, in turn, can be divided into two groups: the generators of $O(N-1)$, and everything else. Let $\tilde{\tau}^j$ equal these remaining generators. The simplest parametrization of this new vacuum is to let the last component of the ϕ^i field develop a nonzero vacuum expectation value: $\phi_N \rightarrow \phi_N + v$, keeping the other ϕ_i the same. However, we will find it convenient to introduce a slightly more complicated parametrization of the N fields within the vector ϕ^i. We will choose instead:

$$\phi = e^{i\xi_j \tilde{\tau}_j / v} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ v + \sigma(x) \end{pmatrix} \tag{10.10}$$

where we have replaced the original N fields ϕ^i with a new set, given by $N-1$ fields ξ_i and by σ.

To see the reason for this particular parametrization, we recall that $O(N)$ has $(1/2)N(N-1)$ generators. The number of generators in $O(N)$, after we have subtracted out the generators of $O(N-1)$, is:

$$\frac{1}{2}N(N-1) - \frac{1}{2}(N-1)(N-2) = N-1 \tag{10.11}$$

Thus, there are $N-1$ generators $\tilde{\tau}^i$ that are not generators of the $O(N-1)$ subgroup.

A particularly useful parametrization for the generators of $O(N)$ is given by a series of delta functions:

$$\begin{aligned} (\tau_{ij})_{kl} &= -i(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) \\ (\tilde{\tau}_{iN})_{kl} &= -i(\delta_{ik}\delta_{Nl} - \delta_{il}\delta_{Nk}) \end{aligned} \tag{10.12}$$

To lowest order, we have $\phi_i = \xi_i$ for $i < N$ and $\phi_N = v + \sigma$, and the action becomes:

$$\begin{aligned} \mathscr{L} &= \frac{1}{2}\left(\partial_\mu \sigma \partial^\mu \sigma + \partial_\mu \xi_i \partial^\mu \xi_i\right) \\ &\quad - \frac{1}{2}m^2(v+\sigma)^2 - \frac{1}{4!}\lambda(v+\sigma)^4 + \text{higher terms} \end{aligned} \tag{10.13}$$

The $N-1$ ξ_i fields have become massless, and σ is now massive but no longer a tachyon. The action is still invariant under a residual symmetry, $O(N-1)$.

We emphasize that the number of massless bosons ξ_i is equal to the number of broken generators $\tilde{\tau}^i$. We call these massless bosons, which signal the spontaneous

breakdown of the theory, *Nambu–Goldstone bosons.*[1–3] These bosons will play a special role in constructing realistic gauge theories of the weak interactions.

In summary, we have:

$$O(N) \to O(N-1) \;\Rightarrow\; N-1 \text{ Nambu–Goldstone bosons} \tag{10.14}$$

that is, the breaking of $O(N)$ symmetry down to $O(N-1)$ symmetry leaves us with $N-1$ Nambu–Goldstone bosons ξ_i, or one boson for each of the broken generators $\tilde{\tau}^i$.

10.2 The Higgs Mechanism

We now have a rule that the number of massless Nambu–Goldstone bosons equals the number of broken generators of the theory. Let us discuss a more general action to see if this result still holds:

$$\frac{1}{2}\partial_\mu \phi \cdot \partial^\mu \phi - V(\phi) \tag{10.15}$$

where ϕ transforms under some representation of a group G, which has N generators.

Let us say that there is a nontrivial vacuum, which we can find by calculating the minimum of the potential:

$$\left.\frac{\delta V}{\delta \phi_i}\right|_{\phi=v} = 0 \tag{10.16}$$

Let $\langle \phi_i \rangle_0 = v_i$ be a solution of this equation that minimizes the potential. We find that this new vacuum is still invariant under a subgroup of G, called H, which has M generators. This means that there are M generators $\bar{L}^a_{ij}$ that leave v_i unchanged; that is, they satisfy:

$$\bar{L}^a_{ij} v_j = 0 \tag{10.17}$$

There are also $N-M$ generators $\tilde{L}^a$ for which $\tilde{L}^a_{ij} v_j \neq 0$. In other words, the generators L^a_{ij} are of two types: $\bar{L}^a_{ij}$, which generate the subgroup H, and $\tilde{L}^a_{ij}$, which are all the remaining generators. The first set of generators annihilate on v_i, by construction, while the second set does not.

Next, we wish to show that expanding around this new vacuum will create $N-M$ massless boson fields. To begin, we define the variation of the scalar field

as:

$$\delta\phi = -i\theta^a L^a \phi \tag{10.18}$$

The potential $V(\phi)$ is invariant under this transformation, so we have:

$$\frac{\delta V}{\delta \phi_i}\delta\phi_i = -i\frac{\delta V}{\delta \phi_i}\theta^a L^a_{ij}\phi_j = 0 \tag{10.19}$$

Since the parameters θ^a are arbitrary, we have N equations:

$$\frac{\delta V}{\delta \phi_i} L^a_{ij}\phi_j = 0 \tag{10.20}$$

Differentiating this, we arrive at:

$$\frac{\delta^2 V}{\delta\phi_i\phi_k} L^a_{ij}\phi_j + \frac{\delta V}{\delta\phi_i} L^a_{ik} = 0 \tag{10.21}$$

Substituting the value of ϕ at the minimum into the previous equation, we find:

$$\left.\frac{\delta^2 V}{\delta\phi_i\delta\phi_k}\right|_{\phi=v} L^a_{ij} v_j = 0 \tag{10.22}$$

If we Taylor expand the potential around the new minimum $\phi_i = v_i$, then the mass matrix can be defined as:

$$V(\phi) = -\frac{1}{2}(M^2)_{ij}(\phi - v)_i(\phi - v)_j + \cdots \tag{10.23}$$

Now let us insert the previous equation for the mass matrix into Eq. (10.22). This gives us:

$$(M^2)_{ij} L^a_{jk} v_k = 0 \tag{10.24}$$

This equation is trivially satisfied if L^a is a generator of the subgroup H, since $\tilde{L}^a_{ij} v_j = 0$. The situation is more interesting, however, when the generator is one of the $N - M$ generators $\tilde{L}^a_{ij}$ of G that are outside H.

For these $N - M$ generators $\tilde{L}^a_{ij}$, Eq. (10.24) is an eigenvalue equation for the matrix M^2. It states that for each of the $N - M$ generators $\tilde{L}^a_{ij}$, there is a zero eigenvalue of the M^2 matrix. Since the eigenvalues of the M^2 matrix give us the mass spectrum of the fields, we have $N - M$ massless bosons in the theory.

Thus, after symmetry breaking there are $N - M$ massless Nambu–Goldstone bosons, one for each broken generator.

In summary:

$$G \to H \;\Rightarrow\; N - M \text{ Nambu–Goldstone bosons} \tag{10.25}$$

A surprising feature of this method, however, arises when we apply the Nambu–Goldstone theorem to gauge theories. In this case, will find that these Nambu–Goldstone bosons are "eaten up" by the gauge particles, converting massless Yang–Mills vector particles into massive ones. This is called the *Higgs–Kibble mechanism.*[4–6]

Furthermore (and this is the key point), the Yang–Mills theory remains renormalizable even after the Higgs mechanism has generated massive vector particles. In other words, this is the long-sought-after mechanism that can render a massive vector theory renormalizable. It is not an exaggeration to say that this discovery, by 't Hooft, changed the landscape of theoretical particle physics.

To see how the Higgs mechanism works, let us begin with a theory of complex scalar particles coupled to Maxwell's theory. The action is:

$$\mathscr{L} = D_\mu \phi^* D^\mu \phi - m^2 \phi^* \phi - \lambda (\phi^* \phi)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{10.26}$$

The coupled system, as before, is invariant under:

$$\begin{aligned} \phi &\to e^{-i\theta(x)} \phi \\ \phi^* &\to e^{+i\theta(x)} \phi^* \\ A_\mu &\to A_\mu - \frac{1}{e} \partial_\mu \theta(x) \end{aligned} \tag{10.27}$$

The action is invariant under $U(1) = SO(2)$.

Now we break this symmetry; the new vacuum is given by:

$$\langle \phi \rangle_0 = v/\sqrt{2} \tag{10.28}$$

We will find it convenient to parametrize the complex field ϕ by introducing two fields σ and ξ:

$$\begin{aligned} \phi &= e^{i\xi/v}(v + \sigma)/\sqrt{2} \\ &= (v + \sigma + i\xi + \cdots)/\sqrt{2} \end{aligned} \tag{10.29}$$

We can now make a gauge rotation on both ϕ and A_μ:

$$\begin{aligned} \phi &\to \phi' = e^{-i\xi/v}\phi = (v + \sigma)/\sqrt{2} \\ A_\mu &\to A'_\mu = A_\mu - \frac{1}{ev} \partial_\mu \xi \end{aligned} \tag{10.30}$$

(We have gauge rotated the ξ field away in the definition of ϕ' and A'_μ.)

The final action now reads:

$$\begin{aligned}\mathscr{L} = & -\frac{1}{4}F'_{\mu\nu}F^{\mu\nu\prime} + \frac{1}{2}\partial_\mu\sigma\,\partial^\mu\sigma + \frac{1}{2}e^2v^2A'_\mu A^{\mu\prime} \\ & + \frac{1}{2}e^2(A'_\mu)^2\sigma(2v+\sigma) - \frac{1}{2}\sigma^2(3\lambda v^2 + m^2) - \lambda v\sigma^3 - \frac{1}{4}\lambda\sigma^4\end{aligned} \tag{10.31}$$

The A'_μ field has become massive. The key point is that the ξ has disappeared completely. In other words, it has been "gauged away," or "eaten up" by the vector field.

The Nambu–Goldstone boson, corresponding to the breaking of $U(1)$ invariance, has disappeared and reappeared in a new guise, as the massive component of the massless vector field. Since $U(1)$ is now broken, there is no longer any gauge symmetry left to prevent the Yang–Mills field from acquiring a mass, and hence it acquires a massive mode (at the expense of the ξ mode, which vanishes). Thus, the new field A'_μ has three components (while a massless vector field has two helicities or transverse polarizations), with one of these fields corresponding to the old ξ field. To see that the A'_μ field has gobbled up the ξ Nambu–Goldstone boson, we note that its definition was: $A'_\mu = A_\mu - (1/ev)\partial_\mu\xi$, which clearly shows that the ξ field has been incorporated into the A'_μ field.

Let us now discuss two slightly more difficult examples of the Higgs mechanism with non-Abelian gauge fields. Then we will see that only some of the gauge fields become massive. First, let us discuss a $SU(2)$ gauge theory coupled to an isovector triplet of scalar fields. We will break this down to $U(1)$, so that only one of the three gauge fields remains massless, while the other two acquire a mass by eating up the Nambu–Goldstone bosons.

The action is:

$$\mathscr{L} = \frac{1}{2}D_\mu\phi^i D^\mu\phi^i - V(\phi^i\phi^i) - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} \tag{10.32}$$

As before, we will choose $V(\phi^i\phi^i)$ such that there is a tachyon in the theory, meaning that we have chosen the wrong vacuum. When we shift to a new vacuum, we find that it is degenerate. We will break this degeneracy by shifting the third component of the isovector:

$$\langle\phi\rangle_0 = \begin{pmatrix} 0 \\ 0 \\ v \end{pmatrix}; \quad \phi = e^{i(\xi_1 L^1 + \xi_2 L^2)/v}\begin{pmatrix} 0 \\ 0 \\ v+\sigma \end{pmatrix} \tag{10.33}$$

where L^i are the generators of $SU(2)$, ξ_i are the Nambu–Goldstone bosons, and we choose the third isospin generator L^3 to be the generator of the unbroken $U(1)$ symmetry.

Next, we make a local gauge transformation on the fields, such that we remove the ξ fields appearing in the previous equation for ϕ:

$$\begin{aligned} \phi &\rightarrow \phi' = \Omega\phi \\ \mathbf{L}\cdot\mathbf{A}_\mu &\rightarrow \mathbf{L}\cdot\mathbf{A}'_\mu = \Omega\mathbf{L}\cdot\mathbf{A}_\mu\Omega^{-1} - \frac{i}{g}(\partial_\mu\Omega)\Omega^{-1} \end{aligned} \tag{10.34}$$

where:

$$\Omega = e^{-i(\xi_1 L^1 + \xi_2 L^2)/v} \tag{10.35}$$

Let us expand the action around the new vacuum. The vector fields now have the following action:

$$-\frac{1}{4}F'_{\mu\nu}F^{\mu\nu\prime} + \frac{1}{2}g^2v^2(A^1_\mu A^{1\mu} + A^2_\mu A^{2\mu}) \tag{10.36}$$

Two of the gauge fields have acquired a mass, but A^3_μ, corresponding to a residual $U(1)$ symmetry, is still massless. Again, the number of Nambu–Goldstone bosons equals the number of broken generators.

The other terms in the action are:

$$\frac{1}{2}\partial_\mu\sigma\,\partial^\mu\sigma - V(v+\sigma) + \text{higher terms} \tag{10.37}$$

The $\xi_{1,2}$ fields, as expected, have disappeared completely by being absorbed into the gauge fields.

Finally, we must now show that the Higgs mechanism works for an arbitrary gauge theory. The important feature that we want to demonstrate is that the number of gauge fields that become massive is equal to the number of broken generators of the gauge group.

We start with a gauge group G, which has N generators, and hence N gauge fields A^a_μ. We also have the real scalar field ϕ transforming under some n-dimensional representation of the group G. We start with the action:

$$\mathscr{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} + \frac{1}{2}(\partial_\mu - igL^aA^a_\mu)\phi\,(\partial^\mu - igL^bA^{\mu b})\phi - V(\phi) \tag{10.38}$$

where L^a is an $n \times n$ representation of the group G, that has N generators.

Let us choose $V(\phi)$ so that symmetry breaking occurs. Let $\phi = v$ be a matrix equation defining the minimum of the potential. We want a vacuum that is invariant under an M-dimensional subgroup H of G. The generators of H, because they leave the vacuum invariant, satisfy $L^a v = 0$.

We now parametrize the scalar field as:

$$\phi = \exp\left(i \sum_{i=1}^{N-M} \xi_i \tilde{L}^i / v\right)(v + \sigma) \tag{10.39}$$

where we sum over the $N - M$ generators that do not correspond to the subgroup H and do not annihilate v. For these generators $\tilde{L}^i v \neq 0$. Also, ξ_i are the Nambu–Goldstone bosons.

The trick is now to choose a gauge transformation which swallows up the ξ fields appearing in the above definition. As before, we choose:

$$\phi \to \phi' = \exp\left(-i \sum_{i=1}^{N-M} \xi_i \tilde{L}^i / v\right)\phi = \Omega\phi \tag{10.40}$$

The Ω in front of ϕ precisely cancels against the Ω^{-1} appearing in front of the parametrization of ϕ, so the ξ_i fields disappear.

Inserting this parametrization into the action, we can collect the terms responsible for the vector meson masses, which is given by:

$$\frac{1}{2} A^i_\mu (M^2)^{ij} A^{j\mu} = \frac{1}{2} \left(g\tilde{L}^i v | g\tilde{L}^j v\right) A^i_\mu A^{j\mu} \tag{10.41}$$

where the brackets represent the matrix contraction or scalar product between two vectors.

Thus, the masses of the gauge fields are given by the eigenvalues of the following matrix:

$$(M^2)^{ij} = g^2 \left(v | \tilde{L}^i \tilde{L}^j v\right) \tag{10.42}$$

There are $N - M$ non-zero eigenvalues to this equation, and hence there are $N - M$ massive vector fields (which have absorbed the remnants of the $N - M$ Nambu–Goldstone bosons).

In retrospect, the Higgs mechanism has a rather simple interpretation. We know that a gauge theory locally invariant under a group G has no mass term in the action for the vector field in perturbation theory. However, if the group G breaks down to a subgroup H via symmetry breaking, then we know that the gauge fields corresponding to the H subgroup must still remain massless. However, the gauge fields corresponding to the broken generators of G are now free to become massive.

Let us now present a more formal, model-independent proof of the Nambu–Goldstone theorem. We begin with the observation that spontaneous symmetry breaking occurs because the vacuum is not invariant under a certain symmetry,

although the action is. In other words, beginning with a symmetry and its conserved current J_μ, we can construct its conserved charge Q such that:

$$Q = \int d^3x J^0; \quad Q|0\rangle \neq 0 \tag{10.43}$$

The essential point is that Q no longer annihilates the vacuum in a broken theory; that is, the vacuum state $|0\rangle$ is not the true vacuum, so it is not annihilated by Q. Current conservation means that the following commutator vanishes:

$$\begin{aligned} 0 &= \int d^3x \left[\partial_\mu J^\mu(x), \phi(0)\right] \\ &= \partial_0 \int d^3x \left[J^0(x), \phi(0)\right] + \int_S d\mathbf{S} \cdot [\mathbf{J}(x), \phi(0)] \end{aligned} \tag{10.44}$$

for some scalar boson field $\phi(0)$. Then we make the assumption that for large enough surfaces S, we can ignore the term on the right-hand side of the equation. Hence:

$$\frac{d}{dt}[Q(t), \phi(0)] = 0 \tag{10.45}$$

Then:

$$\langle 0| [Q(t), \phi(0)] |0\rangle = C \neq 0 \tag{10.46}$$

where C is a nonzero constant. Now insert a complete set of intermediate states inside the commutator. After making a translation, we can write this expression as:

$$\begin{aligned} \sum_n (2\pi)^3 \delta^3(\mathbf{p}_n) \Big[&\langle 0|J^0(0)|n\rangle\langle n|\phi(0)|0\rangle e^{-iE_n t} \\ - \; &\langle 0|\phi(0)|n\rangle\langle n|J^0(0)|0\rangle e^{iE_n t}\Big] = C \neq 0 \end{aligned} \tag{10.47}$$

In general, unless $E_n = 0$, the positive and negative frequency parts cannot possibly cancel, and hence the contributions to the sum cannot give us a constant. In other words, it is impossible to satisfy the previous equation unless the mass of the intermediate states vanishes. (Furthermore, these massless states *must* exist, so that the sum adds up to a non zero value.) Thus, there must be a massless Nambu–Goldstone state in the theory, with the property that:

$$\langle n|\phi(0)|0\rangle \neq 0, \quad \langle 0|J^0(0)|n\rangle \neq 0 \tag{10.48}$$

This completes the proof.

10.3 Weak Interactions

Now that we have investigated the possibility of spontaneous symmetry breaking, we can apply this theory to the weak interactions, where it has enjoyed great success.

To appreciate this breakthrough, we must realize that, from an historical point of view, important progress in the field theory of weak interactions was relatively stagnant for many decades, from the original Fermi action[7] of the 1930s, to the overthrow of parity, to the advent of gauge theories in the 1970s.

Fermi originally tried to explain the decay of the neutron into a proton, an electron, and an anti-neutrino:

$$n \rightarrow p + e + \bar{\nu} \tag{10.49}$$

by postulating the phenomenological Lagrangian:

$$\begin{aligned} \mathscr{L} &= \sum_A \frac{G_F}{\sqrt{2}} (\bar{\Psi}_p \Gamma_A \Psi_n)(\bar{\psi}_e \Gamma^A \psi_\nu) + \text{h.c.} \\ &= \sum_A \frac{G_F}{\sqrt{2}} J^\dagger_{A,\text{had}} J^A_{\text{lept}} + \text{h.c.} \end{aligned} \tag{10.50}$$

(Experimentally, $G_F = (1.16639 \pm 0.00002) \times 10^{-5}$ GeV^{-2}.) This action, from the very start, was known to suffer from a series of diseases. First, the Γ_A matrices could in principle consist of all possible combinations of the 16 Dirac matrices. The lack of precise experimental data for years prevented a decisive determination of the action. It took many decades finally to resolve that the correct combination should be $V - A$,[8-9] rather than scalar or tensor combinations.

The Fermi action also suffered from a fatal theoretical disease: It was nonrenormalizable. Four-fermion interactions must be accompanied by a dimensionful coupling constant (since the dimension of a spinor is 3/2). The Fermi constant thus has the dimensions -2, and hence the theory was nonrenormalizable. Finally, the theory violated unitarity. If we calculate the high energy behavior of the differential cross section of any weak process, we can write (purely by dimensional arguments):

$$\frac{d\sigma}{d\Omega} \rightarrow \frac{G_F^2}{4\pi} s \tag{10.51}$$

(The differential cross section is a pure s wave because the four-fermion interaction takes place at a single point.)

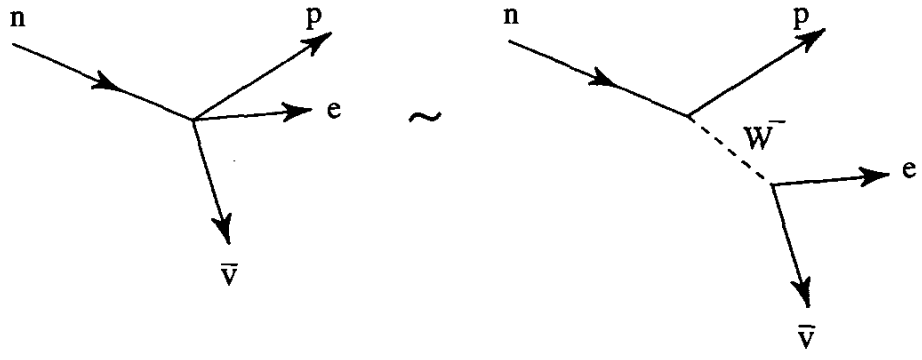


Figure 10.3. If four fermions interact via a massive vector meson, then the exchange of the vector meson, for a large mass, mimics the original Fermi four-fermion interaction.

On the other hand, we know from unitarity that the S matrix for s-wave scattering must obey the law:

$$\frac{d\sigma}{d\Omega} \to \frac{1}{s} \tag{10.52}$$

These two results are obviously in contradiction for high energies. The discrepancy between these two results becomes serious at around $\sqrt{s} \sim 500\,\text{GeV}$.

To extend the Fermi action, physicists tried to emulate the success of the Yukawa theory for the strong interactions. If the Yukawa meson could mediate the strong interactions, then perhaps a massive vector meson could mediate the weak interactions. The obvious proposal was to treat the Fermi action as a byproduct of vector meson exchange (Fig. 10.3):

$$\mathscr{L} = g_W^2(\bar{\Psi}_p \gamma^\mu \Psi_n) \frac{g_{\mu\nu} - k_\mu k_\nu / M_W^2}{k^2 - M_W^2 + i\epsilon} (\bar{\psi}_e \gamma^\nu \psi_\nu) \tag{10.53}$$

This, in turn, gave us a rough determination of the mass of the vector meson:

$$\frac{G_F}{\sqrt{2}} \sim \frac{g_W^2}{M_W^2} \tag{10.54}$$

A rough calculation put the vector meson mass at around 50-100 GeV.

The advantage of the vector meson approach was that it smoothed out the singular behavior of the original Fermi action. However, the massive vector meson theory was still nonrenormalizable. The propagator behaved as $k_\mu k_\nu / M^2 k^2$ in the ultraviolet region; so the higher loop graphs all diverged.

Meanwhile, the experimental confusion in the weak interactions continued with the discovery of three identical sets of lepton pairs, corresponding to the

electron, muon, and τ particles and their corresponding neutrino partners:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}; \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}; \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{10.55}$$

To this day, no one knows why there are three identical copies of lepton families, as well as quark families. The masses of the e, μ, and τ leptons in MeV are .51099906(15), 105.658389(34), and 1784. Their neutrino masses have upper limits given by 10 eV, .27MeV, and 35MeV.

10.4 Weinberg–Salam Model

The Weinberg–Salam model,[10,11] one of the most successful quantum theories besides the original QED, is a curious amalgam of the weak and electromagnetic interactions. Strictly speaking, it is not a "unified field theory" of the weak and electromagnetic interactions, since we must introduce two distinct coupling constants g and g' for the $SU(2)$ and $U(1)$ interactions. Nonetheless, it represents the one of the most important extensions of QED in the past quarter century.

We begin by discussing the $SU(2)$ sector. Observationally, we must incorporate a neutral, left-handed Weyl neutrino along with a Dirac electron, which can be considered to be the sum of left-handed and right-handed Weyl spinors. The left–handed fermions form an isodoublet, consisting of the Weyl neutrino and electron:

$$L \equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \tag{10.56}$$

while the right-handed sector consists of an isosinglet, the right-handed electron:

$$R \equiv (e)_R \tag{10.57}$$

This curious feature, that the electron is split into two parts, with the left- and right-handed sectors transforming differently, is a consequence of the fact that the weak interactions violate parity and are mediated by $V - A$ interactions.

These two lepton sectors transform under $SU(2)$ in different ways:

$$\begin{aligned} L &\rightarrow e^{(i/2)\boldsymbol{\theta}\cdot\boldsymbol{\sigma}} L \\ R &\rightarrow R \end{aligned} \tag{10.58}$$

Now let us examine the transformation of these fields under:

$$\begin{aligned} L &\rightarrow e^{(i/2)\beta} L \\ R &\rightarrow e^{i\beta} R \end{aligned} \tag{10.59}$$

R and L transform slightly differently under the $U(1)$ transformation.

The charge Q of the pair (ν_e, e) is $(0, -1)$, which is almost equal to the eigenvalue of the isospin operator T^3. In fact, the correct formula for the charge is:

$$Q = T^3 + \frac{Y}{2} \tag{10.60}$$

where $T^3 = \pm 1/2$ and $Y = -1$ for the left doublets, and $T^3 = 0$ and $Y = -2$ for the right singlets. Thus, we need both the $SU(2)$ and $U(1)$ sectors to get the charge correctly.

The final action consists of three parts. $\mathscr{L}_1$ is the gauge part; $\mathscr{L}_2$ is the fermionic part; and $\mathscr{L}_3$ is the scalar Higgs sector:

$$\mathscr{L} = \mathscr{L}_1 + \mathscr{L}_2 + \mathscr{L}_3 \tag{10.61}$$

where:

$$\begin{aligned} \mathscr{L}_1 &= -\frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ \mathscr{L}_2 &= i\bar{R}\gamma^\mu D_\mu R + i\bar{L}\gamma^\mu D_\mu L \\ \mathscr{L}_3 &= D_\mu \phi^\dagger D^\mu \phi - m^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \\ &\quad + G_e(\bar{L}\phi R + \bar{R}\phi^\dagger L) \end{aligned} \tag{10.62}$$

where:

$$\begin{aligned} W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W^b_\mu W^c_\nu \\ F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ D_\mu R &= (\partial_\mu + i g' B_\mu) R \\ D_\mu L &= \left[\partial_\mu + (i/2) g' B_\mu - (i/2) g \sigma_i W^i_\mu\right] L \\ D_\mu \phi &= \left[\partial_\mu - (i/2) g \sigma_i W^i_\mu - (i/2) g' B_\mu\right] \phi \end{aligned} \tag{10.63}$$

The scalar multiplet is a complex isodoublet given by:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{10.64}$$

where the doublet has charge $(1, 0)$, which can be given by $Q = T^3 + (1/2)Y$, such that $T^3 = \pm 1/2$ and $Y = 1$.

Symmetry breaking is induced by:

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{10.65}$$

After symmetry breaking, the fields W_μ^a and B_μ recombine and reemerge as the physical photon field A_μ, a neutral massive vector particle Z_μ, and a charged doublet of massive vector particles $W_\mu^\pm$:

$$\begin{aligned} Z_\mu &= \frac{g W_\mu^3 + g' B_\mu}{(g^2 + g'^2)^{1/2}} \equiv \cos\theta_W W_\mu^3 + \sin\theta_W B_\mu \\ A_\mu &= \frac{-g' W_\mu^3 + g B_\mu}{(g^2 + g'^2)^{1/2}} \equiv -\sin\theta_W W_\mu^3 + \cos\theta_W B_\mu \\ W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \pm i W_\mu^2) \end{aligned} \tag{10.66}$$

where the Weinberg angle θ_W is defined via:

$$\begin{aligned} \cos\theta_W &= \frac{g}{(g^2 + g'^2)^{1/2}} \\ \tan\theta_W &= \frac{g'}{g} \end{aligned} \tag{10.67}$$

By examining the mass sector, we can read off the masses of the resulting vector particles:

$$\begin{aligned} M_{W_1}^2 &= M_{W_2}^2 \\ M_Z^2 &= \frac{M_W^2}{\cos^2\theta_W} \\ M_A &= 0 \end{aligned} \tag{10.68}$$

Finally, the electric charge emerges as:

$$e = g \sin\theta_W \tag{10.69}$$

Experimentally, the predictions of the Weinberg–Salam model have been tested to about one part in 10^3 or 10^4. At present, we have the following values for the parameters of the Weinberg–Salam model[12]:

$$\begin{aligned} \sin^2\theta_W &= 0.2325 \pm 0.008 \\ M_Z &= 91.173 \pm 0.020\,\mathrm{GeV/c^2} \\ M_W &= 80.22 \pm 0.26\,\mathrm{GeV/c^2} \end{aligned} \tag{10.70}$$

The Weinberg-Salam model has been one of the outstanding successes of field theory, gradually rivalling the predictive power of QED. The rest of this chapter will be devoted to studying the many consequences of the model.

10.5 Lepton Decay

Let us now use the Weinberg–Salam model to do some simple calculations, such as the decay of the muon or the τ lepton to lowest order. Although the Born term resembles the calculation that one might perform with the old massive vector meson theory, the Weinberg–Salam model allows us to calculate quantum corrections to the massive vector meson theory that we can then compare with experiment.

We are interested in purely leptonic decays, such as:

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \tag{10.71}$$

More generally, we can have:

$$l_a(p_1, s_a) \rightarrow \bar{\nu}_b(p_2) + \nu_a(p_3) + l_b(p_4, s_b) \tag{10.72}$$

where $p_1 = p_2 + p_3 + p_4$ and a and b represent lepton generations. Thus, (l_a, ν_a) and (l_b, ν_b) form two lepton generations.

A straightforward use of the Feynman rules for the Weinberg–Salam model yields, to lowest order:

$$\mathscr{M} = \frac{-ig^2}{8}\left(\bar{u}_3\gamma_\mu(1-\gamma_5)u_1\right)\left(\frac{-g^{\mu\nu}+q^\mu q^\nu/M_W^2}{q^2-M_W^2}\right)\left[\bar{u}_4\gamma_\nu(1-\gamma_5)v_2\right] \tag{10.73}$$

The decay rate $d\omega$ is given by:

$$d\omega = (2\pi)^4\frac{\delta^4(P_i - P_f)|\mathscr{M}|^2 d^3p_2 d^3p_3 d^3p_4}{(2\pi)^9 2E_1 2E_2 2E_3 2E_4} \tag{10.74}$$

We can write $\mathscr{M}$ as follows:

$$|\mathscr{M}|^2 = \frac{g^4}{64M_W^4} L_{\mu\nu} M^{\mu\nu} \tag{10.75}$$

where:

$$\begin{aligned} L_{\mu\nu} &= \text{Tr}\left[(u_3\bar{u}_3)\gamma_\mu(1-\gamma_5)(u_1\bar{u}_1)\gamma_\nu(1-\gamma_5)\right] \\ &= \frac{1}{2}\text{Tr}\left[\not{p}_3\gamma_\mu(1-\gamma_5)(\not{p}_1+m_a)(1+\gamma^5\not{s}_a)\gamma_\nu(1-\gamma_5)\right] \end{aligned} \tag{10.76}$$

where we have used the Gordon identity and have taken M_W to be larger than the momenta q^2 and the mass term $m_a m_b$.

Using the standard identities, we can write:

$$\begin{aligned} L_{\mu\nu} &= \text{Tr}\left[\not{p}_3\gamma_\mu(\not{p}_1+m_a\gamma^5\not{s}_a)\gamma_\nu(1-\gamma_5)\right] \\ &= p_3^\alpha(p_1-m_a s_a)^\beta \text{Tr}\left[\gamma_\alpha\gamma_\mu\gamma_\beta\gamma_\nu(1-\gamma_5)\right] \end{aligned} \tag{10.77}$$

Similarly, $M^{\mu\nu}$ has almost the same structure, so it can be written:

$$M^{\mu\nu} = (p_4 - m_b s_b)_\alpha p_{2\beta} \text{ Tr}\left[\gamma^\alpha\gamma^\mu\gamma^\beta\gamma^\nu(1-\gamma_5)\right] \tag{10.78}$$

so the final result for $\mathscr{M}$ is:

$$|\mathscr{M}|^2 = \frac{g^4}{M_W^4}\left[p_3\cdot(p_4-m_b s_b)p_2\cdot(p_1-m_a s_a)\right] \tag{10.79}$$

Inserting this expression back into the decay rate, we find:

$$d\omega = \frac{g^4(p_4-m_b s_b)^\mu(p_1-m_a s_a)^\nu}{16(2\pi)^5 M_W^4 E_1 E_4} d^3p_4\, I_{\mu\nu} \tag{10.80}$$

where:

$$\begin{aligned} I_{\mu\nu} &= \int \frac{d^3p_2\, d^3p_3\, p_{3\mu}p_{2\nu}\delta^4(p-p_2-p_3)}{E_2E_3} \\ &= \frac{\pi}{6}(g_{\mu\nu}p^2 + 2p_\mu p_\nu) \end{aligned} \tag{10.81}$$

where $p = p_1 - p_4$.

To simplify this calculation, we will take the rest frame of the decaying lepton l_a. Then we have the following rest frame decomposition:

$$\begin{aligned} p_1 &= (m_a, \mathbf{0}) \\ s_a &= (0, \mathbf{s_a}) \\ p_4 &= (E_4, \mathbf{p}_4) \\ s_b &= [\mathbf{p}_4 \cdot \mathbf{s}_b/m_b, \mathbf{s}_b + (\mathbf{p}_4 \cdot \mathbf{s}_b)\mathbf{p}_4/m_b(E_4 + m_b)] \end{aligned} \tag{10.82}$$

Then the differential decay rate becomes:

$$\begin{aligned} d\omega = \frac{g^4 p_4\, dE_4\, d\Omega}{192(2\pi)^4 M_W^4 m_a} \Bigg\{ & \left(m_a^2 + m_b^2 - 2m_a E_4\right) \Bigg[m_a(E_4 - \mathbf{p}_4 \cdot \mathbf{s}_b) \\ & + m_a \left(\mathbf{p}_4 - m_b \mathbf{s}_b - \frac{(\mathbf{p}_4 \cdot \mathbf{s}_b)\mathbf{p}_4}{E_4 + m_b} \right) \cdot \mathbf{s}_a \Bigg] + 2(m_a^2 - m_a E_4 - m_a \mathbf{p}_4 \cdot \mathbf{s}_a) \\ & \times \left[(m_a - E_4)(E_4 - \mathbf{p}_4 \cdot \mathbf{s}_b) + \mathbf{p}_4 \cdot \left(\mathbf{p}_4 - m_b \mathbf{s}_b - \frac{(\mathbf{p}_4 \cdot \mathbf{s}_b)\mathbf{p}_4}{E_4 + m_b} \right) \right] \Bigg\} \end{aligned} \tag{10.83}$$

This formidable expression can be simplified if we let $m_b \sim 0$. Let $\mathbf{n}$ be a unit vector pointing along $\mathbf{p}_4$, such that $\cos\theta = \mathbf{s}_a \cdot \mathbf{n}$. Let $x = E_4/E_4^{\max}$, where $E_4^{\max}$ is the maximum allowed value of E_4, or $m_a/2$. Then we have:

$$d\omega = \frac{g^4 m_a^5}{32 M_W^2 (192\pi)^3} n(x)[1 + \alpha(x)\cos\theta]\,[1 - \mathbf{n} \cdot \mathbf{s}_b]\, \frac{dx\, d\cos\theta\, d\phi}{8\pi} \tag{10.84}$$

where:

$$n(x) = 2x^2(3 - 2x); \qquad \alpha(x) = (1 - 2x)/(3 - 2x) \tag{10.85}$$

Now sum over the spin states of the bth particle, average over the spin states $\mathbf{s}_a$, and integrate over $d\Omega$. We then find:

$$\frac{d\omega}{dx} = \frac{g^4}{32 M_W^4} \frac{m_a^5 n(x)}{192\pi^3} \tag{10.86}$$

Integrating over x, we find:

$$\Gamma = \frac{g^4 m_a^5}{32 M_W^4 (192)\pi^3} = \frac{G_F^2 m_a^5}{192\pi^3} \tag{10.87}$$

We performed this calculation for small m_b. However, the calculation can also be performed by adding in the corrections for small $\epsilon = m_b/m_a$. Power expanding in this variable, we find:

$$\Gamma = \frac{G_F^2 m_a^5}{192\pi^3} F(\epsilon); \quad F(\epsilon) = 1 - 8\epsilon^2 - 24\epsilon^4 \log\epsilon + 8\epsilon^6 - \epsilon^8 + \cdots \tag{10.88}$$

We should point out that this result can be generalized to allow for couplings other than V and A, which allows for a test of the accuracy of the electroweak theory. We could start with the transition probability:

$$\mathscr{M} = \sum_i (\bar{u}_4 \Gamma_i u_1)\left[\bar{u}_3 \Gamma_i (g_i + g_i' \gamma_5) v_2\right] \tag{10.89}$$

where we include all possible 16 Dirac matrices in the transition element, not just $V - A$.

For $m_b = 0$, the calculation of the decay rate is long but very straightforward, and yields[13]:

$$\begin{aligned} d\omega &= \frac{A}{4} \frac{m_a^5 x^2\, dx}{192\pi^3} \frac{d\Omega}{4\pi} \\ &\times \left\{ 6(1-x) + 4\rho\left(\frac{4}{3}x - 1\right) - \xi \cos\theta \left[2(1-x) + 4\delta\left(\frac{4}{3}x - 1\right)\right]\right\} \end{aligned} \tag{10.90}$$

where we have:

$$\begin{aligned} a &= |g_S|^2 + |g_S'|^2 + |g_P|^2 + |g_P'|^2 \\ a' &= 2\mathrm{Re}\,(g_S g_P'^* + g_P g_S'^*) \\ b &= |g_V|^2 + |g_V'|^2 + |g_A|^2 + |g_A'|^2 \\ b' &= -2\mathrm{Re}\,(g_V g_A'^* + g_A g_V'^*) \\ c &= |g_T|^2 + |g_T'|^2 \\ c' &= 2\mathrm{Re}\,(g_T g_T'^*) \end{aligned} \tag{10.91}$$

The Michel parameters are defined as:

$$\begin{aligned} A\rho &= 3b + 6c \\ A\xi &= -3a' - 4b' + 14c' \\ \delta &= (3b' - 6c')/(3a' + 4b' - 14c') \end{aligned} \tag{10.92}$$

where $A = a + 4b + 6c$.

When radiative corrections are included in the calculation, we find, for muon decay:[14]

$$\Gamma = \frac{G_\mu m_\mu^5}{192\pi^3}\left[1 - \frac{8m_e^2}{m_\mu^2}\right]\left[1 + \frac{\alpha}{2\pi}\left(\frac{25}{4} - \pi^2\right) + \frac{3}{5}\frac{m_\mu^2}{M_W^2}\right] \tag{10.93}$$

These and other theoretical calculations have shown good agreement with experiment.

10.6 R_ξ Gauge

Now that we have described the Weinberg–Salam model, we will study, in this and the next section, how to quantize it. We know that massive gauge theories are not renormalizable by a simple analysis of the ultraviolet behavior of their propagators. [These terms diverge as $O(1)$ as k_μ becomes large, which spoils renormalizability.] Unlike the situation in QED, we cannot appeal to the Ward identities to eliminate the troublesome term in the propagator: $k_\mu k_\nu/[M^2(k^2 - m^2)]$.

This makes us wonder how spontaneously broken gauge theories can preserve both unitarity and renormalizability, which seems totally contradictory. On the surface, it seems impossible to preserve both features, which was one reason why massive gauge theories were rejected as a model for the weak interactions.

To see how spontaneously broken gauge theories can be both unitary and renormalizable at the same time, we will use the R_ξ gauge,[15] which has the advantage that it interpolates between two sets of propagators. We will then specialize this to the 't Hooft gauge when we consider the case of the Weinberg–Salam model.

To obtain the R_ξ gauge, we will insert a new term into the action. Let us impose the gauge:

$$F(A_\mu) = a(x) \tag{10.94}$$

on our theory, where $a(x)$ is an arbitrary, real field. We can insert the following term into the path integral:

$$\int Da\, e^{-\frac{i}{2\alpha}\int d^4x\, a^2(x)} \tag{10.95}$$

The path integral, with the new gauge constraint, now becomes:

$$\int DA_\mu^a Da\, \Delta_{FP}\delta\left(F(A_\mu)-a\right)\exp\left[i\int d^4x\left(-\frac{1}{2}a^2+\mathscr{L}(x)\right)\right] \tag{10.96}$$

By performing the path integral over $a(x)$, we find that the Lagrangian is altered by:

$$\mathscr{L}\to\mathscr{L}-\frac{1}{2\alpha}F^TF \tag{10.97}$$

In particular, we will insert the gauge fixing term:

$$F(A_\mu)=\sqrt{\xi}\left(\partial^\mu A_\mu^a\right) \tag{10.98}$$

With this new gauge-fixing term, the action becomes (with $\alpha = 1/\xi$):

$$\begin{aligned}\mathscr{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}-\frac{1}{2\alpha}(\partial_\mu A^\mu)^2\\ &= \frac{1}{2}A^\mu\left[g_{\mu\nu}\partial^2+\left(\alpha^{-1}-1\right)\partial_\mu\partial_\nu\right]A^\nu\end{aligned} \tag{10.99}$$

Inverting this expression and solving for the propagator, we find:

$$D_{\mu\nu}=-\frac{1}{k^2}\left(g_{\mu\nu}-(1-\alpha)\frac{k_\mu k_\nu}{k^2}\right) \tag{10.100}$$

Although the Green's functions for the theory are all gauge dependent (i.e., dependent on the parameter $\alpha = 1/\xi$), we will find that the S matrix elements are all independent of α, which is now seen to be an unphysical artifact of the gauge-fixing procedure.

For various values of α, we have various gauges [see Eq. (4.44)]:

$$\begin{cases}\alpha=1: & \text{Feynman gauge}\\ \alpha=0: & \text{Landau gauge}\end{cases} \tag{10.101}$$

Now let us discuss the massive case, which describes spontaneously broken gauge theories. When the gauge meson develops a mass, the action becomes:

$$\mathscr{L}-\frac{1}{2\alpha}F^TF=\frac{1}{2}A_\mu\left[g^{\mu\nu}\partial^2+\partial^\mu\partial^\nu(\alpha^{-1}-1)+g^{\mu\nu}m^2\right]A_\nu \tag{10.102}$$

To find the Green's function, we need to solve:

$$\left[(\partial^2+m^2)g^{\mu\nu}-\partial^\mu\partial^\nu(1-\alpha^{-1})\right]\Delta(x-y;\alpha)_{\nu\rho}=-\delta^\mu_\rho\delta^4(x-y) \tag{10.103}$$

Inverting this expression, we find that the propagator is given by:

$$\Delta_{\mu\nu}(k,\alpha) = -\left(g_{\mu\nu} - (1-\alpha)\,k_\mu k_\nu \frac{1}{k^2 - \alpha m^2 + i\epsilon}\right)\frac{1}{k^2 - m^2 + i\epsilon} \tag{10.104}$$

(The propagator has a pole at $k^2 = \alpha m^2$, which represents a fictitious particle. This pole is cancelled in the S matrix by other contributions, which preserves unitarity.)

In the limit $\alpha \to 0$, we find:

$$\lim_{\alpha\to 0} \Delta(k,\alpha)_{\mu\nu} = -\frac{g_{\mu\nu} - k_\mu k_\nu/k^2}{k^2 - m^2 + i\epsilon} \tag{10.105}$$

In the ultraviolet limit, this propagator is much better behaved than the usual massive vector propagator; it goes as $O(1/k^2)$, which gives us good power counting behavior in the Feynman graphs. The price we pay for such a propagator, however, is that the theory is not manifestly unitary. If we take the diagonal elements of the propagator, they alternate in sign, an indication that there are longitudinal ghosts.

In the limit $\alpha \to \infty$, however, we have the propagator:

$$\lim_{\alpha\to\infty} \Delta(k,\alpha)_{\mu\nu} = -\frac{g_{\mu\nu} - k_\mu k_\nu/m^2}{k^2 - m^2 + i\epsilon} \tag{10.106}$$

This propagator, by contrast, has very bad convergence properties. For large k, it behaves like a constant, which is disastrous from a power counting point of view. The advantage of this limit, however, from the S matrix point of view, is that it is unitary. The 0, 0 component of the propagator, taken in the rest frame, vanishes.

We now have the strange situation where for $\alpha = 0$, the theory appears renormalizable but not manifestly unitary, but for $\alpha = \infty$, the theory appears unitary, but not manifestly renormalizable.

In summary:

$$\begin{cases} \alpha \to 0: & \text{Renormalizable, not manifestly unitary} \\ \alpha \to \infty: & \text{Unitary, not manifestly renormalizable} \end{cases} \tag{10.107}$$

Although the Green's functions are ξ dependent, we know that the S matrix must be independent of ξ, which is a gauge artifact; therefore the theory is both unitary and renormalizable. Although this argument is not totally rigorous, the R_ξ gauge, because it smoothly interpolates between two gauges, intuitively shows how unitarity and renormalizability are complementary, not contradictory. To see how to apply this to the Weinberg–Salam model, we now define the 't Hooft gauge.

10.7 't Hooft Gauge

One of the most convenient gauges to quantize the Weinberg–Salam model are the 't Hooft gauges,[16] which reveal the close link between unitarity and renormalizability. We begin by analyzing a $O(3)$ gauge theory coupled to a triplet of Higgs bosons:

$$\mathscr{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \frac{1}{2}(\partial_\mu\phi_i + g\epsilon_{ijk}A^j_\mu\phi^k)^2 - V(\phi^i\phi^i) \tag{10.108}$$

where $(\tau^i)_{jk} = -i\epsilon_{ijk}$ form the adjoint representation of $O(3)$.

As before, we will parametrize the Higgs bosons via:

$$\begin{aligned}\phi &= \exp\left(-\frac{i}{v}(\xi_1\tau_1 + \xi_2\tau_2)\right)\begin{pmatrix}0\\0\\v+\eta\end{pmatrix}\\ &= \begin{pmatrix}\xi_2\\-\xi_1\\v+\eta\end{pmatrix} + \text{higher terms}\end{aligned} \tag{10.109}$$

We are replacing the original triplet of Higgs bosons ϕ^1, ϕ^2, ϕ^3 with the set ξ_1, ξ_2, η.

Now let us substitute this new parametrization of the Higgs sector into the original action. After a bit of algebra, we find that the action can be written as the sum of four pieces. The first piece is the usual gauge action plus scalar fields, with a massive η field:

$$\mathscr{L}_1 = -\frac{1}{4}(F^a_{\mu\nu})^2 + \frac{1}{2}\left[(\partial_\mu\xi_1)^2 + (\partial_\mu\xi_2)^2 + (\partial_\mu\eta)^2\right] + M_s^2\eta^2 \tag{10.110}$$

The second term involves a cross-term between the A_μ and ξ_i field that we want to eliminate:

$$\mathscr{L}_2 = M\left(A^1_\mu\partial^\mu\xi_1 + A^2_\mu\partial^\mu\xi_2\right) \tag{10.111}$$

The third term consists of the interactions between the gauge field and the scalar fields:

$$\mathscr{L}_3 = \left(\frac{1}{2}M^2 + g^2v\eta + \frac{1}{2}g^2\eta^2\right)\left(A^1_\mu A^{1\mu} + A^2_\mu A^{2\mu}\right) + g\eta(A^1_\mu\partial^\mu\xi_1 + A^2_\mu\partial^\mu\xi_2) \tag{10.112}$$

The last term contains the scalar self-interactions:

$$\mathscr{L}_4 = -\frac{1}{4}M_s^2 v^2 + \frac{3M_s^2}{2v}\eta^3 + \frac{M_s^2}{4v^2}\eta^4 + \frac{M_s^2}{v}\xi^2\eta + \frac{M_s^2}{2v^2}\xi^2\eta^2 + \frac{M_s^2}{4v^2}\xi^4 \qquad (10.113)$$

where $\xi^2 = \xi_1^2 + \xi_2^2$.

Our goal is to choose a gauge in which the second term $\mathscr{L}_2$ with the gauge–scalar coupling vanishes. To kill this term, we choose the gauge:

$$\mathscr{L}_{GF} = -\frac{1}{2\alpha}\left(\partial_\mu A^{a\mu} - \alpha M \xi_a\right)^2 \qquad (10.114)$$

where $M = gv$. The cross term coming from the gauge fixing kills $\mathscr{L}_3$ after a partial integration.

Now let us write down the propagators for the various fields (disregarding the ghosts):

$$\begin{aligned} A_\mu^{1,2} &\rightarrow -i\frac{g_{\mu\nu} - (1-\alpha)k_\mu k_\nu/(k^2 - \alpha M^2)}{k^2 - M^2 + i\epsilon} \\ A_\mu^3 &\rightarrow i\frac{g_{\mu\nu}}{k^2 + i\epsilon} \\ \xi_{1,2} &\rightarrow \frac{i}{k^2 - \alpha M^2} \\ \eta &\rightarrow \frac{i}{k^2 + 2M^2 + i\epsilon} \end{aligned} \qquad (10.115)$$

In the limit $\alpha \rightarrow \infty$, we have a unitary theory, but one in which manifest renormalizability is lost. For this choice, called the unitary gauge, the spurious pole for the $\xi_{1,2}$ field at αM^2 disappears from the theory, and we are only left with the physical fields propagating in the theory.

In the limit $\alpha \rightarrow 0$, the theory is renormalizable by power counting, but not manifestly unitary. For intermediate values of α, the poles in the propagator of the gauge fields $A_\mu^{1,2}$ at $k^2 = \alpha M^2$ cancel with the poles in the propagators of the $\xi_{1,2}$ field. Thus, it is no contradiction to have a theory which is both unitary and renormalizable.

The point of this exercise is to quantize the Weinberg–Salam model. The Higgs sector is given by:

$$\left|\partial_\mu\phi - \frac{ig'}{2}B_\mu\phi - \frac{ig}{2}\tau^i W_\mu^i\phi\right|^2 - m^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \qquad (10.116)$$

We will choose a parametrization that exchanges the complex doublet of Higgs mesons ϕ_1, ϕ_2, which contains four separate fields, with the four real fields

$\xi_1, \xi_2, \xi_3, \eta$:

$$\begin{aligned}\phi &= \frac{1}{\sqrt{2}}\left(1-\frac{i}{v}\tau^i\xi^i\right)\begin{pmatrix}0\\ v+\eta\end{pmatrix}+\cdots \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix}-i\xi_1/v-\xi_2/v\\ 1-i\xi_3/v\end{pmatrix}(v+\eta)+\cdots\end{aligned} \tag{10.117}$$

Expanding out the action generates a large number of terms. However, the term in which we are interested is the cross term between the gauge and scalar part, which is given by:

$$-M_B B_\mu \partial^\mu \xi_3 + M_W W^i_\mu \partial^\mu \xi_i + \frac{1}{2}\partial_\mu \xi_i \partial^\mu \xi_i \tag{10.118}$$

where we sum over $i = 1, 2$ and $M_B = g'v/2$ and $M_W = gv/2$.

We now choose the 't Hooft gauge so that this cross term is cancelled:

$$\mathscr{L}_{GF} = -\frac{1}{2\alpha}\left(\partial_\mu W^{a\mu} - \alpha M_W \xi_a\right)^2 - \frac{1}{2\alpha}\left(\partial_\mu B^\mu + \alpha M_B \xi_3\right)^2 \tag{10.119}$$

where we sum over $a = 1, 2$.

The cross terms cancel, and we are left with the mass terms:

$$-\frac{1}{2}\alpha M_W^2(\xi_1^2+\xi_2^2) - \frac{1}{2}\alpha M_Z^2 \xi_3^2 \tag{10.120}$$

This can be rewritten in terms of the physical fields:

$$\begin{aligned}W^3_\mu &= -\sin\theta_W A_\mu + \cos\theta_W Z_\mu \\ B_\mu &= \cos\theta_W A_\mu + \sin\theta_W Z_\mu \\ W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \pm W^2_\mu)\end{aligned} \tag{10.121}$$

Now we can write down the new action, and from it extract the Feynman rules for the propagator. The relevant terms are:

$$\begin{aligned}\mathscr{L} &= -\frac{1}{4}(\partial_\mu W^i_\nu - \partial_\nu W^i_\mu)^2 - \frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 \\ &\quad - \frac{1}{2\alpha}\left(\partial^\mu W^i_\mu - \alpha M_W \xi^i\right)^2 - \frac{1}{2\alpha}(\partial_\mu A^\mu)^2 \\ &\quad - \frac{1}{2\alpha}(\partial_\mu Z^\mu + \alpha M_Z \xi^3)^2\end{aligned}$$

$$+\frac{1}{2}M_W^2(W_\mu^1 W^{1\mu}+W_\mu^2 W^{2\mu})+\frac{1}{2}M_Z^2 Z_\mu Z^\mu+\cdots \quad (10.122)$$

The propagators can now be read off from the action:

$$\begin{aligned}
W_\mu^\pm &\rightarrow -i\frac{g_{\mu\nu}-(1-\alpha)k_\mu k_\nu/(k^2-\alpha M_W^2)}{k^2-M_W^2+i\epsilon}\\
A_\mu &\rightarrow \frac{ig_{\mu\nu}}{k^2+i\epsilon}\\
Z_\mu &\rightarrow -i\frac{g_{\mu\nu}-(1-\alpha)k_\mu k_\nu/(k^2-\alpha M_Z^2)}{k^2-M_Z^2+i\epsilon}\\
\xi_{1,2} &\rightarrow \frac{i}{k^2-\alpha M_W^2}\\
\xi_3 &\rightarrow \frac{i}{k^2-\alpha M_W^2}
\end{aligned} \quad (10.123)$$

Again, we have a theory in which unitarity is manifest for $\alpha \rightarrow \infty$ but manifest renormalizability is lost. The fictitious poles vanish, and we are left with a theory defined only in terms of the physical fields. For $\alpha \rightarrow 0$, the theory is renormalizable, but not manifestly unitary. In general, the fictitious poles coming from the propagators of the $\xi_{1,2,3}$ fields cancel against the poles coming from the propagators of the Z_μ and $W_\mu^\pm$ fields, giving us a theory that is both unitary and renormalizable.

10.8 Coleman–Weinberg Mechanism

Although spontaneous symmetry breaking lies at the heart of the Weinberg–Salam model, one of its weaknesses is the arbitrariness of the Higgs potential. This is a serious criticism, since many of the physical parameters depend crucially on the precise form of the Higgs potential.

In principle, one would like to derive the Higgs potential from more fundamental principles, with as few arbitrary parameters as possible.

One interesting approach is the Coleman–Weinberg method,[17] where the Higgs potential is induced by radiative corrections, rather than being inserted by hand. In this approach, we sum over higher-loop graphs to induce an effective potential, which may then produce spontaneous symmetry breaking.

Ideally, one would like to start with a theory that is massless from the very beginning and then induce the mass corrections appearing in the action by radiative corrections. This is called *dimensional transmutation*, where a dimensionless

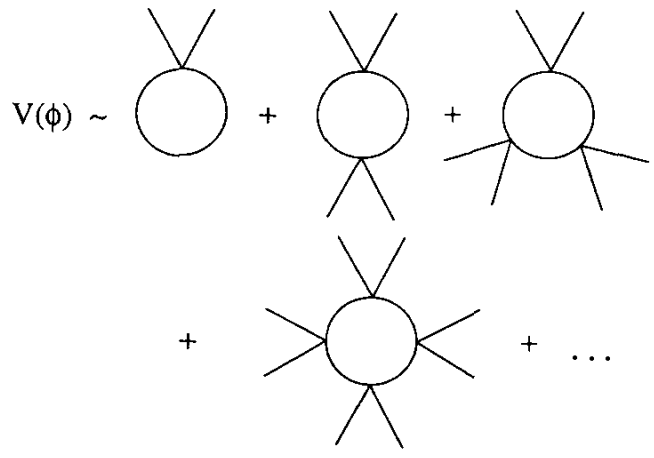


Figure 10.4. The sum over one loop graphs with an arbitrary number of ϕ external lines generates an effective potential, which in turn can induce, under certain circumstances, spontaneous symmetry breaking.

theory trades one of its dimensionless coupling constants for a dimensionful one. This, in turn, gives us the hope of deriving all masses from first principles.

To illustrate this procedure, let us first study the simplest possible example, the ϕ^4 theory. Although this theory is too simple to give us a reliable mechanism to break gauge symmetries, this example reveals the basic principles.

Let us begin with the usual ϕ^4 theory, with the action:

$$L = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 \tag{10.124}$$

(We will eventually take the limit as $m \to 0$ at the end of the calculation.) Our task is to sum over an infinite series of one-loop graphs with an arbitrary number of ϕ^4 vertices attached to it (Fig. 10.4).

After this sum is performed, the net effect of this series is to generate a new, effective action where the potential is nonpolynomial. For example, let us use Feynman's rules to give us the contribution to the single-loop potential. Each single-loop diagram is given by an integral over the internal momentum. Feynman's rules give us the contribution to the single-loop graph with n vertex insertions as:

$$i\frac{(2n)!}{2^n 2n}\int \frac{d^4k}{(2\pi)^4}\left((-i\lambda)\frac{i}{k^2 - m^2 + i\epsilon}\right)^n \tag{10.125}$$

where the symmetry factor must be inserted because there are $(2n)!$ ways to distribute $2n$ particles among the external lines. (We have taken the external lines to have zero momentum. This will be justified later.)

Now we would like to write down an effective action that generates this series at the tree level. To obtain this new, effective potential, we simply multiply the

term with n insertions by ϕ^{2n}, where ϕ from now on will represent the classical value of the field. We must also correct for symmetry factors and then sum. The effective potential now contains the term:

$$i\int \frac{d^4k}{(2\pi)^4}\sum_{n=1}^{\infty}\frac{1}{2n}\left(\frac{(\lambda/2)\phi^2}{k^2-m^2+i\epsilon}\right)^n \tag{10.126}$$

This series, fortunately, can be easily summed by using the Taylor expansion for the function $\log(1+x)$. Let us sum the series (which yields a divergent integral) and then perform the integration by putting in an explicit cutoff Λ:

$$\begin{aligned} V_{\text{eff}} &= \frac{1}{2}m^2\phi^2+\frac{\lambda}{4!}\phi^4+\frac{i}{2}\int\frac{d^4k}{(2\pi)^4}\log\left[1+\frac{\lambda\phi^2/2}{k^2-m^2+i\epsilon}\right] \\ &= \frac{1}{2}m^2\phi^2+\frac{\lambda}{4!}\phi^4+\frac{\Lambda^2}{32\pi^2}\left(m^2+\lambda\phi^2/2\right) \\ &+ \frac{1}{64\pi^2}\left(m^2+\frac{\lambda\phi^2}{2}\right)^2\left[\log\left(\frac{m^2+\lambda\phi^2/2+i\epsilon}{\Lambda^2}\right)-\frac{1}{2}\right] \end{aligned} \tag{10.127}$$

where we have used the summation:

$$\frac{1}{k^2-m^2}+\frac{1}{k^2-m^2}\frac{1}{2}\lambda\phi^2\frac{1}{k^2-m^2}+\cdots=\frac{1}{k^2-(m^2+\lambda\phi^2/2)} \tag{10.128}$$

Since the original theory was renormalizable, this means that, with the addition of counterterms into the action, we can absorb all cutoff-dependent infinities into a renormalization of the parameters of the theory.

We add the following counterterms to the action:

$$V_{\text{eff}}\rightarrow V_{\text{eff}}+\frac{A}{2}\phi^2+\frac{B}{4!}\phi^4 \tag{10.129}$$

and then solve for parameters A, B by making the following definitions of the renormalized parameters:

$$\begin{aligned} m^2 &= \left.\frac{d^2V_{\text{eff}}}{d\phi^2}\right|_{\phi=M} \\ \lambda &= \left.\frac{d^4V_{\text{eff}}}{d\phi^4}\right|_{\phi=M} \end{aligned} \tag{10.130}$$

(We have taken the condition that the classical field $\phi = M$ in order to avoid infrared divergences as we take the limit $m \rightarrow 0$.)

Now let us insert this effective potential (with the counterterms added) and then use these conditions to determine A and B. For small m, the calculation is straightforward and yields:

$$\begin{aligned} V_{\text{eff,r}} &= \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4 \\ &\quad + \frac{1}{64\pi^2}\left[\left(m^2+\frac{\lambda}{2}\phi^2\right)^2 \log\left(\frac{m^2+\lambda\phi^2/2}{m^2}\right)\right. \\ &\quad \left. - \frac{1}{2}\lambda m^2\phi^2 - \frac{25}{24}\lambda^2\phi^4 + \frac{1}{4}\lambda^2\phi^4 \log\left(\frac{2m^2}{\lambda M^2}\right)\right] \end{aligned} \tag{10.131}$$

We now perform our last step, by taking the limit as $m \to 0$:

$$V_{\text{eff,r}} = \frac{1}{4!}\lambda\phi^4 + \frac{\lambda^2\phi^4}{256\pi^2}\left(\log\frac{\phi^2}{M^2} - \frac{25}{6}\right) \tag{10.132}$$

This is our final result. On the surface, it appears that we have accomplished our goal: We have traded a potential (with no mass terms) with a new potential that has a new minimum away from the origin (where a new mass scale has been introduced by radiative corrections).

However, this example has been too simple. The new minimum lies too far from the origin, beyond the reliability of the one-loop potential:

$$\lambda \log\frac{\langle\phi\rangle^2}{M^2} = -\frac{32}{3}\pi^2 + O(\lambda) \tag{10.133}$$

The term on the left is greater than one, so that the loop contribution is larger than that of the tree contribution; so we are outside the region where the single-loop approximation is reliable. We have gained some insight into the use of radiative corrections to drive the minimum of the potential away from the origin, but our example has been too simple, with only one coupling constant.

Next, we couple charged scalars to QED, where we now have enough coupling constants to make the Coleman–Weinberg mechanism work. We start with a new action, with two coupling constants e and λ:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}^2 + \frac{1}{2}\left|(\partial_\mu - ieA_\mu)\phi\right|^2 - \frac{\lambda}{4!}(\phi^*\phi)^2 \tag{10.134}$$

We choose the Landau gauge, where the propagator becomes:

$$i\Delta_{\mu\nu}(k) = -i\frac{g_{\mu\nu} - k_\mu k_\nu/k^2}{k^2 + i\epsilon} \tag{10.135}$$

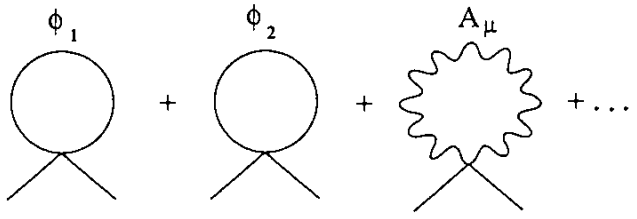


Figure 10.5. These are the only diagrams which contribute to the effective potential in the problem.

We choose this gauge because $k^\mu \Delta_{\mu\nu} = 0$. In this gauge, the only diagrams that contribute to the effective potential are given in Figure 10.5.

The graph in Figure 10.6 does not contribute to the action because k_μ does not couple to the propagator in the Landau gauge. There are thus only three types of diagrams that have to be computed.

We perform the calculation in the same way as before:

1. We sum over each set of diagrams separately in the one-loop approximation by using the power expansion for $\log(1+x)$.
2. We use a cutoff to render the integrals finite.
3. We introduce new parameters into the theory via counterterms.
4. We calculate the value of these new parameters at the classical value of $\phi = M$.

All steps are exactly the same as before; the only difference comes from the value of the coupling constant contribution of each of the three diagrams.

The result is given by:

$$\begin{aligned} V_{\text{eff,r}} &= \frac{1}{4!}\lambda\phi^4 + \frac{\phi^4}{64\pi^2}\left(\log\frac{\phi^2}{M^2} - \frac{25}{6}\right)C \\ C &= (\lambda/2)^2 + (\lambda/2)^2(1/9) + 3e^4 \end{aligned} \tag{10.136}$$

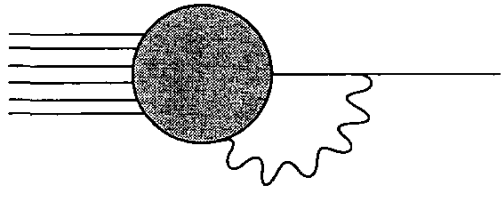


Figure 10.6. This graphs does not contribute because the momentum vector does not couple to the propagator.

[This result is almost identical to Eq. (10.132), except that the last factor of $3e^4$ comes from the trace of the Landau propagator, and the extra $1/3$ ratio between the $\phi_1^2\phi_2^2$ contribution and the ϕ_1^4 contribution comes from different Wick expansion coefficients.]

We now make the assumption that λ is of the order of e^4 (which we will show is self-consistent). Choosing the new minimum to be $\langle\phi\rangle = M$, we find the potential to be:

$$V_{\text{eff,r}} = \frac{1}{4!}\lambda\phi^4 + \frac{3e^4}{64\pi^4}\phi^4\left(\log\frac{\phi^2}{\langle\phi\rangle^2} - \frac{25}{6}\right) \tag{10.137}$$

We know that the slope of the potential at the minimum is related to the mass, which we set to zero:

$$V'(\langle\phi\rangle) = \left(\frac{\lambda}{6} - \frac{11e^4}{16\pi^2}\right)\langle\phi\rangle^3 = 0 \tag{10.138}$$

Solving, we find:

$$\lambda = \frac{33}{8\pi^2}e^4 \tag{10.139}$$

Thus, our assumption $\lambda \sim e^4$ is self-consistent. Moreover, we find a nontrivial constraint between two previously arbitrary coupling constants. We have traded the two dimensionless coupling constants e and λ for a dimensionful parameter $\langle\phi\rangle$ and a dimensionless parameter e. As we said earlier, this is an example of dimensional transmutation. At first, this might seem strange, because the original theory had no mass parameters at all, and yet a new mass parameter seems to have mysteriously entered into our theory.

The origin of this new mass comes from renormalization theory. Even in scale-invariant theories with no mass parameters, renormalization theory introduces a mass parameter because we must perform the subtraction of divergent diagrams at some mass scale μ. Changes in M simply involve a change in the definition of the coupling constant. (This forms the basis of renormalization group theory, which will be discussed in further detail in Chapter 14.)

Finally, with this new value of λ, we can now write:

$$V_{\text{eff}} = \frac{3e^4}{64\pi^2}\phi^4\log\left(\frac{\phi^2}{\langle\phi\rangle^2} - \frac{1}{2}\right) \tag{10.140}$$

This is our final result, which shows that there is indeed a new minimum of the potential away from the origin, as claimed. We also mention that the generalization to gauge theory proceeds as expected, with little change. The only complication

is that new graphs are generated from the interaction Lagrangian that contains the coupling:

$$\phi^\dagger(\mathbf{A}_\mu \cdot \mathbf{A}^\mu)\phi \tag{10.141}$$

As we said earlier, to illustrate this method we have made a number of assumptions; that is, the effective potential in Eq. (10.125) is defined at zero external momenta. Let us now justify this assumption from a more general point of view. To generalize our calculation, we will use the path integral method of effective actions. We recall that $Z(J)$ is the generator of Green's functions, and that $W(J) = e^{iZ}$ is the generator of connected graphs. In Section 8.4, we showed that a Legendre transformation produces the effective potential:

$$\Gamma(\phi) = W(J) - \int d^4x\, J(x)\phi(x) \tag{10.142}$$

We can power expand Γ as:

$$\Gamma(\phi) = \sum_{n=1}^{\infty} \frac{1}{n!} \int d^4x_1 \ldots d^4x_n\, \Gamma^n(x_1, \cdots, x_n)\phi(x_1)\cdots\phi(x_n) \tag{10.143}$$

Each of the $\Gamma^n(x_1, \ldots, x_n)$ is the sum over all one-particle irreducible Feynman graphs. What we are interested in, however, is the effective potential V_{eff}, which is defined by taking the position space expansion of Γ:

$$\Gamma(\phi) = \int d^4x \left[-V(\phi)_{\text{eff}} + \frac{1}{2}(\partial_\mu\phi)^2 Z(\phi) + \cdots \right] \tag{10.144}$$

where the term without any derivatives is defined to be the effective potential. To calculate a manageable expression for the effective potential, we will take the Fourier transform of Γ :

$$\begin{aligned} \Gamma^n(x_1, \ldots, x_n) = \quad & \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_n}{(2\pi)^4} (2\pi)^4 \delta^4(k_1 + \cdots + k_n) \\ & \times e^{i(k_1\cdot x_1 + \cdots k_n \cdot x_n)} \Gamma^n(k_1, k_2, \cdots, k_n) \end{aligned} \tag{10.145}$$

Now let us insert this equation for $\Gamma(\phi)$ as a power expansion in ϕ. Inserting one into the other, we find:

$$\begin{aligned} \Gamma(\phi) \quad = \quad & \sum_n \frac{1}{n!} \int d^4x_1 \ldots d^4x_n \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_n}{(2\pi)^4} \\ & \times\, e^{i(k_1\cdot x_1 + \cdots + k_n \cdot x_n)} \left[\Gamma^n(0, 0, \cdots, 0)\phi(x_1)\phi(x_2)\cdots\phi(x_n) + \cdots \right] \end{aligned}$$

$$= \int d^4x \sum_n \frac{1}{n!} [\Gamma^n(0, 0, \ldots, 0)\phi(x)^n + \cdots] \tag{10.146}$$

In the last step, we have power expanded $\Gamma^n(k_1, \ldots, k_n)$ and have taken only the lowest term. The higher-order terms contribute higher order derivatives of the fields, in which we are not interested.

Now comes the key step, comparing this equation with the power expansion of $\Gamma(\phi)$ in terms of ϕ. Comparing only the lowest-order term (which contains no derivatives of the field ϕ), we can now extract the effective potential:

$$V(\phi)_{\text{eff}} = -\sum_n \frac{1}{n!}\Gamma^n(0, 0, \ldots, 0)\phi^n(x) \tag{10.147}$$

This is the desired expression. It simply says that the effect of summing over the loop expansion produces a series of Feynman diagrams with zero momenta $\Gamma^n(0, 0, \ldots, 0)$, such that they act as the effective potential for a new action. In this way, we can justify all the steps that we made earlier from more intuitive arguments.

The ultimate use of the Coleman–Weinberg method, however, remains unclear, especially since our accelerators have not been able to pin down the Higgs particle and its interactions, other than the fact that its mass must be greater than 90 GeV. At the very least, we must use the Coleman–Weinberg mechanism to calculate radiative corrections to standard spontaneously broken theories to show that radiative corrections do not spoil the breakdown of symmetry. In other words, the Coleman–Weinberg mechanism can erase minima as well as create them in the potential. In this way, we find that the mechanism gives us bounds on the hypothetical mass of the Higgs particle.

For the Weinberg–Salam model, a very straightforward summing of radiative corrections coming from the scalar, fermion, and vector meson loops gives the correction to the potential:

$$V(\phi) = C\phi^4 \log(\phi^2/M^2) \tag{10.148}$$

where:

$$C = \frac{1}{16\pi v^4}\left(3\sum_V m_V^4 + m_\phi^4 - 4\sum_f m_f^4\right) \tag{10.149}$$

where V represents the sum over Z and W bosons. The value of v can be determined by solving for the minimum of the potential:

$$\left.\frac{\partial V}{\partial \phi}\right|_{\phi=v/\sqrt{2}} = 0 \tag{10.150}$$

The mass of the Higgs field then becomes:

$$m_\phi^2 = \frac{1}{2}\frac{\partial^2 V}{\partial \phi^2}\Big|_{\phi=v/\sqrt{2}} = 2v^2\left\{\lambda + C\left[\log\left(\frac{v^2}{2M^2}\right) + \frac{3}{2}\right]\right\} \tag{10.151}$$

Putting everything together, we find:

$$m_\phi > Cv^2 = \frac{1}{16\pi^2 v^4}\sum_V 3m_V^4 = \frac{3\alpha^2(2+\sec^4\theta_W)}{16\sqrt{2}G_F \sin^4\theta_W} \tag{10.152}$$

For the Weinberg–Salam model, we must have:

$$m_\phi > 7.9\,\text{GeV} \tag{10.153}$$

or else the radiative corrections will overwhelm the theory and destabilize the vacuum. This bound is easily met.

Alternatively, one might postulate that the Higgs mechanism is driven entirely by radiative corrections. In this interesting case, we find:

$$m_\phi \sim 11\,\text{GeV} \tag{10.154}$$

(which is experimentally ruled out).

In closing, we should also mention that a broken symmetry may be restored under certain conditions. If we consider a ferromagnet, for example, we know that the Hamiltonian does not select out any preferred direction, but the vacuum state may consist of atoms that are all aligned. However, if we heat the magnet sufficiently, the spins become more disordered until a phase transition occurs. At even higher temperatures, the spin alignment is completely lost, and randomness is restored.

Likewise, in a quantum field theory a spontaneously broken symmetry may also be restored if we place the system in a hot enough environment. This is called *symmetry restoration.* Although the temperature necessary to restore a broken symmetry is extraordinarily high, this is not an academic question. It may have great physical implications if we consider the temperatures found originally near the Big Bang.

Perhaps, at the instant of the Big Bang, a unified theory of all known quantum forces possessed a symmetry large enough to include the strong, weak, and electromagnetic interactions and possibly even gravity. As the temperature of the universe rapidly cooled, the original symmetry broke down in several stages. The gravitational interactions first broke off from the particle interactions, then the GUT symmetry broke down into the $SU(3)\otimes SU(2)\otimes U(1)$ symmetry of the Standard Model, then this group broke down into $SU(3)\otimes U(1)$.

If this general picture is correct, then the study of symmetry restoration gives us a useful tool by which to probe the universe at early cosmological times. This is discussed in more detail in the exercises.

In summary, we have seen how spontaneous symmetry breaking is perhaps the most elegant way in which symmetries can be broken. We retain all the symmetries of the theory in the Lagrangian, but the symmetry is broken via the vacuum state. In particular, spontaneous symmetry breaking allows us to generate a mass for the Yang–Mills theory without spoiling renormalizability. This was the crucial step in creating the Weinberg–Salam model, which successfully unites the electromagnetic interactions with the weak interactions.

In the next chapter, we will discuss how the Yang–Mills theory also forms the basis of the strong interactions, giving us the possibility of splicing all quantum interactions into a single Standard Model.

10.9 Exercises

1. Write down the Lagrangian for a model of Higgs mesons with local $O(N)$ symmetry, broken down to $O(M)$ symmetry, for $N > M$, with the Higgs transforming in the vector representation.

2. Do the same for a Lagrangian of Higgs with local $SU(N)$ symmetry broken down to $SU(M)$ symmetry ($N > M$), with the Higgs transforming in the fundamental representation.

3. Derive the Feynman rules in the 't Hooft gauge for Exercises 1 and 2.

4. Calculate the Coleman–Weinberg potential for self-interacting $O(N)$ mesons in the vector representation.

5. Consider the two-dimensional Gross–Neveu model, with N massless fermions ψ^a. The action is given by:

$$\bar{\psi}^a i \partial\!\!\!/ \psi^a + \frac{g_0}{N}(\bar{\psi}^a \psi^a)^2 \tag{10.155}$$

(This contains a four-fermion interaction, which is nonrenormalizable in four dimensions.) By power counting, is this theory renormalizable in two dimensions? Why? Write down all possible divergent graphs. Show that the action is invariant under a discrete transformation:

$$\psi^a \to \gamma^5 \psi^a, \quad \bar{\psi}^a \to -\bar{\psi}^a \gamma^5 \tag{10.156}$$

6. Show that the Gross–Neveu Lagrangian can be rewritten as:

$$\bar{\psi}^a i\partial\!\!\!/ \psi^a - \frac{N}{2g_0}\sigma^2 + \sigma\bar{\psi}^a\psi^a \tag{10.157}$$

where σ is a scalar field.

7. Examine the one-loop graphs in this theory with external σ legs and an internal fermion loop. Calculate the effective potential $V(\sigma)$ for the σ field by summing over one-loop graphs. Show that it equals:

$$-iV = -i\frac{N}{2g_0}\sigma^2 - \sum_{n=1}^{\infty}\frac{N}{2n}\,\mathrm{Tr}\int\frac{d^2p}{(2\pi)^2}\left(\frac{-p\!\!\!/\sigma}{p^2+i\epsilon}\right)^{2n} \tag{10.158}$$

8. Show that the potential can be written as:

$$\begin{aligned} V &= N\left[\frac{\sigma^2}{2g_0} - \int\frac{d^2p}{(2\pi)^2}\log\left(1+\frac{\sigma^2}{p^2}\right)\right] \\ &= N\left[\frac{\sigma^2}{2g_0} + \frac{1}{4\pi}\sigma^2\left(\log\frac{\sigma^2}{\Lambda} - 1\right)\right] \end{aligned} \tag{10.159}$$

where we have taken a Euclidean integral and cut it off at momentum Λ.

9. Define a renormalization mass M, defined by:

$$\frac{1}{g} \equiv N^{-1}\frac{d^2V}{d\sigma^2}\Big|_M \tag{10.160}$$

Solve for g, and write the potential as:

$$V = N\left[\frac{\sigma^2}{2g} + \frac{1}{4\pi}\sigma^2\left(\log\frac{\sigma^2}{M^2} - 3\right)\right] \tag{10.161}$$

Show that there is a minimum to this potential at a negative value, less than $V(0)$. Show that the theory has spontaneous symmetry breakdown, and that dimensional transmutation has occurred. Which parameter has been exchanged for which parameter?

10. A four-dimensional precursor to the Gross–Neveu model is the Nambu–Jona-Lasinio model,[18] with an interaction Lagrangian given by:

$$g\left[(\bar{\psi}\psi)^2 - (\bar{\psi}\gamma_5\psi)^2\right] \tag{10.162}$$

Show that it is invariant under a global continuous transformation given by a chiral $U(1)\otimes U(1)$. Is the theory renormalizable? Perform the same analysis

as before. Add in an extra σ field, and perform the integration over ψ; that is, sum the fermion bubble graphs. Show that chiral symmetry breaking occurs dynamically; that is, the fundamental action has no scalars, so the symmetry breaking occurs via a pseudoscalar bound state of the fermions.

11. Add in a Maxwell term to make the Nambu–Jona-Lasinio action locally chirally invariant. Show that the massless particles are removed in this case.

12. Although we have mainly discussed symmetry breaking, consider a model where a broken symmetry may be *restored* if we heat the system sufficiently. Consider a theory defined with potential $V = \frac{1}{2}m^2\phi^2 + (\lambda/4!)\phi^4$ for $m^2 < 0$ with a Euclidean metric. Consider the finite-temperature Green's function:

$$G_\beta(x_1, x_2, \cdots x_N) = \frac{\mathrm{Tr}\, e^{-\beta H} T\phi(x_1)\phi(x_2)\cdots\phi(x_N)}{\mathrm{Tr}\, e^{-\beta H}} \tag{10.163}$$

$\beta = 1/kT$, where T is the temperature and k is the Boltzmann constant. Show that the one-loop correction to the potential is given by:

$$V_1 = \frac{1}{2\beta}\sum_n \int \frac{d^3k}{(2\pi)^3} \log\left(-\frac{4\pi^2 n^2}{\beta^2} - \mathbf{k}^2 - M^2\right) \tag{10.164}$$

where $M^2 = m^2 + \frac{1}{2}\lambda\varphi^2$ and where the theory is *periodic* in time and therefore has integral Fourier moments labelled by n.

13. The sum in the previous problem diverges, so use the trick:

$$\begin{aligned} v(E) &\equiv \sum_n \log\left(\frac{4\pi^2 n^2}{\beta^2} + E^2\right) \\ \frac{\partial v(E)}{\partial E} &= \sum_n \frac{2E}{4\pi^2 n^2/\beta^2 + E^2} \end{aligned} \tag{10.165}$$

Using the fact that:

$$\sum_{n=1}^{\infty} \frac{y}{y^2 + n^2} = -\frac{1}{2y} + \frac{1}{2}\pi \coth \pi y \tag{10.166}$$

show that:

$$\begin{aligned} \frac{\partial v(E)}{\partial E} &= 2\beta\left(\frac{1}{2} + \frac{1}{e^{\beta E} - 1}\right) \\ v(E) &= 2\beta\left[(E/2) + \beta^{-1}\log(1 - e^{-\beta E})\right] + \ldots \end{aligned} \tag{10.167}$$

Then show that this implies:

$$V_1 = \int \frac{d^3k}{(2\pi)^3}\left[(E_M/2) + \beta^{-1}\log(1 - e^{-\beta E_M})\right] \tag{10.168}$$

where $E_M^2 = \mathbf{k}^2 + M^2$.

14. Show that, for high temperature (small β), the expression for the potential is:

$$\begin{aligned} V_1 \sim & -\frac{\pi^2}{90\beta^4} + \frac{M^2}{24\beta^2} - \frac{1}{12\pi}\frac{M^3}{\beta} \\ & -\frac{1}{64\pi^2}M^4 \log M^2\beta^2 + \frac{c}{64\pi^2}M^4 + O(M^6\beta^2) \end{aligned} \tag{10.169}$$

where $c \sim 5.41$. Take the second derivative of V_1 with respect to φ. From this, show that the symmetry is restored when:

$$\frac{1}{\beta_c^2} = -\frac{24m^2}{\lambda} \tag{10.170}$$

Calculate the order of magnitude of this temperature for the Weinberg–Salam model. At what temperature is the $SU(2) \otimes U(1)$ symmetry restored? Can these temperatures be found on the earth, in a star, or in the early universe?

15. Prove Eqs. (10.136) and (10.149).

16. For superconductors, assume that there is an attractive force between electrons that forms Cooper pairs. Assume that this many-body system can be described by the Ginzburg–Landau action, which couples a ϕ field to Maxwell's theory:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}^2 + D_\mu\phi D^\mu\phi^* - m^2|\phi|^2 - \lambda|\phi|^4 \tag{10.171}$$

For small enough temperature, spontaneous symmetry breaking occurs at the minimum $|\phi|^2 = -m^2/2\lambda > 0$. Construct the conserved current j_μ. For the static case, calculate the vector current $\mathbf{j}$. Assume that ϕ varies slowly over the medium, then show that this implies London's equation (i.e., $\mathbf{j} = -k^2\mathbf{A}$), where $k^2 = -em^2/2\lambda$.

17. By Ohm's law, we have $\mathbf{E} = R\mathbf{j}$. For the previous problem, using London's equation, show that this means that the resistance is zero. Now take the curl of Ampere's equation. Show that this implies:

$$\nabla^2\mathbf{B} = k^2\mathbf{B} \tag{10.172}$$

so that (in one dimension) $B_z \sim e^{-kx}$. This means that magnetic fields are expelled in a superconductor (Meissner effect), with penetration depth characterized by $1/k$. This implies that spontaneous symmetry breaking has made the Maxwell field massive, with mass k^2.

Chapter 11
The Standard Model

This was a great time to be a high-energy theorist, the period of the famous triumph of quantum field theory. And what a triumph it was, in the old sense of the word: a glorious victory parade, full of wonderful things brought back from far places to make the spectator gasp with awe and laugh with joy.

—S. Coleman

11.1 The Quark Model

The Standard Model, based on the gauge group $SU(3) \otimes SU(2) \otimes U(1)$, is one of the great successes of the gauge revolution. At present, the Standard Model can apparently describe all known fundamental forces (excluding gravity).

The Standard Model is certainly not the final theory of particle interactions. It was created by crudely splicing the electroweak theory and the theory of quantum chromodynamics (QCD). It cannot explain the origin of the quark masses or the various coupling constants. The theory is rather unwieldy and inelegant. However, at present, it seems to be able to explain an enormous body of experimental data. Not only is it renormalizable, it can explain a vast number of results from all areas of particle physics, such as neutrino scattering experiments, hadronic sum rules, weak decays, current algebras, etc. In fact, there is no piece of experimental data that violates the Standard Model.

In this chapter, we will discuss the Standard Model by first reviewing the experimental situation in the 1960s with regard to the quark model. Then we will present compelling evidence, from a wide variety of quarters, that the strong interactions can be described by QCD. Then we will marry QCD to the Weinberg–Salam model to produce the Standard Model.

Although the Standard Model makes the situation seem so clear today, back in the 1960s the experimental situation with the strong interactions was totally confused, with hundreds of "elementary particles" pouring out of our particle accelerators. J. Robert Oppenheimer, in exasperation, said that the Nobel Prize should be given to the physicist who *did not* discover a new particle. Although the Yukawa theory of strong interactions was fully renormalizable, the coupling constant of the strong interactions was large, and hence perturbation theory was unreliable:

$$\frac{g_{\pi N}^2}{4\pi} \sim 14 \tag{11.1}$$

One important observation was that the existence of resonances usually indicated the presence of bound states, so Sakata[1] in the 1950s postulated that the hadrons could be considered to be composite states built out of p, n, and Λ particles. Then Ikeda, Ohnuki, and Ogawa[2] in 1959 made the suggestion that this triplet of particles transformed in the fundamental representation **3** of $SU(3)$. They correctly said that the mesons could be built out bound states of **3** and $\bar{\mathbf{3}}$:

$$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \tag{11.2}$$

However, several of their assignments were incorrect.

In 1961, the correct $SU(3)$ assignments were finally found by Gell-Mann[3,4] and Ne'eman,[5] who postulated that the baryons and mesons could be arranged in what they called the Eightfold Way. Then Gell-Mann[6] and Zweig[7] proposed that these $SU(3)$ assignments could be generated if one postulated the existence of new constituents, called "quarks," which transformed as a triplet **3**. Since all representations of $SU(N)$ can be generated by taking multiple products of the fundamental representation, in this way we could generate all higher representations beginning with the quarks.

The quarks belonged to the fundamental representation of $SU(3)$:

$$\mathbf{3} = q_i = \begin{pmatrix} u \\ d \\ s \end{pmatrix} \tag{11.3}$$

where the quarks were called the "up," "down," and "strange" quark, for historical reasons. The u and the d quarks formed a standard $SU(2)$ isodoublet, but the

addition of the third quark was necessary because it was observed in the 1950s that a new quantum number in addition to isospin was conserved by hadronic processes, called "strangeness." This new quantum number could be explained in terms of $SU(3)$, which is a rank 2 Lie group. Its representations are therefore labeled by two numbers, the third component of isospin T_3 and also a new quantum number Y, called "hypercharge."

The new quantum number of strangeness and hypercharge could be related to each other via the Gell-Mann–Nishijima[8,9] formula:

$$Q = T_3 + \frac{Y}{2} = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix} \tag{11.4}$$

where $Y = B + S$. B is the baryon number, S is the strangeness number, T_3 is the third component of isospin, and Q is the charge.

To fit the known spectrum, the mesons were postulated to be composites of a quark and an antiquark, while the baryons were postulated to be composites of three quarks. Thus, we expect to see the mesons and baryons arranged according to the following tensor product decomposition:

$$\begin{aligned} \text{Meson} &= \mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \\ \text{Baryon} &= \mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \end{aligned} \tag{11.5}$$

The theory predicted that the mesons should be arranged in terms of octets and singlets, while baryons should be in octets as well as decuplets. The fact that this simple picture could arrange the known mesons and baryons in such an elegant picture was remarkable.

In order to reproduce the known charges of the mesons and baryons, it was necessary to give the quarks fractional charges:

$$Q_u = \frac{2}{3}e; \quad Q_d = -\frac{1}{3}e; \quad Q_s = -\frac{1}{3}e \tag{11.6}$$

Since three of them were required to make up a single baryon, this meant that each of them had baryon number $\frac{1}{3}$. We summarize the quantum numbers of the

quarks in the following chart:

	Q	T	T_3	Y	S	B
u	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{3}$
d	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{3}$
s	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$	-1	$\frac{1}{3}$

(11.7)

For the meson spectrum, we can get a rough classification by considering the bound states generated by a $q\bar{q}$ pair. The bound states arrange themselves roughly in the following angular momentum series (similar to the familiar series found in spectroscopy using the notation $^{2S+1}L_J$):

$$^1S_0,\ ^3S_1,\ ^1P_1,\ ^3P_0,\ ^3P_1,\ ^3P_2,\ ^1D_2, \ldots \tag{11.8}$$

The fit between experiment and the predicted bound states of the quark model was exceptionally good. For example, the octet containing the π meson and K meson corresponds to the $^1S_0$ bound state, while the K^* multiplet is part of a $^3S_1$ bound state:

$q_i\bar{q}_j$	$J=0$	$J=1$
$\lvert u\bar{d}\rangle$	$\pi^+(140)$	$\rho^+(770)$
$2^{-1/2}\lvert d\bar{d}-u\bar{u}\rangle$	$\pi^0(135)$	$\rho^0(770)$
$\lvert u\bar{d}\rangle$	$\pi^-(140)$	$\rho^-(770)$
$2^{-1/2}\lvert d\bar{d}+u\bar{u}\rangle$	$\eta(549)$	$\omega(783)$
$\lvert u\bar{s}\rangle$	$K^+(494)$	$K^{*+}(892)$
$\lvert d\bar{s}\rangle$	$K^0(498)$	$K^{*0}(892)$
$\lvert \bar{u}s\rangle$	$K^-(494)$	$K^{*-}(892)$
$\lvert \bar{d}s\rangle$	$\bar{K}^0(498)$	$\bar{K}^{*0}(892)$
$\lvert s\bar{s}\rangle$	$\eta'(958)$	$\phi(1020)$

(11.9)

Similarly, we can also analyze the baryons. The familiar proton and neutron belong to the octet, while the Δ resonance (found in pion–nucleon scattering) belongs to the decuplet:

$q_i q_j q_k$	$J = 1/2$	$J = 3/2$
$\|uuu\rangle$		$\Delta^{++}(1230)$
$\|uud\rangle$	$p(938)$	$\Delta^{+}(1231)$
$\|udd\rangle$	$n(940)$	$\Delta^{0}(1232)$
$\|ddd\rangle$		$\Delta^{-}(1234)$
$\|uus\rangle$	$\Sigma^{+}(1189)$	$\Sigma^{+}(1383)$
$2^{-1/2}\|(ud+du)s\rangle$	$\Sigma^{0}(1192)$	$\Sigma^{0}(1384)$
$\|dds\rangle$	$\Sigma^{-}(1197)$	$\Sigma^{-}(1387)$
$2^{-1/2}\|(ud-du)s\rangle$	$\Lambda(1116)$	
$\|uss\rangle$	$\Xi^{0}(1315)$	$\Xi^{0}(1532)$
$\|dss\rangle$	$\Xi^{-}(1321)$	$\Xi^{-}(1535)$
$\|sss\rangle$		$\Omega^{-}(1672)$

(11.10)

To see how the bound states are constructed, it is sometimes useful to rearrange the meson and baryon matrices according to their quark wave functions. Let us define:

$$
\begin{aligned}
M &= q \otimes \bar{q} = \begin{pmatrix} u\bar{u} & u\bar{d} & u\bar{s} \\ d\bar{u} & d\bar{d} & d\bar{s} \\ s\bar{u} & s\bar{d} & s\bar{s} \end{pmatrix} \\
&= \begin{pmatrix} (2u\bar{u} - d\bar{d} - s\bar{s})/3 & u\bar{d} & u\bar{s} \\ d\bar{u} & (2d\bar{d} - u\bar{u} - s\bar{s})/3 & d\bar{s} \\ s\bar{u} & s\bar{d} & (2s\bar{s} - u\bar{u} - d\bar{d})/3 \end{pmatrix} \\
&+ (1/3)\mathbf{1}(u\bar{u} + d\bar{d} + s\bar{s})
\end{aligned}
\tag{11.11}
$$

In terms of the familiar pseudoscalar mesons, we have the following arrangement of the meson matrix:

$$M = \begin{pmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta & K^0 \\ K^- & \bar{K}^0 & -\frac{2}{\sqrt{6}}\eta \end{pmatrix} \tag{11.12}$$

Likewise, the baryon matrix can be arranged as:

$$B = q \otimes q \otimes q = \begin{pmatrix} \frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda^0 & \Sigma^+ & p \\ \Sigma^- & -\frac{1}{\sqrt{2}}\Sigma^0 + \frac{1}{\sqrt{6}}\Lambda^0 & n \\ \Xi^- & \Xi^0 & -\frac{2}{\sqrt{6}}\Lambda^0 \end{pmatrix} \tag{11.13}$$

To actually perform any calculations with $SU(3)$, we need, of course, an explicit representation of the generators of the algebra. We will choose the standard Gell-Mann representation of $SU(3)$ generators in terms of 3×3 Hermitian matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{11.14}$$

where:

$$\begin{aligned} \mathrm{Tr}(\lambda_i \lambda_j) &= 2\delta_{ij} \\ \left[\frac{\lambda_i}{2}, \frac{\lambda_j}{2}\right] &= i f_{ijk} \frac{\lambda_k}{2} \end{aligned} \tag{11.15}$$

The structure constants are given by:

$$f_{123} = 1; \; f_{147} = -f_{156} = f_{246} = f_{257} = f_{345} = -f_{367} = \frac{1}{2}$$

$$f_{458} = f_{678} = \frac{\sqrt{3}}{2} \tag{11.16}$$

From these commutation relations, we can work out the representations of the group (see Appendix). Because $SU(3)$ is a rank 2 Lie group, we can chart the various representations of the group in a two-dimensional space, plotting the eigenvalues of λ_3 of the $SU(2)$ subgroup against λ_8, which is proportional to the hypercharge. Thus, on a two-dimensional graph (isospin plotted against hypercharge) we can pictorially represent the triplet, antitriplet, octet, decuplet, etc. (Fig. 11.1)

Although hadronic masses are not exactly $SU(3)$ invariant (i.e., the masses of particles within a multiplet vary slightly), it is reasonable to assume that the terms that break $SU(3)$ symmetry should themselves transform covariantly under $SU(3)$. We assume, for example, that the mass term in the Hamiltonian includes a term that breaks the symmetry transforms as λ_8, as hypercharge. We assume the mass term has the form:

$$\bar{\psi}(a + b\lambda_8)\psi + \ldots \tag{11.17}$$

This, in turn, gives us nontrivial relations between the masses of the various particles within a multiplet called the Gell-Mann–Okubo[10] mass relation, which provided experimental verification of the theory. It gives us the mass relation:

$$m_N + m_\Xi = \frac{1}{2}(m_\Sigma + 3m_\Lambda) \tag{11.18}$$

which agrees well with experiment. The left-hand side equals 2.25 GeV experimentally, while the right-hand side equals 2.23 GeV.

For the spin-$3/2^+$ decuplet, we also find the equal-spacing rule:

$$m_\Omega - m_{\Xi^*} = m_{\Xi^*} - m_{\Sigma^*} = m_{\Sigma^*} - m_{N^*} \tag{11.19}$$

The experimental mass differences are 139, 149, and 152 MeV, respectively.

Historically, the prediction of the Ω^- mass from this formula gave a boost to the wide acceptance of $SU(3)$ symmetry.

Because of the success of the $SU(3)$ quark model, attempts were made to generalize this to larger groups. One attempt, merging $SU(3)$ with the $SU(2)$ of spin to create $SU(6)$,[11–13], tried to mix an internal symmetry with a space symmetry. This was possible because $SU(2) \otimes SU(3) \subset SU(6)$. $SU(6)$ had some success in predicting the magnetic moments of baryons. [However, attempts to generalize $SU(6)$ to the relativistic case floundered because of the Coleman–Mandula theorem.]

Attempts were also made to generalize $SU(3)$ to $SU(4)$[14] and beyond by adding more quarks. This approach received experimental vindication with the discovery

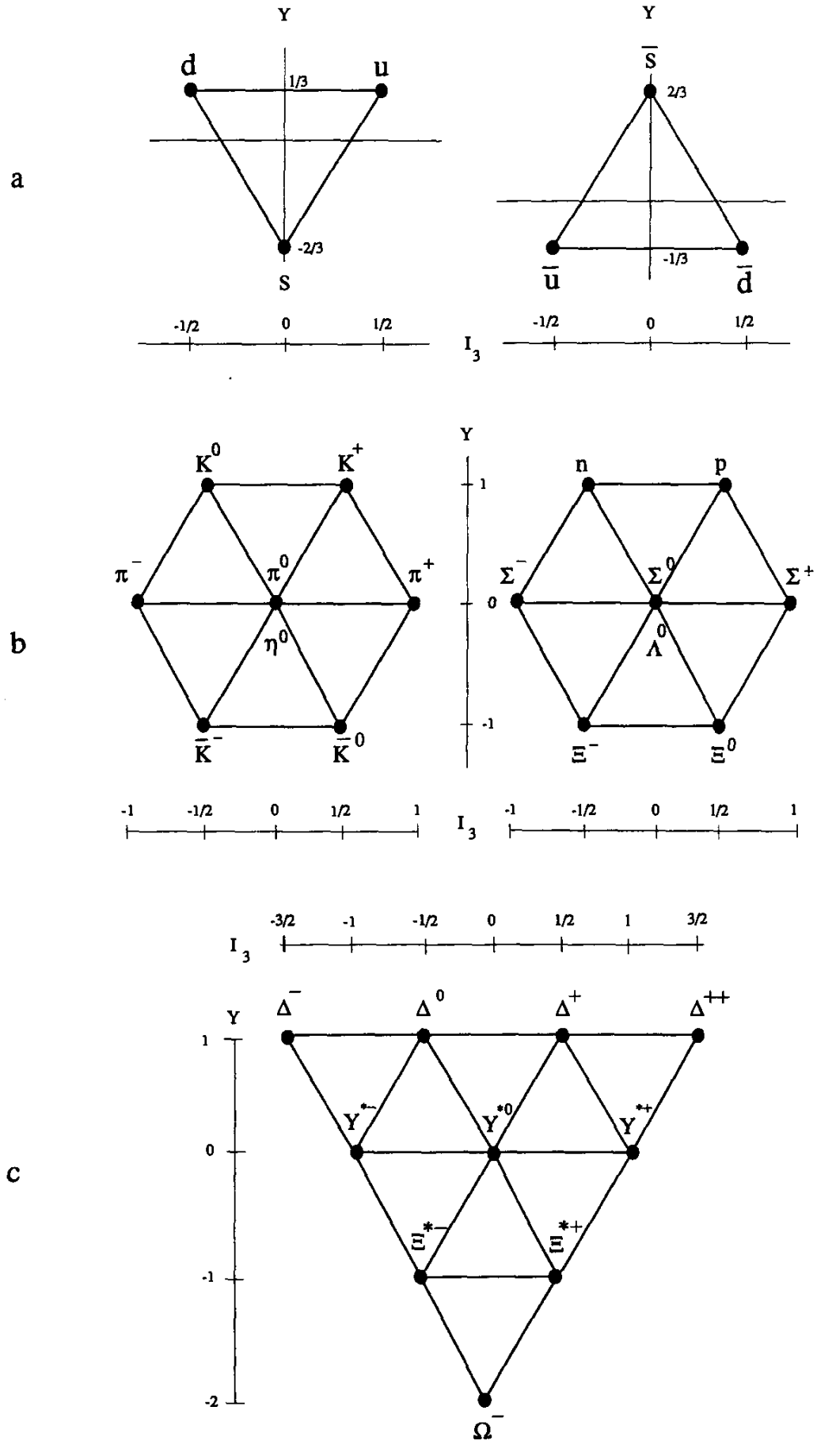


Figure 11.1. When we plot isospin against hypercharge, we can represent the triplet, antitriplet, octet, decuplet, and higher multiplets in simple geometrical patterns.

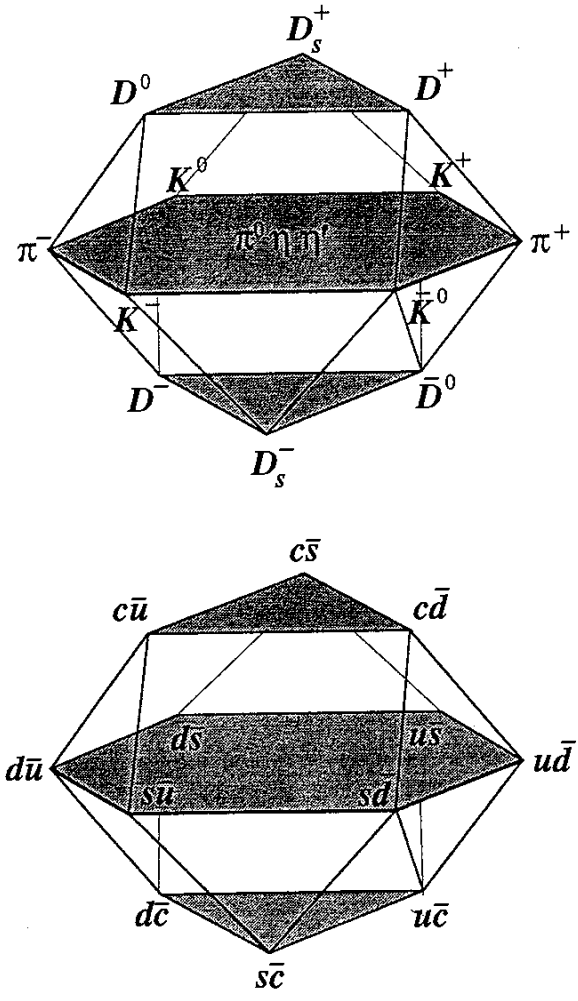


Figure 11.2. With charm plotted on the vertical axis, the quark model gives an excellent fit to the charmed meson multiplet.

of the charmed quark c in 1974 and the bottom quark b in 1977. [However, because the masses of the charmed and bottom quark are so large, global $SU(4)$ and $SU(5)$ are less reliable than $SU(2)$ and $SU(3)$ experimentally.]

For the charmed quark system, the new $q\bar{q}$ and qqq states are given the following group-theoretical assignments:

$$\begin{aligned} \mathbf{4} \otimes \bar{\mathbf{4}} &= \mathbf{15} \oplus \mathbf{1} \\ \mathbf{4} \otimes \mathbf{4} \otimes \mathbf{4} &= \bar{\mathbf{4}} \oplus \mathbf{20} \oplus \mathbf{20} \oplus \mathbf{20} \end{aligned} \tag{11.20}$$

The charmed quark bound states are given the following names for the 0^- and 1^- multiplets (Fig. 11.2):

$q_i\bar{q}_j$	$J=0$	$J=1$
$c\bar{c}$	$\eta_c(2980)$	$J/\Psi(3097)$
$c\bar{d}$	$D^+(1869)$	$(D^*)^+(2010)$
$c\bar{u}$	$D^0(1865)$	$(D^*)^0(2007)$
$\bar{c}u$	$\bar{D}^0(1865)$	$(\bar{D}^*)^0(2007)$
$\bar{c}d$	$D^-(1869)$	$(D^*)^-(2010)$
$c\bar{s}$	$D_s^+(1970)$	$D_s^{*+}(2109)$
$\bar{c}s$	$D_s^-(1970)$	$D_s^{*-}(2109)$

(11.21)

We can also write down an explicit form for the generators of the $SU(4)$ global symmetry. In fact, it is possible to find a simple iterative algorithm to write down the generators of $SU(N)$ almost by inspection. We first notice that we can write the first three λ_i matrices as follows:

$$\lambda_a = \begin{pmatrix} \sigma_a & 0 \\ 0 & 0 \end{pmatrix}; \quad a = 1, 2, 3 \tag{11.22}$$

where this symbolically means that the Pauli spin matrices are placed in the upper left-hand corner for $a = 1, 2, 3$.

Next, λ_{4-7} obey a simple pattern. Along the right column and bottom row, we insert the numbers 1 and 1 (as well as $-i$ and i) symmetrically in all possible slots. Finally, the generator λ_8 has the unit 2×2 matrix in the upper left-hand corner and we choose the last number along the diagonal to make it traceless.

From this algorithm, we can easily write down the generators of $SU(N)$ if we know the generators of $SU(N-1)$. For example, we can now write down the generators of $SU(4)$ almost by inspection. To see this, we place the generators of $SU(3)$ in the upper left-hand corner for $a = 1 - 8$:

$$\lambda_a = \begin{pmatrix} \lambda_a & 0 \\ 0 & 0 \end{pmatrix}; \qquad a = 1, \ldots, 8 \tag{11.23}$$

For the generators λ_{9-14}, we place the pairs of numbers 1, 1 and $-i, i$ symmetrically in the right column and bottom row, while for the last generator λ_{15} we

put the unit 3×3 matrix in the upper left-hand side and make it traceless:

$$\lambda_9 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}; \quad \lambda_{10} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_{11} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}; \quad \lambda_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}$$

$$\lambda_{13} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}; \quad \lambda_{14} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}$$

$$\lambda_{15} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix} \tag{11.24}$$

Similarly, the generators of $SU(5)$ can be constructed in this way. Although less is known about the quark spectrum for mesons containing the bottom quark, all states discovered so far obey the quark model predictions. The lowest lying states include:

$q_i\bar{q}_j$	$J = 0$	$J = 1$
$b\bar{b}$	η_b	$\Upsilon(9460)$
$u\bar{b}$	$B^+(5278)$	$(B^*)^+(5324)$
$d\bar{b}$	$B^0(5278)$	$(B^*)^0(5324)$
$\bar{u}b$	$\bar{B}^0(5278)$	$(\bar{B}^*)^0(5324)$
$\bar{d}b$	$B^-(5278)$	$(B^*)^-(5324)$

(The η_b has not been firmly established, and the charges of the B^* are not yet confirmed.)

Today, the original three quarks have been expanded to six quarks: the up, down, strange, charmed, bottom, and top quark. All but the last have been discovered. The global symmetry group $SU(N)$ for N quarks is now called the "flavor" symmetry.

The rough values for the constituent quark masses in GeV are given by:

$$\begin{aligned} m_u &= 0.35; & m_c &= 1.5 \\ m_d &= 0.35; & m_b &= 5 \\ m_s &= 0.5; & m_t &> 91 \end{aligned} \tag{11.25}$$

These quarks, in turn, can be arranged in three identical families or generations, each having the same quantum numbers: (u, d), (c, s), and (t, b). (The reason why nature should prefer three identical generations of quarks and leptons is one of the great mysteries of subatomic physics.)

Historically, although the quark model had great success in bringing order out of the chaos of the hundreds of resonances found in scattering experiments, it also raised a host of other problems. In fact, each year, even as the successes of the quark model began to pile up, the questions raised by the quark model also began to proliferate. For example, why were the quarks not observed experimentally? Were they real, or were they just a useful mathematical device? And what was the binding force that held the quarks together? For example, some believed that the glue that held the quarks together might be a vector meson; however, to be renormalizable, it had to be massless. But this was impossible, because if it was massless, then it should generate a long-range force, like gravity and electromagnetism, rather than being a short-range force like the strong force.

11.2 QCD

After years of confusion, the theory that has emerged to give us the best understanding of the strong interactions is called QCD, which has the Lagrangian:

$$\mathscr{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \sum_{i=1}^{6} \bar{\psi}_i (i \not{D} - m_i) \psi_i \tag{11.26}$$

where the Yang–Mills field is massless and carries the $SU(3)$ "color" force [not to be confused with the global $SU(3)$ flavor symmetry introduced earlier]. Unlike the electroweak theory, where the gauge group is broken and the Z and W become massive, the color group is unbroken and the gluons remain massless.

The quarks have *two* indices. The i index is taken over the flavors, which labels the up, down, strange, charm, top, and bottom quarks. The flavor index is *not* gauged; it represents a global symmetry. However, the quarks also carry the important local color $SU(3)$ index (which is suppressed here). In other words, quarks come in six flavors and three colors, but only the color index participates in the local gauge symmetry. From the point of view of QCD, the flavor index,

which dominated most of the phenomenology of the 1960s and 1970s, is now relegated to a relatively minor role compared to the color force, which binds the quarks together.

In fact, from the perspective of QCD, we can see the origin of the early phenomenological successes of global $SU(3)$. In the limit of equal quark masses, the QCD action possesses an additional symmetry, global $SU(N)$ symmetry for N flavors of quarks. For the u and d quarks, this is a very good approximation; so $SU(2)$ global symmetry is experimentally seen in the hadron spectrum. The s quark mass, although larger, is still relatively close to the u and d mass when compared to the baryon mass; so we expect flavor $SU(3)$ symmetry to be a relatively good one. However, the masses of the c and b are much larger; so we expect $SU(4)$ or $SU(5)$ flavor symmetry breaking to be quite large. The higher flavor symmetry groups are hence less useful phenomenologically.

Although quarks have never been seen in the laboratory, there is now an overwhelming body of data supporting the claim that QCD is the leading theory of the strong interactions. This large body of theoretical and experimental results and data, accumulated slowly and painfully over the past several decades, can be summarized in the following sections.

11.2.1 Spin-Statistics Problem

According to the spin-statistics theorem, a fermion must necessarily be totally antisymmetric with respect to the interchange of the quantum numbers of its constituents. One long-standing problem, however, was that certain baryon states, such as the **10** representation, which includes the Δ^{++} resonance, were purely symmetric under this interchange, violating the spin-statistics theorem.

For example, the wave function for this resonance is naively given by:

$$\Psi_\Delta = \psi_{SU(3)}\psi_{\text{orbital}}\psi_{\text{spin}} \tag{11.27}$$

This wave function is symmetric under the interchange of any two quarks, which is typical of bosonic, not fermionic, states. To see this, notice that the $SU(3)$ flavor part of the wave function is symmetric, since it is composed of three u quarks, all pointing in the same direction in isospin space, as in (11.10). Also, since the spin of this resonance is 3/2, all three quark spins are pointing spatially in the same direction, so the spin wave function is also symmetric. Finally, the interchange of the quarks in the orbital part yields a factor $(-1)^L$, which is one because $L = 0$ for this resonance. Thus, under an interchange of any two quarks, the wave function picks up a factor of $(+1)(+1)(+1) = 1$, so the overall wave function is symmetric under the interchange of the quarks, which therefore violates the spin-statistics theorem for fermions.

Since physicists were reluctant to abandon the spin-statistics theorem, one resolution of this problem was to postulate the existence of yet another isospin symmetry, a new "color" symmetry, so that the final state could be fully antisymmetric. This was the original motivation behind the Han–Nambu[15] quark model, a precursor of QCD.

11.2.2 Pair Annihilation

The simplest way experimentally to determine the nature of this mysterious new color symmetry is to perform electron–antielectron collision experiments. Pair annihilation creates an off-shell photon, which then decays into various possible combinations. We are interested in the process:

$$e^+ + e^- \rightarrow \gamma \rightarrow \bar{q} + q \rightarrow \text{hadrons} \tag{11.28}$$

This process is highly sensitive to the number of quarks and their charges that appear in the calculation. Using Feynman's rules, we find that the cross section must be proportional to the number of quark colors times the sum of the squares of the quark charges. In practice, it is convenient to divide by the leptonic contribution to the cross section:

$$e^+ + e^- \rightarrow \mu^+ + \mu^- \tag{11.29}$$

By taking the ratio of these two cross sections, we should therefore find the pure contribution of the quark color sector. In particular, we find:

$$R \equiv \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = N_c \sum_{i=1} Q_i^2 \tag{11.30}$$

where N_c is the number of quark colors and Q_i is the charge of each quark.

For low energies, when we excite just the u, d, s quarks, we expect R to equal $3(4 + 1 + 1)/9 = 2$. When we hit the threshold for creating charm–anticharm intermediate states, then this ratio rises to over 4. If we include the u, d, s, c, b quarks, then we have $R = 11/3$. Experimentally, this agrees rather well with experiment, assuming that there are three colors (Fig. 11.3).

11.2.3 Jets

High-energy scattering experiments should be able to knock individual gluons and quarks out of the nucleus. Although they quickly reform into bound states and hence cannot be isolated, they should make a characteristic multiprong event

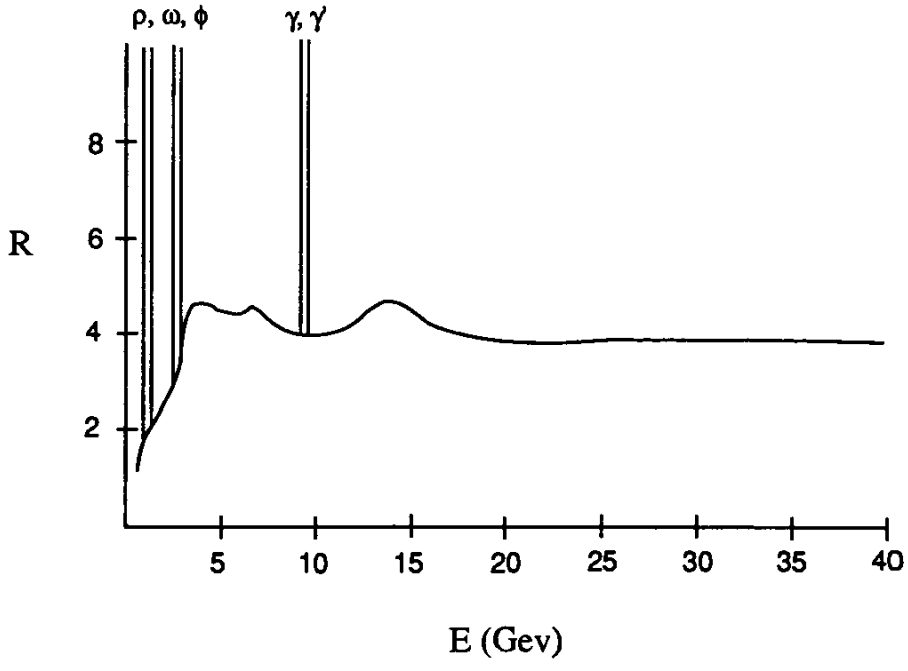


Figure 11.3. The plot of R against energy, which agrees with the value of R found in QCD.

in the scattering apparatus. These multiprong events, as predicted, have been produced in high-energy scattering experiments.

For example, two-prong jets have a characteristic distribution dependence given by $(1 + \cos^2\theta)$, where θ is the angle between the beam and the jets. This is consistent with the process $e^- + e^+ \rightarrow q + \bar{q}$ for spin $\frac{1}{2}$ quarks, as expected. This important topic will be discussed in the next section.

11.2.4 Absence of Exotics

Although the quark model had great success in fitting the known hadrons into $\mathbf{3} \otimes \bar{\mathbf{3}}$ and $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3}$ bound states, it was at a loss to explain why exotic states, such as $\mathbf{3} \otimes \mathbf{3}$, etc., should not be formed as well. Because the original quark model gave no indication of what the binding force was, this question could never be answered within the context of the old quark model.

QCD, however, gives a simple reason why these exotic states are absent. We learned earlier in Section 9.4 that the states of the unbroken Yang–Mills theory are singlets under the gauge group. We notice immediately that $\bar{q}q$ and qqq states are invariant under the color group because they are contracted by constant invariant tensors, like the delta function and the structure constant f_{ijk}. Low-lying exotic states, because they are nonsinglets under the color group, are either absent or will decay into the usual bound states. (Later, when we discuss lattice gauge theory, we will see that QCD can, in principle, give us numerical results to back up this heuristic result.)

11.2.5 Pion Decay

The Feynman diagram for the decay $\pi \rightarrow 2\gamma$ consists of an internal quark triangle loop, with the pion and the gamma rays attached to the three corners of the triangle. Thus, this decay rate is proportional to the sum over all the quarks that occur in this internal triangle loop. By comparing the experimental decay rate of the pion into two gamma rays, we can therefore calculate the number of quark colors. The experimental evidence supports the presence of 3.01 ± 0.08 colors.

11.2.6 Asymptotic Freedom

Historically, it was the discovery of asymptotic freedom that elevated QCD into the leading theory of the strong interactions. Deep inelastic experiments, such as $e + p \rightarrow e +$ anything, showed that the cross sections exhibited scale invariance at high energies; that is, the form factors lost their dependence on certain mass parameters at high energies. This scale invariance could be interpreted to mean that the quark constituents acted as if they were free particles at extremely high energies.

QCD offers a simple explanation of this scale invariance. Using the renormalization group, which will be discussed at length in Chapter 14, one could show that the coupling constant became smaller at high energies, which could explain the reason why the quarks behaved as if they were free. Asymptotic freedom gave a simple reason why the naive quark model, which described complex scattering experiments with free quarks fields, had such phenomenological success.

11.2.7 Confinement

The renormalization group also showed the converse, that the coupling constant should become large at low energies, suggesting that the quarks were permanently bound inside a hadron. This gave perhaps the most convincing theoretical justification that the quarks should be permanently "confined" inside the bound states. Although a rigorous proof that the Yang–Mills theory confines the quarks and gluons has still not been found, the renormalization group approach gives us a compelling theoretical argument that the coupling constant is large enough at small energies to confine the quarks and gluons. If correct, this approach also explains why the massless gluon field does not result in a long-range force, like gravity or the electromagnetic force. Although the range of a massless field is formally infinite, the gluon field apparently "condenses" into a stringlike glue that binds the quarks together at the ends. [In Chapter 15, we will present some compelling (but not rigorous) numerical justification for this picture using lattice gauge theory.]

Thus, with one theory, we are able to interpret two divergent facts, that quarks appear to be confined at low energies but act as if they are free particles at high energy.

The phenomenological success of the quark model, of course, is not exclusively confined to the strong interactions. Quarks also participate in the weak interactions, and have greatly clarified the origins of certain phenomenological models proposed in the 1960s (such as current algebras). By studying the weak currents generated by the quarks, one can find a simple quark model explanation for a number of phenomenologically important results from weak interaction physics, such as are given in the following sections.

11.2.8 Chiral Symmetry

In the limit that the quarks have vanishingly small mass, the QCD action possesses yet another global flavor symmetry, chiral symmetry. For N flavors, the QCD action for massless quarks is invariant under chiral $SU(N) \otimes SU(N)$, generated by:

$$q \rightarrow e^{i\lambda^a\theta^a} q; \quad q \rightarrow e^{i\gamma_5\lambda^a\theta^a} q \tag{11.31}$$

[Actually, the QCD action has the additional chiral symmetry $U(1) \otimes U(1)$ if we drop the λ^a in the previous expression. The first $U(1)$ symmetry gives the usual baryon number conservation. The second chiral $U(1)$ symmetry will be studied in more detail in the next chapter.]

Since the u, d, and s quarks have relatively small mass when compared with the scale of the strong interactions, then QCD has a global chiral symmetry in this approximation:

$$m_u \sim m_d \sim m_s \sim 0 \rightarrow SU(3) \otimes SU(3) \tag{11.32}$$

The $SU(3) \otimes SU(3)$ chiral symmetry of QCD, in turn, allows us to compute a large number of relations and sum rules between different physical processes.

Since chiral symmetry is broken in nature, for small quark masses we can, in fact, view the π meson as the Goldstone boson for broken chiral symmetry. The fact that the π meson has an exceptionally light mass is a good indicator of the validity of chiral symmetry as an approximate symmetry. A large number of successful sum rules for the hadronic weak current, as we shall see, can be written down as a byproduct of the smallness of the pion mass.

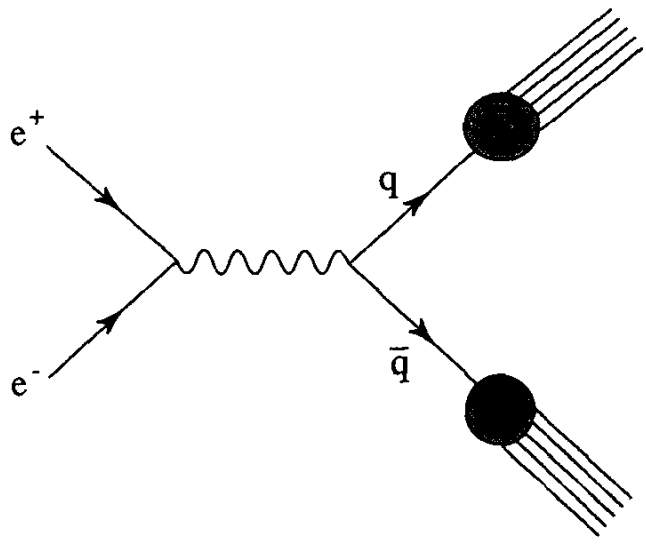


Figure 11.4. Typical Feynman diagrams describing the collision of electrons and positrons, creating jets.

11.2.9 No Anomalies

The theory of leptons given by the Weinberg–Salam model in Chapter 10 is actually fatally flawed by the presence of something called "anomalies," which will be discussed in more detail in the next chapter. These anomalies sometimes arise when a classical symmetry of an action does not survive the process of quantization. In particular, there are certain divergent fermionic triangle graphs that can potentially destroy the Ward–Takahashi identities and hence ruin renormalizability. However, when the quarks are inserted into the Weinberg–Salam model, they also produce anomalies, but of the opposite sign. In fact, the charge assignments of the quarks and leptons in the Standard Model are precisely the ones that cancel the anomaly. The vanishing of the lepton anomalies against the quark anomalies in the Weinberg–Salam model can be seen as one more theoretical justification of the Standard Model.

11.3 Jets

Now that we have completed a broad overview of the theoretical and experimental successes of the Standard Model, let us now focus in detail on some of the specifics. One of the most graphic reasons for supporting QCD is the existence of "jets" in electron–positron collisions. In Figure 11.4, we see some of the typical Feynman diagrams that arise when electrons collide with positrons.

The momentum transfer is so large that the quarks, antiquarks, and gluons in the final states are scattered in different directions. They later regroup into standard hadrons to form a jet-like structure, which has been found experimentally.

We will analyze the process with the following labeling of momenta:

$$e^-(q) + e^+(q') \to q(p) + \bar{q}(p') + g(k) \tag{11.33}$$

with $Q = q + q' = p + p' + k$ and $S = Q^2$.

For the two-jet event (where we drop the final gluon), we can factorize the transition matrix into a leptonic and a hadronic part:

$$|\mathscr{M}|^2 = \frac{1}{S^2} l_{\mu\nu} H^{\mu\nu} \tag{11.34}$$

where:

$$\begin{aligned} l_{\mu\nu} &= \frac{1}{4} \sum_{\text{spins}} \langle 0|J_\mu|e^+e^-\rangle \langle 0|J_\nu|e^+e^-\rangle^* \\ &= \frac{1}{4} \operatorname{Tr}\left(\gamma_\mu \frac{\not{q}}{2E} \gamma_\nu \frac{\not{q}'}{2E}\right) \\ &= \frac{e^2}{4E^2}\left(q_\mu q'_\nu + q_\nu q'_\mu - g_{\mu\nu} q \cdot q'\right) \end{aligned} \tag{11.35}$$

The hadronic part, to lowest order, can also be calculated in the same way:

$$H_{\mu\nu} = \frac{e_f^2}{E^2}\left(p_\mu p'_\nu + p_\nu p'_\mu - g_{\mu\nu} p \cdot p'\right) \tag{11.36}$$

where e_f is the electric charge of the quark flavor index f. Then we have:

$$|\mathscr{M}|^2 = \frac{e^2 e_f^2}{4E^4}(1 + \cos^2\theta) \tag{11.37}$$

where θ is the center-of-mass scattering angle, so the differential and total cross sections for the two-jet process are given by:

$$\begin{aligned} \left(\frac{d\sigma}{d\Omega}\right)_0 &= \frac{\alpha^2}{4S}\left(1 + \cos^2\theta\right) \sum_{f,c} \left(\frac{e_f}{e}\right)^2 \\ \sigma_0 &= \frac{4\pi\alpha^2}{3S} \sum_{f,c} \left(\frac{e_f}{e}\right)^2 \end{aligned} \tag{11.38}$$

This $(1 + \cos^2\theta)$ dependence on the angle θ has actually been seen experimentally in two-jet processes, strengthening our belief in spin-$\frac{1}{2}$ quarks.

For the three-jet event the calculation is a bit more difficult because of the kinematics. The only complication is the hadronic part:

$$\begin{aligned} H_{\mu\nu} &= \sum_{s,s',c,f,t} \langle q\bar{q}g|J_\mu|0\rangle\langle q\bar{q}g|J_\nu|0\rangle^* \\ &= \frac{1}{8p_0p_0'k_0}\sum_{f,t} e_f^2 g^2 \,\mathrm{Tr}\Big[\not{p}\Big(\not{\epsilon}\frac{1}{p\cdot k}(\not{p}+\not{k})\gamma_\mu \\ &\quad - \gamma_\mu\frac{1}{p'\cdot k}(\not{p}+\not{k})\not{\epsilon}\Big) \\ &\quad \times \not{p}'\Big(\gamma_\nu\frac{1}{p\cdot k}(\not{p}+\not{k})\not{\epsilon} - \not{\epsilon}\frac{1}{p'\cdot k}(\not{p}'+\not{k})\gamma_\nu\Big)\Big] \end{aligned} \tag{11.39}$$

where t is the gluon polarization. This then becomes:

$$\begin{aligned} H_{\mu\nu} &= \frac{2}{p_0p_0'k_0}\sum_f e_f^2 g^2 \\ &\quad \times \Big(\frac{p\cdot p'}{(p\cdot k)(p'\cdot k)}\big(2[p,p']_{\mu\nu} + [k, p+p']_{\mu\nu}\big) \\ &\quad + \frac{1}{p\cdot k}\big([p', p+k]_{\mu\nu} - [p,p]_{\mu\nu}\big) \\ &\quad + \frac{1}{p'\cdot k}\big([p, p'+k]_{\mu\nu} - [p',p']_{\mu\nu}\big)\Big) \end{aligned} \tag{11.40}$$

where we have defined $[p,q]_{\mu\nu} \equiv p_\mu q_\nu + q_\mu p_\nu - g_{\mu\nu}p\cdot q$. Then the transition element:

$$|\mathscr{M}|^2 = \left(\frac{1}{4E^2}\right)^2 l_{\mu\nu}H^{\mu\nu} \tag{11.41}$$

becomes:

$$\begin{aligned} |\mathscr{M}|^2 &= \frac{1}{16E^6p_0p_0'k_0}\sum_f (ee_fg)^2 \\ &\quad \times \Big(\frac{p\cdot p'}{(p\cdot k)(p'\cdot k)}\Big[2(p\cdot q)(p'\cdot q') + 2(p\cdot q')(p'\cdot q) \\ &\quad - 2(k\cdot q)(k\cdot q') + (Q\cdot q)(Q\cdot k)\Big] \\ &\quad + \frac{1}{p\cdot k}\Big[(Q\cdot q)(Q\cdot p') - 2(p\cdot q)(p\cdot q') - 2(p'\cdot q)(p'\cdot q')\Big] \end{aligned}$$

$$+\frac{1}{p'\cdot k}\Big[(Q\cdot q)(Q\cdot p)-2(p\cdot q)(p\cdot q')-2(p'\cdot q)(p'\cdot q')\Big]\Big) \tag{11.42}$$

In practice, it is difficult to tell which jet emerged from which constituent. Therefore, let us number the jets. Let $p_{i\mu}$ represent the momentum of the ith jet. We define the variables:

$$x_i \equiv p_{i0}/E \tag{11.43}$$

where $p_3 = k$. Let us also choose the center-of-mass frame, so that: $q = (E, \mathbf{q})$ and $q' = (E, -\mathbf{q})$.

Then the kinematics gives us:

$$x_1 + x_2 + x_3 = 2 \tag{11.44}$$

The differential cross section is then given by:

$$d\sigma = \frac{\pi}{(8\pi^3)^2}\int |\mathscr{M}|^2 d^3p\, d^3p'\, d^3k\, \delta^4\left(p + p' + k - Q\right) \tag{11.45}$$

Inserting the value of $\mathscr{M}$ into the cross section, we get:

$$\begin{aligned} d\sigma &= \frac{2\alpha^2}{3E^2}\frac{g^2}{4\pi}dx_1\, dx_2 \sum_f \left(\frac{e_f}{e}\right)^2 \\ &\quad \times \left(\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)} + \text{perm}\right) \end{aligned} \tag{11.46}$$

The experimental data is sometimes analyzed in terms of a quantity called "thrust," which can be defined as:

$$T = \frac{2\left(\sum_i' p_{||i}\right)_{\max}}{\sum_i |p_i|} \tag{11.47}$$

where the prime in the numerator means we sum over all particles in just one hemisphere. The thrust is a good variable because it varies from $T = \frac{1}{2}$ (isotropy) to $T = 1$ (perfect jet behavior). In Figure 11.5, we see some typical three-jet events found at PETRA compared to the theoretical prediction. The agreement with the experimental data is excellent.

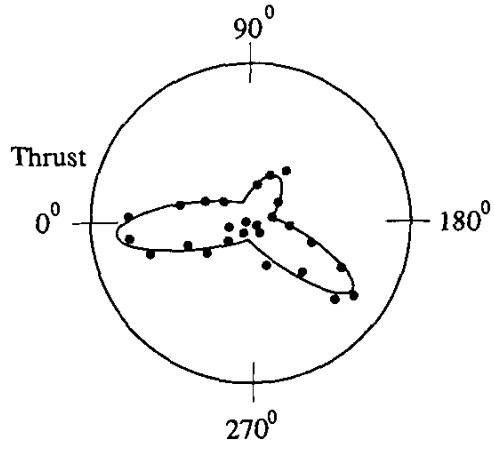


Figure 11.5. Experimentally, jets are found in collision processes, confirming the prediction of QCD (solid line).

11.4 Current Algebra

In the 1960s, before the rise of gauge theories and the Weinberg–Salam model, the original four-fermion Fermi action was a useful phenomenological guide that could account for some of the qualitative features of the weak interactions.

The hadrons participate in the weak interactions. For example, the hadrons are mostly unstable and decay via the weak interactions. The β decay of the neutron is the classic example of the weak interaction of the hadrons. Mimicking the success of the simple Fermi action, attempts were made to describe the hadronic weak interactions by postulating that the action was the product of two currents. The phenomenology of the weak interactions in the 1960s was dominated by something called *current algebra*,[16,17] which postulated the commutation relations of the currents among themselves. By taking their matrix elements, one could therefore derive sum rules that linked different physical processes. Although there was no understanding why the effective action should be the product of two currents, or why they should obey a chiral algebra, these current algebra relations agreed rather well with experiment. From the perspective of the Standard Model, however, we can see the simple origin of the current algebras.

To be specific, let us analyze the weak interactions of the quarks. We will insert quarks into the Weinberg–Salam model, arranging them into standard $SU(2)$ doublets and $U(1)$ singlets, which gives us the Glashow–Weinberg–Salam model. We now pair off the three generations of quarks u, d, c, s, t, and b with the three generations of leptons:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \begin{pmatrix} u \\ d \end{pmatrix}_L ; \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \begin{pmatrix} c \\ s \end{pmatrix}_L ; \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \begin{pmatrix} t \\ b \end{pmatrix}_L \tag{11.48}$$