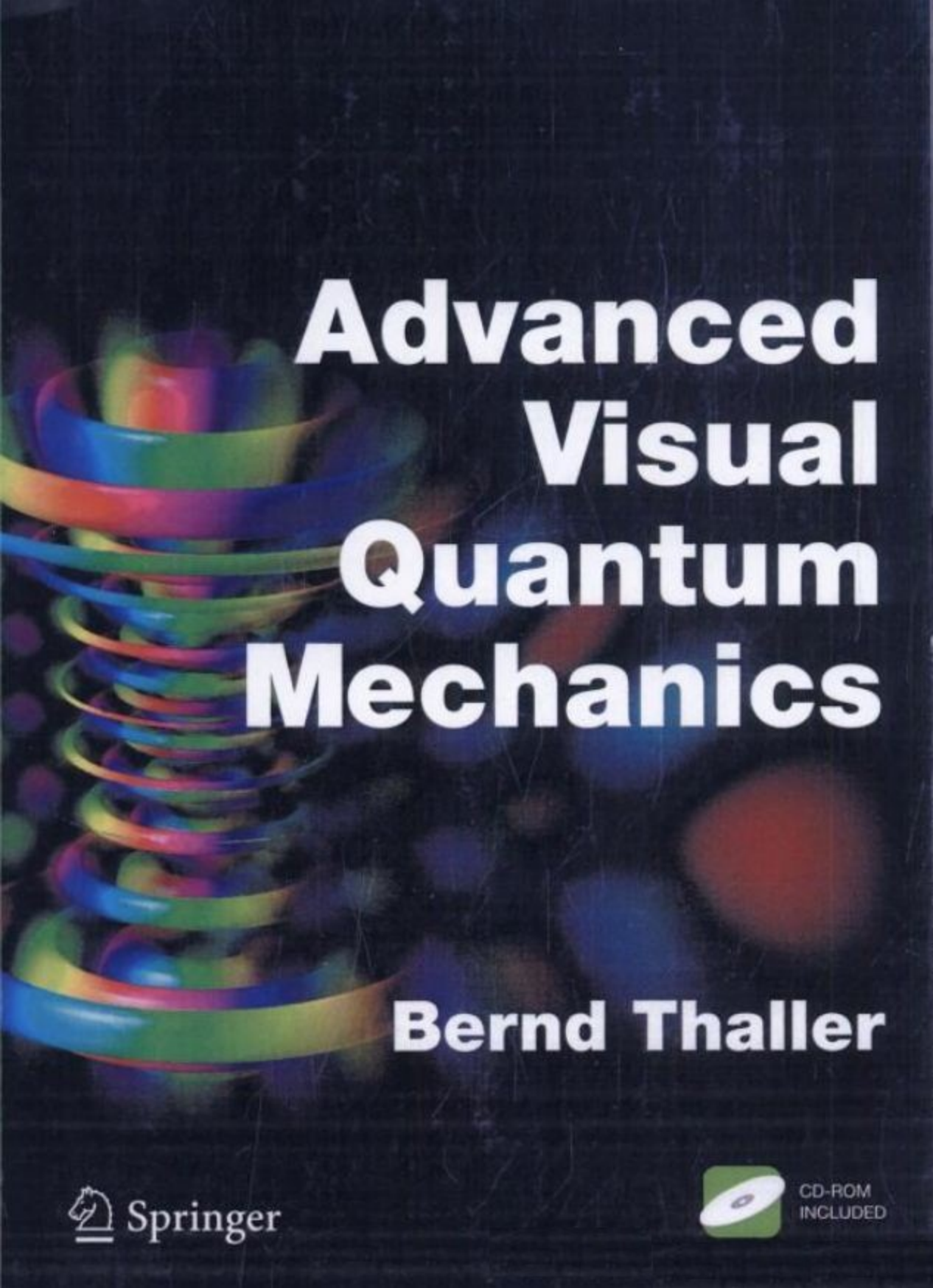
Advanced Visual Quantum Mechanics

Bernd Thaller



Springer

CD-ROM INCLUDED

Advanced Visual Quantum Mechanics

Bernd Thaller

Advanced Visual Quantum Mechanics

With 103 Illustrations



INCLUDES CD-ROM

Springer

Bernd Thaller
Institute for Mathematics and Scientific Computing
University of Graz
A-8010 Graz
Austria
bernd.thaller@uni-graz.at

Library of Congress Cataloging-in-Publication Data
Thaller, Bernd, 1956-
Advanced visual quantum mechanics / Bernd Thaller
p. cm.
Includes bibliographical references and index
ISBN 0-387-20777-5 (acid-free paper)
1. Quantum theory. 2. Quantum theory--Computer simulation. I. Title.

QC174.12.T45 2004
530.12--dc22 2003070771

Mathematica® is a registered trademark of Wolfram Research, Inc.
QuickTime™ is a registered trademark of Apple Computer, Inc., registered in the United States and other countries. Used by licence.
Macromedia and Macromedia® Director™ are registered trademarks of Macromedia, Inc., in the United States and other countries.

ISBN 0-387-20777-5 Printed on acid-free paper.

© 2005 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (HAM)

9 8 7 6 5 4 3 2 1 SPIN 10945810

springeronline.com

Preface

Advanced Visual Quantum Mechanics is a systematic effort to investigate and to teach quantum mechanics with the aid of computer-generated animations. But despite its use of modern visualization techniques, it is a conventional textbook of (theoretical) quantum mechanics. You can read it without a computer, and you can learn quantum mechanics from it without ever using the accompanying CD-ROM. But, the animations will greatly enhance your understanding of quantum mechanics. They will help you to get the intuitive feeling for quantum processes that is so hard to obtain from the mathematical formulas alone.

A first book with the title *Visual Quantum Mechanics* ("Book One") appeared in the year 2000. The CD-ROM for Book One earned the European Academic Software Award (EASA 2000) for outstanding innovation in its field. The topics covered by Book One mainly concerned quantum mechanics in one and two space dimensions. *Advanced Visual Quantum Mechanics* ("Book Two") sets out to present three-dimensional systems, the hydrogen atom, particles with spin, and relativistic particles. It also contains a basic course of quantum information theory, introducing topics like quantum teleportation, the EPR paradox, and quantum computers. Together, the two volumes constitute a fairly complete course on quantum mechanics that puts an emphasis on ideas and concepts and satisfies some modest requirements of mathematical rigor. Nevertheless, Book Two is fairly self-contained. References to Book One are kept to a minimum so that anyone with a basic training in quantum mechanics should be able to read Book Two independently of Book One. Appendix A includes a short synopsis of quantum mechanics as far as it was presented in Book One.

The CD-ROM included with this book contains a large number of QuickTime movies presented in a multimedia-like environment. The movies illustrate the text, add color, a time-dimension, and a certain level of interactivity. The computer-generated animations will help you to explore quantum mechanics in a systematic way. The point-and-click interface gives you quick and easy access to all the movies and lots of background information. You need no special computer skills to use the software. In fact, it is no more

difficult than surfing the Internet. You are not required to produce simulations by yourself. The general idea is that you should first think about quantum mechanics and not about computers. The movies provide some phenomenological background. They will train and enhance your intuition, and the desire to understand the movies should motivate you to learn the (sometimes nasty, sometimes elegant) theory.

Computer visualizations are particularly rewarding in quantum mechanics because they allow us to depict objects and events that cannot be seen by other means. However, one has to be aware of the fact that the animations depict the mathematical objects describing reality, not reality itself. Usually, one needs some explanation and interpretation to understand the visualizations. The visualization method used here makes extensive use of color. It displays all essential information about the quantum state in an intuitive way. Watching the numerous animations will thus create an intuitive feeling for the behavior of quantum systems—something that is hardly achieved just by solving the Schrödinger equation mathematically. I would even say that the movies allow us to see the whole subject in a new way. In any case, the "visual approach" had a great influence on the selection of topics as well as on the style and the level of the presentation. For example, *Visual Quantum Mechanics* puts an emphasis on quantum dynamics, because a movie adds a natural time-dimension to an illustration. Whereas other textbooks stop when the eigenfunctions of the Hamiltonian are obtained, this book will go on to discuss dynamical effects.

It depends on the situation, but also on the personality of the student or of the teacher, how the movies are used. In some cases, the movies are certainly useful to stimulate the student's interest in some phenomenon. The animation thus serves to motivate the development of the theory. In other cases, it is, perhaps, more appropriate to show a movie confirming the theory by an example. Personally, I present the movies by video projection as a supplement to an introductory course on quantum mechanics. I talk about the movies in a rather informal way, and soon the students start asking interesting questions that lead to fruitful discussions and deeper explanations. Often, the movies motivate students to study related topics on their own initiative.

One could argue that in advanced quantum mechanics, visualizations are not very useful because the student has to learn abstract notions and that he or she should think in terms of linear operators, Hilbert spaces, and so on. It is certainly true that a solid foundation of these subjects is indispensable for a deeper understanding, and you will have occasion to learn much about the mathematical theory from this text. But, I claim that despite a good training in the abstract theory, you can still gain a lot from the visualizations.

Talking about my own experience, I found that I learned much, even about simple systems, when I prepared the movies for *Visual Quantum Mechanics*. For example, having done research on the mathematical aspects of the Dirac equation for several years, I can claim to have a good background concerning the quantum mechanical abstractions in this field. But nevertheless, I was not able to predict how a wave packet performing a "Zitterbewegung" would appear until I started to do some visualizations of that phenomenon. Moreover, when one tries to understand the visualizations one often encounters phenomena, that one is able to explain with the theory, but that one simply hasn't thought of before. The main thing that you can gain from the visualizations is a good feeling for the behavior of solutions of the quantum mechanical equations.

Though the CD-ROM presents a few simple interactive simulations in the chapter about qubits, the overwhelming content consists of prefabricated movies. A true computer simulation, that is, a live computation of some process, would of course allow a higher degree of interactivity. The reader would have more flexibility in the choice of parameters and initial conditions. But in many cases, this approach is forbidden because of the insufficient speed of present-day computers. Moreover, in order to produce a useful visualization, one has to analyze the physical system very carefully. For every situation, one has to determine the scale of space and time and suitable ranges of the parameters where something interesting is going to happen. In quantum mechanics, the number of possibilities is very large, and if one chooses the wrong parameter values, it is very likely that nothing can be seen that is easily interpreted or that shows some effect in an interesting way. Therefore, I would not recommend to learn basic quantum mechanics by doing time-consuming computer simulations.

Producing simulations and designing visualizations can, however, bring enormous benefit to the advanced student who is already familiar with the foundations of quantum mechanics. Many of the animations on the CD-ROM were done with the help of Mathematica. With the exception of the Mathematica software, all the necessary tools for producing similar results are provided on the CD-ROM: The source code for all movies, Mathematica packages both for the numerical solution of the Schrödinger equation and for the graphical presentation of the results, and OpenGL-based software for the three-dimensional visualization of wave functions. My recommendation is to start with some small projects based on the examples provided by the CD-ROM. It should not be difficult to modify the existing Mathematica notebooks by slightly varying the parameters and initial conditions, and then watching and interpreting the results. You could then proceed to look for other examples of quantum systems that might be good for a

physically or mathematically interesting visualization. When you produce a visualization, often some natural questions about the system will arise. This makes it necessary to learn more about the system (or about quantum mechanics), and by knowing the system better, you will produce better visualizations. When the visualization finally becomes useful, you will understand the system almost perfectly. This is "learning by doing", and it will certainly enhance your understanding of quantum mechanics, as the making of this book helped me to understand quantum mechanics better. Be warned, however, that personal computers are still too slow to perform simulations of realistic quantum mechanical processes within a reasonable time. Many of the movies provided with this book typically took several hours to generate.

Concerning the mathematical prerequisites, I tried to keep the two books on an introductory level. Hence, I tried to explain all the mathematical methods that go beyond basic courses in calculus and linear algebra. But, this does not mean that the content of the book is always elementary. It is clear that any text that sets out to explain quantum phenomena must have a certain level of mathematical sophistication. Here, this level is occasionally higher than in other introductions, because the text should provide the theoretical background for the movies. Doing visualizations is more than just obtaining numerical solutions. A surprising amount of mathematical know-how is in fact necessary to prepare an animation. Without presenting too many unnecessary details, I tried to include just what I thought was necessary to produce the movies. My approach to teaching quantum mechanics thus makes no attempt to trivialize this subject. The animations do not replace mathematical formulas. But in order to facilitate the approach for the beginner, I marked some of the more difficult sections as "special topics" and placed the symbol $\boxed{\Psi}$ in front of paragraphs intended for the mathematically interested reader. These parts may be skipped at first reading.

Though the book thus addresses students and scientists with some background in mathematics, the movies (together with the movies of Book One) can certainly be used in front of a wider audience. The success, of course, depends on the style of the presentation. I myself have had the occasion to use the movies in lectures for high-school students and for scientifically interested people without any training in higher mathematics. Based on this experience, I hope that the book together with CD-ROM will have broader applications than each could have if used alone.

According to its subtitle, Book Two can be divided roughly into three parts: atomic physics (Chapters 1–3), quantum information theory (Chapters 4–6), and relativistic quantum mechanics (Chapters 7, 8). This division, however, should not be taken too seriously. For example, Chapter 4 on

qubits completes the discussion of spin-1/2 particles in Chapter 3 and serves at the same time as an introduction to quantum information theory. Chapter 5 discusses composite quantum systems by combining topics relevant for quantum information theory (for example, two-qubit systems) with topics relevant for atomic physics (for example, addition of angular momenta).

Together, Book One and Book Two cover a wide range of the standard quantum physics curriculum and supplement it with a series of advanced topics. For the sake of completeness, some important topics have been included in the form of several appendices: the perturbation theory of eigenvalues, the variational method, adiabatic time evolution, and formal scattering theory. Though most of these matters are very well suited for an approach using lots of visualizations and examples, I simply had neither time nor space (the CD-ROM is full) to elaborate on these topics as I would have liked to do. Therefore, these appendices are rather in the style of an ordinary textbook on advanced theoretical physics. I would be glad if this material could serve as a background for the reader's own ventures into the field of visualization. If there should ever be another volume of Visual Quantum Mechanics, it will probably center on these topics and on others like the Thomas-Fermi theory, periodic potentials, quantum chaos, and semiclassical quantum mechanics, just to name a few from my list of topics that appear to be suitable for a modernized approach in the style of Visual Quantum Mechanics.

This book has a home page on the internet with URL

`http://www.uni-graz.at/imawww/vqm/`

An occasional visit to this site will inform you about software upgrades, printing errors, additional animations, etc.

Acknowledgements

I would like to thank my son Wolfgang who quickly wrote the program "QuantumGL" when it turned out that the available software wouldn't serve my purposes. Thanks to Manfred Liebmann, Gerald Roth, and Reinhold Kainhofer for help with *Mathematica*-related questions. I am very grateful to Jerry Batzel who read large parts of the manuscript and gave me valuable hints to improve my English. This book owes a lot to Michael A. Morrison. He studied the manuscript very carefully, made a large number of helpful comments, asked lots of questions, and eliminated numerous errors. Most importantly, he kept me going with his enthusiasm. Thanks, Michael. Financial support from Steiermärkische Landesregierung, from the University of Graz, and from Springer-Verlag is gratefully acknowledged.

Graz, January 2004 *Bernd Thaller*

Contents

Preface v

Chapter 1. Spherical Symmetry 1
1.1. A Note on Symmetry Transformations 2
1.2. Rotations in Quantum Mechanics 7
1.3. Angular Momentum 12
1.4. Spherical Symmetry of a Quantum System 17
1.5. The Possible Eigenvalues of Angular-Momentum Operators 21
1.6. Spherical Harmonics 26
1.7. Particle on a Sphere 34
1.8. Quantization on a Sphere 38
1.9. Free Schrödinger Equation in Spherical Coordinates 44
1.10. Spherically Symmetric Potentials 50

Chapter 2. Coulomb Problem 57
2.1. Introduction 58
2.2. The Classical Coulomb Problem 61
2.3. Algebraic Solution Using the Runge-Lenz Vector 66
2.4. Algebraic Solution of the Radial Schrödinger Equation 70
2.5. Direct Solution of the Radial Schrödinger Equation 84
2.6. Special Topic: Parabolic Coordinates 91
2.7. Physical Units and Dilations 96
2.8. Special Topic: Dynamics of Rydberg States 105

Chapter 3. Particles with Spin 113
3.1. Introduction 113
3.2. Classical Theory of the Magnetic Moment 115
3.3. The Stern-Gerlach Experiment 118
3.4. The Spin Operators 123
3.5. Spinor-Wave Functions 127
3.6. The Pauli Equation 134
3.7. Solution in a Homogeneous Magnetic Field 138
3.8. Special Topic: Magnetic Ground States 142
3.9. The Coulomb Problem with Spin 146

Chapter 4. Qubits 157
4.1. States and Observables 158
4.2. Measurement and Preparation 162
4.3. Ensemble Measurements 167
4.4. Qubit Manipulations 171
4.5. Other Qubit Systems 181
4.6. Single-Particle Interference 189
4.7. Quantum Cryptography 197
4.8. Hidden Variables 200
4.9. Special Topic: Qubit Dynamics 204

Chapter 5. Composite Systems 211
5.1. States of Two-Particle Systems 212
5.2. Hilbert Space of a Bipartite System 216
5.3. Interacting Particles 221
5.4. Observables of a Bipartite System 223
5.5. The Density Operator 227
5.6. Pure and Mixed States 233
5.7. Preparation of Mixed States 238
5.8. More About Bipartite Systems 244
5.9. Indistinguishable Particles 250
5.10. Special Topic: Multiparticle Systems with Spin 256
5.11. Special Topic: Addition of Angular Momenta 259

Chapter 6. Quantum Information Theory 271
6.1. Entangled States of Two-Qubit Systems 272
6.2. Local and Nonlocal 278
6.3. The Einstein-Podolsky-Rosen Paradox 281
6.4. Correlations Arising from Entangled States 285
6.5. Bell Inequalities and Local Hidden Variables 290
6.6. Entanglement-Assisted Communication 300
6.7. Quantum Computers 305
6.8. Logic Gates 307
6.9. Quantum Algorithms 316

Chapter 7. Relativistic Systems in One Dimension 323
7.1. Introduction 324
7.2. The Free Dirac Equation 325
7.3. Dirac Spinors and State Space 327
7.4. Plane Waves and Wave Packets 332
7.5. Subspaces with Positive and Negative Energies 339
7.6. Kinematics of Wave Packets 343
7.7. Zitterbewegung 347

7.8. Special Topic: Energy Representation and Velocity Space 359
7.9. Relativistic Invariance 365

Chapter 8. The Dirac Equation 377
8.1. The Dirac Equation 378
8.2. Relativistic Covariance 382
8.3. Classification of External Fields 389
8.4. Positive and Negative Energies 392
8.5. Nonrelativistic Limit and Relativistic Corrections 402
8.6. Spherical Symmetry 410
8.7. The Dirac-Coulomb Problem 417
8.8. Relativistic Hydrogen Atom 426

Appendix A. Synopsis of Quantum Mechanics 433

Appendix B. Perturbation of Eigenvalues 443

Appendix C. Special Topic: Analytic Perturbation Theory 455

Appendix D. Variational Method 461

Appendix E. Adiabatic and Geometric Phases 467

Appendix F. Formal Scattering Theory 475

Appendix G. Books 491

Appendix H. Movie Index 493
0. Introduction 493
1. Spherical Symmetry 494
2. Coulomb Problem 495
3. Spin 498
4. Qubits 499
5. Composite Systems 500
6. Relativistic Systems 501

List of Symbols 505

Index 511

Chapter 1

Spherical Symmetry

Chapter summary: In the first book of *Visual Quantum Mechanics*, we considered mainly one- and two-dimensional systems. Now we turn to the investigation of three-dimensional systems. This chapter is devoted to the very important special case of systems with spherical symmetry.

In the presence of spherical symmetry, the Schrödinger equation has solutions that can be separated into a product of a radial part and an angular part. In this chapter, all possible solutions of the equation for the angular part will be determined once and for all.

We start by discussing symmetry transformations in general. In quantum mechanics, all symmetry transformations may be realized by unitary or antiunitary operators. We define the unitary transformations corresponding to rotations of a particle in $\mathbb{R}^3$. Their self-adjoint generators are the components of the orbital angular momentum $\mathbf{L}$. We describe the angular-momentum commutation relations and discuss their geometrical meaning.

A quantum system is called invariant under a given symmetry transformation if the Hamiltonian commutes with the corresponding unitary operator. A particle moving under the influence of a potential $V(\mathbf{x})$ is a spherically symmetric system (invariant under rotations) if the potential function depends only on the distance r from the origin. Spherical symmetry implies the conservation of the angular momentum and determines the structure of the eigenvalue spectrum of the Hamiltonian (degeneracy). The square L^2 and any component L_k of the angular momentum can be diagonalized simultaneously with the Hamiltonian of a spherically symmetric system. The structure of the common system of eigenvectors can essentially be derived from the angular-momentum commutation relations. In general, the possible eigenvalues of the angular-momentum operators are characterized by integer and half-integer quantum numbers. It turns out, however, that only integer quantum numbers occur in case of the orbital angular momentum.

The eigenvalues and eigenfunctions (spherical harmonics) of the orbital angular momentum are then determined explicitly. The spherical harmonics are the energy eigenfunctions of a particle whose configuration space is a sphere (rigid rotator). The rigid rotator can serve as a simple model for a diatomic molecule in its vibrational ground state.

The restriction of the eigenvalue problem to an angular-momentum eigenspace reduces the Schrödinger equation to an ordinary differential equation. We conclude the chapter with a brief discussion of this so-called radial Schrödinger equation.

1.1. A Note on Symmetry Transformations

1.1.1. Rotations as symmetry transformations

Consider a physical system S in three-dimensions, for example, a few particles moving under the influence of mutual and external forces. The state of S is described with respect to a given coordinate system I in terms of suitably chosen coordinates $\mathbf{x} \in \mathbb{R}^3$. We remind the reader of the following basic assumption.

Homogeneity and isotropy of space:

No point and no direction in $\mathbb{R}^3$ is in any way physically distinguished. Therefore, the behavior of physical systems should not depend on the location of the experimenter's lab or its orientation in space (*principle of relativity*).

In order to test the isotropy of space, we can perform an experiment with the physical system S in the coordinate system I and then repeat the experiment in a rotated coordinate system I'. This can be done in several different ways (see Fig. 1.1).

(1) **Rotate the system and the observer**. This procedure consists in rotating the whole experimental setup: the system S (the particles, the external forces, the devices for preparing the initial state) and the observer (the measurement devices). The isotropy of space means that with respect to the rotated frame of reference I', the system behaves exactly as it did in I. The mathematical description is exactly the same as before. The only difference is that the coordinates now refer to the new coordinate frame I'.

(2) **Rotate the system but not the observer** (*active transformation*). Now the rotated physical system has to be described by an observer in the old coordinate frame I. The motion of the system S will look different, and the observer has to change the mathematical description (in particular, the numerical values of the coordinates). From the point of view of the observer, the rotation changes the state of the system. Hence, the rotation corresponds to a transformation T in the state space of S. We say that the transformation T is a *representation* of the rotation in the state space of the system.

(3) **Rotate the observer but not the system** (*passive transformation*). This procedure is equivalent to procedure (2), but in the mathematical description, T has to be replaced by the inverse transformation. This can be seen as follows: With respect to the new coordinates in I' (that is, from the point of view of a rotated observer), the states of the physical system

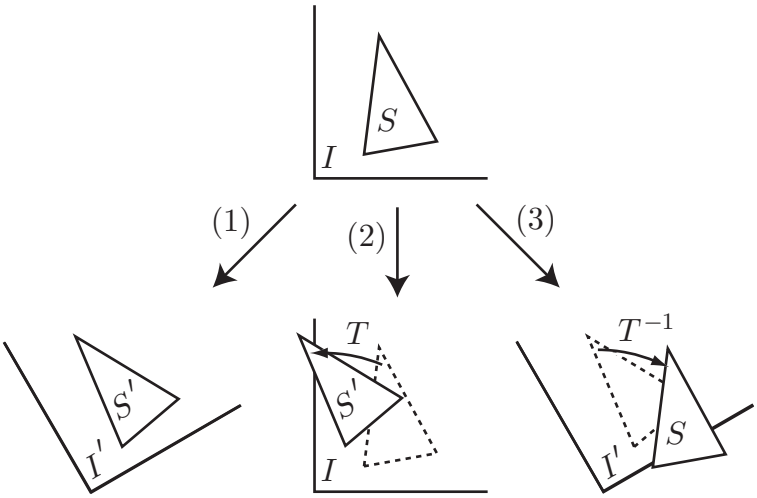


FIGURE 1.1. Symmetry transformations of a physical system. (1) Both the physical system S and the frame of reference I are transformed. The behavior of the system and the mathematical description remain unchanged (principle of relativity). (2) The system is transformed with respect to a fixed coordinate frame I. The states of the system undergo a transformation T. (3) The frame I is transformed, the system is left unchanged. T^{-1} maps the states in I to the states in I'.

appear to be transformed by a mapping T'. Now we can perform an active transformation by T, as described in (2), and we end up with situation (1): Both the physical system and the observer are rotated to I', and by the principle of relativity the behavior in I' is the same as it was in I. Hence, the transformation T' followed by the transformation T gives the identity. A similar argument applies to T followed by T'. We conclude that $T' = T^{-1}$.

In the following, we prefer the "active" point of view expressed in (2). We choose a fixed coordinate system and perform rotations with the objects. Let us assume that an experiment changes the system's initial state A to a certain final state B (with respect to the coordinate frame I). The rotated system has the initial state $A' = T(A)$ (again with respect to I). Repeating the experiment with the rotated system changes its state into B'. What is the relation between the final states B' and B? The principle of relativity states that the rotation does not change the physical laws that govern the system, that is, the mechanism relating the initial and the final state. Hence,

the same relation that holds for the initial states must also hold for the final states: $B' = T(B)$.

If the properties of the system depend on its orientation, then some additional influence would alter the transition to the final state, and B' would in general be different from $T(B)$. The same is true if not everything that is relevant to the behavior of the system is transformed in the same way. For example, one rotates the particles but not the external fields. In this case, the system is subject to a changed external influence, and the final state B' of the rotated system will differ from the rotated final state $T(B)$ of the original system.

The discussion above applies not only to rotations but also to other transformations of the system. In general, a symmetry transformation need not be related to geometry (an example is the exchange of two identical particles, see Section 5.9). Let us try to give a general (but somewhat vague) definition of a symmetry transformation.

A *symmetry transformation* of a physical system is an invertible transformation T that can be applied to all possible states of the system such that all physical relations among the states remain unchanged.

The mathematical description of a symmetry transformation T depends on how the states are described in a physical theory. The next section shows how symmetry transformations are implemented in quantum mechanics.

1.1.2. Symmetry transformations in quantum mechanics

Quantum states are usually described in terms of vectors in a Hilbert space $\mathfrak{H}$. But the correspondence between vectors and states is not one-to-one. For a given vector ψ, all vectors in the one-dimensional subspace (*ray*)

$$[\psi] = \{\lambda\psi \mid \lambda \in \mathbb{C}\} \tag{1.1}$$

represent the same state. Hence, the mathematical objects corresponding to the physical states are rays rather than vectors.

The set of states:

A quantum state of a physical system is a one-dimensional subspace $[\psi]$ of the Hilbert space $\mathfrak{H}$ of the system. The set of all possible quantum states will be denoted by $\hat{\mathfrak{H}}$,

$$\hat{\mathfrak{H}} = \{[\psi] \mid \psi \in \mathfrak{H}\} \tag{1.2}$$

In linear algebra, the set of one-dimensional subspaces of a linear space is called a *projective space*.

EXERCISE 1.1. *If ψ and $\lambda\psi$ both represent the same state, and if ϕ and $\mu\phi$ both represent some other state, why do $\psi + \phi$ and $\lambda\psi + \mu\phi$ in general represent different states?*

In quantum mechanics, all experimentally verifiable predictions can be formulated in terms of transition probabilities. The *transition probability* from a state $[\phi]$ to a state $[\psi]$ is defined by

$$P([\phi]\to[\psi]) = |\langle\psi, \phi\rangle|^2 = P([\psi]\to[\phi]), \tag{1.3}$$

where ϕ and ψ are arbitrary unit vectors in $[\phi]$ and $[\psi]$, respectively. Transition probabilities may be regarded as the basic physically observable relations among quantum states.

Hence, the basic requirement for a symmetry transformation is that the transition probability between any two states should be the same as between the corresponding transformed states.

Definition:

A *symmetry transformation* in quantum mechanics is a transformation of rays that preserves transition probabilities. More precisely, a map $T : \hat{\mathfrak{H}} \to \hat{\mathfrak{H}}$ is a symmetry transformation if it is one-to-one and onto and satisfies

$$P(T[\phi]\to T[\psi]) = P([\phi]\to[\psi]) \quad \text{for all states } [\phi] \text{ and } [\psi]. \tag{1.4}$$

1.1.3. Realizations of symmetry transformations

Instead of working with rays, it is more convenient to describe symmetry transformations in terms of the vectors in the underlying Hilbert space. Consider, for example, a unitary or antiunitary[1] operator U in the Hilbert space $\mathfrak{H}$. The operator U induces a ray transformation in a very natural way. To this purpose, choose a vector ψ representing the state $[\psi]$ and define the ray transformation $\hat{U}$ associated with the operator U by

$$\hat{U}[\psi] = [U\psi]. \tag{1.5}$$

$\hat{U}$ transforms the ray $[\psi]$ into the one-dimensional subspace spanned by the vector $U\psi$.

[1] An *antiunitary operator* A is a one-to-one map from $\mathfrak{H}$ onto $\mathfrak{H}$ which is antilinear, that is, $A(\alpha\psi + \beta\phi) = \overline{\alpha}\, A(\psi) + \overline{\beta}\, A(\phi)$, and satisfies $\langle A\psi, A\phi\rangle = \langle\phi, \psi\rangle$, whereas a unitary transformation U is linear and satisfies $\langle U\psi, U\phi\rangle = \langle\psi, \phi\rangle$.

EXERCISE 1.2. *Show that it follows from the linearity or antilinearity of U that the definition* (1.5) *does not depend on the chosen representative ψ.*

A unitary operator U leaves the scalar product invariant, and hence the corresponding ray transformation $\hat{U}$ must be a symmetry transformation. The same is true for an antiunitary operator, which does not change the absolute value of the scalar product.

The following famous theorem due to Eugene P. Wigner states that unitary and antiunitary operators are in fact the only ways to realize symmetry transformations.

Theorem of Wigner:

Every symmetry transformation T in $\hat{\mathfrak{H}}$ is of the form

$$T = \hat{U}, \quad \text{where } U \text{ is either unitary or antiunitary in } \mathfrak{H}. \tag{1.6}$$

Two operators U_1 and U_2 representing the same symmetry transformation differ at most by a phase factor,

$$U_1 = e^{\mathrm{i}\theta} U_2, \quad \text{for some } \theta \in [0, 2\pi). \tag{1.7}$$

In particular, U_1 and U_2 are either both unitary or both antiunitary.

Ψ The investigation and classification of the possible symmetry transformations has played an important role in mathematical physics. For example, according to the special theory of relativity, a relativistic system must admit the Lorentz transformations as symmetry transformations. It must be possible to implement all (proper orthochronous) Lorentz transformations as unitary operators in the corresponding Hilbert space. This imposes some restrictions on the possible choices of Hilbert spaces and scalar products for relativistic systems. In fact, the theory of group representations allows one to classify all possible relativistic wave equations and their associated Hilbert spaces (scalar products).

1.1.4. Invariance of a physical system

A symmetry transformation of the states also induces a similarity transformation of the linear operators in the Hilbert space of a physical system. Let U be a unitary or antiunitary operator representing a given symmetry transformation. Assume that two vectors ϕ and ψ are related by the equation $\phi = A\psi$, where A is a linear operator. After the symmetry transformation, the transformed states are related by

$$U\phi = UA\psi = UAU^{-1}U\psi. \tag{1.8}$$

Here, we have inserted the operator $U^{-1}U = 1$ (unitarity condition). Hence, the corresponding relation between the transformed vectors $U\phi$ and $U\psi$ is given by the linear operator UAU^{-1}. We see that after applying the symmetry transformation, an operator A has to be replaced by the operator UAU^{-1}.

EXERCISE 1.3. *Prove that UAU^{-1} is self-adjoint, whenever A is self-adjoint and U is unitary or antiunitary.*

EXERCISE 1.4. *Explain in what sense the expectation value of an observable is invariant under symmetry transformations.*

Sometimes an observable might be unchanged by a given symmetry transformation. In such a case the operator is said to be invariant. This is often very useful information about the system. A physical system for which the Hamiltonian operator itself is invariant is said to possess a symmetry or invariance.

Definition:

A physical system is *invariant* under a symmetry transformation U (or *symmetric* with respect to U) if the Hamiltonian H of the system has the property

$$H = UHU^{-1}. \tag{1.9}$$

The symmetry transformation U is called an *invariance transformation* of the system represented by H. Invariance transformations are usually very helpful for the solution of the Schrödinger equation. In this chapter, we want to investigate systems that are invariant under rotations (spherically symmetric). But first we have to describe the unitary operators corresponding to rotations, and their self-adjoint generators, the angular-momentum operators.

1.2. Rotations in Quantum Mechanics

1.2.1. Rotation of vectors in $\mathbb{R}^3$

Rotations in the three-dimensional space $\mathbb{R}^3$ are described by orthogonal 3×3 matrices with determinant $+1$. You are perhaps familiar with the following matrix that rotates any vector through an angle α about the x_3-axis of a fixed coordinate system

$$\mathbf{R}(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \alpha \in \mathbb{R}. \tag{1.10}$$

There are similar matrices for rotations about the other coordinate axes. An arbitrary rotation can be most intuitively characterized by a rotation vector $\boldsymbol{\alpha} = \alpha \mathbf{n}$, where α specifies the angle of the rotation, and the unit vector $\mathbf{n}$ gives the axis (here the sense of the rotation is determined by the right-hand rule). We consider only rotation angles α with $-\pi < \alpha \leq \pi$ because the angles $\alpha + 2\pi k$ (with k an integer) may be identified with α. Moreover, a rotation through a negative angle about the axis $\mathbf{n}$ is the same as a rotation through a positive angle about the axis defined by $-\mathbf{n}$. Hence, it is sufficient to consider rotation angles α in the interval $[0, \pi]$.

The elements of the 3×3 rotation matrix $\mathbf{R}(\boldsymbol{\alpha})$ are given by

$$\mathbf{R}(\boldsymbol{\alpha})_{ik} = \delta_{ik} \cos\alpha + n_i \, n_k \, (1 - \cos\alpha) - \sum_{m=1}^{3} \epsilon_{ikm} \, n_m \, \sin\alpha. \tag{1.11}$$

Here, we have used the Kronecker delta symbol δ_{ik} and the totally antisymmetric tensor ϵ_{ikm}, which are defined by

$$\delta_{ik} = \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k. \end{cases} \tag{1.12}$$

$$\epsilon_{ikm} = \begin{cases} 1, & \text{if } (i,k,m) \text{ is a cyclic permutation of } (1,2,3), \\ -1, & \text{for other permutations,} \\ 0, & \text{else.} \end{cases} \tag{1.13}$$

Any rotation matrix has determinant 1 and is orthogonal, that is, the transposed matrix is equal to the inverse:

$$\mathbf{R}(\boldsymbol{\alpha})^{\top} = \mathbf{R}(\boldsymbol{\alpha})^{-1}. \tag{1.14}$$

EXERCISE 1.5. *Show that* (1.35) *can be written as*

$$[L_j, L_k] = \mathrm{i}\hbar \sum_{m=1}^{3} \epsilon_{jkm} \, L_m. \tag{1.15}$$

EXERCISE 1.6. *Show that an orthogonal transformation leaves the Euclidean scalar product invariant.*

EXERCISE 1.7. *What sort of transformation is described by an orthogonal matrix with determinant* -1*?*

EXERCISE 1.8. *Verify that the matrices* $\mathbf{R}(\alpha)$ *given by* (1.10) *form a commutative group under matrix multiplication. In particular:*

$$\mathbf{R}(0) = \mathbf{1}_3, \qquad \mathbf{R}(\alpha)\,\mathbf{R}(\beta) = \mathbf{R}(\alpha + \beta), \quad \alpha, \beta \in \mathbb{R}. \tag{1.16}$$

EXERCISE 1.9. *Prove that rotations around different axis in general do not commute.*

EXERCISE 1.10. *Verify that* (1.11) *reduces to* (1.10) *for* $\mathbf{n} = (0, 0, 1)$.

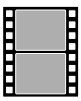
EXERCISE 1.11. *Prove the following formulas for the Kronecker delta and the totally antisymmetric tensor:*

$$\sum_m \epsilon_{klm}\, \epsilon_{ijm} = \delta_{ki}\, \delta_{lj} - \delta_{kj}\, \delta_{li}, \tag{1.17}$$

$$\sum_{l,m} \epsilon_{klm}\, \epsilon_{ilm} = 2\delta_{ki}, \tag{1.18}$$

$$\sum_{k,l,m} \epsilon_{klm}\, \epsilon_{klm} = 6. \tag{1.19}$$

Ψ The set of all rotation matrices $\mathbf{R}(\boldsymbol{\alpha})$ forms a (non-commutative) group. In particular, the composition of any two rotations is again a rotation. Mathematically, the composition of rotations is described by the product of the corresponding rotation matrices. The elements of the rotation group can be characterized by their coordinates $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. The set of all possible coordinates $\boldsymbol{\alpha}$ forms a sphere with radius π in $\mathbb{R}^3$. Note that the matrix elements depend smoothly (analytically) on the parameters $\boldsymbol{\alpha}$. Such a group is called a Lie group. It is a group and a differentiable manifold at the same time. The rotation group is denoted by $SO(3)$, which means "special orthogonal group in three dimensions" ("special" refers to the fact that the determinant is $+1$). The sphere with radius π in $\mathbb{R}^3$ is a useful coordinate space for the rotation group. Every element of the rotation group is uniquely labeled by a rotation vector inside or on that sphere. The sphere is an image of the group manifold. It has unusual topological properties because two points on the surface of the sphere that are connected by a diameter correspond to the same group element (why?) and have to be identified.



CD 1.1 explores the rotation group. The group manifold is visually represented by the coordinate sphere. Any rotation is visualized by the rotation vector $\boldsymbol{\alpha}$ and by the orientation of a rectangular box to which the rotation is applied. The movies show how the orientation of the box changes as the rotation vector moves through the group manifold on straight lines or on closed circles. As a topological space, the group manifold is not simply connected: there are closed orbits that cannot be continuously deformed into a point.

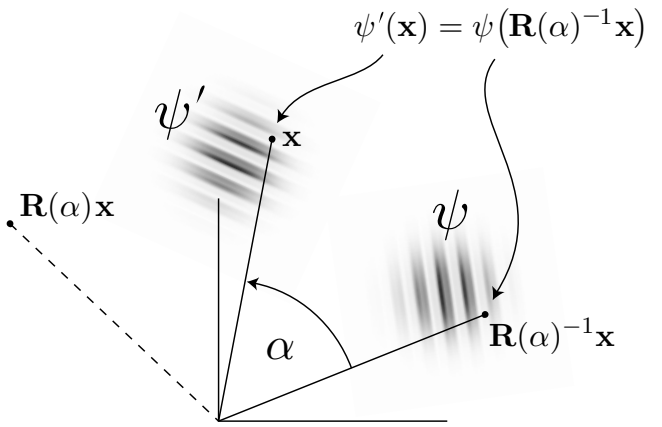


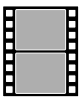
FIGURE 1.2. A rotation $\mathbf{x} \to \mathbf{R}(\alpha)\mathbf{x}$ maps a wave function ψ to $\psi' = U(\alpha)\psi$. The value of the rotated function ψ' at a point $\mathbf{x}$ is given by the value of ψ at the point $\mathbf{R}(\alpha)^{-1}\mathbf{x}$.

1.2.2. Rotation of wave functions

The wave functions considered here are complex-valued functions of the space variable $\mathbf{x}$. Such a function can be rotated by applying a linear operator $U(\boldsymbol{\alpha})$ defined by:

$$\big(U(\boldsymbol{\alpha})\,\psi\big)(\mathbf{x}) = \psi\big(\mathbf{R}(\boldsymbol{\alpha})^{-1}\mathbf{x}\big). \tag{1.20}$$

Here, $\mathbf{R}(\boldsymbol{\alpha})$ is the rotation matrix defined in (1.11). Figure 1.2 explains why we use the inverse rotation matrix in the argument of the function we want to rotate. The operator $U(\boldsymbol{\alpha})$ acts on wave function by a rotation in the literal sense. That is, the "cloud" of complex values that represents the wave function simply gets rotated according to the rotation vector $\boldsymbol{\alpha}$.



The rotations of a box in CD 1.1 can also be interpreted as the rotation of a wave function. Just take the box as an isosurface of some square-integrable wave function ψ, or as the outline of the characteristic function of the box-shaped region. The action of $U(\boldsymbol{\alpha})$ on the wave function ψ just appears as the action of the ordinary rotation $\mathbf{R}(\boldsymbol{\alpha})$ on the box.

For any rotation $\boldsymbol{\alpha}$, the operators $U(\boldsymbol{\alpha})$ are unitary in the Hilbert space $L^2(\mathbb{R}^3)$. The rotations around a fixed axis form a so-called *one-parameter strongly continuous unitary group*. Consider, for example, the rotations about the x_3-axis (see Exercise 1.8). The rotation vector is of the form $\boldsymbol{\alpha} = (0, 0, \alpha)$ with $-\pi \le \alpha \le \pi$. We write $U(\boldsymbol{\alpha}) = U(\alpha)$ and extend the

definition of U to arbitrary real arguments by $U(\alpha \pm 2\pi) = U(\alpha)$. Then we find for all real numbers α and β,

$$U(\alpha)^\dagger = U(\alpha)^{-1} = U(-\alpha), \quad U(0) = \mathbf{1}, \quad U(\alpha)\, U(\beta) = U(\alpha + \beta). \quad (1.21)$$

We refer to Appendix A.6 and to Book One for more details about unitary groups and their self-adjoint generators.

EXERCISE 1.12. *Let $U(\alpha)$, $\alpha \in \mathbb{R}$, describe the rotations around the x_3-axis in space. Using Exercise 1.8, prove that these operators form a unitary group.*

EXERCISE 1.13. *For differentiable functions ψ, and for operators $U(\alpha)$ as in the previous exercise, show that*

$$\frac{\partial}{\partial \alpha} \big(U(\alpha)\, \psi\big)(\mathbf{x})\Big|_{\alpha=0} = \Big(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}\Big)\, \psi(\mathbf{x}) = -\mathrm{i}\, L_3\, \psi(\mathbf{x}). \quad (1.22)$$

Exercise 1.13 above shows that the operator

$$L_3 = \mathrm{i}\hbar\Big(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}\Big) \quad (1.23)$$

is the generator of rotations around the x_3-axis. The operator L_3 is the third component of the angular-momentum operator $\mathbf{L}$ defined in Book One (see also (1.30) below).

If ψ is a differentiable wave function (in the domain of L_3), then its dependence on the angle of rotation can be described by the differential equation

$$\mathrm{i}\hbar \frac{\partial}{\partial \alpha}\, \psi(\mathbf{x}, \alpha) = L_3\, \psi(\mathbf{x}, \alpha). \quad (1.24)$$

This equation is completely analogous to the Schrödinger equation for the time evolution. We can write

$$U(\alpha) = \exp\big(-\frac{\mathrm{i}}{\hbar} L_3\, \alpha\big). \quad (1.25)$$

Similar results hold for the rotations about the x_1- and x_2-axes and the components L_1 and L_2 of the angular momentum.

The components L_1, L_2, and L_3 of the angular-momentum operator $\mathbf{L}$ are the infinitesimal generators of the rotations about the x_1, x_2, and x_3-axis.

1.3. Angular Momentum

1.3.1. Angular momentum in classical mechanics

An observable that is intimately connected with rotations—both in classical and in quantum mechanics—is angular momentum. A classical particle that is at the point $\mathbf{x}$ with momentum $\mathbf{p}$ has angular momentum

$$\mathbf{L} = \mathbf{x} \times \mathbf{p} = \begin{pmatrix} x_2 p_3 - x_3 p_2 \\ x_3 p_1 - x_1 p_3 \\ x_1 p_2 - x_2 p_1 \end{pmatrix}. \tag{1.26}$$

The angular-momentum vector is always perpendicular to the plane spanned by the position vector $\mathbf{x}$ and the momentum vector $\mathbf{p}$. In the classical Hamiltonian formalism, the angular momentum generates the canonical transformations describing the rotations of the system. The angular momentum is a constant of motion whenever the equation of motion is invariant under rotations. (This is a special case of *Noether's theorem*.)



CD 1.2 shows the classical angular momentum in various situations with spherical symmetry: circular motion (see also Figure 1.3), motion along a straight line, and the Coulomb motion. The angular momentum vector is perpendicular to the plane of motion and is conserved whenever the coordinate origin coincides with the center of spherical symmetry.

Exercise 1.14. *A classical particle moves with constant velocity on a straight line. Show that its angular momentum is constant in time.*

Exercise 1.15. *A classical particle with mass* m *performs a circular motion around the coordinate origin, as in Figure 1.3. Show that its angular momentum has the value*

$$L = I\omega, \tag{1.27}$$

where $I = \mathrm{m}r^2$ *is the moment of inertia,* r *is the radius of the circle, and* ω *is the angular velocity.*

Exercise 1.16. *Show that the kinetic energy of the particle in the previous exercise can be written as*

$$E = \frac{1}{2\mathrm{m}} \frac{L^2}{r^2} = \frac{L^2}{2I}. \tag{1.28}$$

1.3.2. Angular momentum in quantum mechanics

One can define the angular momentum in quantum mechanics as the operator corresponding to the classical expression (1.26) via the usual substitution rule. According to this heuristic rule, the transition to quantum mechanics

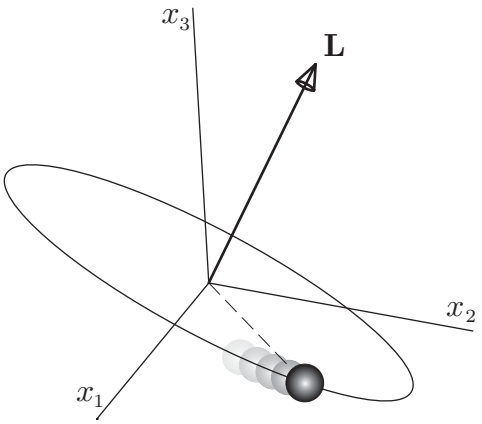


FIGURE 1.3. The angular-momentum vector for a particle moving with constant angular speed on a circle with center at the origin is a conserved quantity. Its magnitude is the product of the radius and the linear momentum, its direction is perpendicular to the plane of motion and determined by the right-hand rule: You are looking in the direction of L, when the motion is clockwise. (CD 1.5.4 is an animated version of this figure.)

is made by substituting linear operators acting on wave functions for the classical quantities $\mathbf{p}$ and $\mathbf{x}$. The classical momentum $\mathbf{p}$ is replaced by the differential operator $\mathbf{p} = -\mathrm{i}\hbar\nabla$, and the position $\mathbf{x}$ is replaced by the operator of multiplication with $\mathbf{x}$,

$$x_i \longrightarrow \text{multiplication by } x_i, \qquad p_i \longrightarrow -\mathrm{i}\hbar\,\frac{\partial}{\partial x_i}. \tag{1.29}$$

An application of this rule leads to the *angular-momentum operator*[2]

$$\mathbf{L} = -\mathrm{i}\hbar\,\mathbf{x}\times\nabla = \mathbf{x}\times\mathbf{p}, \tag{1.30}$$

which is perhaps familiar from Book One. This observable is also called the *orbital angular momentum* in order to distinguish it from other types of angular momentum (to be described later). The components of the "vector operator" $\mathbf{L}$ contain products of position and momentum operators, for example, $L_1 = x_2p_3 - x_3p_2$. The order of the position and momentum operators does not matter here, because x_i and p_j commute for $i \neq j$, and therefore the substitution rule is unambiguous (as explained in Book One).

[2]Usually, we denote the quantum mechanical operators by the same letter as the corresponding classical quantities.

As a generalization of the results in Section 1.2.2, we obtain the following connection between the angular momentum $\mathbf{L}$ and the unitary operators $U(\boldsymbol{\alpha})$ describing rotations in quantum mechanics:

Rotations about a fixed axis:

With a given unit vector $\mathbf{n}$, define for an arbitrary wave function ψ the rotated wave function

$$\psi(\mathbf{x}, \alpha) = U(\alpha\mathbf{n})\,\psi(\mathbf{x}) = \psi\big(\mathbf{R}(\alpha\mathbf{n})^{-1}\mathbf{x}\big) \tag{1.31}$$

(rotation through the angle α about the axis defined by the unit vector $\mathbf{n}$). If ψ is differentiable, then it satisfies the equation

$$\mathrm{i}\hbar\,\frac{\partial}{\partial\alpha}\,\psi(\mathbf{x}, \alpha) = \mathbf{n}\cdot\mathbf{L}\;\psi(\mathbf{x}, \alpha). \tag{1.32}$$

The self-adjoint operator $\mathbf{n}\cdot\mathbf{L}$ is thus the generator of the rotations about a fixed axis, and the unitary group can be written as

$$U(\alpha\mathbf{n}) = \exp\big(-\frac{\mathrm{i}}{\hbar}\,\alpha\,\mathbf{n}\cdot\mathbf{L}\big). \tag{1.33}$$

1.3.3. Commutation relations of the angular-momentum operators

The individual components of the angular momentum $\mathbf{L}$ do not commute. Instead, we find, by an explicit calculation, the following result.

Angular-momentum commutation relations:

The three components of the angular-momentum operator

$$\mathbf{L} = -\mathrm{i}\hbar\,\mathbf{x}\times\nabla \tag{1.34}$$

satisfy the *angular-momentum commutation relations*

$$[L_1, L_2] = \mathrm{i}\hbar L_3, \quad [L_2, L_3] = \mathrm{i}\hbar L_1, \quad [L_3, L_1] = \mathrm{i}\hbar L_2. \tag{1.35}$$

As a consequence of the angular-momentum commutation relations, it is impossible to prepare a state where the values of all three components can be predicted with arbitrary accuracy. The product of the uncertainties of two components is related to the expectation value of the third component, as you can see from Eq. (A.12) in Appendix A. Hence, you have to be very cautious when you try to depict the angular momentum as an arrow as in

classical mechanics. Closely related to the angular momentum in a state ψ is the vector

$$\mathbf{L}_{\rm av} = (\langle L_1\rangle_\psi, \langle L_2\rangle_\psi, \langle L_3\rangle_\psi), \tag{1.36}$$

whose components are the *expectation values* of the three angular-momentum operators. This vector describes a statistical property of an ensemble of quantum systems. For an individual system, the components of $\mathbf{L}$ simply do not have sharp values simultaneously.

EXERCISE 1.17. *Determine the commutation relations between the components of the angular-momentum operator and the components of the position and momentum operators,*

$$[L_3, x_1] = \mathrm{i}\hbar\, x_2, \quad \textit{etc.} \tag{1.37}$$

EXERCISE 1.18. *Compute the angular-momentum commutation relations from the result of the previous exercise, using the algebraic rules for commutators, in particular,* $[A, BC] = B[A, C] + [A, B]C$.

EXERCISE 1.19. *Prove the operator identities*

$$\mathbf{p}\cdot\mathbf{L} = 0, \quad \mathbf{x}\cdot\mathbf{L} = 0, \tag{1.38}$$

$$\mathbf{L}\times\mathbf{L} = \mathrm{i}\hbar\,\mathbf{L}, \tag{1.39}$$

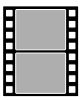
$$[\mathbf{n}\cdot\mathbf{L}, \mathbf{v}] = \mathrm{i}\hbar\,\mathbf{v}\times\mathbf{n}, \tag{1.40}$$

where $\mathbf{n}$ *is a unit vector and* $\mathbf{v}$ *is any of the operators* $\mathbf{x}$, $\mathbf{p}$, *or* $\mathbf{L}$.

The angular-momentum commutation relations are deeply connected with the properties of the rotation group. This is the topic of the next section.

1.3.4. The meaning of the angular-momentum commutation relations

The reason that the components of the angular momentum do not commute lies in the local structure of the group of rotations. It is an elementary observation, that two rotations about different axes do not commute.



CD 1.3.1 shows that the final orientation of a body depends on the order of the rotations applied to it.

Let us now consider the noncommutativity of small rotations. We denote by $\mathbf{R}_x(\alpha)$, $\mathbf{R}_y(\alpha)$, and $\mathbf{R}_z(\alpha)$ the matrices describing rotations about the x-, y-, and z-axis, respectively. The noncommutativity of the rotations about different axis means, for example, that

$$\mathbf{R}_x(\alpha)\mathbf{R}_y(\alpha) \neq \mathbf{R}_y(\alpha)\mathbf{R}_x(\alpha). \tag{1.41}$$

Now, consider the following matrix

$$\mathbf{M}(\alpha) = \mathbf{R}_x(\alpha)\,\mathbf{R}_y(\alpha) - \mathbf{R}_z(\alpha^2)\,\mathbf{R}_y(\alpha)\,\mathbf{R}_x(\alpha). \tag{1.42}$$

The matrix $\mathbf{M}(\alpha)$ describes the difference between two operations. Each operation is a composition of rotations. We insert the explicit expressions for the rotation matrices and compute the matrix product. Thus, we obtain the explicit form of the matrix $\mathbf{M}(\alpha)$ by a little calculation (made easy with the help of a computer algebra system). Expanding the matrix elements of $\mathbf{M}(\alpha)$ in power series with respect to α (around $\alpha = 0$), we obtain

$$\mathbf{M}(\alpha) = \frac{\alpha^3}{2!} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix} + \mathrm{O}(\alpha^4). \tag{1.43}$$

What does this result mean? It means that whenever α is small, then $\mathbf{M}(\alpha)$ is very small. For small angles, the operations in (1.42) are thus comparable:

$$\mathbf{R}_x(\alpha)\,\mathbf{R}_y(\alpha) \approx \mathbf{R}_z(\alpha^2)\,\mathbf{R}_y(\alpha)\,\mathbf{R}_x(\alpha), \tag{1.44}$$
up to terms of order α^3.



CD 1.3.2 shows the difference between the final orientations of a body to which rotations about the x- and y-axes are applied in different order. If the angle α is small enough, then the final orientations differ only by a rotation about the z-axis through an angle α^2.

This property of the rotation group now must also be true for the rotations performed on wave functions. Hence, there has to be a relation analogous to (1.44) between the unitary groups generated by the angular-momentum operators L_1, L_2, and L_3. For small α we expect, by analogy with (1.44), something like

$$\mathrm{e}^{-\mathrm{i}L_1\alpha}\,\mathrm{e}^{-\mathrm{i}L_2\alpha} = \mathrm{e}^{-\mathrm{i}L_3\alpha^2}\,\mathrm{e}^{-\mathrm{i}L_2\alpha}\,\mathrm{e}^{-\mathrm{i}L_1\alpha} + \text{“a small correction”}. \tag{1.45}$$

(We choose units with $\hbar = 1$ in order to simplify the notation.) Formally, we can approximate the exponential functions by the lowest-order terms of the power series

$$\mathrm{e}^{-\mathrm{i}L_1\alpha} = \mathbf{1} - \mathrm{i}L_1\alpha - \frac{1}{2}\,L_1^2\,\alpha^2 + \ldots, \tag{1.46}$$

and similarly for L_2 and L_3. We insert these expansions into (1.45) and multiply everything out. Assuming that α is small, we keep only the terms

up to the order α^2. After cancellation of the terms that are linear in α, the right and left sides of (1.45) become

$$-L_1 L_2\, \alpha^2 = -(\mathrm{i}L_3 + L_2 L_1)\, \alpha^2 \alpha^2 + \mathrm{O}(\alpha^3). \tag{1.47}$$

We conclude that (1.45) is accurate for small α up to terms of order α^3 if and only if the generators L_1, L_2, and L_3 satisfy the commutation relation

$$[L_1, L_2] = \mathrm{i}L_3. \tag{1.48}$$

The angular-momentum commutation relations are an unavoidable consequence of the noncommutativity of rotations.

Ψ The power series expansion of the exponential function converges in the operator norm if the generator is a bounded operator. In the Hilbert space $L^2(\mathbb{R}^3)$, the angular-momentum operators are unbounded, and the expansion (1.46) makes sense only on a dense set of so-called analytic vectors. We omitted these details here for the sake of a short heuristic argument. But the above derivation of the commutation relations is rigorous for unitary representations in finite dimensional Hilbert spaces. See, for example, Section 4.4.2.

1.4. Spherical Symmetry of a Quantum System

1.4.1. Conservation of angular momentum

A physical system with Hamiltonian H is called *invariant under rotations* or *spherically symmetric* whenever H commutes with the unitary rotation operators $U(\boldsymbol{\alpha}) = \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{L}/\hbar)$ defined in (1.20), that is, whenever

$$U(\boldsymbol{\alpha})\, H\, U(\boldsymbol{\alpha})^{-1} = H, \quad \text{for all angles } \boldsymbol{\alpha} = \alpha\mathbf{n}. \tag{1.49}$$

For the quantum systems considered in this book, H commutes with rotations whenever H commutes with the generators of rotations, the angular-momentum operators:

$$[H, L_k] = 0, \quad \text{for } k = 1, 2, 3, \text{ or simply } \quad [H, \mathbf{L}] = 0. \tag{1.50}$$

In the same way that H does not change under rotations, the components of $\mathbf{L}$ do not change under the time evolution,

$$e^{-\mathrm{i}Ht/\hbar}\, \mathbf{L}\, e^{\mathrm{i}Ht/\hbar} = \mathbf{L}, \tag{1.51}$$

that is, the angular momentum is a conserved quantity, a constant of motion.

In classical mechanics, the close connection between symmetries and conservation laws is known as Noether's theorem. Classically, as well as quantum mechanically, the physical quantity that is conserved during the time evolution of a spherically symmetric system is the angular momentum.

Ψ As usual, it is understood that a commutation relation like (1.50) holds on a suitable dense domain that is left invariant by the operators H and L_k. One may take, for example, the set $\mathcal{S}(\mathbb{R}^3)$ of rapidly decreasing smooth functions introduced in Book One, Section 7.7.1. We also note that the commutativity of unbounded self-adjoint operators is actually defined by the commutativity of the unitary groups (see also Book One, Section 6.11). The relation $[H, L_k] = 0$ on a dense domain implies the commutativity only if additional conditions are met, which are usually satisfied for the systems considered here.

1.4.2. Spherically symmetric potentials

If the Hamiltonian is of the particular form $H = H_0 + V(\mathbf{x})$, where H_0 is the free-particle Hamiltonian and $V(\mathbf{x})$ is a potential, then the physical system is spherically symmetric whenever the potential V is spherically symmetric. A potential $V(\mathbf{x})$ is spherically symmetric if it does not depend on the direction of $\mathbf{x}$, but only on the distance of the point $\mathbf{x}$ from the origin. We write

$$V(\mathbf{x}) = V(r), \quad \text{where} \quad r = \sqrt{x_1^2 + x_2^2 + x_3^2} = |\mathbf{x}|. \tag{1.52}$$

Thus, a potential is spherically symmetric if its isosurfaces (the surfaces over which V is constant) are concentric spheres around the origin. The line of action of the corresponding force field

$$\mathbf{F}(\mathbf{x}) = -\nabla V(r) = -\Big(\frac{d}{dr} V(r)\Big)\, \frac{\mathbf{x}}{r} \tag{1.53}$$

always passes through the coordinate origin (see Fig. 1.4), and the strength of the force does not depend on the direction of $\mathbf{x}$.

Ψ To a mathematician, Eq. (1.52) constitutes a slight abuse of notation, which is, however, very common in physics. Two different functions are denoted by the same letter V (one function depends on the three variables $\mathbf{x} = (x_1, x_2, x_3)$, the other is a function of the single variable r). In physics, the notation often emphasizes the physical quantity and not the explicit function describing its dependence on other quantities.

The most important example of a spherically symmetric potential is the Coulomb potential. It describes the electrostatic energy of an electron in the field of an atomic nucleus. The Schrödinger equation for this system will be solved in Chapter 2.

In the presence of spherical symmetry, the Schrödinger equation can be simplified by the separation of variables technique. This technique seeks a solution in the form of a product of three functions, one depending on the radial variable r and the others on angular variables ϑ and φ. In that way,

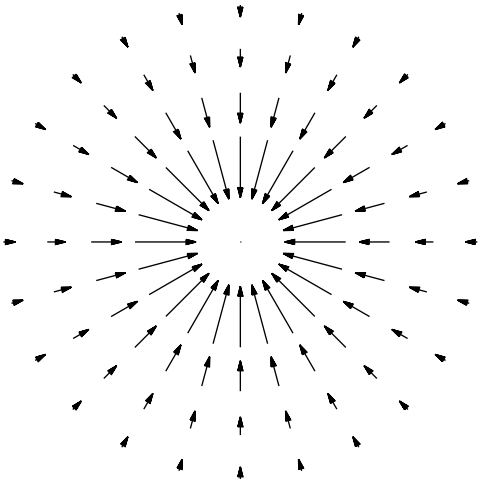


FIGURE 1.4. Example of a spherically symmetric force field. The line of action always passes through the coordinate origin.

the eigenvalue equation (a partial differential equation) splits into three ordinary differential equations, one for each variable. The spherically symmetric potential $V(r)$ enters only the equation for the radial part of the wave function, whereas the equations involving the angular variables ϑ and φ are the same for all systems with spherical symmetry. In this chapter, we are going to determine the possible solutions of the angular equations once and for all.



CD 1.4 presents three-dimensional views of an attractive harmonic oscillator force and a repulsive Coulomb force.

EXERCISE 1.20. *Show that the Hamiltonian for a particle in a spherically symmetric potential commutes with all components of the angular-momentum operator.*

1.4.3. Symmetry and degeneracy

A major step in the solution of the Schrödinger equation is to determine whether the Hamiltonian operator admits eigenstates. An eigenstate or eigenvector of H is a nonzero square-integrable function ψ for which there exists a number E (called an eigenvalue) such that

$$H\,\psi(\mathbf{x}) = E\,\psi(\mathbf{x}). \tag{1.54}$$

An eigenvector of the energy operator leads immediately to a stationary solution of the Schrödinger equation

$$\psi(\mathbf{x}, t) = e^{-\mathrm{i}Et/\hbar}\,\psi(\mathbf{x}), \tag{1.55}$$

for which the time-dependence is only a phase factor (with absolute value = 1), so that predictions of physical properties do not depend on time.

It is important to note that the invariance under a symmetry transformation may be related to a degeneracy of eigenvalues. An eigenvalue E is called *degenerate* if there are several linearly independent eigenvectors belonging to that eigenvalue. The subspace spanned by all these eigenvectors is called the *eigenspace* belonging to that eigenvalue. The dimension of the eigenspace is called the *degree of degeneracy* by physicists and the *multiplicity* by mathematicians.

Even if a Hamiltonian operator H is invariant under a symmetry transformation U, an eigenvector ψ need not be invariant. However, if ψ is an eigenvector of H, belonging to the eigenvalue E, then the transformed vector $U\psi$ is again an eigenvector of H belonging to the same eigenvalue. This can be seen as follows:

$$HU\psi = (UHU^{-1})U\psi = UH\psi = UE\psi = EU\psi. \tag{1.56}$$

Hence, the eigenspace of E is invariant under the transformation U, that is, $U\psi$ is in that eigenspace whenever ψ is.

Next, we consider the eigenvector ψ belonging to a non-degenerate eigenvalue of H. An eigenvalue is non-degenerate if the corresponding eigenspace is one-dimensional. A symmetry transformation U that leaves the eigenspaces of H invariant must turn ψ into a vector $U\psi$ in the same (one-dimensional) eigenspace. Hence, $U\psi$ is simply a multiple of ψ, and we may write $U\psi = \lambda\psi$ with some complex number λ. But $|\lambda| = 1$ because U is unitary. An eigenstate belonging to a non-degenerate eigenvalue is invariant (up to a phase factor). In the case of spherical symmetry this means that for non-degenerate energies the corresponding eigenfunctions are spherically symmetric.

Likewise, the eigenspaces of the angular-momentum operators (the *angular-momentum subspaces*) are invariant under the time evolution generated by a spherically symmetric Hamiltonian H. The operator H leaves the eigenspace of each of the angular-momentum operators invariant. It can be a major simplification to solve the eigenvalue problem for H in an eigenspace of the angular-momentum operators. Thus, our next task is the investigation of the possible angular-momentum eigenvalues and the associated eigenspaces. This is done in the next section.

1.5. The Possible Eigenvalues of Angular-Momentum Operators

In this section, we present a purely algebraic approach to the solution of the eigenvalue problem for the angular-momentum operators. For simplicity, we work with units where $\hbar$ has the numerical value 1. We consider a set of three symmetric operators J_1, J_2, and J_3 that satisfy the commutation relations

$$[J_1, J_2] = \mathrm{i}J_3, \quad [J_2, J_3] = \mathrm{i}J_1, \quad [J_3, J_1] = \mathrm{i}J_2. \tag{1.57}$$

A lot can be learned by studying these relations. Because of these relations, we cannot hope to find simultaneous eigenvectors belonging to nonzero eigenvalues.

Ψ Indeed, assume that ψ is a simultaneous eigenvector of, say, J_1 and J_2. Let $J_1\psi = m_1\psi$ and $J_2\psi = m_2\psi$. Then we find immediately that

$$\mathrm{i}J_3\psi = [J_1, J_2]\psi = (m_1m_2 - m_2m_1)\psi = 0.$$

Hence,

$$J_3(J_1 + \mathrm{i}J_2)\psi = [J_3, J_1]\psi + \mathrm{i}[J_3, J_2]\psi = (\mathrm{i}J_2 - J_1)\psi = (\mathrm{i}m_2 - m_1)\psi,$$

and

$$J_3(J_1 + \mathrm{i}J_2)\psi = J_3(m_1 + \mathrm{i}m_2)\psi = (m_1 + \mathrm{i}m_2)J_3\psi = 0.$$

Hence, $\mathrm{i}m_2 - m_1 = 0$, and because the eigenvalues of symmetric operators are always real, this implies that $m_1 = m_2 = 0$. We conclude that there are no nontrivial simultaneous eigenvectors belonging to nonzero eigenvalues.

The square of the angular-momentum vector $\mathbf{J} = (J_1, J_2, J_3)$, that is, the operator

$$J^2 = J_1^2 + J_2^2 + J_3^2 \tag{1.58}$$

commutes with all components J_k,

$$[J^2, J_k] = 0, \quad \text{for } k = 1, 2, 3. \tag{1.59}$$

Hence, we could try to find simultaneous eigenvectors for the operator J^2 and any one of the components, say J_3. We are going to prove the following theorem.

THEOREM 1.1. *Assume that there is a simultaneous eigenvector of the commuting operators J^2 and J_3. Then the eigenvalue of J^2 is $j(j+1)$ where j is one of the numbers $0, \frac{1}{2}, 1, \frac{3}{2}, 2 \ldots$. Moreover, there are $2j+1$ eigenvectors $\psi_{j,m}$ of J_3, such that*

$$J^2\,\psi_{j,m} = j(j+1)\,\psi_{j,m}, \qquad J_3\,\psi_{j,m} = m\,\psi_{j,m}, \tag{1.60}$$

for $m = -j, -j+1, \ldots, j-1, j$.

PROOF. We first define the operators

$$J_\pm = J_1 \pm \mathrm{i} J_2 \tag{1.61}$$

and note the following commutation properties

$$[J_3, J_\pm] = \pm J_\pm, \qquad [J^2, J_\pm] = 0. \tag{1.62}$$

The operators $J_\pm$ are not self-adjoint, but formally adjoint to each other,

$$J_+^\dagger = J_-, \qquad J_-^\dagger = J_+. \tag{1.63}$$

Their products can be expressed in terms of J^2 and J_3,

$$J_+ J_- = J^2 - J_3^2 + J_3, \qquad J_- J_+ = J^2 - J_3^2 - J_3. \tag{1.64}$$

Now, let us assume that there exists a simultaneous eigenvector ψ_m^λ belonging to eigenvalues λ for J^2 and m for J_3:

$$J^2\, \psi_m^\lambda = \lambda\, \psi_m^\lambda, \qquad J_3\, \psi_m^\lambda = m\, \psi_m^\lambda. \tag{1.65}$$

We can always multiply the eigenvector with a suitable complex constant, and therefore we may assume that ψ_m^λ is normalized,

$$\|\psi_m^\lambda\|^2 = \langle \psi_m^\lambda, \psi_m^\lambda \rangle = 1. \tag{1.66}$$

Next, consider the state $\psi_+ = J_+ \psi_m^\lambda$. Whenever ψ_+ is not the zero vector, it is an eigenstate of J_3 belonging to the eigenvalue $m+1$. This follows from the commutation property (1.62):

$$\begin{aligned} J_3\, \psi_+ &= J_3 J_+\, \psi_m^\lambda = (J_+ J_3 + [J_3, J_+])\, \psi_m^\lambda = (J_+ J_3 + J_+)\, \psi_m^\lambda \\ &= (J_+ m + J_+)\, \psi_m^\lambda = (m+1)\, J_+\, \psi_m^\lambda \\ &= (m+1)\, \psi_+. \end{aligned} \tag{1.67}$$

The vector ψ_+ is still an eigenvector of J^2 belonging to the same eigenvalue λ, because J_+ commutes with J^2:

$$J^2\, \psi_+ = J^2 J_+\, \psi_m^\lambda = J_+ J^2\, \psi_m^\lambda = J_+\, \lambda\, \psi_m^\lambda = \lambda\, J_+\, \psi_m^\lambda = \lambda\, \psi_+. \tag{1.68}$$

An analogous observation holds for the state $J_- \psi_m^\lambda$. Either this vector is the zero vector, or it is a simultaneous eigenstate with eigenvalues $m-1$ for J_3 and λ for J^2. In order to determine the norm of the vectors $J_\pm \psi_m^\lambda$, we perform the following calculation

$$\begin{aligned} \|J_\pm \psi_m^\lambda\|^2 &= \langle J_\pm\, \psi_m^\lambda\, ,\, J_\pm\, \psi_m^\lambda \rangle \\ &= \langle \psi_m^\lambda\, ,\, J_\mp\, J_\pm\, \psi_m^\lambda \rangle && \text{by (1.63)} \\ &= \langle \psi_m^\lambda\, ,\, (J^2 - J_3^2 \mp J_3)\, \psi_m^\lambda \rangle && \text{by (1.64)} \\ &= \lambda - m^2 \mp m. \end{aligned} \tag{1.69}$$

From $\|J_\pm \psi_m^\lambda\|^2 \geq 0$ it follows immediately that

$$\lambda \geq m^2 \pm m = m(m \pm 1). \tag{1.70}$$

Moreover, we find that $J_+ \psi_{j,m} = 0$ if and only if $\lambda - m^2 - m = 0$, that is,

$$J_+ \psi_m^\lambda = 0 \quad \text{if and only if} \quad \lambda = m(m+1), \tag{1.71}$$

and similarly,

$$J_- \psi_m^\lambda = 0 \quad \text{if and only if} \quad \lambda = m(m-1). \tag{1.72}$$

Now it is easy to determine the possible values for λ and m. Whenever we have a simultaneous eigenvector of J^2 and J_3 with eigenvalues λ and m, we can also find eigenvectors belonging to eigenvalues $m+1$, $m+2$, and so forth, of J_3. These eigenvectors are obtained by successive application of the *ladder operator* J_+. All new eigenvectors belong to the same eigenvalue λ of J^2. This process of creating new eigenvectors will stop as soon as we reach a maximal value of m, say $m_{\max}$ for which $J_+ \psi_{m_{\max}}^\lambda = 0$, or equivalently, for which

$$\lambda = m_{\max}(m_{\max} + 1). \tag{1.73}$$

It is crucial to observe that such a maximal value $m_{\max}$ must exist: Otherwise, we could raise the eigenvalue of J_3 indefinitely until the inequality (1.70) would be violated, thus giving a contradiction.

Similarly, using the ladder operator J_-, we can lower the eigenvalue m until we reach a minimum value $m_{\min}$ for which we must have

$$\lambda = m_{\min}(m_{\min} - 1). \tag{1.74}$$

Combining Eqs. (1.73) and (1.74) we find

$$(m_{\max} - m_{\min} + 1)(m_{\max} + m_{\min}) = 0. \tag{1.75}$$

Here, because of $m_{\max} \geq m_{\min}$, only the second factor can be zero, that is, $m_{\min} = -m_{\max}$. Because we can get from $m_{\min}$ to $m_{\max}$ in integer steps (by applying the operator J_+ to the corresponding eigenvectors), we find that $m_{\max} - m_{\min} = 2m_{\max}$ must be a non-negative integer. Writing $m_{\max} = j$ we find that the only allowed values of j are $0, \frac{1}{2}, 1, \frac{3}{2}, 2$, and so forth. From (1.73) we see that $\lambda = j(j+1)$.

Finally, write $\psi_{j,m}$ instead of ψ_m^λ. □

Figure 1.5 visualizes the spectrum of possible simultaneous eigenvalues of J^2 and J_3 according to Theorem 1.1.

THEOREM 1.2. *For a fixed j, all the $2j+1$ eigenvalues of J_3 have the same multiplicity k (which might be infinite). The eigenspace of J^2 belonging to the eigenvalue $j(j+1)$ is therefore $k(2j+1)$-dimensional. This space is invariant under the action of the operators J_1, J_2, and J_3.*

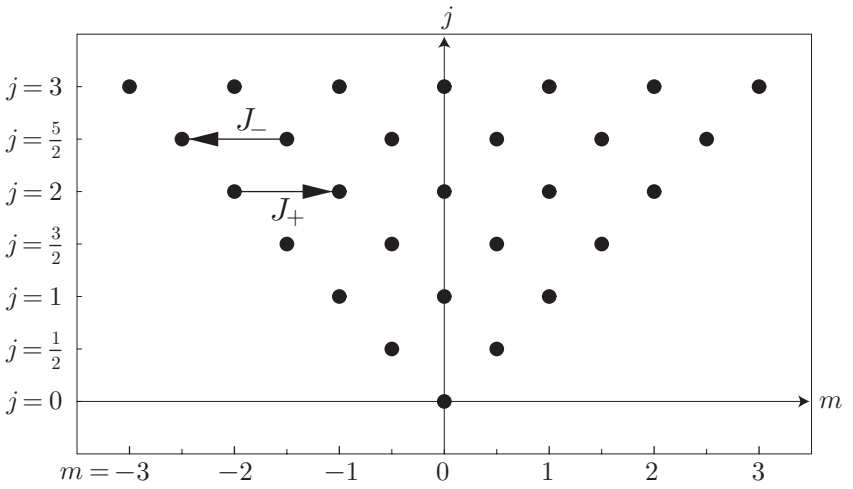


FIGURE 1.5. The possible values (j, m) for the operators J^2 and J_3 (with $j \leq 3$). Each point represents a simultaneous eigenvector of J^2 and J_3. The ladder operators $J_\pm$ let you jump from one point to the next in the horizontal direction, that is, within an eigenspace of J^2.

PROOF. Assume that there are two orthogonal vectors $\psi^{(1)}_{j,m}$ and $\psi^{(2)}_{j,m}$ both of which belong to the eigenvalues $j(j+1)$ of J^2 and m of J_3. Assume $m < j$. Then $J_+ \psi^{(1)}_{j,m}$ and $J_+ \psi^{(2)}_{j,m}$ are both nonzero vectors, and

$$\begin{aligned}\langle J_+ \psi^{(1)}_{j,m}\,,\, J_+ \psi^{(2)}_{j,m}\rangle &= \langle \psi^{(1)}_{j,m}\,,\, J_- J_+ \psi^{(2)}_{j,m}\rangle \\ &= \langle \psi^{(1)}_{j,m}\,,\, (J^2 - J_3^2 - J_3)\, \psi^{(2)}_{j,m}\rangle \\ &= \big(j(j+1) - m^2 - m\big)\langle \psi^{(1)}_{j,m}\,,\, \psi^{(2)}_{j,m}\rangle = 0. \end{aligned} \tag{1.76}$$

And, similarly, for $m > -j$ we find

$$\langle J_- \psi^{(1)}_{j,m}\,,\, J_- \psi^{(2)}_{j,m}\rangle = 0. \tag{1.77}$$

You can see that the orthogonality is preserved by the ladder operators J_+ and J_-. Hence, if there are precisely k orthogonal states for some eigenvalue m, then there are precisely k orthogonal states for all $m = -j, -j+1, \ldots, j-1, j$. For a given j there are $2j+1$ different values of m. Because the eigenvectors belonging to different eigenvalues of a symmetric operator are orthogonal, eigenvectors with different eigenvalues m are orthogonal. Therefore, the subspace spanned by all the eigenvectors belonging to the eigenvalue $j(j+1)$ of the operator J^2 is $k(2j+1)$-dimensional. This

eigenspace is clearly left invariant by the operators J^2, J_3, and $J_\pm$ (an operator leaves a subspace invariant if the operator maps any vector of this subspace to a vector in the same subspace). Hence, the eigenspace of J^2 is also left invariant by the operators J_1 and J_2, which can be written as linear combinations of J_+ and J_-. □

Our considerations in this section have shown that the possible eigenvalues of the angular momentum are characterized by *angular-momentum quantum numbers* j and m that can have integer and half-integer values. In the next section, we will find that for the *orbital* angular momentum $\mathbf{L} = \mathbf{x} \times \mathbf{p}$ only integer values can occur. Angular-momentum operators with half-integer quantum numbers are nevertheless important for describing the spin of elementary particles (see Chapter 3).

EXERCISE 1.21. *Verify the commutation relations*

$$[J^2, J_k] = 0, \quad \text{for } k = 1, 2, 3. \tag{1.78}$$

EXERCISE 1.22. *Define the three* 2×2*-matrices*

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.79}$$

These matrices are called the Pauli matrices. They are symmetric and hence define self-adjoint operators in the Hilbert space $\mathbb{C}^2$*. Verify that the operators* $S_j = (1/2)\sigma_j$*,* $j = 1, 2, 3$*, satisfy the angular-momentum commutation relations* (1.57)*. Show that the only eigenvalue of* $S^2 = S_1^2 + S_2^2 + S_3^2$ *is* $s(s+1)$ *with* $s = 1/2$*.*

EXERCISE 1.23. *Show that if the operators* J_1*,* J_2*,* J_3 *satisfy the commutation relations* $[J_1, J_2] = \mathrm{i}\hbar J_3$*, and so forth, then the possible eigenvalues of* J^2 *are* $\hbar^2 j(j+1)$ *with* $j = 0, \frac{1}{2}, 1, \ldots$*, and for each* j *the possible eigenvalues of* J_3 *are* $\hbar m$ *with* $m = -j, -j+1, \ldots, j$*.*

EXERCISE 1.24. *Let* $\phi_{j,j}$ *be a simultaneous eigenvector of* J^2 *and* J_3*, with the maximal* $m = j$*. Assume that* $\phi_{j,j}$ *is normalized,* $\|\phi_{j,j}\| = 1$*. Define*

$$\phi_{j,m-1} = \frac{1}{\sqrt{j(j+1) - m(m-1)}}\, J_- \,\phi_{j,m}, \quad m = j, j-1, \ldots, -j+1. \tag{1.80}$$

Show that the vectors $\phi_{j,m}$*,* $m = -j, -j+1, \ldots, j$ *are normalized simultaneous eigenvectors of* J^2 *and* J_3*.*

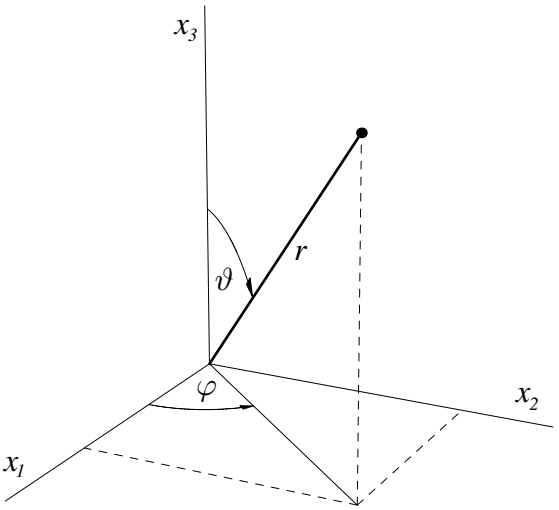


FIGURE 1.6. Spherical coordinates on $\mathbb{R}^3$. Instead of giving the Cartesian coordinates (x_1, x_2, x_3), we can also specify the position of a point in $\mathbb{R}^3$ by spherical coordinates (r, ϑ, φ). See also CD 1.5.

1.6. Spherical Harmonics

1.6.1. Spherical coordinates

In order to determine the eigenvalues and eigenfunctions of the orbital angular-momentum operators L^2 and L_3, it is convenient to express them as differential operators in spherical coordinates. On $\mathbb{R}^3 \setminus \{0\}$ (three-dimensional space without the origin) we can introduce spherical coordinates (r, ϑ, φ) as in Figure 1.6.

In a spherical coordinate system, the position of a point is specified by its distance r from the origin, its *polar angle* ϑ and its *azimuthal angle* φ. The Cartesian coordinates can be expressed in terms of the spherical coordinates as follows

$$\begin{aligned} x_1(r, \vartheta, \varphi) &= r \, \sin\vartheta \, \cos\varphi, \\ x_2(r, \vartheta, \varphi) &= r \, \sin\vartheta \, \sin\varphi, \\ x_3(r, \vartheta, \varphi) &= r \, \cos\vartheta. \end{aligned} \tag{1.81}$$

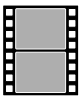
It is often necessary to invert this formula, that is, to express the spherical coordinates of a point in terms of its Cartesian coordinates

$$\begin{aligned} r(x_1, x_2, x_3) &= |\mathbf{x}| \in (0, \infty), \\ \vartheta(x_1, x_2, x_3) &= \arccos(x_3/|\mathbf{x}|) \in [0, \pi], \\ \varphi(x_1, x_2, x_3) &= \arctan(x_1, x_2) \in (-\pi, \pi]. \end{aligned} \tag{1.82}$$

Here, the function arctan of two variables is defined as

$$\arctan(x_1, x_2) \equiv \pi\, \theta(-x_1)\, \mathrm{sgn}(x_2) + \arctan(x_2/x_1). \tag{1.83}$$

(This definition has to be extended by continuity for $x_1 \to 0$ and $x_2 \neq 0$.) Here, $\mathrm{sgn}(x)$ is the sign of x, and θ is the step function ($\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x \leq 0$).



CD 1.5.1 is an animated view of a point in a Cartesian coordinate system similar to Figure 1.6. CD 1.5.2–4 deal with the uniform motion of a free particle and the circular motion of the rigid rotator and discuss the description of these systems in terms of spherical coordinates.

At each point in $\mathbb{R}^3 \setminus \{0\}$, we can define the unit vectors in the directions of the spherical coordinate lines (these are the curves on which two of the three spherical coordinates are held fixed)

$$\begin{aligned} \mathbf{e}_r &= (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta) = \frac{\mathbf{x}}{r}, \\ \mathbf{e}_\vartheta &= (\cos\vartheta\cos\varphi, \cos\vartheta\sin\varphi, -\sin\vartheta) = \frac{\partial \mathbf{e}_r}{\partial\vartheta}, \\ \mathbf{e}_\varphi &= (-\sin\varphi, \cos\varphi, 0) = \frac{1}{\sin\vartheta}\frac{\partial \mathbf{e}_r}{\partial\varphi}. \end{aligned} \tag{1.84}$$

Exercise 1.25. *Using* (1.84)*, verify the following formulas:*

$$\begin{aligned} &\mathbf{e}_r \cdot \mathbf{e}_\vartheta = 0, \quad && \mathbf{e}_r \cdot \mathbf{e}_r = 1, \quad && \mathbf{e}_r \times \mathbf{e}_\vartheta = \mathbf{e}_\varphi, \\ &\mathbf{e}_\vartheta \cdot \mathbf{e}_\varphi = 0, \quad && \mathbf{e}_\vartheta \cdot \mathbf{e}_\vartheta = 1, \quad && \mathbf{e}_\vartheta \times \mathbf{e}_\varphi = \mathbf{e}_r, \\ &\mathbf{e}_\varphi \cdot \mathbf{e}_r = 0, \quad && \mathbf{e}_\varphi \cdot \mathbf{e}_\varphi = 1, \quad && \mathbf{e}_\varphi \times \mathbf{e}_r = \mathbf{e}_\vartheta. \end{aligned} \tag{1.85}$$

At each point, the three unit vectors thus form a right-handed, orthonormal system.

Now, if $\psi(x_1, x_2, x_3)$ is a wave function in Cartesian coordinates, then the same function in spherical coordinates is obtained just by inserting the expressions (1.81) into the arguments of ψ

$$\phi(r, \vartheta, \varphi) = \psi\big(x_1(r, \vartheta, \varphi), x_2(r, \vartheta, \varphi), x_3(r, \vartheta, \varphi)\big) \tag{1.86}$$

The collection of formulas is completed by giving the expression for the gradient operator ∇ in spherical coordinates

$$\hat{\nabla} = \mathbf{e}_r \frac{\partial}{\partial r} + \frac{1}{r}\left(\mathbf{e}_\vartheta \frac{\partial}{\partial \vartheta} + \mathbf{e}_\varphi \frac{1}{\sin\vartheta}\frac{\partial}{\partial \varphi}\right). \tag{1.87}$$

This formula has to be understood as follows. The gradient of a function can either be evaluated in Cartesian coordinates or in spherical coordinates. At some point (x_1, x_2, x_3) corresponding to (r, ϑ, φ) we have

$$\nabla\psi(x_1, x_2, x_3) = \hat{\nabla}\phi(r, \vartheta, \varphi), \tag{1.88}$$

with ϕ and ψ being related as in (1.86).



The spherical coordinate space is a three-dimensional space where the coordinate axes describe the r-, ϑ-, and φ-coordinates of a point in $\mathbb{R}^3$. CD 1.6 visualizes the familiar examples of linear and circular motion, which look rather unfamiliar in the spherical coordinate space.



The transition to spherical coordinates, that is, the mapping $U : \psi \to \phi$ defined in (1.86) is a unitary transformation from the Hilbert space $L^2(\mathbb{R}^3)$ to the Hilbert space $L^2([0,\infty) \times S^2, dV)$. Here, S^2 denotes the two-dimensional surface of the unit sphere, and $dV = r^2 \sin\vartheta\, dr\, d\vartheta\, d\varphi$ is the volume element in spherical coordinates. The points in $[0,\infty) \times S^2$ have the coordinates (r, ϑ, φ), and integration has to be done with respect to the volume element in spherical coordinates. It follows from the usual rules of variable substitution in an integral that

$$\int_{\mathbb{R}^3} |\psi(x_1, x_2, x_3)|^2\, d^3x = \int_0^\infty \int_{S^2} |\phi(r, \vartheta, \varphi)|^2\, r^2 \sin\vartheta\, dr\, d\vartheta\, d\varphi. \tag{1.89}$$

This means that the norm of ψ in $L^2(\mathbb{R}^3)$ is equal to the norm of ϕ in $L^2([0,\infty) \times S^2, dV)$ (this is the unitarity of U). As a consequence, the operators ∇ and $\hat{\nabla}$ are unitarily equivalent, that is, $\hat{\nabla} = U\nabla U^{-1}$. In the following, we always put a hat on an operator in spherical coordinates in order to indicate that it acts on functions $\phi(r, \vartheta, \varphi)$.

1.6.2. Angular momentum in spherical coordinates

With the results of the previous section, it is easy to derive the expressions of the angular-momentum operators in spherical coordinates. Using the formulas from Exercise 1.25, we obtain

$$\hat{\mathbf{L}} = r\mathbf{e}_r \times (-\mathrm{i}\hbar\hat{\nabla}) = \mathrm{i}\hbar\left(\mathbf{e}_\vartheta \frac{1}{\sin\vartheta}\frac{\partial}{\partial\varphi} - \mathbf{e}_\varphi \frac{\partial}{\partial\vartheta}\right), \tag{1.90}$$

$$\hat{L}^2 = -\frac{\hbar^2}{\sin\vartheta}\,\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) - \frac{\hbar^2}{\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2}. \tag{1.91}$$

Again, the hat (ˆ) simply indicates that the operator acts on wave functions in spherical coordinates. We have

$$\mathbf{L}\psi(x_1, x_2, x_3) = \hat{\mathbf{L}}\phi(r, \vartheta, \varphi), \tag{1.92}$$

where ϕ is the function ψ expressed in spherical coordinates as in (1.86). In particular, the operators L^2 and $\hat{L}^2$ have the same eigenvalues (as well as L_3 and $\hat{L}_3$). The angular-momentum operators fulfill the angular-momentum commutation relations. Hence, Theorem 1.1 in Section 1.5 shows that the possible eigenvalues of $\hat{L}^2$ are among the numbers $\hbar^2\ell(\ell+1)$, where ℓ is a non-negative integer or half-integer. For each eigenvalue of $\hat{L}^2$, the third component has the eigenvalues $\hbar m$ with $m = -\ell, -\ell+1, \ldots, \ell$.

The expression for L_3 in spherical coordinates is particularly simple. Just insert the third Cartesian component of $\mathbf{e}_\vartheta$ and $\mathbf{e}_\varphi$ (see Eq. (1.84)) into (1.90):

$$\hat{L}_3 = -\mathrm{i}\hbar\,\frac{\partial}{\partial\varphi}. \tag{1.93}$$

This expression is already familiar from the two-dimensional situation described in Book One, Section 8.8.

Now we can see that $\hat{L}_3$ cannot have half-integer eigenvalues m. The domain of the differential operator $\hat{L}_3$ consists of continuous functions. As a function of the azimuthal angle φ, any eigenfunction of $\hat{L}_3$ must therefore be a periodic function:

$$\phi(r, \vartheta, \varphi + 2\pi) = \phi(r, \vartheta, \varphi). \tag{1.94}$$

Denoting the eigenvalue of $\hat{L}_3$ by m, the eigenvalue equation reads

$$\hat{L}_3\phi = -\mathrm{i}\hbar\,\frac{\partial}{\partial\varphi}\,\phi = \hbar m\,\phi \tag{1.95}$$

so that the φ-dependence of ϕ must be described by $\exp(\mathrm{i}m\varphi)$, which is periodic with period 2π if and only if m is an integer.

The considerations in Section 1.5 thus also exclude the possibility that L^2 has half-integer eigenvalues. Only the numbers $\hbar^2\ell(\ell+1)$ with integer ℓ can occur as eigenvalues of L^2. Below, we are going to show that simultaneous eigenfunctions of L^2 and L_3 indeed exist for all non-negative integers ℓ. Hence, we obtain the following result:

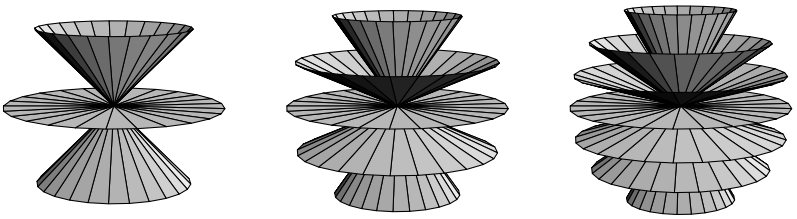


FIGURE 1.7. The absolute value $|\mathbf{L}| = \sqrt{L^2}$ of the angular momentum can only take the values $\hbar\sqrt{\ell(\ell+1)}$, with ℓ being a non-negative integer. For a given ℓ, the component of the angular momentum in an arbitrary direction (here taken as the vertical direction) is also quantized and can only have the values $-\hbar\ell, -\hbar(\ell+1), \ldots + \hbar\ell$. In a classical picture, the angular momentum vector (if measured with respect to a certain direction) thus lies on certain cones.

Eigenvalues of the orbital angular momentum:

The eigenvalues of the operator L^2 are precisely the numbers $\hbar^2\ell(\ell+1)$, where ℓ is a non-negative integer. For each ℓ, the operator L_3 has the eigenvalues $\hbar m$, where $m = -\ell, -\ell+1, \ldots, \ell$.

A classical vector $\mathbf{L}$ for which the values of L^2 and L_3 are restricted to the eigenvalues above would have to lie on certain cones which are described in Figure 1.7. In quantum mechanics, this picture should not be taken seriously because the same result would be obtained for the possible values of L_1 and L_2 (or the component of $\mathbf{L}$ in an arbitrary direction).

In spherical coordinates, the operators $\hat{L}_3$ and $\hat{L}^2$ only act on the angular variables. Hence, we may try a separation of the variables by writing

$$\phi(r, \vartheta, \varphi) = \frac{1}{r}\, f(r)\, \chi(\vartheta, \varphi). \tag{1.96}$$

The factor $1/r$ has been introduced for "cosmetic reasons." Later, it will simplify some formulas involving the radial part of the wave function.

It has to be noted that most wave functions $\phi(r, \vartheta, \varphi)$ cannot be written in the form of a product of an r-dependent part and a part depending only on the angular variables. But it turns out that the set of wave functions of the type (1.96) is large enough to contain an orthonormal basis of the

Hilbert space $L^2(\mathbb{R}^3)$. Therefore, any wave function has an expansion

$$\phi(r,\vartheta,\varphi) = \sum_{k=0}^{\infty} \frac{1}{r}\, f_k(r)\, \chi_k(\vartheta,\varphi), \tag{1.97}$$

the sum being convergent in the Hilbert space norm.

The wave function (1.96) consists of a radial part $f(r)/r$ and an angular part $\chi(\vartheta,\varphi)$. The norm is given by (see also (1.89))

$$\begin{aligned} \|\phi\|^2 &= \int |\phi(r,\vartheta,\varphi)|^2\, r^2 dr\, \sin\vartheta\, d\vartheta\, d\varphi \\ &= \int_0^\infty |f(r)|^2\, dr \int_{S^2} |\chi(\Omega)|^2\, d\Omega. \end{aligned} \tag{1.98}$$

Here, Ω is shorthand for the angular coordinates (ϑ,φ). The set of these coordinates forms the sphere S^2 (the two-dimensional surface of the unit sphere in three dimensions), and $d\Omega = \sin\vartheta\, d\vartheta\, d\varphi$ denotes the area element on this sphere.

Hence, the part $f(r)$ of the wave function in spherical coordinates is square-integrable on $[0,\infty)$, and the angular part is a square integrable function on the sphere S^2 (it belongs to the Hilbert space $L^2(S^2)$).

The application of the operators $\hat{L}^2$ and $\hat{L}_3$ only affects the angular part χ. It is thus sufficient to look for angular eigenfunctions

$$\hat{L}^2\, \chi_\ell^m(\vartheta,\varphi) = \hbar^2 \ell(\ell+1)\, \chi_\ell^m(\vartheta,\varphi), \quad \hat{L}_3\, \chi_\ell^m(\vartheta,\varphi) = \hbar m\, \chi_\ell^m(\vartheta,\varphi). \tag{1.99}$$

Because we know already that m must be an integer, we only have to look for solutions with integer ℓ. These can be determined, in principle, as follows. According to the proof of Theorem 1.1, we first look for a solution of

$$(\hat{L}_1 + \mathrm{i}\hat{L}_2)\chi(\vartheta,\varphi) = 0, \tag{1.100}$$

which can be written as

$$\mathrm{i}\, \cos\vartheta\, \frac{\partial\chi}{\partial\varphi} + \sin\vartheta\, \frac{\partial\chi}{\partial\vartheta} = 0. \tag{1.101}$$

It is easily checked that for each $\ell = 0, 1, 2, 3, \ldots$

$$\chi(\vartheta,\varphi) = N_\ell\, e^{\mathrm{i}\ell\varphi} \left(\sin\vartheta\right)^\ell \tag{1.102}$$

is a solution for this equation (where N_ℓ is a normalization constant). For each ℓ we can now obtain $2\ell+1$ eigenfunctions of $\hat{L}^2$ simply by differentiating, that is, by applying the differential operator $\hat{L}_1 - \mathrm{i}\hat{L}_2$ to the solution above. This procedure (which was described in Theorem 1.1) yields, one after another, the eigenfunctions of L_3 for $m = \ell, \ell-1, \ldots, -\ell$. It is convenient

to normalize these eigenfunctions by requiring

$$\int_{S^2} |\chi(\vartheta,\varphi)|^2 \, d\Omega = 1. \tag{1.103}$$

In that way, one obtains the *spherical harmonics* $Y_\ell^m(\vartheta,\varphi)$, which are described in more detail in the next section.

Eigenfunctions of the orbital angular momentum:

In spherical coordinates, the normalized simultaneous eigenfunctions of the angular-momentum operators $\hat{L}^2$ and $\hat{L}_3$ are the spherical harmonics $Y_\ell^m(\vartheta,\varphi)$,

$$\hat{L}^2 \, Y_\ell^m(\vartheta,\varphi) = \hbar^2 \, \ell(\ell+1) \, Y_\ell^m(\vartheta,\varphi), \qquad \ell = 0,1,2,3,\ldots, \tag{1.104}$$

$$\hat{L}_3 \, Y_\ell^m(\vartheta,\varphi) = \hbar m \, Y_\ell^m(\vartheta,\varphi), \qquad -\ell \le m \le \ell. \tag{1.105}$$

As an example, Figure 1.8 visualizes the spherical harmonic Y_4^1.



CD 1.7 visualizes the spherical harmonics Y_ℓ^m for $\ell \le 6$ by various methods. The spherical harmonics are complex-valued functions depending on two angles, hence the most natural visualization uses a color density plot on the surface of the sphere (either shown as a globe in three dimensions, or represented by a two-dimensional map).

1.6.3. Special topic: Properties of the spherical harmonics

The spherical harmonics are usually[3] defined by

$$Y_\ell^m(\vartheta,\varphi) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} \, e^{\mathrm{i}m\varphi} \, P_\ell^m(\cos\vartheta) \qquad \text{for } 0 \le m \le \ell, \tag{1.106}$$

$$Y_\ell^m(\vartheta,\varphi) = (-1)^m \, \overline{Y_\ell^{-m}(\vartheta,\varphi)} \qquad \text{for } -\ell \le m \le 0. \tag{1.107}$$

Here, the functions $P_\ell^m(z)$ are the *associated Legendre functions*,

$$P_\ell^m(z) = \frac{(-1)^m}{2^\ell \, \ell!} \, (1-z^2)^{m/2} \, \frac{d^{\ell+m}}{dz^{\ell+m}} (z^2-1)^\ell, \tag{1.108}$$

defined for $-1 \le z \le 1$ and integers $0 \le m \le \ell$. The function $P_\ell^0(z) = P_\ell(z)$ is called the *Legendre polynomial* of degree ℓ.

[3] In various books, these definitions can differ by phase factors. We follow the conventions made in *Mathematica*, $Y_\ell^m(\vartheta,\varphi) \equiv$ `SphericalHarmonicY`$[\ell, m, \vartheta, \varphi]$, and $P_\ell^m(z) \equiv$ `LegendreP`$[\ell, m, z]$.

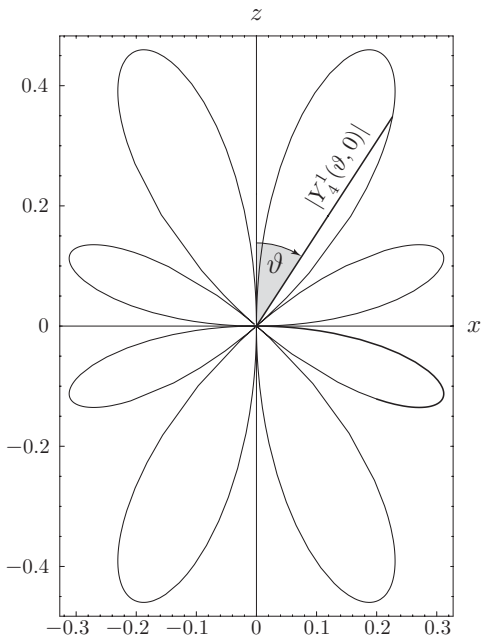


FIGURE 1.8. The absolute value of the spherical harmonics $Y_\ell^m(\vartheta,\varphi)$ depends only on ϑ. Therefore, we can visualize them by plotting a curve in the x_1x_3-plane where at each angle ϑ, the distance from the origin is $|Y_\ell^m(\vartheta,0)|$. This is shown here for $\ell = 4$ and $m = 1$. More examples and three-dimensional visualizations can be found on the CD-ROM; see CD 1.7.

Spherical harmonics with different indices are orthogonal, because they are eigenfunctions belonging to different eigenvalues of a self-adjoint operator, and because of their normalization we simply have

$$\int_{S^2} \overline{Y_\ell^m(\vartheta,\varphi)}\, Y_{\ell'}^{m'}(\vartheta,\varphi)\, d\Omega = \delta_{\ell\ell'}\, \delta_{mm'}. \tag{1.109}$$

The spherical harmonics are either symmetric or antisymmetric under a reflection through the origin,

$$Y_\ell^m(\pi-\vartheta,\varphi+\pi) = (-1)^\ell\, Y_\ell^m(\vartheta,\varphi). \tag{1.110}$$

Finally, we note the addition theorem:

$$\frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} \overline{Y_\ell^m(\vartheta,\varphi)}\, Y_\ell^m(\vartheta',\varphi') = P_\ell(\cos\alpha). \tag{1.111}$$

Here, α is the angle between the directions (ϑ, φ) and (ϑ', φ'). Denoting the corresponding unit vectors by $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$, we have $\cos\alpha = \boldsymbol{\omega} \cdot \boldsymbol{\omega}'$.

In Cartesian coordinates (x, y, z), with $r = (x^2+y^2+z^2)^{1/2}$, the spherical harmonics are (for $m \geq 0$)

$$Y_\ell^m(x,y,z) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\Big(\frac{x+\mathrm{i}y}{\sqrt{x^2+y^2}}\Big)^m P_\ell^m\Big(\frac{z}{r}\Big). \tag{1.112}$$

For example,

$$Y_0^0 = \frac{1}{\sqrt{4\pi}}, \quad Y_1^0 = \sqrt{\frac{3}{4\pi}}\,\frac{z}{r}, \quad Y_1^{\pm1} = \mp\sqrt{\frac{3}{8\pi}}\,\frac{x \pm \mathrm{i}y}{r}. \tag{1.113}$$

1.7. Particle on a Sphere

1.7.1. Classical particle on a sphere

Here, we consider a quantum mechanical system that is only able to perform a rotational motion. Imagine a particle with mass[4] m whose motion is restricted to the surface of a sphere. There are no other forces. This particle has two degrees of freedom, its position on the sphere is given by two angular coordinates, the longitude and the latitude, described by the angles ϑ and φ. A classical particle subject to these constraints will move along the geodesics of the sphere (great circles) with constant (angular) velocity. Its kinetic energy is given by

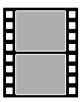
$$E = \frac{1}{2}\mathrm{m}v^2 = \frac{1}{2}\mathrm{m}(r\omega)^2 = \frac{1}{2}I\omega^2, \tag{1.114}$$

where we have introduced the angular speed $\omega = v/r$ and the moment of inertia

$$I = \mathrm{m}r^2. \tag{1.115}$$

The angular momentum of a particle moving with velocity v on a sphere of radius r is $L = \mathrm{m}vr$, or $v = L/\mathrm{m}r$. Insert this into the expression for the energy to obtain

$$E = \frac{L^2}{2\mathrm{m}r^2} = \frac{L^2}{2I}. \tag{1.116}$$



CD 1.2.1, CD 1.5.4, and CD 1.6.3 are all visualizations of the classical motion at a fixed distance from the coordinate origin.

[4]It is general custom to denote the mass of the particle and the eigenvalue of L_3 by the same letter. Usually there is little danger of confusion. To be on the safe side, we denote the mass by the roman letter m and use the italic letter m for the eigenvalue of L_3.

1.7.2. The rigid rotator

The particle on a sphere is a mathematical model for the rigid rotator. Consider two point-like particles of mass m_1 and m_2 that are connected by a weightless, rigid rod of length r. Denote by $\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$ the vector pointing from m_1 to m_2. The position of the center of mass is given by

$$\mathbf{X} = \frac{\mathrm{m}_1\mathbf{x}_1 + \mathrm{m}_2\mathbf{x}_2}{\mathrm{m}_1 + \mathrm{m}_2} \tag{1.117}$$

The distances of the particles from the center of mass are therefore

$$r_1 = |\mathbf{x}_1 - \mathbf{X}| = \frac{\mathrm{m}_2\, r}{\mathrm{m}_1 + \mathrm{m}_2}, \quad r_2 = |\mathbf{x}_2 - \mathbf{X}| = \frac{\mathrm{m}_1\, r}{\mathrm{m}_1 + \mathrm{m}_2}. \tag{1.118}$$

If we are only interested in the internal motion of the two-particle system, we can choose a coordinate system that has its origin at the center of mass. In this coordinate system, the moment of inertia of the rotator is thus

$$I = \mathrm{m}_1 r_1^2 + \mathrm{m}_2 r_2^2 = \frac{\mathrm{m}_1\mathrm{m}_2}{\mathrm{m}_1 + \mathrm{m}_2}\, r^2. \tag{1.119}$$

This is precisely the moment of inertia of a particle with mass

$$\mu = \frac{\mathrm{m}_1\mathrm{m}_2}{\mathrm{m}_1 + \mathrm{m}_2} \tag{1.120}$$

at a distance r from the origin. Thus, we can replace the two-particle system with an effective one-particle system. This particle has the so-called *reduced mass* μ and the fixed distance r from the origin, that is, it is a particle on a sphere with radius r.

This is a good model for a diatomic molecule. It consists of two atoms that can oscillate along the line connecting their centers and rotate around the center of mass. The vibrational motion is much faster than the rotational motion. So, the vibrations belong to much higher energies. Quantum mechanically, oscillation states have a quantized energy. As long as the diatomic molecule is in its vibrational ground state, it is a rigid rotator.

1.7.3. Transition to quantum mechanics

In order to define the quantum mechanical Hamiltonian for a particle that is constrained to the surface of a sphere, we start with the classical expression (1.116) for the energy. From this, the Hamiltonian operator of the rigid rotator is obtained by replacing L^2 with the quantum mechanical angular-momentum operator. Note that r, the radius of the sphere, is treated as a fixed parameter. Using the expression (1.91) for the angular momentum in spherical coordinates, we arrive at

$$\hat{H} = \frac{1}{2\mathrm{m}\, r^2}\, \hat{L}^2 = \frac{1}{2I}\, \hat{L}^2. \tag{1.121}$$

The operator $\hat{L}^2$ has a discrete spectrum of eigenvalues, therefore the same is true for the energy of the rotator.

Eigenvalues of the rigid rotator:

A particle with mass m on a sphere with radius r can only have the energies

$$E_\ell = \frac{\hbar^2}{2\mathrm{m}r^2}\,\ell(\ell+1), \quad \ell = 0, 1, 2, 3, \ldots. \tag{1.122}$$

Each eigenvalue E_ℓ has the multiplicity $2\ell+1$, that is, there are $2\ell+1$ orthogonal wave functions all belonging to the eigenvalue E_ℓ:

$$\psi_{\ell,m}(\vartheta,\varphi) = \frac{1}{r}\,Y_\ell^m(\vartheta,\varphi), \quad m = -\ell, -\ell+1, \ldots, \ell. \tag{1.123}$$

It follows from (1.109) that

$$\int_{S_r^2} \left|\psi_{\ell,m}(\vartheta,\varphi)\right|^2 r^2\,\sin\vartheta\,d\vartheta\,d\varphi = 1. \tag{1.124}$$

The factor $1/r$ in (1.123) thus guarantees that the eigenfunctions are normalized on the sphere with radius r. We conclude that the eigenfunctions form an orthonormal basis in the Hilbert space $L^2(S_r^2)$ of square integrable functions on the sphere of radius r.

We want to stress that $\hat{L}^2$ is part of any single-particle Hamiltonian, because it appears in the angular part of the kinetic energy in spherical coordinates (see (1.150) below). For a particle constrained to the surface of a sphere, the operator $\hat{L}^2$ plays the same role as the operator $-\hbar^2\Delta$ for a particle in $\mathbb{R}^3$: It is proportional to the kinetic energy and it generates the free time evolution. (In fact, $\hat{L}^2/\hbar^2$ is the so-called Laplace-Beltrami operator of the sphere.) The action of the free time evolution $\exp(-\mathrm{i}\hat{L}^2 t/2I)$ on a sphere will be discussed next.

1.7.4. Dynamics of the rigid rotator

The time evolution of an eigenfunction of the rigid rotator Hamiltonian is rather trivial. If $\psi_{\ell,m}(\vartheta,\varphi)$ is an eigenfunction belonging to the eigenvalue E_ℓ, then the function

$$\psi_{\ell,m}(t,\vartheta,\varphi) = \exp(-\mathrm{i}E_\ell\,t)\,\psi_{\ell,m}(\vartheta,\varphi) \tag{1.125}$$

is a solution of the time-dependent Schrödinger equation. The ground state of the rigid rotator has energy $E_0 = 0$, and hence the corresponding wave function does not depend on time at all.

The fact that the Schrödinger equation is linear means that we can consider linear combinations of eigenfunctions. For example, for the initial function $\psi_0(\vartheta,\varphi) = \sum_{\ell,m} c_{\ell,m}\,\psi_{\ell,m}(\vartheta,\varphi)$, the time evolution is

$$\psi(t,\vartheta,\varphi) = \sum_{\ell,m} c_{\ell,m}\,\exp(-\mathrm{i}E_\ell\,t)\,\psi_{\ell,m}(\vartheta,\varphi) \tag{1.126}$$

and this clearly gives us the solution for any initial function that is square integrable on the sphere, because the set of functions $\psi_{\ell,m}$ forms an orthonormal basis in $L^2(S^2_r)$.

Given any initial function ψ_0, the expansion coefficient $c_{\ell,m}$ can be found by an integration,

$$c_{\ell,m} = \int_{S^2_r} \overline{\psi_{\ell,m}(\vartheta,\varphi)}\,\psi_0(\vartheta,\varphi)\,r^2\,d\Omega \tag{1.127}$$

(with $d\Omega = \sin\vartheta\,d\vartheta\,d\varphi$). It is remarkable that all nonzero energies $E_\ell = \ell(\ell+1)/2I$ are integer multiples of $E_1 = 1/I$, hence all time-dependent factors $\exp(-\mathrm{i}E_\ell\,t)$ have the same basic time period. Hence, any state of the rigid rotator is periodic in time.

For the unit sphere ($r = 1$), we summarize our results in the following box.

Time evolution of the rigid rotator:

For a particle with mass m = 1 on the unit sphere S^2 the time evolution of any square-integrable initial function $\psi_0(\vartheta,\varphi)$ is given by the formula

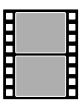
$$\psi(t,\vartheta,\varphi) = \sum_{\ell,m} c_{\ell,m}\,\exp(-\mathrm{i}E_\ell\,t)\,Y^m_\ell(\vartheta,\varphi) \tag{1.128}$$

with

$$c_{\ell,m} = \int_{S^2} \overline{Y^m_\ell(\vartheta,\varphi)}\,\psi_0(\vartheta,\varphi)\,d\Omega. \tag{1.129}$$

The time evolution of the wave function is periodic in time,

$$\psi(t+T,\vartheta,\varphi) = \psi(t,\vartheta,\varphi) \quad \text{with period} \quad T = 2\pi. \tag{1.130}$$



CD 1.9–CD 1.12 is a collection of several movies showing various time-dependent states of the rigid rotator. CD 1.13 visualizes the time evolution of initially well-localized (Gaussian) states. In CD 1.14, the initial state has roughly the shape of the letter Ψ. This illustrates that we can indeed compute the time evolution of any square-integrable initial function on the sphere.

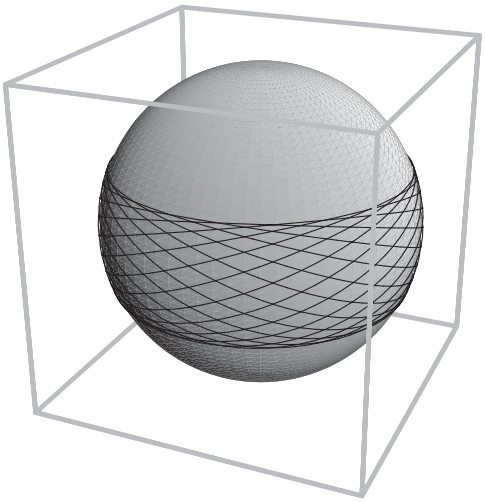


FIGURE 1.9. Possible trajectories of classical particles on a sphere. For all trajectories, the angular-momentum vector has $|\mathbf{L}| = \sqrt{\ell(\ell+1)}$, with $\ell = 6$ and $L_3 = 6$.

1.8. Quantization on a Sphere

1.8.1. Comparison of classical and quantum probability densities

Figure 1.9 shows a set of classical trajectories of particles for which the angular momenta have a fixed absolute value $|\mathbf{L}| = L$ and a fixed value of the third component L_3. The admissible positions of classical particles with these angular momenta form the classically allowed region on the sphere. This region is determined by the polar angle of the classical angular-momentum vector. We denote the polar angle of $\mathbf{L}$ by θ in order to distinguish it from the polar angle ϑ of the particle's position vector, see Figure 1.10.

Assuming $L_3 = m$, we find that the polar angle of $\mathbf{L}$ is

$$\theta = \arccos\Big(\frac{m}{L}\Big) \tag{1.131}$$

and hence the classically allowed region for the position of the particle is between the polar angles

$$\vartheta_{\min} = \pi/2 - \theta, \qquad \vartheta_{\max} = \pi/2 + \theta. \tag{1.132}$$

All the classical trajectories are grand circles of the sphere. The particles move on these circles with a constant angular velocity. When a particle circles around the sphere, the polar angle ϑ performs a periodic motion

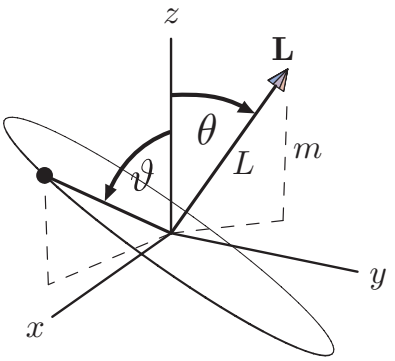


FIGURE 1.10. The polar angle $\theta = \arccos(m/L)$ of the angular momentum and the time-dependent polar angle ϑ of the position vector for a classical particle on a sphere.

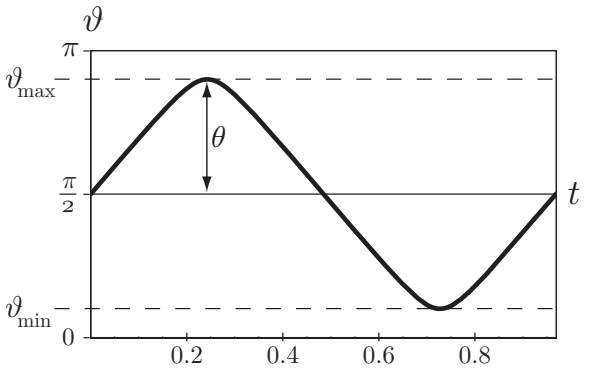


FIGURE 1.11. Polar angle ϑ as a function of time according to (1.133) for $L^2 = 42$ and $m = 2$.

between $\vartheta_{\min}$ and $\vartheta_{\max}$. It is given by

$$\vartheta(t) = \arccos\Big(-\frac{1}{L}\sqrt{L^2 - m^2}\,\sin(L\,t)\Big). \tag{1.133}$$

An example is shown in Figure 1.11.

The angular-momentum vector is normal to the plane of the circle, hence the angle θ defined in (1.131) describes the inclination of that plane. If m is small compared to L, the inclination is large. For $m = 0$, the classical motion is on a circle through the poles of the sphere. For $m = L$ (which is impossible quantum mechanically), the circle of the classical motion is the equator of the sphere.

We want to determine the classical probability density as a function of the coordinate ϑ. This can be done in complete analogy to the calculation for the harmonic oscillator in Book One, Section 7.2.3. The probability density that a particle will be found in an interval $d\vartheta$ around an angle ϑ is equal to the fraction of a period that is spent in the interval. We first calculate the time $dt(\vartheta)$ spent in an infinitesimal interval $d\vartheta$ at the point ϑ. During one period, the trajectory goes twice through ϑ, hence we multiply the time by 2 to obtain the total time spent in $d\vartheta$ during one period. The time needed to complete a period is $T = 2\pi/L$. Hence, the probability of finding the particle in $d\vartheta$ is

$$p(\vartheta)\,d\vartheta = \frac{L}{2\pi}\,2\,dt(\vartheta) = \frac{L}{\pi}\,\frac{dt(\vartheta)}{d\vartheta}\,d\vartheta. \tag{1.134}$$

Hence, one needs to invert the function $\vartheta(t)$ and differentiate with respect to ϑ. It is sufficient to invert the function $\vartheta(t)$ on a part of the time interval during which the trajectory goes through the point ϑ under consideration. Thus, we write

$$t(\vartheta) = -\frac{1}{L}\,\arcsin\Big(\frac{L}{\sqrt{L^2-m^2}}\,\cos\vartheta\Big). \tag{1.135}$$

The function $t(\vartheta)$ is inverse to $\vartheta(t)$ on the interval $(T/4, 3T/4)$. Inserting (1.135) into (1.134), we finally obtain for the classical position probability density

$$p(\vartheta) = \frac{L\,\sin\vartheta}{\pi\,\sqrt{L^2\,(\sin\vartheta)^2 - m^2}}. \tag{1.136}$$

This function is shown as a black line in Figure 1.12 for $m = 8$ and $L = \sqrt{\ell(\ell+1)}$ with $\ell = 24$.

We want to compare the classical position probability density $p(\vartheta)$ with the corresponding quantum mechanical density. With the eigenfunction $\psi_{\ell,m}$ defined in (1.123), we obtain

$$\int_{B\subset S_r^2} \big|\psi_{\ell,m}(\vartheta,\varphi)\big|^2\,r^2\,\sin\vartheta\,d\vartheta\,d\varphi \tag{1.137}$$

as the probability that a particle is found in a region B on S_r^2 (the surface of the sphere with radius r). Let us do the integration over the angle φ because we are interested in the ϑ-position irrespective of the φ-position. This gives only a factor 2π because the absolute value of $\psi_{\ell,m} = (1/r)\,Y_\ell^m$ does not depend on φ. Hence, $2\pi\,|\psi_{\ell,m}(\vartheta,\varphi)|^2\,r^2\,\sin\vartheta\,d\vartheta$ is the probability that a particle on the sphere is in an infinitesimal circular strip of width $d\vartheta$ around the polar angle ϑ. Inserting the definition (1.123) gives for the position probability density as a function of ϑ the expression

$$p_{\rm qm}(\vartheta) = 2\pi\,\sin\vartheta\,|Y_\ell^m(\vartheta,0)|^2. \tag{1.138}$$

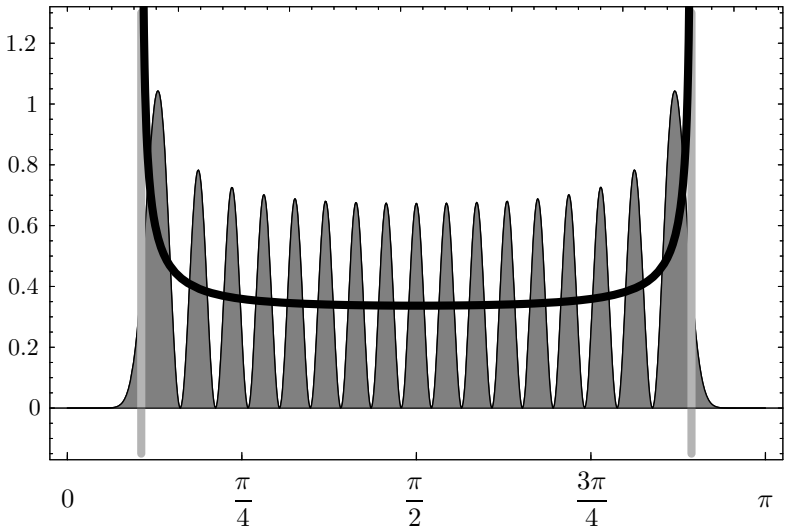
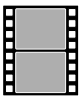


FIGURE 1.12. Classical versus quantum position probability densities as functions of the polar angle ϑ for the rigid rotator. The quantum probability (filled curve) oscillates around the classical probability (black line). The quantum wave function decreases rapidly outside the classically allowed region whose borders are indicated by the vertical gray lines.

Figure 1.12 compares the quantum and classical probability densities as functions of ϑ.



Similar to Figures 1.9, 1.11, and 1.12, but for various values of L_3, the images in CD 1.8 depict the relations between the classically allowed region of the rigid rotator, the classical ϑ-oscillation, and the quantum probability density as a function of ϑ.

1.8.2. Special topic: Curvilinear coordinates

It is worthwhile to consider the transition from classical to quantum mechanics for the rigid rotator in more detail. The classical kinetic energy of a particle with mass $\mathrm{m} = 1$ is given by

$$T(t) = \frac{1}{2}\sum_{i=1}^{3} \dot{x}_i(t)^2. \tag{1.139}$$

Assuming that the motion of the particle takes place on the surface of the unit sphere, we insert (1.81), that is, $x_1(t) = \sin\big(\vartheta(t)\big)\,\cos\big(\varphi(t)\big)$ and so forth,

and obtain an expression for the kinetic energy in spherical coordinates, namely

$$T(t) = \frac{1}{2}\Big(\dot\vartheta(t)^2 + \big(\sin\vartheta(t)\big)^2\,\dot\varphi(t)^2\Big). \tag{1.140}$$

If there are no external forces, the kinetic energy is a constant of motion. Another conserved quantity is the angular momentum. In fact, one finds that the above expression for the kinetic energy is just $T = L^2/2$. In spherical coordinates, we obtain a particularly simple expression for the third component of the angular momentum,

$$L_3(t) = x_1(t)\,\dot x_2(t) - \dot x_1(t)\,x_2(t) = \big(\sin\vartheta(t)\big)^2\,\dot\varphi(t). \tag{1.141}$$

Denoting the constant value of L_3 by m, we may eliminate the angular velocity $\dot\varphi$ and express the kinetic energy solely in terms of ϑ

$$T = \frac{1}{2}\Big(\dot\vartheta(t)^2 + \frac{m^2}{\sin^2\vartheta(t)}\Big). \tag{1.142}$$

This looks like the total energy of a particle with mass 1 in one dimension (coordinate ϑ), with $\dot\vartheta^2/2$ being the kinetic energy and $V(\vartheta) = \frac{1}{2}m^2/\sin^2\vartheta$ being the potential. This potential confines the motion to the interval $(0,\pi)$. Indeed, the equation of motion for the ϑ coordinate is obtained as

$$\ddot\vartheta = \frac{m^2}{\sin^2\vartheta}\,\cot\vartheta \tag{1.143}$$

and the solutions of this equation are given by (1.133) for arbitrary values of L and m.

It is tempting to try a transition to quantum mechanics by applying the standard substitution rule $p_\vartheta = \dot\vartheta \to -\mathrm{i}\partial/\partial\vartheta$ to the classical expression (1.142). This would lead, however, to the wrong equation. The correct equation for the ϑ-coordinate was obtained earlier, when we derived the expression for the Laplacian operator in spherical coordinates. It reads

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(\vartheta,t) = \frac{1}{2}\Big(-\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\,\sin\vartheta\,\frac{\partial}{\partial\vartheta} + \frac{m^2}{\sin^2\vartheta}\Big)\,\psi(\vartheta,t). \tag{1.144}$$

Obviously, the standard substitution rule does not work in spherical coordinates. The reason is that unlike the situation in Cartesian coordinates, the components $\dot\vartheta$ and $\dot\varphi$ of the velocity in curvilinear coordinates do not independently contribute to the total kinetic energy.

Let us just state how to obtain the operator for kinetic energy in general curvilinear coordinates. We denote the coordinates of a system with n degrees of freedom by $q^1, q^2, \ldots, q^n$. The classical kinetic energy in these

coordinates is an expression of the form

$$T = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij}(q^1, \dots, q^n)\, \dot{q}^i\, \dot{q}^j \tag{1.145}$$

with a positive definite symmetric matrix g_{ij} that is defined at every point $(q^1, \dots, q^n)$ of the configuration space. The matrix $g = (g_{ij})$ (the *metric tensor*) defines a *Riemannian metric* on the configuration space, which thus becomes a *Riemannian manifold*. Let us denote the components of the inverse matrix g^{-1} by g^{ij}. Then, the operator for kinetic energy in the coordinates q^j is given by

$$H_0 = -\frac{1}{2}\,\Delta = -\frac{1}{2}\,\frac{1}{\sqrt{\det g}} \sum_{i,j} \frac{\partial}{\partial q^i} \Big(g^{ij}\, \sqrt{\det g}\, \frac{\partial}{\partial q^j} \Big). \tag{1.146}$$

This operator is called the *Laplace-Beltrami operator* on the Riemannian manifold with metric g.

The volume element in curvilinear coordinates is $\sqrt{\det g}\, dq^1\, dq^2 \cdots dq^n$, hence the Hilbert space consists of functions that are square-integrable in the sense that

$$\int |\psi(q^1, \dots q^n)|^2\, \sqrt{\det g}\; dq^1 \cdots dq^n < \infty. \tag{1.147}$$

As an example, we consider the unit sphere with $q^1 = \vartheta$, and $q^2 = \varphi$. From (1.140) we find

$$g = \begin{pmatrix} 1 & 0 \\ 0 & (\sin\vartheta)^2 \end{pmatrix}. \tag{1.148}$$

For the inverse matrix we have $g^{11} = 1$, $g^{22} = 1/(\sin\vartheta)^2$, $g^{12} = g^{21} = 0$. Because of the relation $\sqrt{\det g} = \sin\vartheta$, (1.146) reduces to the well-known expression (see (1.91))

$$H_0 = -\frac{1}{2} \left(\frac{1}{\sin\vartheta}\, \frac{\partial}{\partial\vartheta} \Big(\sin\vartheta \frac{\partial}{\partial\vartheta} \Big) + \frac{1}{\sin^2\vartheta} \frac{\partial^2}{\partial\varphi^2} \right) \tag{1.149}$$

in the Hilbert space of square-integrable function with respect to the volume element $\sin\vartheta\, d\vartheta\, d\varphi$.

A transition to a non-Cartesian coordinate system is also necessary to describe the influence of a gravitational field according to the general theory of relativity.

1.9. Free Schrödinger Equation in Spherical Coordinates

1.9.1. Solutions of the radial equation

For the free-particle Hamiltonian H_0, it is easy to obtain an expression in spherical coordinates by using $\hat{\Delta} = \hat{\nabla}^2$ and the expression (1.87) for the gradient:

$$\hat{H}_0 = -\frac{\hbar^2}{2\mathrm{m}}\hat{\Delta} = \frac{\hbar^2}{2\mathrm{m}}\Big(-\frac{d^2}{dr^2} - \frac{2}{r}\frac{d}{dr}\Big) + \frac{1}{2\mathrm{m}}\frac{1}{r^2}\hat{L}^2. \tag{1.150}$$

The first part, which involves only derivatives with respect to r, describes the kinetic energy of the radial motion. The second part, which contains the operator $\hat{L}^2$ given by (1.91), is the kinetic energy of the angular motion. $\hat{L}^2$ is a partial differential operator that involves only derivatives with respect to ϑ and φ. Hence, $\hat{L}^2$ commutes with the radial kinetic energy and hence with $\hat{H}_0$. The kinetic energy operator $\hat{H}_0$ also commutes with $\hat{L}_3$, because $\hat{L}_3 = -\mathrm{i}d/d\varphi$ commutes with $\hat{L}^2$ and with expressions depending on r. This proves that the operator of kinetic energy is invariant under rotations:

$$[\hat{H}_0, \hat{L}^2] = 0. \tag{1.151}$$

The same commutation relation holds for the operators in Cartesian coordinates, which are related to the operators in spherical coordinates by a unitary transformation. Hence, $[H_0, L^2] = 0$. Any eigenspace of L^2 is left invariant by H_0. If ψ_ℓ is an eigenvector of L^2, then $H_0\psi_\ell$ is an eigenvector of L^2 belonging to the same eigenvalue. We can thus restrict the operator H_0 to an arbitrary eigenspace of L^2. Using spherical coordinates, we see immediately that this restriction reduces the partial differential operator $\hat{H}_0$ to the ordinary differential operator

$$\mathbf{h}_{0\ell} = \frac{\hbar^2}{2\mathrm{m}}\Big(-\frac{d^2}{dr^2} - \frac{2}{r}\frac{d}{dr} + \frac{\ell(\ell+1)}{r^2}\Big), \tag{1.152}$$

which is called the *free radial Schrödinger operator*. Here, the angular kinetic energy appears in the form of a potential energy $\ell(\ell+1)/r^2$. This term is called the *centrifugal potential energy* or *centrifugal barrier*, because it has the effect of a repulsive force in the radial direction.

In order to solve the Schrödinger equation in spherical coordinates we could use the trial function

$$\phi(r, \vartheta, \varphi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{r} f_{\ell m}(r)\, Y_\ell^m(\vartheta, \varphi) \tag{1.153}$$

and find immediately that $\hat{H}_0\phi = E\phi$ holds if $f_{\ell m}$ is a solution of the *radial Schrödinger equation*

$$\frac{\hbar^2}{2\mathrm{m}}\Big(-\frac{d^2}{dr^2}+\frac{\ell(\ell+1)}{r^2}\Big)\,f(r) = E\,f(r). \tag{1.154}$$

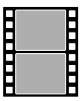
Note that the factor $1/r$ in (1.153) makes the first-order term $(2/r)\,d/dr$ disappear. The condition for square-integrability of the wave function ψ (see (1.98)) would mean that each $f_{\ell m}$ has to be square integrable in the radial Hilbert space $L^2([0,\infty)$, that is,

$$\int_0^\infty |f_{\ell m}(r)|^2\,dr < \infty. \tag{1.155}$$

But we know that the free-particle Schrödinger equation has no square-integrable solutions with sharp energy. Instead, the solutions of the stationary Schrödinger equation are plane waves with arbitrary momentum $\mathbf{k}$. We expect a similar behavior for the solutions of the radial Schrödinger equation. Hence, we expect that there exist bounded solutions for $E > 0$. The differential equation (1.154) is a Bessel equation. As a differential equation of second order it has two linearly independent solutions. These solutions are called *Riccati-Bessel functions*. With $k = \sqrt{2mE/\hbar^2}$ (positive square root, $E > 0$), the solutions are

$$\hat{\jmath}_\ell(kr) = \sqrt{\frac{\pi}{2}\,kr}\,J_{\ell+1/2}(kr), \qquad \hat{n}_\ell(kr) = \sqrt{\frac{\pi}{2}\,kr}\,N_{\ell+1/2}(kr) \tag{1.156}$$

where J_ν and N_ν are the Bessel function and Neumann functions of order ν. The function $\hat{\jmath}_\ell$ is regular for $r \to 0$ whereas $\hat{n}_\ell$ is singular for $\ell > 0$.



An interactive plot of the Riccati-Bessel function $\hat{\jmath}_\ell(r)$ in CD 1.15 allows investigation of the dependence on ℓ.

1.9.2. Special Topic: Properties of the Riccati-Bessel functions

The real-valued functions $\hat{\jmath}_\ell(z)$ and $\hat{n}_\ell(z)$ are for $z > 0$ solutions of the equation

$$-\frac{d^2y(z)}{dz^2}+\frac{\ell(\ell+1)}{z^2}y(z) - y(z) = 0. \tag{1.157}$$

We have

$$\hat{\jmath}_0(z) = \sin z, \quad \hat{n}_0(z) = \cos z, \tag{1.158}$$

and the Riccati-Bessel functions of higher order can be computed from

$$\hat{j}_\ell(z) = -(-z)^{\ell+1} \Big(\frac{1}{z}\,\frac{d}{dz}\Big)^\ell \Big(\frac{1}{z}\,\hat{j}_0(z)\Big), \tag{1.159}$$

$$\hat{n}_\ell(z) = -(-z)^{\ell+1} \Big(\frac{1}{z}\,\frac{d}{dz}\Big)^\ell \Big(\frac{1}{z}\,\hat{n}_0(z)\Big). \tag{1.160}$$

Their limiting behavior for small z is given by

$$\hat{j}_\ell(z) = \frac{2^\ell \ell!}{(2\ell+1)!}\, z^{\ell+1} \big(1 + \mathrm{O}(z^2)\big), \quad \text{as } z \to 0, \tag{1.161}$$

$$\hat{n}_\ell(z) = \frac{(2\ell)!}{2^\ell \ell!}\, z^{-\ell} \big(1 + \mathrm{O}(z^2)\big), \quad \text{as } z \to 0. \tag{1.162}$$

In scattering theory, one often defines the Riccati-Hankel functions

$$\hat{h}^\pm_\ell(z) = \hat{n}_\ell(z) \pm \mathrm{i}\hat{j}_\ell(z) = \mathrm{e}^{\pm \mathrm{i}(z - \ell\pi/2)} \big(1 + \mathrm{O}(1/z)\big), \quad \text{as } z \to \infty. \tag{1.163}$$

Further details about the Riccati-Bessel functions can be found in the book [**1**], where the notation $\hat{\jmath}(z) = zj(z)$, and $\hat{n}(z) = zy(z)$ is used.

1.9.3. Special Topic: Expanding the plane wave

The plane waves $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$ are important solutions of the free-particle Schrödinger equation, despite the fact that they are not square-integrable. Here, we show that plane waves have an expansion like (1.153). The radial part of a plane wave in the subspace with angular-momentum quantum number ℓ is just the Riccati-Bessel function $\hat{j}_\ell(kr)$. This result is important for the applications to stationary scattering theory with spherically symmetric potentials.

Representation of plane waves:

The stationary plane wave $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$ has the following expansion in terms of spherical harmonics:

$$\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} = \frac{1}{r} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} c_{\ell,m}(\mathbf{k})\, \hat{j}_\ell(kr)\, Y_\ell^m(\vartheta,\varphi), \tag{1.164}$$

where (r,ϑ,φ) are the spherical coordinates of $\mathbf{x} \in \mathbb{R}^3$. The coefficients in this expansion are given by

$$c_{\ell,m}(\mathbf{k}) = \frac{4\pi \mathrm{i}^\ell}{k}\, \overline{Y_\ell^m(\vartheta',\varphi')}\,. \tag{1.165}$$

Here, (k,ϑ',φ') are the spherical coordinates of $\mathbf{k} \in \mathbb{R}^3$.

The proof of the expansion formula uses properties of the Legendre polynomials $P_\ell(x)$ and the addition theorem for spherical harmonics (1.111). The

Legendre polynomials form an orthogonal set in the Hilbert space of square-integrable functions on the interval $-1 \le x \le 1$. According to Abramowitz-Stegun [**1**], the Legendre polynomials satisfy

$$\int_{-1}^{1} P_\ell(x)\, P_{\ell'}(x)\, dx = \frac{2}{2\ell+1}\, \delta_{\ell\ell'}. \tag{1.166}$$

From this, we conclude that the functions

$$f_\ell(x) = \sqrt{\frac{2\ell+1}{2}}\, P_\ell(x) \tag{1.167}$$

form an orthonormal basis in the Hilbert space $L^2([-1,1])$.

For any wave number $\mathbf{k}$, the plane wave can be rewritten as

$$\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} = \mathrm{e}^{\mathrm{i}kr\cos\alpha} = \mathrm{e}^{\mathrm{i}krx}, \quad \text{with } x = \cos\alpha \in [-1,1]. \tag{1.168}$$

α is the angle between the vectors $\mathbf{k}$ and $\mathbf{x}$. As a function of x (for fixed k and r), the function $\exp(\mathrm{i}krx)$ belongs to the Hilbert space $L^2([-1,1])$. Hence, it can be expanded in the Legendre polynomials in the usual way,

$$\mathrm{e}^{\mathrm{i}krx} = \sum_\ell c_\ell(kr)\, f_\ell(x), \qquad c_\ell(kr) = \int_{-1}^{1} f_\ell(x)\, \mathrm{e}^{\mathrm{i}krx}\, dx. \tag{1.169}$$

The integral involving a Legendre polynomial and an exponential function can be evaluated analytically. It is best to consult a good book (see, for example, [**1**], Section 10.1.14), where we find

$$\int_{-1}^{1} P_\ell(x)\, \mathrm{e}^{\mathrm{i}qx}\, dx = \frac{2\mathrm{i}^\ell}{q}\, \hat{\jmath}_\ell(q), \quad \ell = 0, 1, 2, \ldots. \tag{1.170}$$

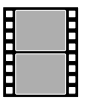
With the help of this formula, we find

$$c_\ell(kr) = \sqrt{2(2\ell+1)}\, \frac{\mathrm{i}^\ell}{kr}\, \hat{\jmath}_\ell(kr) \tag{1.171}$$

and hence

$$\mathrm{e}^{\mathrm{i}krx} = \frac{1}{kr} \sum_{\ell=0}^{\infty} (2\ell+1)\, \mathrm{i}^\ell\, \hat{\jmath}_\ell(kr)\, P_\ell(x) \quad \text{or} \tag{1.172}$$

$$\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} = \frac{1}{kr} \sum_{\ell=0}^{\infty} (2\ell+1)\, \mathrm{i}^\ell\, \hat{\jmath}_\ell(kr)\, P_\ell(\cos\alpha). \tag{1.173}$$



The expansion formula (1.173) is visualized in CD 1.16. The missing square integrability of a plane wave manifests itself in the fact that the series (1.173) does not converge in the usual sense. With increasing ℓ, the individual summands do not get smaller (with respect to the norm in $L^2(\mathbb{R}^3)$), but they contribute only in regions increasingly far away from the origin.

We may insert the summation rule (1.111) for spherical harmonics into (1.173). We choose the angles (ϑ', φ') in the direction of $\mathbf{k}$, which has an angle α with the direction of (ϑ, φ) (= the direction of $\mathbf{x}$). This leads to the result

$$\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} = \frac{4\pi}{kr} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \mathrm{i}^{\ell}\, \hat{\jmath}_{\ell}(kr)\, \overline{Y_{\ell}^{m}(\vartheta', \varphi')}\, Y_{\ell}^{m}(\vartheta, \varphi), \tag{1.174}$$

which is identical with (1.164).

1.9.4. Special topic: Spherical harmonics and the Fourier transformation

Consider a function $\psi(\mathbf{x})$ that can be separated into a radial part and an angular part,

$$\psi(\mathbf{x}) = \frac{1}{r}\, f(r)\, Y_{\ell}^{m}(\vartheta, \varphi), \tag{1.175}$$

where (r, ϑ, φ) are the spherical coordinates of the point $\mathbf{x} \in \mathbb{R}^3$. Obviously, ψ is an eigenfunction of the angular-momentum operators L^2 and L_3. We want to find the Fourier transform $\hat{\psi}(\mathbf{k})$. To that end, we insert (1.175) into the formula for the Fourier transform

$$\hat{\psi}(\mathbf{k}) = \Big(\frac{1}{2\pi}\Big)^{3/2} \int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\, \psi(\mathbf{x})\, d^3x \tag{1.176}$$

$$= \Big(\frac{1}{2\pi}\Big)^{3/2} \int_0^{\infty} \int_{S^2} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\, \frac{1}{r}\, f(r)\, Y_{\ell}^{m}(\vartheta, \varphi)\, r^2 dr\, d\Omega \tag{1.177}$$

For the plane wave $\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}$, we substitute the expansion in terms of spherical harmonics (the complex conjugate of (1.174)),

$$\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} = \frac{4\pi}{kr} \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} (-\mathrm{i})^{\ell'}\, \hat{\jmath}_{\ell'}(kr)\, Y_{\ell'}^{m'}(\vartheta', \varphi')\, \overline{Y_{\ell'}^{m'}(\vartheta, \varphi)}. \tag{1.178}$$

Using the orthonormality of the spherical harmonics, Eq. (1.109), we can perform the integration over the angles ϑ and φ. This removes the sum over ℓ' and m', because only the term with $\ell' = \ell$ and $m' = m$ gives a contribution. The short calculation gives

$$\hat{\psi}(\mathbf{k}) = \Big(\frac{1}{2\pi}\Big)^{3/2} \int_0^{\infty} \frac{4\pi}{kr}\, (-\mathrm{i})^{\ell}\, \hat{\jmath}_{\ell}(kr)\, Y_{\ell}^{m}(\vartheta', \varphi')\, \frac{1}{r}\, f(r)\, r^2 dr \tag{1.179}$$

$$= \frac{1}{k}\, (-\mathrm{i})^{\ell} \Big(\frac{2}{\pi}\Big)^{1/2} \int_0^{\infty} f(r)\, \hat{\jmath}_{\ell}(kr)\, dr\, Y_{\ell}^{m}(\vartheta', \varphi') \tag{1.180}$$

$$= \frac{1}{k}\, h(k)\, Y_{\ell}^{m}(\vartheta', \varphi'), \tag{1.181}$$

where $(k, \vartheta', \varphi')$ are the spherical coordinates of $\mathbf{k} \in \mathbb{R}^3$. We have found that the Fourier transform $\hat{\psi}(k)$ is again an eigenfunction of L^2 and L_3 with the same quantum numbers ℓ and m. The Fourier transform maps an angular momentum eigenspace into itself.

The function h is connected to f via an integral transformation that is known as the *Hankel transformation*,

$$h(k) = (-\mathrm{i})^\ell \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty f(r)\, \hat{\jmath}_\ell(kr)\, dr. \tag{1.182}$$

The Hankel transformation is the "radial Fourier transformation" in the angular-momentum subspace belonging to the quantum number ℓ. The inverse Hankel transformation can be derived from the formula for the inverse Fourier transformation. It is given by

$$f(r) = \mathrm{i}^\ell \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty h(k)\, \hat{\jmath}_\ell(kr)\, dk. \tag{1.183}$$

Like the Fourier transformation on $L^2(\mathbb{R}^3)$, the Hankel transformation extends to a unitary transformation on $L^2([0,\infty))$,

$$\int_0^\infty |f(r)|^2\, dr = \int_0^\infty |h(k)|^2\, dk. \tag{1.184}$$

The quantum mechanical interpretation of f and h is straightforward. The function f is the radial position probability amplitude and h gives the radial momentum distribution:

$$\int_a^b |f(r)|^2\, dr \quad \text{is the probability of finding the particle in a distance between } a \text{ and } b \text{ from the origin.}$$

$$\int_a^b |h(k)|^2\, dk \quad \text{is the probability that the absolute value of the momentum is between } a \text{ and } b.$$

One should keep in mind that the relation between the radial momentum distribution and the radial position distribution depends on the angular-momentum quantum number ℓ. Let us finally put the main result into a box.

Radial-angular separation and the Fourier transform:

The Fourier transform of $\frac{1}{r}\, f(r)\, Y_\ell^m(\vartheta, \varphi)$ is given by $\frac{1}{k}\, h(k)\, Y_\ell^m(\vartheta', \varphi')$, where h is the Hankel transform of f. The Fourier transform thus leaves each angular-momentum subspace invariant.

An arbitrary function in $L^2(\mathbb{R}^3)$ has an expansion of the form

$$\psi(\mathbf{x}) = \frac{1}{r} \sum_{\ell,m} f_{\ell,m}(r)\, Y_\ell^m(\vartheta, \varphi), \tag{1.185}$$

and the result above can be applied, because the Fourier transformation is linear.

1.10. Spherically Symmetric Potentials

1.10.1. The structure of the eigenvalue spectrum

A spherically symmetric potential $V(\mathbf{x}) = V(r)$ is the operator of multiplication with a function that depends only on the distance $r = |\mathbf{x}|$ from the origin. Because the angular-momentum operators involve derivatives only with respect to the angular coordinates ϑ and φ, and because V does not depend on the angles, we have the commutation relation

$$[V(r), \mathbf{L}] = 0. \tag{1.186}$$

The angular momentum also commutes with the kinetic energy, and hence the quantum mechanical system described by the Hamiltonian $H = H_0 + V$ is rotationally invariant:

$$[H, \mathbf{L}] = 0. \tag{1.187}$$

The stationary Schrödinger equation in the angular-momentum subspace with quantum numbers ℓ, m is

$$\frac{\hbar^2}{2\mathrm{m}}\Big(-\frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{r^2}\Big)\, f(r) + V(r)\, f(r) = E\, f(r). \tag{1.188}$$

This is the radial Schrödinger equation with a potential. Clearly, this equation does not depend on the eigenvalue m of L_3, because a spherically symmetric Hamiltonian H does not contain L_3.

The eigenvalues of the three-dimensional Schrödinger equation

$$H\psi = \Big(-\frac{\hbar^2}{2\mathrm{m}}\,\Delta + V(|\mathbf{x}|)\Big)\psi = E\,\psi \tag{1.189}$$

are those numbers $E_{\ell;n_r}$ for which (1.188) has a nonzero square-integrable solution in the Hilbert space $L^2([0,\infty))$. Like the radial equation, the eigenvalues $E_{\ell;n_r}$ are independent of the quantum number m. The number n_r is called the *radial quantum number*. It just serves as a label for the different eigenvalues of the radial equation. We denote the radial eigenfunction belonging to the eigenvalue $E_{\ell;n_r}$ by $f_{\ell;n_r}(r)$.

Using the radial eigenfunctions, we define eigenfunctions of H belonging to the eigenvalues $E_{\ell;n_r}$ by

$$\psi_{\ell,m;n_r}(\mathbf{x}) = \frac{1}{r} f_{\ell;n_r}(r)\, Y_\ell^m(\vartheta,\varphi), \tag{1.190}$$

where (r,ϑ,φ) are the spherical coordinates of $\mathbf{x}$, and m is any of the eigenvalues of L_3. The functions $\psi_{\ell,m;n_r}$ are simultaneous eigenstates of H, L^2, and L_3, belonging to the eigenvalues $E_{\ell;n_r}$, $\ell(\ell+1)$, and m, respectively. For every possible energy $E_{\ell;n_r}$ and every possible ℓ we have $2\ell+1$ linearly independent eigenfunctions $\psi_{\ell,m;n_r}$, with $m = -\ell, -\ell+1, \ldots, +\ell$. Hence, the degree of degeneracy of the eigenvalue $E_{\ell;n_r}$ is at least $2\ell+1$. Experimentally, the states belonging to different eigenvalues of L_3 can only be distinguished in the presence of magnetic fields. Hence, m is often called the *magnetic quantum number*.

Eigenvalue problem with spherical symmetry:

The simultaneous eigenfunctions of L^2, L_3, and the spherically symmetric Schrödinger operator $H = p^2/2m + V(r)$ are given by

$$\psi_{\ell,m;n_r}(\mathbf{x}) = \frac{1}{r} f_{\ell;n_r}(r)\, Y_\ell^m(\vartheta,\varphi), \tag{1.191}$$

where $f_{\ell;n_r}$ is the n_r-th eigenfunction of the *radial Schrödinger operator*

$$\mathbf{h}_\ell = \frac{\hbar^2}{2\mathrm{m}}\left(-\frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{r^2}\right) + V(r). \tag{1.192}$$

The corresponding eigenvalues $E_{\ell;n_r}$ do not depend on the quantum number m. Hence, the multiplicity of each eigenvalue of H is at least $2\ell+1$.

The radial Schrödinger operator $\mathbf{h}_\ell$ is a self-adjoint differential operator in $L^2([0,\infty))$. We can order its eigenvalues according to their size,

$$E_{\ell;n_r} < E_{\ell,n_r+1}, \quad n = 0,1,2,\ldots. \tag{1.193}$$

The quantum number n_r counts the number of zeros of the radial eigenfunction $f_{\ell;n_r}(r)$ in the interval $(0,\infty)$. The ground state with angular momentum ℓ has no zero at all (except at $r=0$), the first excited state has precisely one zero, and so forth. To every eigenvalue $E_{\ell;n_r}$ belongs a unique eigenfunction $f_{\ell;n_r}$. Hence, the eigenvalues of $\mathbf{h}_\ell$ are non-degenerate (but not the eigenvalues of H).

Any linear combination of the eigenvectors $\psi_{\ell,m;n_r}$ belonging to some eigenvalue E of H is again an eigenvector belonging to the same eigenvalue E,

$$\psi_{\ell;n_r}(\mathbf{x}) = \sum_{m=-\ell}^{\ell} c_m \, \psi_{\ell,m;n_r}(\mathbf{x}), \quad \text{where } c_m \in \mathbb{C}. \tag{1.194}$$

In particular, the eigenfunctions

$$\psi^+_{\ell,|m|;n_r}(\mathbf{x}) = \frac{1}{\sqrt{2}} \big(\psi_{\ell,m;n_r}(\mathbf{x}) + (-1)^m \, \psi_{\ell,-m;n_r}(\mathbf{x})\big), \tag{1.195}$$

$$\psi^-_{\ell,|m|;n_r}(\mathbf{x}) = \frac{\mathrm{i}}{\sqrt{2}} \big(\psi_{\ell,m;n_r}(\mathbf{x}) - (-1)^m \, \psi_{\ell,-m;n_r}(\mathbf{x})\big) \tag{1.196}$$

are real-valued. (Note that the radial part $f_{\ell;n_r}$ can be chosen to be real because it is a solution of a real differential equation vanishing for $r \to 0$ and $r \to \infty$). These eigenfunctions are called *real orbitals*. A real orbital is an eigenfunction of the energy and of L^2, but not of L_3.

The states with angular momenta $\ell = 0, 1, 2$, and 3 are sometimes also denoted by the letters s, p, d, and f. This is the spectroscopic notation. The letters have the following historical meaning: s = simple, p = principal, d = diffuse, f = fundamental. Higher quantum numbers are then denoted alphabetically by g, h, and so forth.

Ψ As a differential equation of second order, (1.188) has two linearly independent solutions for every energy E, say $u(E, r)$ and $v(E, r)$. Typically, these solutions behave as $u(E, r) \sim r^{\ell+1}$ and $v(E, r) \sim r^{-\ell}$ for small r, see also (1.161) and (1.162). A physically correct solution is one for which the corresponding function (1.190) is in the domain of the self-adjoint operator H. The mathematical theory tells us that these functions have to be bounded and continuous on $\mathbb{R}^3$, at least for physically meaningful (that is, not too singular) potential functions V. Therefore, $(1/r)\, f_{\ell;n_r}(r)$ has to remain bounded, as $r \to 0$. Only the solution u has this property.[5] This solution is called the regular solution of (1.188). It is distinguished by the boundary condition $u(r) \to 0$, as $r \to 0$. The solution $u(E, r)$ turns out to be square-integrable only for exceptional values of the parameter E, that is, for the eigenvalues $E = E_{\ell;n_r}$. In these cases, the solution $f_{\ell;n_r}(r) = u(E_{\ell;n_r}, r)$ also vanishes at infinity.

[5]For $\ell = 0$, the second solution v is also bounded and hence square-integrable in a neighborhood of $r = 0$. Hence, the condition of square-integrability alone is not sufficient to select the physically correct solution.

1.10.2. The vibrating rotator: A model of a diatomic molecule

In Section 1.7.2, we stated that the rigid rotator is an approximate model for a diatomic molecule in a vibrational ground state. Here, we are going to refine the model and take into account the vibrational degree of freedom. We assume that the two atoms are bound together as a result of some complicated interaction between the two nuclei and all their electrons. The configuration with the minimal potential energy will have the two nuclei at some distance r_0 from each other. The effective force between the nuclei would be repulsive at closer distances $r < r_0$ and attractive at larger distances $r > r_0$. It is evident that the details of the effective interaction can be very complicated because many particles are involved. We are going to discuss a very simple model that nevertheless shows many essential features of diatomic molecules. This potential is called *Kratzer's molecular potential* (see Fig. 1.13). It is given by

$$V(r) = V_0 \left(\frac{r_0^2}{r^2} - 2\,\frac{r_0}{r} \right), \quad V_0 > 0, \quad r_0 > 0, \tag{1.197}$$

where r is the internuclear distance. This function is spherically symmetric, and it has a minimum for $r = r_0$, with $V(r_0) = V_0$. It is clear that this model is very unphysical both at very small distances and at very large distances. At small distances, the model potential describes a very strong repulsive potential with a $1/r^2$-singularity (too strong for the Coulomb repulsion of the positively charged nuclei which is proportional to $1/r$). For large r, the dominating term in $V(r)$ is an attractive Coulomb potential $-2V_0 r_0/r$. But at large distances, the two atoms are separated completely. As they are neutral, the effective force between them will be much weaker than the attractive Coulomb force described by the model. But nevertheless, the model is not bad within a certain range of energies. It describes a two-particle system that is able to perform rotations and simultaneous oscillations around an equilibrium distance. Moreover, Kratzer's model is convenient because the Schrödinger equation with this potential can be solved exactly.

We use the center-of-mass coordinates described in Section 1.7.2 and insert for m the reduced mass μ of the two atoms. The radial Schrödinger equation (1.188) with the potential (1.197) reads

$$\frac{\hbar^2}{2\mu}\Big(-\frac{d^2}{dr^2} + \frac{\ell(\ell+1)+a}{r^2} \Big)\, f(r) - \frac{2\,V_0\,r_0}{r}\, f(r) = E\, f(r), \tag{1.198}$$

where we used the abbreviation

$$a = \frac{2\mu\, r_0^2}{\hbar^2}\, V_0. \tag{1.199}$$

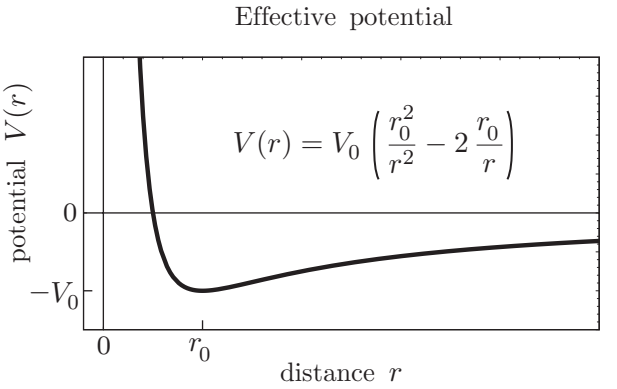


FIGURE 1.13. A simple model for the effective potential between the two nuclei in a diatomic molecule.

By the variable substitution

$$f(r) = g(x), \quad x = \frac{a}{r_0}\, r, \tag{1.200}$$

the radial Schrödinger equation is transformed into

$$\frac{1}{2}\Big(-\frac{d^2}{dx^2} + \frac{\ell(\ell+1)+a}{x^2} \Big)\, g(x) - \frac{1}{x}\, g(x) = \frac{1}{2aV_0}\, E\, g(x). \tag{1.201}$$

Writing

$$\epsilon = \frac{E}{2aV_0} \quad \text{and} \quad \lambda = -\frac{1}{2} + \left(\Big(\ell + \frac{1}{2}\Big)^2 + a \right)^{1/2} \tag{1.202}$$

such that $\lambda(\lambda+1) = \ell(\ell+1) + a$, we arrive at

$$\frac{1}{2}\Big(-\frac{d^2}{dx^2} + \frac{\lambda(\lambda+1)}{x^2} \Big)\, g(x) - \frac{1}{x}\, g(x) = \epsilon\, g(x). \tag{1.203}$$

This equation is formally identical to the radial Schrödinger equation for the Coulomb problem (except that the parameter λ need not be an integer). We are going to solve this equation in the next chapter (see Section 2.5.2, Eq. (2.99)). Here, we just quote the result (2.62) that (1.203) has the eigenvalues

$$\epsilon_{\lambda;n_r} = -\frac{1}{2(n_r + \lambda + 1)^2}, \quad n = 0, 1, 2, 3, \ldots. \tag{1.204}$$

Inserting the definitions of λ and ϵ, we obtain the energy-eigenvalues for this problem as

$$E_{\ell;n_r} = -\frac{2\mu r_0^2}{\hbar^2}\, V_0^2 \left(n_r + \frac{1}{2} + \left(\Big(\ell + \frac{1}{2}\Big)^2 + \frac{2\mu r_0^2}{\hbar^2}\, V_0 \right)^{1/2} \right)^{-2}. \tag{1.205}$$

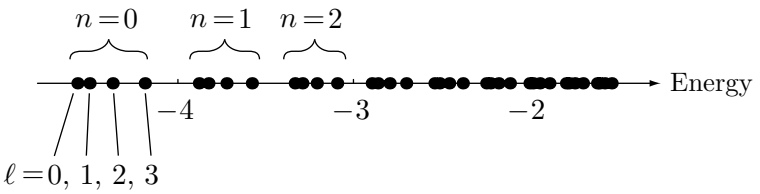


FIGURE 1.14. Energy eigenvalues arising from Kratzer's molecular potential with $V_0 = 5$ and $r_0 = 5$. The graph shows the eigenvalues with quantum numbers $n = n_r \leq 8$ and $\ell \leq 3$.

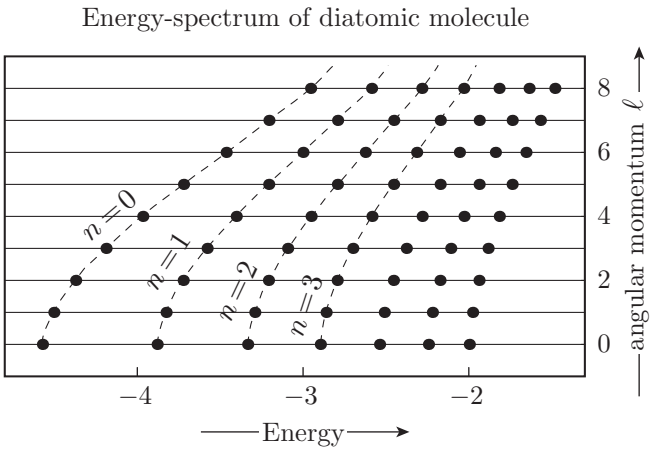


FIGURE 1.15. Two-dimensional diagram of the eigenvalue spectrum according to Kratzer's model of a diatomic molecule ($V_0 = 5$ and $r_0 = 5$), showing the quantum numbers $n = n_r \leq 8$ and $\ell \leq 8$.

The quantum number n_r enumerates the eigenvalues of the radial Schrödinger operator. It thus describes the vibration around the classical equilibrium at r_0. The quantum number ℓ describes the rotational state as usual. We expect that the formula (1.205) is a good approximation for a diatomic molecule if the radial oscillations are not too large. This is the case for small values of the quantum number n_r, where the corresponding eigenstates are localized sufficiently close to the classical equilibrium at $r = r_0$. Similar restrictions will apply to the quantum number ℓ.

Figures 1.14 and Figure 1.15 show the energies $E_{\ell;n_r}$ for small values of n_r and ℓ, assuming that $a \gg 1$ (as it is the case for most molecules). Obviously,

Figure 1.14 would become too cluttered if more than a few eigenvalues are shown. In Figure 1.15, we use a two-dimensional plot where the vertical position of an eigenvalue indicates the quantum number ℓ. For our choice of parameter values, the energy-difference between two vibrational levels $E_{\ell,n_r+1} - E_{\ell;n_r} \approx \mathrm{O}(1/a)$ is much larger than the spacing $E_{\ell+1;n_r} - E_{\ell;n_r}$ of the rotational levels, which is of order $1/a^2$. Hence, it needs much more energy to bring the system from $n_r = 0$ to $n_r = 1$ than to change the angular-momentum quantum number. If the energies involved are not too high, the quantum number n_r remains unchanged, and the system behaves like a rigid rotator.

Chapter 2

Coulomb Problem

Chapter summary: Perhaps the most important success of quantum mechanics is the explanation of the internal electronic structure of atoms on the basis of general physical laws. The Schrödinger equation with a Coulomb potential, although a crude model of a "real" hydrogen atom, can describe its properties with high accuracy and explains the spectroscopic observations that have puzzled physicists at the beginning of the 20th century.

Already the classical Coulomb problem has interesting aspects. We discuss, in particular, the conservation of the Runge-Lenz vector, which implies that the form and the orientation of the classical orbits (Kepler ellipses) is constant in time and depends only on the initial conditions. In quantum mechanics, the Runge-Lenz vector can be used for an elegant algebraic computation of the energy levels given in Section 2.3.

The following sections are devoted to various approaches to the solution of the Coulomb problem. As the Coulomb potential is spherically symmetric, we may apply the results of the previous chapter in order to reduce the problem to the solution of ordinary differential equations in the angular-momentum subspaces. In Section 2.4, we use a factorization method to solve the radial Coulomb problem in an essentially algebraic way. This approach leads to a solution of the Coulomb problem via a system of simultaneous eigenfunctions of H, L^2, and L_3 and exhibits clearly the structure of the energy spectrum, in particular the high degeneracy of the eigenvalues. In this approach, the ℓ-degeneracy appears as a consequence of a supersymmetry of the radial Schrödinger operators, which is similar to the one observed for the harmonic oscillator in Book One.

We continue by outlining the traditional approach the Coulomb problem, which consists in solving the radial Schrödinger equation in terms of special functions. In Section 2.5.1, we present a solution of the Coulomb problem in two dimensions which, naturally, plays an important role for the visualizations accompanying this book. Finally, we present a method for solving the Schrödinger equation in parabolic coordinates. The separation in parabolic coordinates is important for the investigation of the Stark effect (hydrogen atom in a constant electric field).

In the major part of this chapter, we use dimensionless units in all calculations. This simplifies mathematical derivations, yields beautiful formulas, and is by no means a restriction of generality. A simple scaling transformation gives all information about the eigenfunctions and eigenvalues in SI units (or in any other system of units). Section 2.7 is devoted to the actual physical dimensions of the

hydrogen atom. Moreover, we describe general scaling transformations (dilations) as unitary transformations in the Hilbert space, because they are an important tool in mathematical physics. In this connection, we present the virial theorem in Section 2.7.4.

Concerning the dynamics of states in the subspace of bound states, we investigate the behavior of simple superpositions in Section 2.4.5 and the dynamics of circular Rydberg states in Section 2.8. Circular Rydberg states are highly excited states that move in the close neighborhood of classical circular orbits. The long lifetime of these states and their quasiclassical behavior make them an interesting research topic in atomic physics.

2.1. Introduction

2.1.1. Classical models of atoms

An atom is often described as a small planetary system where the gravitational force is replaced by the electrostatic attraction. Most of the mass of the atom is concentrated in a nucleus with positive charge Ze, where Z is the *atomic number* (the number of protons in the nucleus). In a neutral atom, Z negatively charged electrons move around the positively charged nucleus like planets around the sun. The simplest atom is the hydrogen atom, which consists of a single proton as the nucleus and a single electron.

This classical model of the atom as a system of moving charges dates back to Rutherford (1911), who could explain in this way the scattering of α-particles (doubly ionized He^{++} nuclei) by the atoms of a thin gold foil. In the framework of classical physics, the planetary system model has some problems, and many of the facts observed in experiments cannot be explained in that way. For example, a classical planetary system consisting of the sun and a single planet is planar. But hydrogen atoms (in their ground state) appear as spherical objects with a radius of about 10^{-10} m. The main difficulty is perhaps the following: The classical elliptic orbit of a charged particle in the electrostatic field of the nucleus should not be stable, because the acceleration of a charge would cause an electromagnetic radiation leading to a continuous loss of energy. The atomic radius should therefore decrease continuously, and the emitted radiation would have a continuous range of frequencies. Indeed, energetically excited atoms do emit radiation. But contrary to the classical expectation, the atom only emits radiation with a few discrete frequencies. Atoms in their ground state do not emit radiation at all.

The investigation of the radiation emitted or absorbed by atoms or molecules is called spectroscopy. The spectrum is the intensity of the emitted radiation as a function of the frequency. It is the main source of experimental information about atoms. A gas of atomic hydrogen emits light with

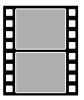
frequencies ν_{nm} described by *Balmer's formula*

$$\nu_{nm} = R_{\mathrm{H}}\, c \left(\frac{1}{n^2} - \frac{1}{m^2} \right). \tag{2.1}$$

Here, c is the speed of light, n and m are positive integers with $m > n$, and the factor $R_{\mathrm{H}} \approx 1.09678 \times 10^7\,\mathrm{m}^{-1}$ is called the *Rydberg constant for hydrogen*.

In 1913, these frequencies have been explained by Niels Bohr by assuming that the electrons in the field of the nucleus can only have the energies $-R_{\mathrm{H}}\, hc/n^2$, where n is a positive integer, and h is Planck's constant.[1] A photon is emitted whenever an electron makes a transition from a level with energy $-R_{\mathrm{H}}\, hc/m^2$ to a level with lower energy $-R_{\mathrm{H}}\, hc/n^2$. The energy of the photon is the difference in energy between the two levels. According to an observation of Einstein, the energy of a photon is related to its frequency by $E = h\nu$. From this the formula (2.1) follows immediately. Independent experimental evidence that the internal energy of atoms can only have discrete values came in particular through experiments with mercury atoms conducted by Franck and Hertz in 1914.

In retrospect, given deBroglie's hypothesis relating the energy of a particle to the frequency of a wave, it appears very natural to describe the discrete energies of electrons bound to an atomic nucleus by the possible frequencies of a wave function. In the 1920s it was already well-known that under certain circumstances, the frequencies of waves can only assume discrete values (for example, the solutions of the classical wave equation in the presence of boundary conditions). In 1926, Erwin Schrödinger found the wave equation that gave the right frequencies and hence energies for the hydrogen atom. Solving the Schrödinger equation for hydrogen is the topic of this section.



CD 2.2.2 is a short introduction to the spectroscopy of hydrogen atoms. The radiation emitted by the atoms consists of photons that carry away the energy set free when an electron makes a transition from an excited state to a state with lower energy. This is a process whose complete description requires quantum electrodynamics.

2.1.2. Transitions between eigenstates

From elementary quantum mechanics you certainly know physical systems, for which the Schrödinger equation has discrete energy levels. See, for example, the harmonic oscillator presented in Book One. Soon we are going to find the energy levels of the hydrogen atom by solving the Schrödinger

[1] Actually, in Bohr's model, the Rydberg constant for hydrogen R_{H} is replaced by the Rydberg constant $R_\infty \approx 1.097 \times 10^7\,\mathrm{m}^{-1}$. Bohr obtained this value under the assumption of an infinite nuclear mass. At Bohr's time, R_∞ was in perfect agreement with the experimental results.

equation with an electrostatic Coulomb potential. As you certainly know, the discrete energy levels correspond to stationary states. According to the theory, stationary states are absolutely stable. Once an electron is in a stationary state, it remains there forever. In reality, however, electrons make transitions due to unavoidable perturbations. Even in the complete absence of external fields, a spontaneous decay of excited states will be stimulated by the zero-point fluctuations of the vacuum field (a quantum-electrodynamical effect).

As explained in the previous section, the spectroscopic information about the energies of an atom is obtained from the transitions between energy levels. These transitions are associated with the emission or absorption of photons, and these photons can be observed. Obviously, the Schrödinger equation with a time-independent potential (from which we obtain the energy levels) cannot describe the physical effect that tells us about these energy levels. The description of the emission or absorption of photons requires quantum electrodynamics. This theory combines the (relativistic) quantum theory of electrons with the quantum theory of electromagnetic fields. Unfortunately, quantum electrodynamics is beyond the scope of this book. For now, the stationary solutions of the Schrödinger equation will have to suffice.

It should be mentioned, however, that in the framework of "ordinary" quantum mechanics, transitions between stationary states can be modeled with time-dependent perturbation theory. One starts with an unperturbed system for which the Hamiltonian has eigenvalues E_n with corresponding eigenvectors ψ_n. If the initial state $\psi(0)$ of the electron is one of the eigenstates ψ_n, then its time evolution is just given by $\psi(t) = \exp(-\mathrm{i}E_n t)\,\psi_n$. Hence, the transition probability from $\psi(t)$ to another eigenstate ψ_m will be zero for all times,

$$|\langle\psi_m, \psi(t)\rangle|^2 = |\langle\psi_m, \psi_n\rangle|^2 = 0 \quad \text{for } m \neq n, \tag{2.2}$$

because two eigenstates belonging to different energies are orthogonal. This is why we call the eigenstates of the Hamiltonian "stationary states."

In practice, it is impossible to keep a physical system isolated. There will always be some (weak) external influence. In a simple mathematical model, we may describe the external influence by an additional, generally time-dependent potential that is added to the Hamiltonian of the unperturbed system. The whole system is then represented by a time-dependent Hamiltonian. The energy eigenstates of the unperturbed Hamiltonian are not stationary with respect to the time evolution of the perturbed system. If the initial state is $\psi(0) = \psi_n$, then the state $\psi(t)$ will depend on time in a nontrivial way, and the transition probability between $\psi(t)$ and ψ_m (with $m \neq n$) will, in general, be nonzero for $t \neq 0$.

From the remarks above, it should be clear that the Schrödinger equation with an electrostatic Coulomb potential is just a simplified mathematical model of the real hydrogen atom. A more realistic model would have to take into account in particular the following phenomena, ordered according to their importance:

(1) The spin of the electron (see Chapter 3).
(2) The interaction between the spin magnetic moment and the magnetic moment due to the orbital motion (spin-orbit interaction).
(3) Relativistic kinematics.
(4) Nuclear contributions like the interaction with the magnetic moment of the nucleus and the finite size of the nucleus.
(5) Vacuum fluctuations of the electromagnetic fields.
(6) Vacuum polarization.

These effects lead to corrections of the energy levels obtained from the Schrödinger equation and to a splitting of otherwise degenerate levels. The effects 1 to 3 describe the so-called *fine structure* of the hydrogen spectrum. They are successfully dealt with by the Dirac equation. The magnetic moment of the nucleus is the origin of the *hyperfine structure*. The finite size of the nucleus is important, in particular, for hydrogenic atoms with large atomic number Z. These effects are usually incorporated via perturbation theory. The effects 5 and 6 can be described with the help of quantum electrodynamics. The corresponding correction of the eigenvalues is called the *Lamb shift*.

Let us now turn to the problem of understanding the hydrogen atom as it is described by the Schrödinger equation with a Coulomb potential. By exploiting the symmetries of the system, it is possible to find all stationary states with analytic (that is, non-numerical) methods. These symmetries are already present at the level of classical physics. Most obvious is the fact that the Coulomb potential is spherically symmetric, that is, invariant under rotations. But there is an additional symmetry that results in the conservation of the so-called Runge-Lenz vector.

2.2. The Classical Coulomb Problem

2.2.1. The Coulomb force

Using the international system of units (SI), the Coulomb law for the magnitude of the force between two charges q_1 and q_2 separated by a distance r is written as

$$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}. \tag{2.3}$$

Here, ε_0 is the *permittivity of the vacuum*. The unit of the charge in the SI is the coulomb (C), defined as ampere×second. The charge of an electron is denoted by $-e$, where e is the *elementary charge*

$$e = 1.6021773 \times 10^{-19}\,\mathrm{C}. \tag{2.4}$$

In atomic physics, it is most common to describe the Coulomb force in the Gaussian system (cgs-system), where the Coulomb law reads

$$F = \frac{q_1 q_2}{r^2}. \tag{2.5}$$

The Gaussian system uses length, mass, and time as basic dimensions with the units centimeter (cm), gram (g), and second (s), respectively. The unit of charge in the Gaussian system is defined to be that charge which exerts a force of $1\,\mathrm{g\,cm\,s^{-1}} = 1\,\mathrm{dyn}$ on a charge of the same size at a distance of 1 cm. The elementary charge is $e = 4.8032 \times 10^{-10}\,\mathrm{g}^{1/2}\,\mathrm{cm}^{3/2}\,\mathrm{s}^{-1}$.

A hydrogenic atom is a one-electron atom or ion, for example, H, He^+, Li^{2+}, and so forth. It consists of an atomic nucleus and an electron that interacts with the nucleus by the electrostatic Coulomb attraction. The nucleus contains Z protons and has a total charge $+Ze$. Hence, the magnitude of the force on an electron at a distance r from the nucleus is given by

$$F = \frac{Ze^2}{4\pi\varepsilon_0\, r^2} = \frac{\gamma}{r^2}. \tag{2.6}$$

The constant $\gamma = Ze^2/4\pi\varepsilon_0$ is the *Coulomb coupling constant*. This abbreviation makes us fairly independent of the chosen system of units. If you plan to use the Gaussian system instead of the SI, you just have to use the definition $\gamma = Ze^2$ in the following (with e in Gaussian units).

In our model of the hydrogenic atom, the nucleus is a point mass. We assume that the dimensions of the nucleus can be neglected compared to the dimensions of the atom. If the nuclear charge Z is not too large, this assumption is indeed satisfied with high accuracy.

The hydrogenic atom is a two-particle system, but we are going to describe it in the framework of the single-particle formalism. The motion of the atom as a whole is ignored, and the nucleus is just a fixed center-of-force at the coordinate origin. Nevertheless, as in Section 1.7.2, we can take into account the finite mass of the nucleus simply by describing the configuration of the atom in terms of the relative coordinate vector $\mathbf{x} = \mathbf{x}_{\mathrm{electron}} - \mathbf{x}_{\mathrm{nucleus}}$ and by replacing the electron mass with the reduced mass

$$\mu = \frac{\mathrm{m}_{\mathrm{nucleus}}\,\mathrm{m}_e}{\mathrm{m}_{\mathrm{nucleus}} + \mathrm{m}_e}. \tag{2.7}$$

Here, $\mathrm{m_e}$ is the mass of the electron,

$$\mathrm{m}_e = 9.10938 \times 10^{-31}\,\mathrm{kg}. \tag{2.8}$$

$\mathrm{m}_{\mathrm{nucleus}}$ is the mass of the nucleus. The nucleus of a hydrogen atom consists of just one proton having about 1836 times the mass of the electron,

$$\mathrm{m}_p = 1.67262 \times 10^{-27}\,\mathrm{kg}. \tag{2.9}$$

In this case the reduced mass is

$$\mu = \frac{1836}{1837}\,\mathrm{m}_e = 9.10443 \times 10^{-31}\,\mathrm{kg}. \tag{2.10}$$

The nucleus of a hydrogenic atom is much heavier than the electron. Hence, the reduced mass μ is roughly equal to the mass of the electron. In Bohr's model of the hydrogen atom, it is even assumed that the nucleus has an infinite mass, hence $\mu = \mathrm{m}_e$.

The Coulomb force on the electron is the negative gradient of a potential energy $V(\mathbf{x})$,

$$\mathbf{F}(\mathbf{x}) = -\nabla V(\mathbf{x}), \quad V(\mathbf{x}) = -\frac{\gamma}{|\mathbf{x}|}. \tag{2.11}$$

The energy of a classical particle in a Coulomb field is given by the kinetic energy $p^2/2\mathrm{m}$ plus the potential energy due to the Coulomb attraction:

$$H = \frac{|\mathbf{p}|^2}{2\mathrm{m}} - \frac{\gamma}{|\mathbf{x}|}. \tag{2.12}$$

Here and in the following, we simply write m for the mass. If you want to take into account the finite mass of the nucleus, simply set $\mathrm{m} = \mu$. If you want to work within Bohr's model, use $\mathrm{m} = \mathrm{m}_e$ instead.

2.2.2. Classical motion

The energy $H = H(\mathbf{p}, \mathbf{x})$ of a particle moving in a Coulomb field is given in (2.12) as a function of the position $\mathbf{x}$ and the momentum $\mathbf{p}$ of the particle. $H(\mathbf{p}, \mathbf{x})$ is the classical Hamiltonian function. It is a conserved quantity, that is, for the motion $(\mathbf{p}(t), \mathbf{x}(t))$ of the classical particle we have

$$H\big(\mathbf{p}(t), \mathbf{x}(t)\big) = E \quad (\text{independent of } t). \tag{2.13}$$

The angular momentum $\mathbf{L} = \mathbf{x} \times \mathbf{p}$ is another constant of motion, because the potential is spherically symmetric. It follows that the motion takes place in the plane orthogonal to $\mathbf{L}$, because the position vector $\mathbf{x}$ of the particle satisfies $\mathbf{L} \cdot \mathbf{x} = 0$.



The movie CD 1.2.4 shows the classical Coulomb motion with the conserved angular momentum vector. CD 2.1.1 gives another example. Note that the classical motion takes place in a plane. In contrast, the quantum mechanical solutions make use of all three space dimensions (the ground state is even spherically symmetric).

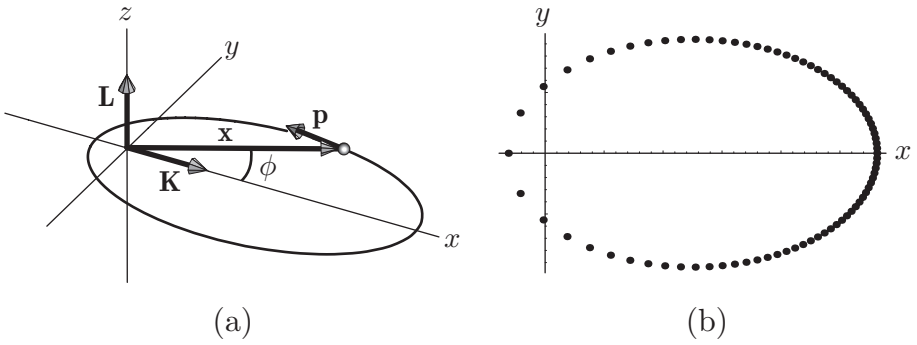


FIGURE 2.1. (a) Classical trajectory and the conserved quantities **L** (angular momentum) and **K** (Runge-Lenz vector). (b) Stroboscopic image of the motion of the particle. The spots show the position of the particle after equal time steps.

It has been known since the times of Lagrange that the Coulomb system has additional conserved quantities. As a consequence, all bounded orbits in the Coulomb field are closed. After some time, the particle returns to its initial position, that is, all bounded motions are periodic in time. Among all spherically symmetric potentials, only the Coulomb potential and the harmonic oscillator have this property.

As you certainly know, a bounded orbit in a Coulomb field has the shape of an ellipse. This is Kepler's first law for the motion of planets around the sun. The form and the orientation of this ellipse is constant in time and depends only on the initial conditions. Hence, the vector that points in the direction along the major axis and whose length is the eccentricity of the ellipse is a conserved quantity. This vector is known as the Runge-Lenz vector (see Fig. 2.1). It is defined by

$$\mathbf{K} = \mathbf{L} \times \mathbf{p} + \mathrm{m}\gamma \frac{\mathbf{x}}{|\mathbf{x}|}. \tag{2.14}$$

The three components of the Runge-Lenz vector are all constants of motion. Together with the energy and the three components of the angular-momentum operator, this makes seven constants of motion that cannot possibly be all independent of each other (the phase space of the Coulomb problem is only six-dimensional).

EXERCISE 2.1. *Prove that the vector* **K** *is always orthogonal to* **L***, that is,* $\mathbf{K} \cdot \mathbf{L} = 0$*. Moreover, show that the length of* **K** *is determined by the energy* H *and the angular momentum* **L** *as*

$$K^2 = 2\mathrm{m}\, H\, L^2 + \mathrm{m}^2\gamma^2. \tag{2.15}$$

The shape of the Coulomb orbit is easily determined from the Runge-Lenz vector. We just compute the scalar product between $\mathbf{K}$ and $\mathbf{x}$:

$$\mathbf{K} \cdot \mathbf{x} = |\mathbf{K}|\,|\mathbf{x}|\,\cos\phi = -L^2 + \mathrm{m}\,\gamma\,|\mathbf{x}|, \tag{2.16}$$

where ϕ is the angle between $\mathbf{K}$ and $\mathbf{x}$. Hence, we find

$$|\mathbf{x}| = r = \frac{L^2}{\mathrm{m}\,\gamma - |\mathbf{K}|\,\cos\phi}. \tag{2.17}$$

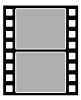
The distance r as a function of the angle ϕ is always finite if $K^2 \leq \mathrm{m}^2\gamma^2$, because in this case the denominator is always nonzero. In view of Eq. (2.15), this can only happen for orbits with negative energies $E < 0$. Negative energies can only occur if $\gamma > 0$, that is, if the Coulomb potential is attractive.

Equation (2.17) can be written as

$$r(\phi) = \frac{d}{1 - \epsilon\,\cos\phi}, \tag{2.18}$$

with $d = L^2/\mathrm{m}\gamma$ and $\epsilon = |\mathbf{K}|/\mathrm{m}\gamma$. This is the focal equation of a conic section. For $\epsilon < 1$, the focal equation describes an ellipse. The constant ϵ is called the *eccentricity*, and d is the *parameter* of the ellipse. The distance r from the origin and the angle ϕ between position vector and Runge-Lenz vector can be interpreted as polar coordinates in the two-dimensional plane of motion. Figure 2.2 shows the ellipse and the geometric meaning of the various constants. The Runge-Lenz vector always points from the origin (focus) to the center of the ellipse. In the special case of a circular orbit, we have $|\mathbf{K}| = 0$ (the focus coincides with the center).

Orbits with positive energies are unbounded. The distance r of the particle can become infinite if for certain values of ϕ the denominator of Eq. (2.18) vanishes. For this we must have $\epsilon\cos\phi = 1$, which can only be the case for $\epsilon \geq 1$ or $K^2 \geq \mathrm{m}^2\gamma^2$. In view of (2.15), this can only happen for $H(\mathbf{p}, \mathbf{x}) = E \geq 0$. For $E = 0$ (that is, $K^2 = \mathrm{m}^2\gamma^2$ or $\epsilon = 1$) the orbit is a parabola, and for $H > 0$ the orbit is a hyperbola.



CD 2.1.1 shows the motion in a Coulomb field and the vectors $\mathbf{K}$ and $\mathbf{L}$, which determine the shape of the elliptic orbit. CD 2.1.2 explores the dependence of the Coulomb orbit on the initial condition for the momentum. The influence of the coupling constant γ is shown in CD 2.1.3. Finally, CD 2.1.4 shows a collection of orbits with the same energy.

EXERCISE 2.2. *Consider a bounded circular orbit. Show that the time needed for one revolution is given by*

$$T = \frac{2\pi L^3}{\mathrm{m}\,\gamma^2}. \tag{2.19}$$

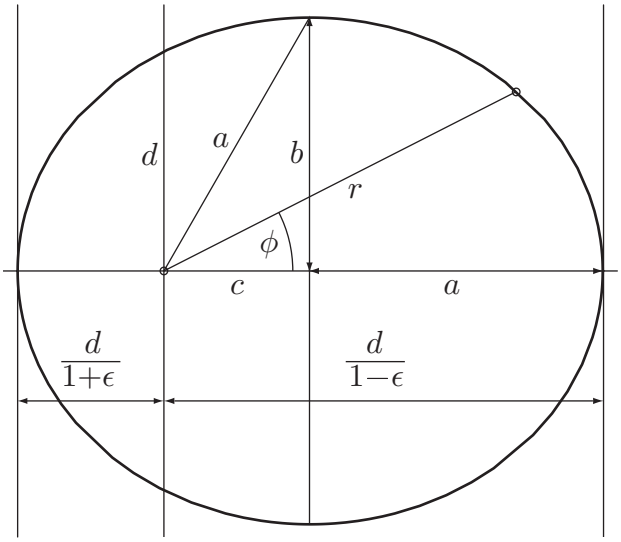


FIGURE 2.2. Any bounded and hence periodic orbit of a particle in a Coulomb potential is an ellipse.

2.3. Algebraic Solution Using the Runge-Lenz Vector

2.3.1. The Coulomb problem in quantum mechanics

The quantum mechanical wave equation is the Schrödinger equation

$$\mathrm{i}\,\hbar\,\frac{d}{dt}\,\psi(\mathbf{x},t) = H\,\psi(\mathbf{x},t), \tag{2.20}$$

where H is the Hamiltonian operator describing the energy of the system. The Hamiltonian H is obtained from the classical Hamiltonian function (2.12) by replacing the variables x_i and p_i with linear operators in the Hilbert space of the system. With Eq. (1.29) we obtain from Eq. (2.12) the Hamiltonian operator of the hydrogen atom

$$H = -\frac{\hbar^2}{2\mathrm{m}}\,\Delta - \frac{\gamma}{|\mathbf{x}|}. \tag{2.21}$$

In order to simplify the following formulas and calculations, we assume that $\mathrm{m} = 1$, $\hbar = 1$, and $\gamma = 1$ (attractive Coulomb potential). It will be shown below (Section 2.7) that this can always be achieved by a suitable scaling transformation of the space and time coordinates. Hence, it entails no loss

of generality to solve the Schrödinger equation in the simplified form

$$\mathrm{i}\,\frac{d}{dt}\,\psi(\mathbf{x},t) = \Big(-\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|}\Big)\,\psi(\mathbf{x},t). \tag{2.22}$$

This is the *Schrödinger equation in dimensionless units*. It is easy to get the solution in SI units (or in any other system of units) from the solution in dimensionless units (see Section 2.7).

As usual, the general solution of (2.22) in the subspace of bound states is found by first solving the eigenvalue problem

$$H\,\psi(\mathbf{x}) = \Big(-\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|}\Big)\,\psi(\mathbf{x}) = E\,\psi(\mathbf{x}). \tag{2.23}$$

We will do so by exploiting the conservation laws already found for the classical problem.

2.3.2. The Runge-Lenz vector in quantum mechanics

The classical expression (2.14) for the Runge-Lenz vector has to be symmetrized in order to find its quantum mechanical analog, because the operators $\mathbf{L}$ and $\mathbf{p}$ do not commute.

Exercise 2.3. *Prove the following formula for the vector product of the operators* $\mathbf{L}$ *and* $\mathbf{p}$*:*

$$(\mathbf{L}\times\mathbf{p})^\dagger = -\mathbf{p}\times\mathbf{L} = \mathbf{L}\times\mathbf{p} - 2\mathrm{i}\mathbf{p}. \tag{2.24}$$

In view of (2.24), we define the quantum mechanical Runge-Lenz vector as

$$\mathbf{K} = \frac{1}{2}\,\big(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}\big) + \frac{\mathbf{x}}{|\mathbf{x}|}. \tag{2.25}$$

Each component K_1, K_2, and K_3 of $\mathbf{K}$ is formally symmetric and can thus give rise to a self-adjoint operator.

Exercise 2.4. *Prove the identity*

$$\mathrm{i}[L^2,\mathbf{p}] = \mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}. \tag{2.26}$$

Using the formula obtained in the previous exercise, we see that the Runge-Lenz vector can likewise be defined as

$$\mathbf{K} = \frac{\mathrm{i}}{2}\,[L^2,\mathbf{p}] + \frac{\mathbf{x}}{|\mathbf{x}|}. \tag{2.27}$$

This form is often more useful when one tries to compute commutation relations.

The three components of the Runge-Lenz vector are conserved during the time evolution. Indeed, one can verify that $\mathbf{K}$ commutes with the Coulomb Hamiltonian H defined in (2.23),

$$[H, \mathbf{K}] = 0. \tag{2.28}$$

The relations between $\mathbf{K}$, $\mathbf{L}$, and H essentially carry over from classical physics. Indeed, from $\mathbf{L} \cdot \mathbf{x} = \mathbf{L} \cdot \mathbf{p} = 0$ one finds the relations

$$\mathbf{K} \cdot \mathbf{L} = \mathbf{L} \cdot \mathbf{K} = 0. \tag{2.29}$$

Moreover, a little calculation proves

$$K^2 = \mathbf{K} \cdot \mathbf{K} = 2H(L^2 + 1) + 1, \tag{2.30}$$

which is the analog of Eq. (2.15). It is also useful to note the following commutation relations:

$$[L_j, K_k] = \mathrm{i}\epsilon_{jkm} K_m, \qquad [K_j, K_k] = -2\mathrm{i}\epsilon_{jkm} H L_m. \tag{2.31}$$

In the next section, I am going to show that these algebraic relations between H, $\mathbf{K}$, and $\mathbf{L}$ are sufficient to determine the eigenvalues of H.

2.3.3. The eigenvalues of H

In the following, we restrict ourselves to that part of the state space which is spanned by the eigenfunctions of H. This is the subspace of bound states, a Hilbert-subspace of $L^2(\mathbb{R}^3)$. On this subspace, we define the two vector-operators $\mathbf{J}^+$ and $\mathbf{J}^-$ by

$$\mathbf{J}^{\pm} = \frac{1}{2}\Big(\mathbf{L} \pm \frac{1}{\sqrt{-2H}}\,\mathbf{K}\Big). \tag{2.32}$$

This definition requires a few comments. It follows from general considerations using the virial theorem (see Section 2.7.4 below) that the stationary states all belong to negative eigenvalues E_n of H. Hence, the operator $1/\sqrt{-2H}$ is well defined on the subspace of bound states because of the minus under the square-root. Any state in this subspace is a linear combination of stationary states. Because H (and likewise any function of H) commutes with $\mathbf{K}$, the action of $\mathbf{J}^{\pm}$ on such a linear combination is given by

$$\mathbf{J}^{\pm}\big(\sum_n c_n \psi_n\big) = \sum_n c_n \frac{1}{2}\Big(\mathbf{L} \pm \frac{1}{\sqrt{2|E_n|}}\,\mathbf{K}\Big)\psi_n. \tag{2.33}$$

It is crucial to observe the commutation relations of the operators $\mathbf{J}^{\pm}$. The calculation gives

$$[J_j^+, J_k^+] = \mathrm{i}\epsilon_{jkm}\, J_m^+, \quad [J_j^-, J_k^-] = \mathrm{i}\epsilon_{jkm}\, J_m^-, \quad [J_j^+, J_k^-] = 0. \tag{2.34}$$

We see that $\mathbf{J}^+$ and $\mathbf{J}^-$ fulfill the commutation relations of two commuting angular-momentum operators. This gives us occasion to apply our knowledge concerning the possible eigenvalues of angular-momentum operators (see Section 1.5). Because of

$$(\mathbf{J}^+)^2 = (\mathbf{J}^-)^2 = -\frac{1}{2}\Big(1 + \frac{1}{2H}\Big) \tag{2.35}$$

the two operators $(\mathbf{J}^+)^2$ and $(\mathbf{J}^-)^2$ have the same eigenvalues. Moreover, these eigenvalues are related to the eigenvalues of H. On the subspace of bound states we have

$$H = -\frac{1}{2}\,\frac{1}{1 + 4(\mathbf{J}^+)^2}. \tag{2.36}$$

We know already from Theorem 1.1 that because of the commutation relations (2.34), the operator $(\mathbf{J}^+)^2$ can only have the eigenvalues $j(j+1)$ with $j = 0, 1/2, 1, 3/2, \ldots$. Thus, because of (2.36), H can only have the eigenvalues

$$E_n = -\frac{1}{2}\,\frac{1}{1 + 4j(j+1)} = -\frac{1}{2}\,\frac{1}{(2j+1)^2} = -\frac{1}{2n^2}, \tag{2.37}$$

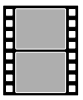
where $n = 2j + 1 = 1, 2, 3, \ldots$ is called the *principal quantum number* of the Coulomb problem.

The orbital angular momentum $\mathbf{L}$ can only have the eigenvalues $\ell(\ell+1)$ with ℓ being a non-negative integer. But in an eigenspace of H belonging to a fixed n, the quantum number ℓ can only have the values $0, 1, \ldots, n-1$. One can see this by using (2.30) to express the eigenvalues of K^2 in terms of n and ℓ. If $\psi_{n,\ell}$ is a simultaneous eigenvector of both H and L^2, then

$$K^2\,\psi_{n,\ell} = \Big(1 - \frac{1}{2n^2}\big(\ell(\ell+1) + 1\big)\Big)\,\psi_{n,\ell}, \tag{2.38}$$

where ℓ is a non-negative integer (see Section 1.6.2). Because the eigenvalue of K^2 must be non-negative, we find that $\ell \leq n - 1$.

For a given j, there are $n^2 = (2j+1)^2$ different pairs (m_j^+, m_j^-) of eigenvalues of J_3^+ and J_3^-. Therefore, the eigenvalue $-1/2n^2$ of H is n^2-fold degenerate and the corresponding orthogonal eigenvectors may be distinguished by the eigenvalues of J_3^+ and J_3^-.



CD 2.2.1 shows the eigenvalues of H, indicating the degree of degeneracy of each E_n by the possible eigenvalues of the angular-momentum operators L^2 and L_3.

So far we have only shown that the eigenvalues of the Coulomb Hamiltonian are among the numbers $-1/2n^2$, with n a positive integer. It remains to prove by an explicit construction of the eigenvectors that all these numbers indeed do occur. (This will also prove that $(\mathbf{J}^+)^2$ has both integer and

half-integer eigenvalues j.) The joint eigenfunctions of H, $(\mathbf{J}^+)^2$, J_3^+, and J_3^- are obtained if one solves the Coulomb problem in parabolic coordinates (see Section 2.6 below). Because of its importance, we are going to present several methods for solving the Coulomb problem. In the next section, we give a solution by exploiting the spherical symmetry.

Ψ The Coulomb Hamiltonian is invariant with respect to rotations in three dimensions. The three-dimensional rotation group $SO(3)$ (special orthogonal transformations in three dimensions) is a three-dimensional Lie group that has the angular-momentum operators as generators. For the Coulomb system, we have in fact found two independent angular-momentum algebras. These six angular-momentum operators generate a six-dimensional symmetry group that can be described as a product of two rotation groups. The symmetry group of the Coulomb problem is therefore isomorphic to the group $SO(4) \cong SO(3) \times SO(3)$, which can be interpreted as a group of rotations in a four-dimensional space. We note that because of slightly changed commutation relations in the case of positive energies, the symmetry group of the Coulomb Hamiltonian in the positive-energy subspace (scattering states) is $SO(3,1)$ (the Lorentz group). This has nothing to do with relativity (in relativistic quantum mechanics, the generators of $SO(3,1)$ have a different physical meaning).

2.4. Algebraic Solution of the Radial Schrödinger Equation

2.4.1. Factorization of the radial Coulomb problem

The Hamiltonian operator for an attractive Coulomb potential

$$H = -\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|} \tag{2.39}$$

is spherically symmetric. In an eigenspace of the angular-momentum operators, the eigenvalue equation $H\psi = E\psi$ becomes the ordinary differential equation

$$\mathbf{h}_\ell\, f(r) = E\, f(r), \qquad \text{where} \quad \mathbf{h}_\ell = \frac{1}{2}\Big(-\frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{r^2}\Big) - \frac{1}{r}. \tag{2.40}$$

From this we conclude immediately (see also Section 1.10.1):

> The Coulomb Hamiltonian H has the eigenvalue E whenever one of the radial Hamiltonians $\mathbf{h}_\ell$ has E as an eigenvalue.

Given a solution $f(r)$ of the radial Coulomb problem (2.40), we obtain a solution of $H\psi = E\psi$ by setting

$$\psi(\mathbf{x}) = \frac{1}{r} f(r)\, Y_\ell^m(\vartheta, \varphi). \tag{2.41}$$

Here, as usual, (r, ϑ, φ) are the spherical coordinates of $\mathbf{x}$. The function $f(r)$ may be interpreted as a radial position probability amplitude (see Section 1.9.4).



CD 2.3 presents a collection of images of the radial eigenfunctions $g(r) = f(r)/r$, comparing them with the corresponding solutions of the two-dimensional Coulomb problem (see below).

We present a method of solving (2.40) which will be very similar to our method of solving the harmonic oscillator problem in Book One. This method works because of the higher symmetry of the Coulomb problem. The crucial step toward the solution of the eigenvalue problem is to note that the radial Hamiltonian $\mathbf{h}_\ell$ in Eq. (2.40) can be factorized, that is, we can write it as the product of two first-order differential operators. Define the operator

$$A_\ell^- = \frac{1}{\sqrt{2}} \Big(\frac{d}{dr} + \frac{\ell + 1}{r} - \frac{1}{\ell + 1} \Big) \tag{2.42}$$

and the (formally) adjoint operator

$$A_\ell^+ = \frac{1}{\sqrt{2}} \Big(-\frac{d}{dr} + \frac{\ell + 1}{r} - \frac{1}{\ell + 1} \Big) \tag{2.43}$$

for all values $\ell = 0, 1, 2, 3, \ldots$ of the angular-momentum quantum number. A little calculation shows that

$$A_\ell^-\, A_\ell^+ = \frac{1}{2} \Big(-\frac{d}{dr} + \frac{\ell(\ell + 1)}{r^2} \Big) - \frac{1}{r} + \frac{1}{2(\ell + 1)^2}. \tag{2.44}$$

Up to a constant term, this expression is just the radial Hamiltonian operator

$$A_\ell^-\, A_\ell^+ = \mathbf{h}_\ell + \frac{1}{2(\ell + 1)^2}. \tag{2.45}$$

The product of the operators A_ℓ^- and A_ℓ^+ in reverse order is easily calculated. It gives

$$A_\ell^+\, A_\ell^- = \mathbf{h}_{\ell+1} + \frac{1}{2(\ell + 1)^2}. \tag{2.46}$$

The importance of this observation lies in the following simple fact about the product of operators:

Whenever the product of two operators AB has a nonzero eigenvalue E with an associated eigenvector f, then the product BA has the same eigenvalue, and the corresponding eigenvector is Bf.

We called this fact the *spectral supersymmetry* of AB and BA. The proof is the same as the one given in Book One, Section 7.7.1, for $B = A^\dagger$.

EXERCISE 2.5. *Let $ABf = Ef$ for two operators A and B and some eigenvalue $E \neq 0$. Show that $g = Bf$ is a nonzero eigenvector of the operator BA belonging to the eigenvalue E.*

The two equations (2.45) and (2.46) will allow us to relate eigenvalues and eigenvectors of $\mathbf{h}_\ell$ for different values of the angular-momentum quantum number ℓ.

2.4.2. The ground state of the radial Schrödinger operator

First of all, we note that it is quite easy to obtain an eigenfunction of $\mathbf{h}_\ell$ for arbitrary $\ell \geq 0$ just by requiring

$$A_\ell^+ f = 0, \quad \text{that is,} \quad -\frac{df(r)}{dr} + \frac{\ell+1}{r} f(r) = \frac{1}{\ell+1} f(r). \tag{2.47}$$

This differential equation of first order has the solution

$$f_{\ell;0}(r) = r^{\ell+1} \, \mathrm{e}^{-r/(\ell+1)}, \tag{2.48}$$

which is square-integrable on $0 \leq r < \infty$ (but not yet normalized). Because $A_\ell^- A_\ell^+ f_{\ell;0} = 0$, we find from Eq. (2.54) that

$$\mathbf{h}_\ell \, f_{\ell;0} = -\frac{1}{2(\ell+1)^2} \, f_{\ell;0}. \tag{2.49}$$

We note that $A_\ell^- f = 0$ has no square integrable solution for any ℓ, because any solution of this equation is proportional to $r^{-(\ell+1)} \, \mathrm{e}^{r/(\ell+1)}$. Hence, we cannot obtain eigenfunctions of $\mathbf{h}_{\ell+1}$ by solving $A_\ell^- f = 0$, as one might think after inspecting Eq. (2.46).

The radial Coulomb Hamiltonian $\mathbf{h}_\ell$ has for each ℓ an eigenvalue

$$E_{\ell;0} = -\frac{1}{2(\ell+1)^2}. \tag{2.50}$$

Moreover, $E_{\ell;0}$ is the lowest possible eigenvalue of $\mathbf{h}_\ell$.

PROOF. It remains to prove that $E_{\ell;0}$ is the lowest possible eigenvalue for a given ℓ. Let

$$\langle f, g\rangle = \int_0^\infty \overline{f(r)}\, g(r)\, dr \tag{2.51}$$

be the scalar product in $L^2([0,\infty), dr)$. Then we have

$$\langle f, A_\ell^-\, A_\ell^+ f\rangle = \langle A_\ell^+ f, A_\ell^+ f\rangle = \|A_\ell^+ f\|^2 \geq 0 \tag{2.52}$$

for all functions f that are differentiable, square-integrable, and satisfy the boundary conditions $f(r) \to 0$, as r goes to 0 or to ∞. We conclude that $A_\ell^-\, A_\ell^+$ cannot have any negative eigenvalues, because otherwise the scalar product above would be negative for the corresponding eigenfunction. Hence, $f_{\ell;0}$ already belongs to the lowest possible eigenvalue 0 of $A_\ell^-\, A_\ell^+$. Because of (2.54), $f_{\ell;0}$ belongs to the lowest possible eigenvalue of $\mathbf{h}_\ell$. □

The eigenvalue $E_{\ell;0}$ of $\mathbf{h}_\ell$ is, of course, also an eigenvalue of the three-dimensional Coulomb Hamiltonian (2.39). For the three-dimensional problem, the degree of degeneracy of this eigenvalue is at least $2\ell + 1$, because all the functions

$$\phi_{\ell,m;0}(r,\vartheta,\varphi) = \frac{1}{r}\, f_{\ell;0}(r)\, Y_\ell^m(\vartheta,\varphi), \quad m = -\ell, -\ell+1, \ldots, \ell \tag{2.53}$$

are eigenfunctions of the energy operator $\hat{H}$ in spherical coordinates (and, simultaneously, eigenfunctions of $\hat{L}^2$ and $\hat{L}_3$). This is the normal degree of degeneracy that we expect in spherically symmetric situations, because the angular-momentum quantum number m does not occur in the radial Schrödinger equation. In fact, it will turn out that the degree of degeneracy of the Coulomb eigenvalues is much higher than that.

2.4.3. Excited states of the radial Schrödinger operator

We may rewrite (2.45) and (2.46) as

$$\mathbf{h}_\ell = A_\ell^-\, A_\ell^+ + E_{\ell;0} \tag{2.54}$$

$$\mathbf{h}_\ell = A_{\ell-1}^+\, A_{\ell-1}^- + E_{\ell-1;0} \quad (\text{for } \ell \geq 1). \tag{2.55}$$

For $\ell \geq 1$ we conclude from the second identity that the ground state $f_{\ell;0}$ of $\mathbf{h}_\ell$ is also an eigenvector of the operator $A_{\ell-1}^+\, A_{\ell-1}^-$:

$$A_{\ell-1}^+\, A_{\ell-1}^-\, f_{\ell;0}(r) = (E_{\ell;0} - E_{\ell-1;0})\, f_{\ell;0}(r). \tag{2.56}$$

The corresponding eigenvalue is obviously nonzero, and therefore we can introduce the function

$$f_{\ell-1;1}(r) = A_{\ell-1}^-\, f_{\ell;0}(r) \tag{2.57}$$

which in view of the spectral supersymmetry (Exercise 2.5) is a solution of

$$A^-_{\ell-1}\, A^+_{\ell-1}\, f_{\ell-1;1}(r) = (E_{\ell;0} - E_{\ell-1;0})\, f_{\ell-1;1}(r). \tag{2.58}$$

Using (2.54) with ℓ replaced by $\ell - 1$, this can be rewritten as

$$\mathbf{h}_{\ell-1}\, f_{\ell-1;1}(r) = E_{\ell;0}\, f_{\ell-1;1}(r). \tag{2.59}$$

Thus, the function $f_{\ell-1;1}(r)$ is a solution of the radial Schrödinger equation with angular momentum $\ell - 1$. It belongs to the eigenvalue $E_{\ell-1;1} = E_{\ell;0}$ of the operator $\mathbf{h}_{\ell-1}$.

According to the result of the previous section, $\mathbf{h}_{\ell-1}$ has $E_{\ell-1;0} = -\frac{1}{2}\ell^{-2}$ as the lowest possible eigenvalue. The energy $E_{\ell-1;1} = -\frac{1}{2}(\ell+1)^{-2} > E_{\ell-1;0}$ therefore belongs to the first excited eigenstate of $\mathbf{h}_{\ell-1}$.

EXERCISE 2.6. *Show that* $\mathbf{h}_{\ell-1}$ *can have no eigenvalue between* $E_{\ell-1;0}$ *and* $E_{\ell-1;1}$*, because then* $E_{\ell;0} = E_{\ell-1;1}$ *would not be the lowest possible eigenvalue of* $\mathbf{h}_\ell$*.*

The next step is to note the following result:

> Assume $\ell \geq 1$. Whenever $\mathbf{h}_\ell$ has an eigenvector f belonging to an eigenvalue $E > E_{\ell;0}$, then $\mathbf{h}_{\ell-1}$ has the same eigenvalue E and the corresponding eigenvector of $\mathbf{h}_{\ell-1}$ is given by $A^-_{\ell-1} f$.

PROOF. Eq. (2.55) implies that $E - E_{\ell-1;0}$ is an eigenvalue of $A^+_{\ell-1}A^-_{\ell-1}$. It follows from spectral supersymmetry that $A^-_{\ell-1}A^+_{\ell-1}$ has the same eigenvalue and that the corresponding eigenvector is $A^-_{\ell-1} f$. Substituting $\ell - 1$ for ℓ in (2.54) shows $A^-_{\ell-1}A^+_{\ell-1} = \mathbf{h}_{\ell-1} - E_{\ell-1;0}$, and this operator has the eigenvalue $E - E_{\ell-1;0}$. Hence, the operator $\mathbf{h}_{\ell-1}$ has the eigenvalue E. □

As long as $\ell - 1 \geq 1$, we can thus iterate the process that led to the first excited level of $\mathbf{h}_{\ell-1}$. It follows that the second eigenvalue of $\mathbf{h}_{\ell-2}$ is

$$E_{\ell-2;2} = E_{\ell_1;1} = E_{\ell;0} = -\frac{1}{2}(\ell+1)^{-2} \quad \text{for } \ell \geq 2. \tag{2.60}$$

The corresponding eigenvector is obtained as

$$f_{\ell;2} = A^-_{\ell-2}\, f_{\ell;1} = A^-_{\ell-2}\, A^-_{\ell-1}\, f_{\ell;0} \tag{2.61}$$

from the ground state $f_{\ell;0}$ of $\mathbf{h}_\ell$.

Now it is also clear that we can obtain all eigenvectors and eigenvalues of the radial Hamiltonian with fixed index ℓ from the ground states of the radial Hamiltonian with some higher index. For example, the function $f_{\ell;2}$,

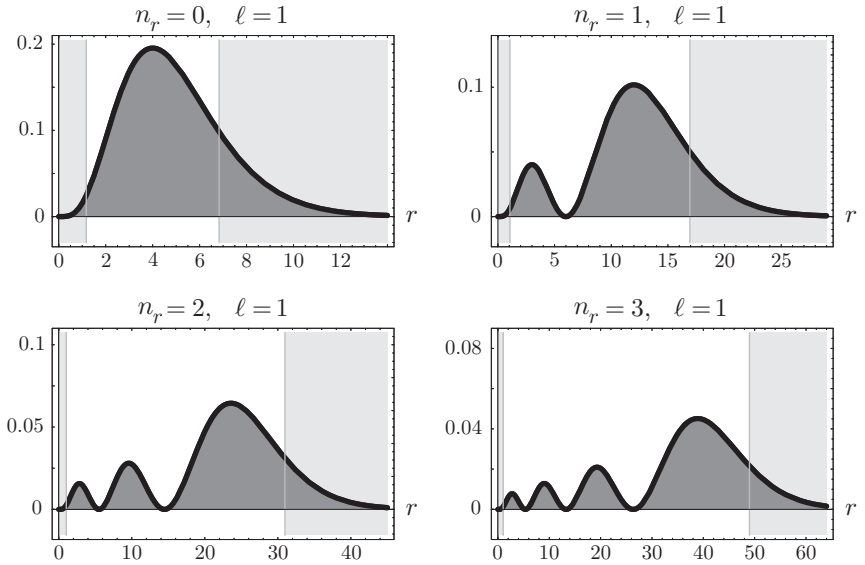


FIGURE 2.3. Radial position probability densities $|f_{\ell;n_r}(r)|^2$ for various values of n_r and $\ell = 1$. The gray regions are the classically forbidden regions.

which belongs to the eigenvalue $E_{\ell,2} = -\frac{1}{2}(\ell+3)^{-2}$, is obtained from the ground state of $\mathbf{h}_{\ell+2}$:

$$f_{\ell+2;0} = r^{\ell+3}\,\mathrm{e}^{-r/(\ell+3)},$$

by applying first the differential operator $A^-_{\ell+1}$ and then A^-_ℓ. As an example, Figure 2.3 shows the first four eigenvectors for $\ell = 1$. It is very remarkable that in order to obtain these functions, it is not necessary to solve a differential equation, except a very simple first-order equation for $f_{\ell;0}$. We would also like to emphasize the following facts:

(1) The eigenvalues of $\mathbf{h}_\ell$ except $E_{\ell;0}$ also occur as the ground state energy of another radial Schrödinger operator with higher angular momentum.
(2) All eigenvalues of all the radial Schrödinger operators with $\ell > 0$ are already contained in the eigenvalue spectrum of the radial Schrödinger operator with $\ell = 0$.
(3) The lowest possible energy of the Coulomb system is $E_{0;0} = -1/2$. The corresponding eigenstate has $\ell = 0$.

We collect our results in the following box:

Eigenvalues of the radial Coulomb Hamiltonian:

For each ℓ, the radial Coulomb Hamiltonian $\mathbf{h}_\ell$ has infinitely many eigenvalues

$$E_{\ell;n_r} = -\frac{1}{2(n_r+\ell+1)^2} \qquad n_r, \ell = 0, 1, 2, \dots . \tag{2.62}$$

For any $n_r \geq 1$ the corresponding eigenvectors $f_{\ell;n_r}$ of $\mathbf{h}_\ell$ can be obtained by a successive application of the operators A_ℓ^- defined in (2.42) from the ground state $f_{\ell+n_r;0}$ of the radial Hamiltonian $\mathbf{h}_{\ell+n_r}$:

$$f_{\ell;n_r}(r) = A_\ell^- \, A_{\ell+1}^- \cdots A_{\ell+n_r-1}^- \, f_{\ell+n_r-1;0}(r). \tag{2.63}$$

In order to obtain the explicit r-dependence of the radial eigenfunctions, one has to compute all the differentiations in (2.63). But still, the eigenfunctions are not normalized. Hence, one has to find a suitable constant $c_{\ell;n_r}$ from the condition

$$\int_0^\infty |c_{\ell;n_r} \, f_{\ell;n_r}(r)|^2 \, dr = 1. \tag{2.64}$$

We omit the somewhat tedious calculation of $c_{\ell;n_r}$, because in (2.103) and (2.104) below we are going to derive an explicit formula for the normalized eigenfunctions with another method. In Figures 2.3 and 2.4, we show a few plots of the radial position probability densities $|f_{\ell;n_r}(r)|^2$ obtained from the normalized radial eigenfunctions.

EXERCISE 2.7. *Find an explicit expression for $f_{\ell;2}$ according to* (2.63).

EXERCISE 2.8. *The function $\phi_{0,0;0}$ defined by* (2.53) *is the eigenfunction with the lowest possible energy $E_{0;0} = -1/2$ (the ground-state energy). Find a constant c such that $\int_0^\infty |c f_{0;0}(r)|^2 \, dr = 1$. As a consequence, show that $c\,\phi_{0,0;0}$ has norm 1 in $L^2(\mathbb{R}^3)$.*

2.4.4. Quantum numbers of the Coulomb problem

The normalized eigenfunctions of the three-dimensional Coulomb Hamiltonian in spherical coordinates have the form

$$\frac{1}{r} \, c_{\ell;n_r} \, f_{\ell;n_r}(r) \, Y_\ell^m(\vartheta, \varphi). \tag{2.65}$$

But while the eigenfunctions of the radial Schrödinger equations depend on n_r and on ℓ, we found that the eigenvalues depend only on the sum $n_r + \ell$. Therefore, it is natural to introduce the *principal quantum number*

$$n = n_r + \ell + 1, \tag{2.66}$$

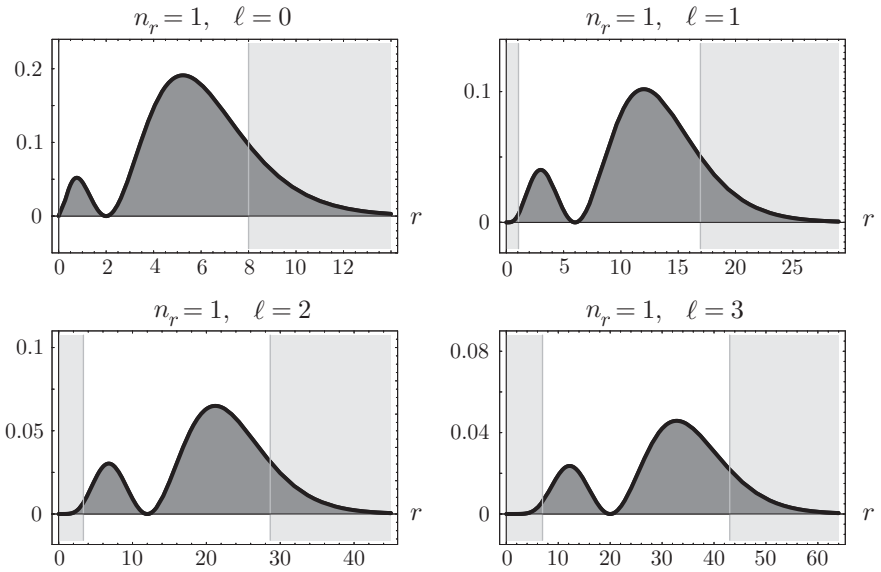


FIGURE 2.4. Radial position probability densities $|f_{\ell;n_r}(r)|^2$ for $n_r = 1$ and various values of ℓ.

which labels the possible bound-state energies in accordance with Eq. (2.37),

$$E_n = -\frac{1}{2n^2}, \qquad n = n_r + \ell + 1 = 1, 2, 3, \ldots. \tag{2.67}$$

The quantum number n_r is called the *radial quantum number*. Figure 2.5 shows the Coulomb spectrum in a similar way as Figure 1.15 shows the spectrum of the vibrating rotator.

The principal quantum number should also be used to label the Coulomb eigenfunctions. Thus, we define the normalized radial eigenfunctions

$$f_{n,\ell}(r) = c_{\ell;n_r}\, f_{\ell;n_r}(r) \quad \text{with } n_r = n - \ell - 1, \tag{2.68}$$

and write the Coulomb eigenfunction in Cartesian coordinates as

$$\psi_{n,\ell,m}(\mathbf{x}) = \frac{1}{r}\, f_{n,\ell}(r)\, Y_\ell^m(\vartheta, \varphi), \tag{2.69}$$

As always, (r, ϑ, φ) are the spherical coordinates of $\mathbf{x}$.

The $\psi_{n,\ell,m}$ are the simultaneous eigenstates of the Coulomb Hamiltonian H and of the angular-momentum operators L^2 and L_3. The corresponding eigenvalues are E_n, $\ell(\ell+1)$, and m.

The radial quantum number $n_r = n - \ell - 1$ counts the number of nodes (zeros) of the radial part $f_{n,\ell}(r)$ of the wave function. It is a quite general

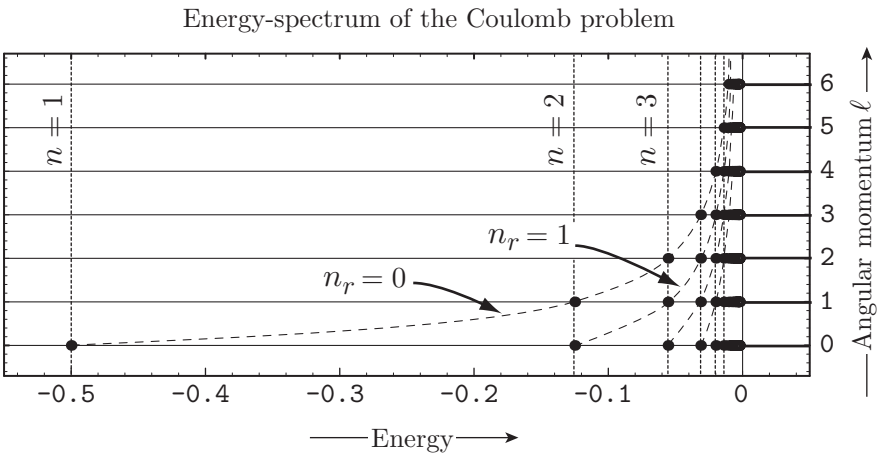


FIGURE 2.5. The structure of the eigenvalue spectrum of the Coulomb problem (in dimensionless units). Eigenstates belonging to different angular quantum numbers have the same energy.

phenomenon that the n-th eigenstate of a radial Schrödinger operator has just n zeros (nodes).

The Coulomb eigenvalues are highly degenerate:

- We have the m-degeneracy according to spherical symmetry (as described in Section 1.10.1): This means that in each angular-momentum subspace with quantum number ℓ, we can find $2\ell+1$ orthogonal eigenstates with the same energy. These eigenstates belong to different eigenvalues of L_3.
- Moreover, we have a certain ℓ-degeneracy according to the "higher symmetry" of the Coulomb problem: For a given n, we find eigenstates with the same energy in every angular-momentum subspace with $\ell \leq n-1$. These eigenstates belong to different eigenvalues of L^2.

The ℓ-degeneracy is caused by the spectral supersymmetry described in Section 2.4.1. It implies that the n radial eigenfunctions $f_{\ell;n_r}$ with quantum numbers

$$(\ell; n_r) = (0, n-1), (1; n-2), \ldots, (n-1; 0) \tag{2.70}$$

all have the same energy (although they are solutions of different radial Schrödinger equations). The total degree of degeneracy of the eigenvalue E_n

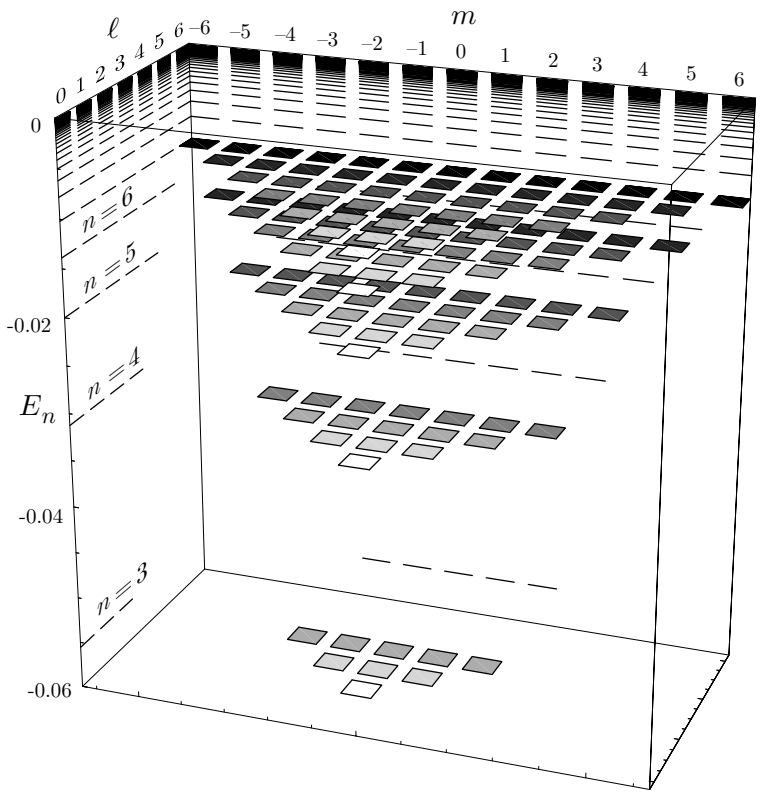
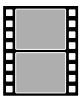


FIGURE 2.6. A part of the energy spectrum of the hydrogen atom showing the high degeneracy. The little squares represent the states with principal quantum numbers n between 3 and 7. For each n there are n^2 states with the same energy. They can be distinguished by the quantum numbers ℓ and m.

is therefore given by

$$\sum_{\ell=0}^{n-1}(2\ell+1) = n^2. \tag{2.71}$$

This structure of the energy spectrum is depicted in Figure 2.6.



CD 2.3.1 is similar to Figure 2.6 and allows one to adjust the coupling constant γ, that is, the strength of the attractive Coulomb potential. CD 2.3.3 presents an analogous image for the two-dimensional Coulomb problem.



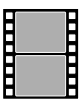
The Coulomb eigenfunctions are complex wave functions defined on the three-dimensional space. In CD 2.12–2.14, we present some examples of eigenstates $\psi_{n,\ell,m}$ and discuss possible visualization methods using colored isosurfaces or slice planes. We explain in detail how to recognize the quantum numbers n_r, ℓ, and m in the images.

According to the traditional terminology, which has its origin in the language of spectroscopists, one denotes the energy levels and angular-momentum quantum numbers by letters. The states with orbital angular momentum $\ell = 0$ are called s-waves (s = sharp), $\ell = 1$ are p-waves (p = principal), $\ell = 2$ are d-waves (d = diffuse), and then alphabetically f, g, h, and so forth, for $\ell = 3, 4, 5, \ldots$. The spectroscopic notation, which is still found in many books, denotes the states with quantum numbers n and ℓ by nx with $x = s, p, d, \ldots$ according to the angular momentum. Hence, the ground state is $1s$ and the first excited states are $2s$ and $2p$. There are three states $2p$ with $m = -1, 0, 1$, and all of them have the same energy as $2s$. Very often, one attaches the quantum number m as a subscript and writes, for example, $2p_1$ for a $2p$ electron with $m = 1$.

The states with the same principal quantum number n are said to belong to the same energy shell. Hence, the n-th energy shell is just the n^2-dimensional eigenspace of the Coulomb Hamiltonian belonging to the eigenvalue E_n. The energy shells are denoted by the letters $K, L, M, \ldots$ for $n = 1, 2, 3, \ldots$.

Figure 2.7 shows the position probability densities $|\psi_{n,\ell,m}(\mathbf{x})|^2$ of a few Coulomb eigenfunctions. Each plot shows a square region in the xz-plane with the origin at the center. The bounding box of each image has a side length of 50 in dimensionless units. The position probability densities of all eigenfunctions are symmetric with respect to rotations about the z-axis.

The three-dimensional shape of some Coulomb eigenfunctions is shown in Figure 2.8. All these images belong to the quantum numbers $n = 3$ and $\ell = 2$. The first row shows isosurfaces of the position probability density $|\psi_{n,\ell,m}|^2$ with (a) $m = 0$, (b) $m = \pm 1$, and (c) $m = \pm 2$. The remaining images in Figure 2.8 show *real orbitals* similar to (1.195) and (1.196). The real orbitals are linear combinations of $\psi_{n,\ell,m}$ and $\psi_{n,\ell,-m}$ and are proportional to the real and imaginary parts of $\psi_{n,\ell,m}$. They are eigenfunctions of H and L^2, but not of L_3. Here (d) and (e) belong to $|m| = 1$ whereas (f) and (g) belong to $|m| = 2$.



CD 2.15 is a gallery of the first 20 eigenfunctions and the associated real orbitals. The visualizations show density plots as well as phase-colored isosurfaces. We also mention that the first 49 eigenfunctions of the two-dimensional Coulomb problem are presented in CD 2.5.

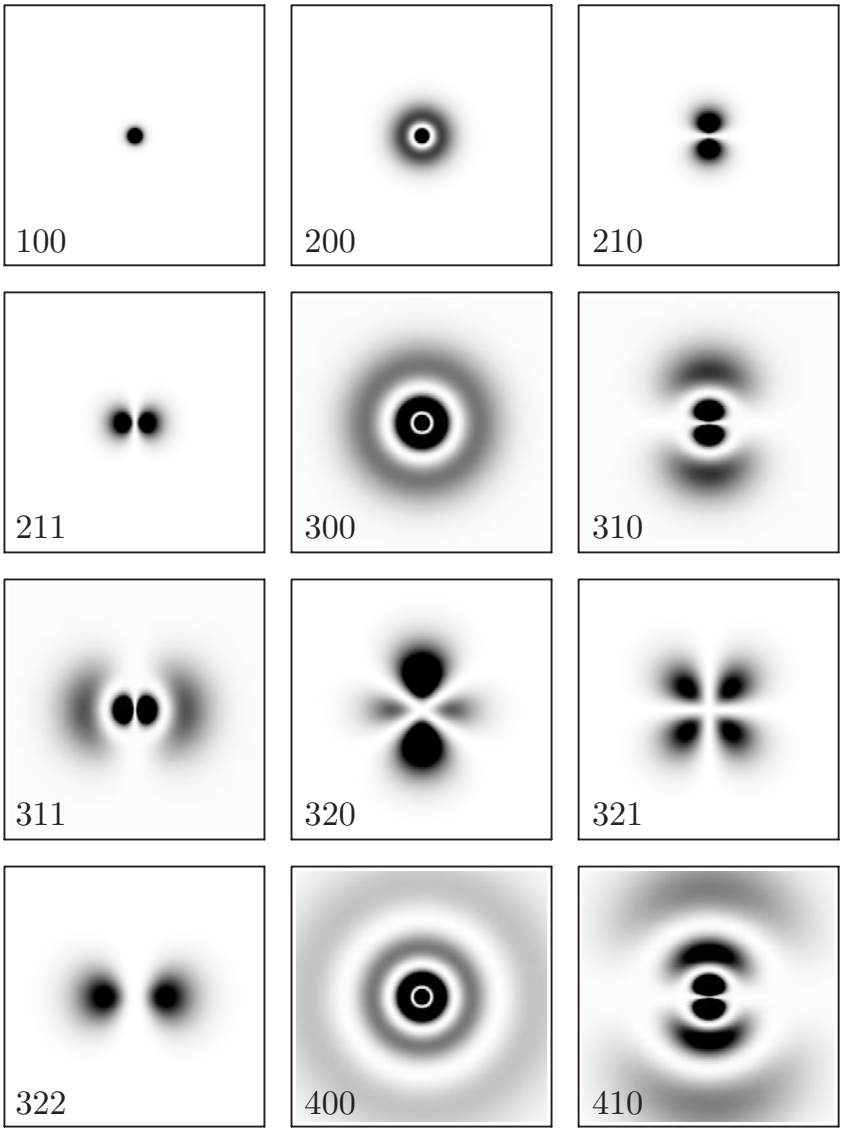


FIGURE 2.7. Position probability densities of Coulomb eigenfunctions with the quantum numbers $n\ell m$ as indicated.

In a real hydrogen atom, the ℓ-degeneracy is partially removed by relativistic corrections. An energy level also splits into different levels when perturbations like electric or magnetic fields are applied. For example, a magnetic field removes the m-degeneracy (the independence of the energy

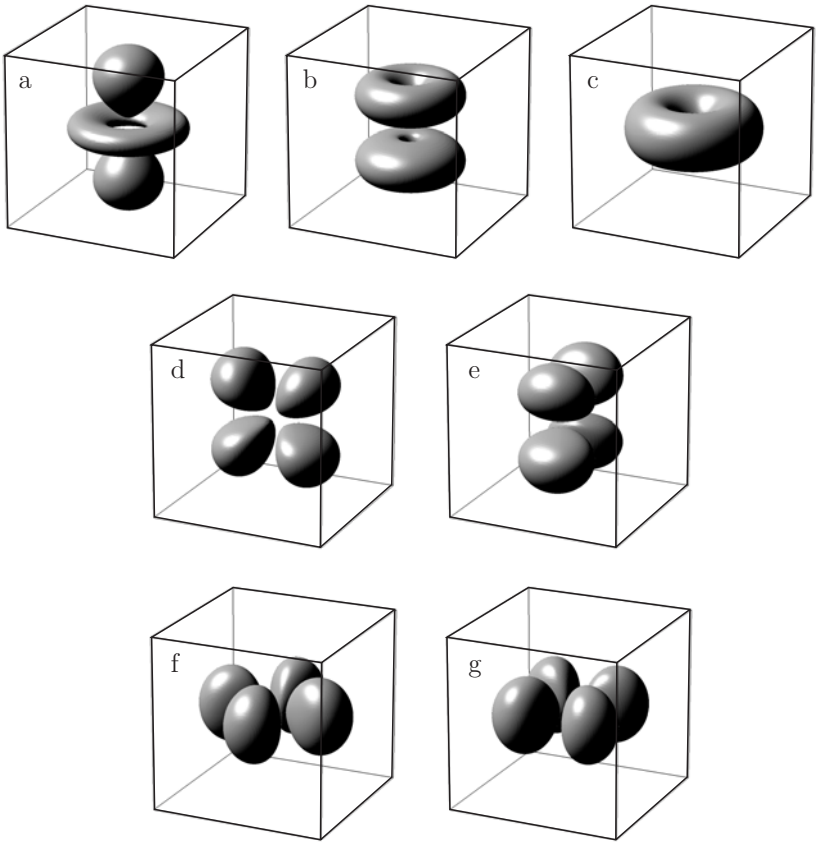
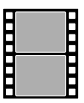


FIGURE 2.8. Isosurfaces of the position probability density show the geometry of some eigenfunctions of the Coulomb problem with $n = 3$, $\ell = 2$ and (a) $m = 0$, (b) $m = \pm 1$, and (c) $m = \pm 2$. The remaining images are isosurfaces of real orbitals.

on the eigenvalue m of L_3). Therefore, the quantum number m is often called the *magnetic quantum number*.

2.4.5. Time evolution of simple superpositions



The time evolution of an eigenstate of H is described by the phase factor $\exp(-\mathrm{i}Et)$, where E is the energy of the eigenstate. We show a few examples in CD 2.6, CD 2.7, and CD 2.16 (in three dimensions). In the following, we deal with more complicated forms of the time evolution, as shown for example in CD 2.8.1 and CD 2.9.

Any superposition of two or more states with different energies shows a nontrivial time-dependence. In this section, we consider the superposition of just two eigenfunctions. In Section 2.8 below, we investigate special superpositions of many eigenstates showing a quasiclassical behavior.

A superposition of two eigenfunctions with energies E_1 and E_2 depends on time according to

$$\begin{aligned}\psi(r,\vartheta,\varphi,t) = c_1\,\psi_{n_1,\ell_1,m_1}(r,\vartheta,\varphi)\,\exp(-\mathrm{i}E_{n_1}t)\\ + c_2\,\psi_{n_2,\ell_2,m_2}(r,\vartheta,\varphi)\,\exp(-\mathrm{i}E_{n_2}t).\end{aligned} \tag{2.72}$$

If all states are normalized, then $|c_1|^2$ is the probability that in a measurement of the energy of an electron in the state ψ, the value E_1 would be obtained, and $|c_2|^2$ is the probability for E_2. Using the formula (2.69) for $\psi_{n,\ell,m}$ together with expression (1.106) for Y_ℓ^m, we find that for $m_1 \neq m_2$, the superposition (2.72) has the form

$$\begin{aligned}\psi(r,\vartheta,\varphi,t) &= g_1(r,\vartheta)\,\mathrm{e}^{\mathrm{i}m_1\varphi}\,\mathrm{e}^{-\mathrm{i}E_1t} + g_2(r,\vartheta)\,\mathrm{e}^{\mathrm{i}m_2\varphi}\,\mathrm{e}^{-\mathrm{i}E_2t}\\ &= \mathrm{e}^{\mathrm{i}\lambda(\varphi,t)}\Big(g_1 + g_2\,\mathrm{e}^{\mathrm{i}(m_2-m_1)(\varphi-\omega t)}\Big),\end{aligned} \tag{2.73}$$

with

$$\lambda(\varphi,t) = m_1\,\varphi - E_1\,t, \qquad \omega = \frac{E_2-E_1}{m_2-m_1}. \tag{2.74}$$

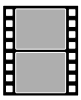
Let us denote the position probability density of the initial state by

$$\rho(r,\vartheta,\varphi) = |\,g_1(r,\vartheta) + g_2(r,\vartheta)\,\mathrm{e}^{\mathrm{i}(m_2-m_1)\varphi}\,|^2. \tag{2.75}$$

According to (2.73), the time evolution of this quantity is given by

$$|\psi(r,\vartheta,\varphi,t)|^2 = \rho(r,\vartheta,\varphi-\omega t). \tag{2.76}$$

We find that the time evolution of the position distribution is just a rotation about the z-axis with a constant angular speed ω.



In a superposition of two eigenfunctions, a radial oscillation can only occur for $m_1 = m_2$. CD 2.8 shows an example in two dimensions. In three dimensions, CD 2.17.2 and CD 2.17.4 show pure radial oscillations within a single angular-momentum subspace ($\ell_1 = \ell_2$). CD 2.17.1 and CD 2.17.3 show additional oscillations in the direction of the polar angle ϑ, because $\ell_1 \neq \ell_2$.



CD 2.18 and CD 2.19 present superpositions of two states with different magnetic quantum numbers $m_1 \neq m_2$. In that case, there are no radial oscillations nor ϑ-oscillations. The isosurface of the position probability rotates like a rigid body about the z-axis. The phases and hence the colors depend on time with $\exp(\mathrm{i}\lambda(\varphi,t))$ according to Eq. (2.73), but the shape remains constant. CD 2.9 shows rotating states in two dimensions.

2.5. Direct Solution of the Radial Schrödinger Equation

2.5.1. Solution of the Coulomb problem in two dimensions

In this section, we shall outline the conventional approach to the Coulomb problem by solving the radial Schrödinger equation directly. We do this here for two space dimensions, because many of the visualizations on the CD show two-dimensional systems. The radial equation arising from the three-dimensional problem will be treated along similar lines in the next section.

In classical physics, the Coulomb problem in two dimensions is equivalent to the Coulomb problem in three dimensions, because the motion takes place in a plane anyway. In quantum mechanics, however, the energy levels of the two-dimensional Coulomb problem are are always below the corresponding energies in three dimensions. A Coulomb potential binds stronger in two-dimensions. Moreover, two-dimensional systems have a lower degree of m-degeneracy.

We look for a solution of the eigenvalue equation $H\psi = E\psi$ by separating the angular variable φ from the radial variable r,

$$\phi(r,\varphi) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\mathrm{i}m\varphi}\,g(r). \tag{2.77}$$

Here, $m = \pm\ell$ with $\ell = 0, 1, 2, 3, \ldots$. The quantum number m is the eigenvalue of the angular-momentum operator L. In two dimensions, L is a scalar quantity. It may be interpreted as the third component of an angular-momentum vector $\mathbf{L} = (0, 0, L)$ for a particle whose motion is confined to the xy-plane. In polar coordinates and with $\hbar = 1$, we have

$$L = -\mathrm{i}\,\frac{\partial}{\partial\varphi}. \tag{2.78}$$

This is the same expression as for the operator L_3 in three dimensions, and hence we obtain the same result for the eigenvalues. The function $\mathrm{e}^{\mathrm{i}m\varphi}/\sqrt{2\pi}$ in (2.77) is the normalized eigenfunction of L belonging to the eigenvalue m.

The quantum number $\ell = |m|$ is the eigenvalue of the operator $|L|$. Hence, in two dimensions, the square L^2 of the angular momentum has the eigenvalues ℓ^2 (unlike $\ell(\ell+1)$ in three dimensions).

The separation of variables leads to the radial Schrödinger equation for a two-dimensional system. This equation has been derived already in Book One (Eq. (8.145) in Section 8.8). For the Coulomb potential $V(r) = -1/r$ we obtain

$$\frac{1}{2}\left(-\frac{d^2}{dr^2} - \frac{1}{r}\,\frac{d}{dr} + \frac{\ell^2}{r^2}\right) g - \frac{1}{r}\,g = E\,g. \tag{2.79}$$

We have $\ell^2 = m^2$, and therefore the radial equation does not depend on the sign of m. The radial equation has two linearly independent solutions on $(0, \infty)$ that can easily be found using a computer algebra system like Mathematica. Only one of these solutions remains finite for $r \to 0$ (it is called the *regular solution*). The other solution is unbounded for $r \to 0$. In the context of quantum mechanics, we are only interested in the solution that is regular at 0.



It has been pointed out in Section 1.10.1 that the reason for choosing the regular solution has little to do with the requirement of square-integrability. Indeed, for $\ell = 0$, the singularity of the irregular solution is so mild that it is also square-integrable. In this case, the square-integrability alone gives no sufficient criterion to select one of the two linearly independent solutions. The choice of the regular solution comes from the observation that any eigenfunction of the Coulomb Hamiltonian H must be in the domain of H. It can be shown that the domain of this second-order differential operator contains only continuous functions. Any singularity of the radial function $g(r)$ at $r = 0$ leads to a discontinuity of ψ and hence ψ would not be in the domain of H.

For all E, the radial equation has the regular solution

$$g(r) = N\, \mathrm{e}^{-\sqrt{-2E}\, r}\, r^{\ell}\, {}_1F_1\big(\ell + \tfrac{1}{2} - \tfrac{1}{\sqrt{-2E}}\,,\, 2\ell + 1\,,\, 2\sqrt{-2E}\, r\big). \tag{2.80}$$

Here, ${}_1F_1$ is the confluent hypergeometric function, and N is an arbitrary constant. We are going to choose N to normalize the radial wave function (see (2.93) below).

The properties of this well-known special function are described in great detail, for example, in [**1**]. Here, we just quote the definition:

$${}_1F_1(a, b, r) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n}\, \frac{r^n}{n!}, \tag{2.81}$$

where the expressions

$$(a)_0 = 1, \quad (a)_n = a\,(a+1)\,\cdots\,(a+n-1) \tag{2.82}$$

are sometimes called the Pochhammer symbols.[2]

EXERCISE 2.9. *Evaluate* ${}_1F_1(0, b, r)$ *and* ${}_1F_1(a, a, r)$. *Examine the definition of* ${}_1F_1(a, b, r)$ *in the case that* b *is a negative integer.*

[2]Our definition of the confluent hypergeometric function agrees with the definition of the *Mathematica* function `Hypergeometric1F1`$[a, b, r]$. The Pochhammer symbols are implemented by `Pochhammer`$[a, n]$

EXERCISE 2.10. *With the help of the definition* (2.81)*, show that the function* ${}_1F_1(a,b,r)$ *is a polynomial in* r *of degree* n*, whenever* $a=-n$ *is a negative integer or zero (and provided that* b *is not a negative integer).*

In a sense, the hypergeometric function can be seen as a generalization of the exponential function $\mathrm{e}^r = {}_1F_1(a,a,r)$. Whenever a and b are not equal to negative integers, the power series above is convergent (even for complex r). However, in these cases the function diverges, as $r\to\infty$. So the only hope that $f(r)$ is square-integrable at infinity is that the first argument in the hypergeometric function is a negative integer or zero. In this case, the hypergeometric function becomes a polynomial in r and the exponential factor $\exp(-\sqrt{-2E}\,r)$ lets the function $f(r)$ go to zero very fast, as r goes to infinity. This is, however, only true if $E<0$, otherwise the exponential factor would just oscillate. Hence, we have to find numbers $E_{\ell;n_r}<0$ such that

$$\ell+\frac{1}{2}-\frac{1}{\sqrt{-2E_{\ell;n_r}}}=-n_r,\quad \text{with } n_r=0,1,2,\dots\,. \tag{2.83}$$

It is easy to solve this equation for $E_{\ell;n_r}$,

$$E_{\ell;n_r}=-\frac{2}{(2n_r+2\ell+1)^2},\qquad n_r,\ell=0,1,2,\dots\,. \tag{2.84}$$

The eigenvalues $E_{\ell;n_r}$ only depend on the sum $n_r+\ell$. Hence, it makes sense to introduce the principal quantum number $n=n_r+\ell+1$ also in the two-dimensional Coulomb problem. n can take the values $1,2,3,\dots$. Then one can write

$$E_{\ell;n_r}=E_n=-\frac{1}{2\,(n-1/2)^2}. \tag{2.85}$$

Inserting these numbers into the expression (2.80) for $g(r)$, we find for each $n_r=n-\ell-1$ a square-integrable solution of the radial Schrödinger equation with angular-momentum quantum number ℓ. With the abbreviation

$$\kappa_n=\sqrt{-2E_n}=\frac{1}{n-1/2} \tag{2.86}$$

we write these solutions as

$$g_{n,\ell}(r)=N_{n,\ell}\;\mathrm{e}^{-\kappa_n r}\,(2\,\kappa_n\,r)^\ell\;{}_1F_1\big(\ell+1-n,2\ell+1,2\,\kappa_n\,r\big). \tag{2.87}$$

The factor $(2\,\kappa_n)^\ell$ has been extracted from the normalization constant $N_{n,\ell}$ for cosmetic reasons.

The set of eigenfunctions of the two-dimensional Coulomb Hamiltonian H (in polar coordinates) is thus given by

$$\phi_{n,\ell,m}(r,\varphi)=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\mathrm{i}m\varphi}\,g_{n,\ell}(r), \tag{2.88}$$

with $\ell = |m|$, $m = 0, \pm 1, \pm 2, \ldots$, and $n = 1, 2, \ldots$. Because of the simple connection between m and ℓ, it is sufficient to label the eigenfunction with two indices, that is, we define

$$\phi_{n,m}(r, \varphi) = \phi_{n,|m|,m}(r, \varphi). \tag{2.89}$$

For a given principal quantum number n, the eigenfunctions $\phi_{n,m}$ with

$$m = \pm(n-1), \pm(n-2), \ldots \pm 1, 0$$

all have the same energy. Therefore, the degree of degeneracy of the eigenvalue E_n of H is only $2n-1$, instead of n^2 in three dimensions. In two dimensions, the m-degeneracy is lower, because L_3 can only have two different values for a given ℓ (instead of $2\ell + 1$ in three dimensions).



CD 2.2.3 is an interactive image presenting the energy spectrum of the two-dimensional Coulomb problem and its dependence on the strength of the Coulomb force. The radial eigenfunctions are shown in CD 2.3, and the first 49 eigenfunctions $\psi_{n,m}(r, \varphi)$ defined in (2.89) are visualized in CD 2.5. CD 2.4 compares radial oscillations in two and three dimensions.

We still have to determine the constants $N_{n,\ell}$ in (2.87). We require that the eigenfunctions all have norm 1. This is needed, in particular, for the eigenfunction expansion. For the norm of $\phi_{n,m}(r, \phi)$ we find

$$\|\phi_{n,m}\|^2 = \frac{1}{2\pi} \int_0^\infty \Big(\int_0^{2\pi} |g_{n,\ell}(r)\, \mathrm{e}^{\mathrm{i}m\varphi}|^2 \, d\varphi \Big) r \, dr \tag{2.90}$$

$$= \int_0^r r \, |g_{n,\ell}(r)|^2 \, dr \quad (\text{with } \ell = |m|). \tag{2.91}$$

Hence, we may determine the constant $N_{n,\ell}$ from the condition that

$$\int_0^r r \, |g_{n,\ell}(r)|^2 \, dr = 1. \tag{2.92}$$

A long and tedious calculation gives

$$N_{n,\ell} = \frac{1}{(2\ell)!} \left(2\, \kappa_n^3 \, \frac{(n+\ell-1)!}{(n-\ell-1)!} \right)^{1/2}. \tag{2.93}$$

Quite frequently, the solution is expressed in terms of generalized Laguerre polynomials. The generalized Laguerre polynomials can be defined as[3]

$$\mathcal{L}_n^{(m)}(x) = \frac{(n+m)!}{n!\, m!} \, {}_1F_1(-n, m+1, x). \tag{2.94}$$

[3]The generalized Laguerre polynomials are implemented by *Mathematica* as `LaguerreL`$[n, m, x]$. They have to be distinguished from the associated Laguerre polynomials $L_n^m(x) = (-1)^m \, \mathcal{L}_{n-m}^{(m)}(x)$, which are also used frequently.

Hence, the solution of the radial Coulomb problem can also be written as

$$g_{n,\ell}(r) = K_{n,\ell}\, \mathrm{e}^{-\kappa_n r}\, (2\kappa_n\, r)^\ell\, \mathcal{L}^{(2\ell)}_{n-\ell-1}\big(2\kappa_n\, r\big), \tag{2.95}$$

where we used the abbreviation

$$K_{n,\ell} = \Big(2\,\kappa_n^3\, \frac{(n-\ell-1)!}{(n+\ell-1)!}\Big)^{1/2}. \tag{2.96}$$

We want to emphasize that the set of eigenfunctions of H does not form a complete orthonormal basis in $L^2(\mathbb{R}^2)$. The eigenfunctions only span the subspace of bound states. Hence, for a given square-integrable function ψ, the expression

$$\psi_{\text{bound}} = \sum_{n,m} \langle \psi_{n,m}, \psi \rangle\, \psi_{n,m} \tag{2.97}$$

does not represent ψ (as it would be the case, for example, for the harmonic oscillator). Instead, the sum above gives the part of ψ which is in the subspace of bound states. The mapping $\psi \to \psi_{\text{bound}}$ is an orthogonal projection operator.



CD 2.11 shows the effect of projecting Gaussian wave packets onto the subspace of bound states. The time evolution leaves the subspace of bound states invariant. In the absence of perturbations, a bound state remains a bound state forever. This follows from the conservation of energy (see also Eq. (2.199) below).

EXERCISE 2.11. *Can you find a transformation $g(r) \to f(r)$ such that $f(r)$ satisfies the radial equation*

$$\frac{1}{2}\Big(-\frac{d^2 f}{dr^2} + \frac{\ell^2}{r^2}\, f\Big) - \frac{1}{r}\, f = E\, f\ ? \tag{2.98}$$

2.5.2. Solution of the Coulomb problem in three dimensions

It is worthwhile to list the formulas corresponding to the results in the previous section also for the three-dimensional case. For a particle in a Coulomb field in three dimensions, the radial Schrödinger equation becomes

$$\frac{1}{2}\Big(-\frac{d^2 f}{dr^2} + \frac{\ell(\ell+1)}{r^2}\, f\Big) - \frac{1}{r}\, f = E\, f. \tag{2.99}$$

We remind the reader that $\ell(\ell+1)$ is the eigenvalue of the operator L^2. The angular-momentum quantum number ℓ is a non-negative integer, $\ell = 0, 1, 2, \ldots$. The radial equation does not depend on the eigenvalue m of L_3, which for a given ℓ can have all integer values between $-\ell$ and $+\ell$.

The radial Schrödinger equation (2.99) has the regular solution

$$f(r) = N\,\mathrm{e}^{-\sqrt{-2E}\,r}\,r^{\ell+1}\,{}_1F_1\big(\ell - \tfrac{1}{\sqrt{-2E}}\,,\,2\ell+2\,,\,2\sqrt{-2E}\,r\big). \tag{2.100}$$

For negative E, this solution becomes square-integrable if the first argument of the confluent hypergeometric function is a non-positive integer, denoted by $-n_r$, where n_r is the radial quantum number. The first argument of ${}_1F_1$ equals $-n_r$ whenever E is equal to

$$E_{\ell;n_r} = -\frac{1}{2(n_r+\ell+1)^2}, \qquad n_r = 0,1,2,\ldots. \tag{2.101}$$

In terms of the principal quantum number $n = n_r + \ell + 1$, we obtain the familiar formula (2.67) for E_n. Inserting these values into (2.100) and writing

$$\sqrt{-2E_n} = \frac{1}{n} \tag{2.102}$$

gives the following family of solutions:

$$f_{n,\ell}(r) = N_{n,\ell}\;\mathrm{e}^{-r/n}\,(2\,r/n)^{\ell+1}\,{}_1F_1\big(\ell+1-n, 2\ell+2, 2\,r/n\big), \tag{2.103}$$

with $\ell = 0,1,2,\ldots$, and $n = 1,2,3,\ldots$. Choosing the normalization constants

$$N_{n,\ell} = \frac{1}{(2\ell+1)!\,n}\,\Big(\frac{(n+\ell)!}{(n-\ell-1)!}\Big)^{1/2} \tag{2.104}$$

guarantees that

$$\int_0^r |f_{n,\ell}(r)|^2\,dr = 1. \tag{2.105}$$

Hence, the three-dimensional solution

$$\psi_{n,\ell,m}(\mathbf{x}) = \frac{1}{r}\,f_{n,\ell}(r)\,Y_\ell^m(\vartheta,\varphi) \tag{2.106}$$

is also normalized, $\|\psi_{n,\ell,m}\| = 1$. Here, (r,ϑ,φ) are the spherical coordinates of $\mathbf{x}$.

The following box describes the normalized radial eigenfunction in terms of the generalized Laguerre polynomials.

Normalized solutions of the radial Schrödinger equation:

For each $n = 1,2,3,\ldots$ and $\ell = 0,1,\ldots n-1$, the radial Schrödinger equation (2.99) has the normalized eigenfunction

$$f_{n,\ell}(r) = \frac{1}{n}\,\Big(\frac{n_r!}{(n+\ell)!}\Big)^{1/2}\,\mathrm{e}^{-r/n}\,\Big(\frac{2\,r}{n}\Big)^{\ell+1}\,\mathcal{L}_{n_r}^{(2\ell+1)}\Big(\frac{2\,r}{n}\Big), \tag{2.107}$$

with $n_r = n-\ell-1$. It belongs to the eigenvalue $E_n = -1/(2n^2)$. Here, $\mathcal{L}_n^{(m)}(x)$ is the generalized Laguerre polynomial defined in (2.94). The function $f_{n,\ell}$ has precisely n_r zeros in $0 < r < \infty$.



CD 2.3 shows a gallery of some radial eigenfunctions according to the two- and three-dimensional Coulomb problems. CD 2.4 shows radial oscillations (the time evolution of superpositions of two radial eigenfunctions in the same angular-momentum subspace).

EXERCISE 2.12. *In order to solve the radial Schrödinger equation* (2.99)*, write the radial wave function as*

$$f(r) = r^{\ell+1}\,\mathrm{e}^{-\kappa r}\,u(r), \quad \text{with } \kappa = \sqrt{-2E}. \tag{2.108}$$

and show that the function u has to be a solution of the differential equation

$$\frac{d^2}{dr^2}\,u + 2\Big(\frac{\ell+1}{r} - \kappa\Big)\,\frac{d}{dr}\,u + 2\Big(\frac{1}{r} - \kappa\,\frac{\ell+1}{r}\Big)\,u = 0. \tag{2.109}$$

EXERCISE 2.13. *Perform a variable substitution in* (2.109)*. Set $\rho = 2\kappa r$ and define $v(\rho) = u(r)$. Show that v has to be a solution of the equation*

$$\frac{d^2}{d\rho^2}\,v + \Big(\frac{2\ell+2}{\rho} - 1\Big)\,\frac{d}{d\rho}\,v - \Big(\ell + 1 - \frac{1}{\kappa}\Big)\,\frac{1}{\rho}\,v = 0. \tag{2.110}$$

Compare this with Kummer's differential equation

$$\frac{d^2}{d\rho^2}\,v + \Big(\frac{b}{\rho} - 1\Big)\,\frac{d}{d\rho}\,v - \frac{a}{\rho}\,v = 0. \tag{2.111}$$

EXERCISE 2.14. *Find a solution to Kummer's equation* (2.111) *by setting*

$$v(\rho) = \sum_{k=0}^{\infty} c_k\,\rho^k. \tag{2.112}$$

Insert the power series into the equation and write everything as a single power series in ρ. Note that the coefficient of each ρ^k has to vanish. Use this observation to show that the following recursion relation holds for the coefficients in (2.112)*:*

$$c_{k+1}\,(b+k)(k+1) = c_k(a+k), \quad k = 0, 1, 2, \ldots. \tag{2.113}$$

EXERCISE 2.15. *Starting with a given c_0, iterate* (2.113) *in order to obtain all c_k. Show that (for $b > 0$)*

$$c_k = c_0\,\frac{a}{b}\,\frac{a+1}{b+1}\cdots\frac{a+k-1}{b+k-1}\,\frac{1}{k!}. \tag{2.114}$$

Set $c_0 = 1$ and compare the result with (2.81) *and* (2.82)*.*

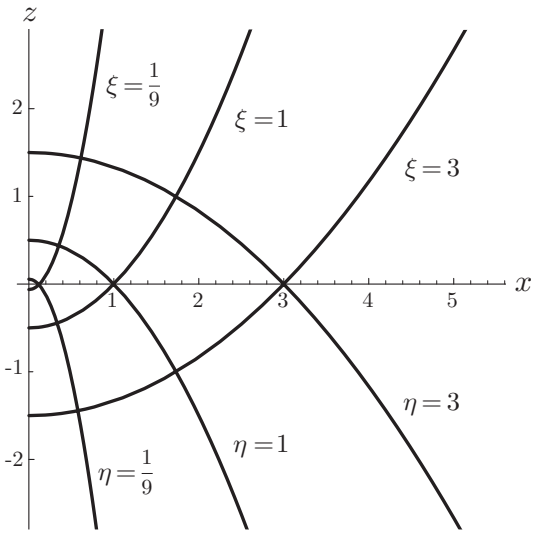


FIGURE 2.9. The coordinate curves $\xi = \text{const.}$, $\eta = \text{const.}$ of the parabolic coordinate system in the half-plane where $\varphi = 0$ are parabolas.

2.6. Special Topic: Parabolic Coordinates

The Schrödinger equation for the hydrogen atom also separates in parabolic coordinates. This observation is important because it remains true in the presence of an additional constant electric field in the z-direction (*Stark effect*).

The parabolic coordinates of a point (x, y, z) in $\mathbb{R}^3$ are given by (ξ, η, φ), where

$$\begin{aligned} \xi &= r + z = r(1 + \cos\vartheta), & x &= \sqrt{\xi\eta}\,\cos\varphi, \\ \eta &= r - z = r(1 - \cos\vartheta), & y &= \sqrt{\xi\eta}\,\sin\varphi, \\ \varphi &= \varphi, & z &= (\xi - \eta)/2. \end{aligned} \tag{2.115}$$

Here, (r, ϑ, φ) are the spherical coordinates of the point (x, y, z). The coordinates ξ and η are non-negative, and φ is the familiar azimuthal angle. The parabolic coordinate system is an orthogonal, right-handed system. Hence, the coordinate curves meet at right angles. Figure 2.9 shows a few coordinate curves in the half-plane where $\varphi = 0$ (the xz-plane). The coordinate surfaces with constant ξ or η are obtained by rotating the corresponding parabolas of Figure 2.9 about the z-axis. In Figure 2.10, we see the coordinate surfaces for $\xi = 1$ and $\eta = 1$.

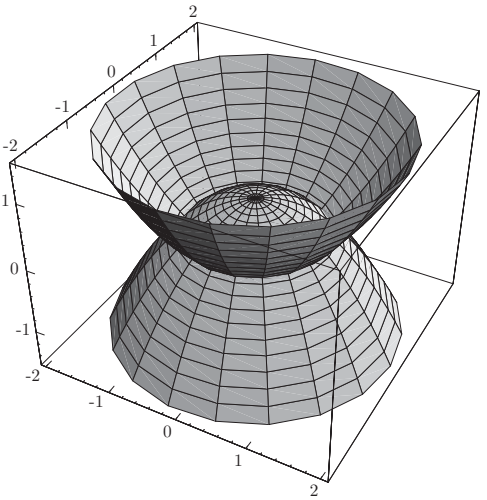


FIGURE 2.10. The surfaces with $\xi = 1$ (upper paraboloid) and $\eta = 1$ (lower paraboloid) are obtained by rotating the corresponding parabolas of Figure 2.9 about the z-axis.

The unit vectors along the coordinate lines are given by

$$\mathbf{e}_\xi = \sqrt{\frac{\eta}{\xi+\eta}}\begin{pmatrix}\cos\varphi\\ \sin\varphi\\ \sqrt{\xi/\eta}\end{pmatrix}, \quad \mathbf{e}_\eta = \sqrt{\frac{\xi}{\xi+\eta}}\begin{pmatrix}\cos\varphi\\ \sin\varphi\\ -\sqrt{\eta/\xi}\end{pmatrix}, \quad \mathbf{e}_\varphi = \begin{pmatrix}-\sin\varphi\\ \cos\varphi\\ 0\end{pmatrix}.$$

The gradient in parabolic coordinates reads

$$\nabla = 2\mathbf{e}_\xi \sqrt{\frac{\xi}{\xi+\eta}}\,\frac{\partial}{\partial\xi} + 2\mathbf{e}_\eta \sqrt{\frac{\eta}{\xi+\eta}}\,\frac{\partial}{\partial\eta} + \mathbf{e}_\varphi \frac{1}{\sqrt{\xi\eta}}\,\frac{\partial}{\partial\varphi}. \tag{2.116}$$

From this it is not difficult to obtain the expression for the Coulomb Hamiltonian (2.39) in parabolic coordinates

$$H = -\frac{2}{\xi+\eta}\left(\frac{\partial}{\partial\xi}\Big(\xi\frac{\partial}{\partial\xi}\Big) + \frac{\partial}{\partial\eta}\Big(\eta\frac{\partial}{\partial\eta}\Big)\right) - \frac{1}{2\xi\eta}\,\frac{\partial^2}{\partial\varphi^2} - \frac{2}{\xi+\eta}. \tag{2.117}$$

The stationary Schrödinger equation $(H - E)\psi = 0$ can be separated by writing

$$\psi(\xi,\eta,\varphi) = N\, f(\xi)\, g(\eta)\, \mathrm{e}^{\mathrm{i}m\varphi}, \tag{2.118}$$

where m is the magnetic quantum number. For this wavefunction, the Schrödinger equation in parabolic coordinates can be rearranged to give the equation

$$\left(-2\,\frac{\partial}{\partial\xi}\Big(\xi\frac{\partial f(\xi)}{\partial\xi}\Big)+\frac{m^2}{2\xi}\,f(\xi)-f(\xi)-E\xi\,f(\xi)\right)g(\eta)$$
$$+\left(-2\,\frac{\partial}{\partial\eta}\Big(\eta\frac{\partial g(\eta)}{\partial\eta}\Big)+\frac{m^2}{2\eta}\,g(\eta)-g(\eta)-E\eta\,g(\eta)\right)f(\xi)=0. \tag{2.119}$$

This is an expression of the form

$$v(\xi)\,g(\eta)+u(\eta)\,f(\xi)=0, \tag{2.120}$$

which is supposed to hold for all $\xi\geq 0$ and $\eta\geq 0$. This can only be the case if

$$v(\xi)=c\,f(\xi),\quad u(\eta)=-c\,g(\eta). \tag{2.121}$$

Here, c is called the separation constant. We end up with two almost identical ordinary differential equations:

$$-2\,\frac{d}{d\xi}\Big(\xi\frac{df(\xi)}{d\xi}\Big)+\frac{m^2}{2\xi}\,f(\xi)-(1+c+E\xi)\,f(\xi)=0, \tag{2.122}$$

$$-2\,\frac{d}{d\eta}\Big(\eta\frac{dg(\eta)}{d\eta}\Big)+\frac{m^2}{2\eta}\,g(\eta)-(1-c+E\eta)\,g(\eta)=0. \tag{2.123}$$

The following solutions are bounded near the coordinate origin:

$$f(\xi)=\mathrm{e}^{-\sqrt{-2E}\,\xi/2}\,\xi^{|m|/2}\,{}_1F_1\Big(\frac{1+|m|}{2}-\frac{1+c}{2\sqrt{-2E}}\,,1+|m|,\sqrt{-2E}\,\xi\Big), \tag{2.124}$$

$$g(\eta)=\mathrm{e}^{-\sqrt{-2E}\,\eta/2}\,\eta^{|m|/2}\,{}_1F_1\Big(\frac{1+|m|}{2}-\frac{1-c}{2\sqrt{-2E}}\,,1+|m|,\sqrt{-2E}\,\eta\Big). \tag{2.125}$$

The hypergeometric functions are polynomials if the first argument is an integer ≤ 0. In this case, we can expect that the solutions are square-integrable (because of the exponential damping factor). Hence, we obtain the equations (quantization conditions)

$$\frac{1+|m|}{2}-\frac{1+c}{2\sqrt{-2E}}=-n_1, \tag{2.126}$$

$$\frac{1+|m|}{2}-\frac{1-c}{2\sqrt{-2E}}=-n_2, \tag{2.127}$$

with $n_1,n_2\geq 0$. Subtracting these equations leads to

$$\frac{c}{\sqrt{-2E}}=n_1-n_2 \tag{2.128}$$

and adding leads to

$$\frac{1}{\sqrt{-2E}} = n_1 + n_2 + |m| + 1. \tag{2.129}$$

From this, we get immediately that the allowed energies are

$$E = -\frac{1}{2n^2}, \quad n = n_1 + n_2 + |m| + 1, \tag{2.130}$$

and the separation constant becomes

$$c = \frac{n_1 - n_2}{n}. \tag{2.131}$$

For given integers $n \geq 1$ and m with $|m| \leq n-1$, we see that the quantum numbers n_1 and n_2 take all values between 0 and $n_{\max} = n - |m| - 1$. There are $n - |m|$ choices of n_1 and n_2. For given n, the quantum number m can have the $2n-1$ values between $-n+1$ and $n-1$. Hence, there are

$$\sum_{m=-n+1}^{n-1} n - |m| = n^2 \tag{2.132}$$

different solutions belonging to the energy $E = -1/(2n^2)$. This is the result for the degeneracy of the eigenspace obtained earlier. Hence, the solutions found in this way span the whole corresponding eigenspace.

The solutions in parabolic coordinates are

$$\psi_{n_1,n_2,m}(\xi,\eta,\varphi) = N_{n_1,n_2,m}\, f_{n_1,m}(\xi/n)\, f_{n_2,m}(\eta/n)\, \exp(\mathrm{i}m\varphi) \tag{2.133}$$

with $n = n_1 + n_2 + |m| + 1$ denoting the principal quantum number and with

$$f_{k,m}(\zeta) = \mathrm{e}^{-\zeta/2}\, \zeta^{|m|/2}\, {}_1F_1(-k, 1+|m|, \zeta). \tag{2.134}$$

The function $f_{k,m}$ is a regular solution of the equation

$$2\,\frac{\partial}{\partial\zeta}\left(\zeta\,\frac{\partial}{\partial\zeta}\right) f(\zeta) = \frac{m^2}{2\zeta}\, f(\zeta) - \left(2k + 1 + |m| - \frac{\zeta}{2}\right) f(\zeta). \tag{2.135}$$

We choose the constant

$$N_{n_1,n_2,m} = \frac{(-1)^m}{n^2\,|m|!^2}\,\sqrt{\frac{(n_1+|m|)!\,(n_2+|m|)!}{\pi\, n_1!\, n_2!}} \tag{2.136}$$

so that the functions defined in Eq. (2.133) are normalized:

$$\int_0^\infty d\xi \int_0^\infty d\eta \int_0^{2\pi} d\varphi\, \tfrac{1}{4}\,(\xi+\eta)\,|\psi_{n_1,n_2,m}(\xi,\eta,\varphi)|^2 = 1. \tag{2.137}$$

The factor $\frac{1}{4}\,(\xi+\eta)$ (the Jacobi determinant of the coordinate transformation) describes the volume element in parabolic coordinates

$$d^3x \equiv \tfrac{1}{4}(\xi+\eta)\, d\xi\, d\eta\, d\varphi. \tag{2.138}$$

We note that the functions $\psi_{n_1,n_2,m}$ are not only eigenfunctions of the Hamiltonian, but simultaneously eigenfunctions of L_3. The operator L_3 has in parabolic coordinates the familiar form

$$L_3 = -\mathrm{i}\frac{\partial}{\partial\varphi} \tag{2.139}$$

and therefore

$$L_3\,\psi_{n_1,n_2,m} = m\,\psi_{n_1,n_2,m}. \tag{2.140}$$

The third component of the Runge-Lenz vector in parabolic coordinates is given by

$$K_3 = \frac{\xi-\eta}{\xi+\eta} - \frac{2\eta}{\xi+\eta}\,\frac{\partial}{\partial\xi}\Big(\xi\frac{\partial}{\partial\xi}\Big) + \frac{2\xi}{\xi+\eta}\,\frac{\partial}{\partial\eta}\Big(\eta\frac{\partial}{\partial\eta}\Big) + \frac{\xi-\eta}{2\xi\eta}\,\frac{\partial^2}{\partial\varphi^2}. \tag{2.141}$$

Using (2.135), it is not difficult to compute $K_3\psi_{n_1,n_2,m}$. A little calculation shows that $\psi_{n_1,n_2,m}$ is an eigenfunction of the operator K_3:

$$K_3\,\psi_{n_1,n_2,m} = \frac{n_1-n_2}{n}\,\psi_{n_1,n_2,m}. \tag{2.142}$$

The parabolic eigenfunctions are therefore eigenfunctions of the operators

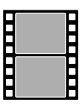
$$J_3^{\pm} = \frac{1}{2}\Big(L_3 \pm \frac{1}{\sqrt{-2H}}K_3\Big) \tag{2.143}$$

which we defined in (2.32). We have

$$J_3^{\pm}\,\psi_{n_1,n_2,m} = \frac{1}{2}\,\big(m \pm (n_1-n_2)\big)\,\psi_{n_1,n_2,m} = m_j^{\pm}\,\psi_{n_1,n_2,m}. \tag{2.144}$$

Because of (2.35), it is clear that the eigenfunctions of H are also eigenfunctions of the operators $(\mathbf{J}^{\pm})^2$. We conclude that the parabolic eigenfunctions are joint eigenfunctions of the complete set of commuting observables $H, (\mathbf{J}^{\pm})^2, J_3^{\pm}$. One can obtain these eigenfunctions also by the algebraic method using the Runge-Lenz vector and the algebra of pseudo-angular-momentum operators (Sections 2.3 and 1.5).

Exercise 2.16. *What is the shape of the nodal surfaces (= set of zeros) of the functions $\psi_{n_1,n_2,m}$?*



CD 2.25 shows a gallery of the eigenfunctions arising from the separation of the Coulomb problem in parabolic coordinates.

2.7. Physical Units and Dilations

2.7.1. The Coulomb problem in physical units

Because of its great importance, we want to solve the Schrödinger equation for the hydrogen atom with all physical constants in a given physical system of units for mass, length, time, and charge. We denote the space and time coordinates measured in the physical units by $\hat{x}$ and $\hat{t}$. The Schrödinger equation for the Coulomb problem in physical units has the form

$$\mathrm{i}\,\hbar\,\frac{\partial}{\partial \hat{t}}\,\phi(\hat{\mathbf{x}},\hat{t}) = -\frac{\hbar^2}{2\mathrm{m}}\,\hat{\Delta}\phi(\hat{\mathbf{x}},\hat{t}) - \frac{\gamma}{|\hat{\mathbf{x}}|}\,\phi(\hat{\mathbf{x}},\hat{t}). \tag{2.145}$$

Here, $\hat{\Delta}$ is the Laplace operator where the derivatives are taken with respect to the coordinates $\hat{x}_i$. The coupling constant γ is, for example, $\gamma = Ze^2$ (in the Gaussian system) or $\gamma = Ze^2/4\pi\epsilon_0$ (in the international system, with e measured in coulomb = ampere×second). As usual, e is the elementary charge, m is the mass of the particle, and $\hbar$ is Planck's constant. All these constants have to be expressed in the given system of units. Some values and dimensions in the SI are given in Table 2.1.

So far, we have obtained solutions only for the simpler equation

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(\mathbf{x},t) = -\frac{1}{2}\,\Delta\psi(\mathbf{x},t) - \frac{1}{|\mathbf{x}|}\,\psi(\mathbf{x},t). \tag{2.146}$$

What does this tell about the solutions of Schrödinger equation in physical units? You will see that the relation between the equations (2.146) and

TABLE 2.1. Important physical constants in terms of SI base units.

Physical constant	Value in SI units
elementary charge	$e = 1.602176 \times 10^{-19}\,\mathrm{A\,s}$
electron mass	$\mathrm{m}_e = 9.10938 \times 10^{-31}\,\mathrm{kg}$
proton mass	$\mathrm{m}_p = 1.672622 \times 10^{-27}\,\mathrm{kg}$
speed of light	$c = 2.99792458 \times 10^{8}\,\mathrm{m\,s^{-1}}$
Planck's constant	$\hbar = 1.054572 \times 10^{-34}\,\mathrm{kg\,m^2\,s^{-1}}$
permittivity of vacuum	$\epsilon_0 = 8.8541878 \times 10^{-12}\mathrm{A^2\,kg^{-1}m^{-3}\,s^4}$
coupling constant for hydrogen	$\gamma_0 = 2.30708 \times 10^{-28}\,\mathrm{kg\,m^3\,s^{-2}}$
Bohr radius	$a_0 = 5.2917721 \times 10^{-11}\,\mathrm{m}$
atomic time unit	$t_0 = 0.242 \times 10^{-16}\,\mathrm{s}$
atomic energy unit (hartree)	$E_h = 4.35974 \times 10^{-18}\,\mathrm{kg\,m^2\,s^{-2}}$
Rydberg constant	$R_\infty = 1.0973731534 \times 10^{7}\,\mathrm{m^{-1}}$
fine structure constant	$\alpha = 7.297352 \times 10^{-3} = 1/137.036$

(2.145) is simply given by a change of units. It is always possible to choose a system of units where $\hbar$, m, and the Coulomb coupling constant γ all have the numerical value 1. These units will be called *dimensionless units*. Equation (2.146) is the Schrödinger equation in dimensionless units.

After a change of units, the new value for a physical quantity is obtained from the value in the old units by multiplication with a conversion factor. If the equations (2.146) and (2.145) are indeed related by a change of units, we can express $\hat{\mathbf{x}}$ (which is measured, say, in meters) in terms of $\mathbf{x}$ (in dimensionless units) by

$$\hat{\mathbf{x}} = a\,\mathbf{x}, \quad \text{with } a > 0, \tag{2.147}$$

where the conversion factor a will be chosen appropriately. Because $|\mathbf{x}| = 1$ corresponds to $|\hat{\mathbf{x}}| = a$, we find that a gives the numerical value of the dimensionless length unit in meters.

In a similar way, we change the unit of time by multiplication with a conversion factor b,

$$\hat{t} = b\,t, \quad \text{with } b > 0, \tag{2.148}$$

where $\hat{t}$ gives the time in seconds and b gives the dimensionless time unit in seconds.

The coordinate transformation $\mathbf{x} \to \hat{\mathbf{x}}$, $t \to \hat{t}$ induces a transformation of wave functions according to

$$\psi(\mathbf{x}, t) = \phi(\hat{\mathbf{x}}, \hat{t}), \quad \text{with } \hat{\mathbf{x}} = a\mathbf{x},\ \hat{t} = bt. \tag{2.149}$$

Next, we have to figure out how to convert the Schrödinger equation. We can use the chain rule to obtain

$$\frac{\partial}{\partial t}\,\psi(\mathbf{x}, t) = \Big(\frac{\partial}{\partial \hat{t}}\,\phi(\hat{\mathbf{x}}, \hat{t})\Big)\,\frac{d\hat{t}}{dt} = b\,\frac{\partial}{\partial \hat{t}}\,\phi(\hat{\mathbf{x}}, \hat{t}). \tag{2.150}$$

A similar calculation can be done for the derivatives with respect to the components of $\mathbf{x}$:

$$\frac{\partial}{\partial x_i}\,\psi(\mathbf{x}, t) = a\,\frac{\partial}{\partial \hat{x}_i}\,\phi(\hat{\mathbf{x}}, \hat{t}). \tag{2.151}$$

If we assume that $\psi(\mathbf{x}, t)$ satisfies Schrödinger's equation (2.146) in dimensionless units, we find immediately that $\phi(\hat{\mathbf{x}}, \hat{t})$ satisfies the equation

$$\mathrm{i}\,b\,\frac{\partial}{\partial \hat{t}}\,\phi(\hat{\mathbf{x}}, \hat{t}) = -a^2\,\frac{1}{2}\,\hat{\Delta}\phi(\hat{\mathbf{x}}, \hat{t}) - a\,\frac{1}{|\hat{\mathbf{x}}|}\,\phi(\hat{\mathbf{x}}, \hat{t}). \tag{2.152}$$

By a suitable choice of a, we can achieve that the coefficients in front of $\hat{\Delta}$ and of $1/|\hat{\mathbf{x}}|$ have the same ratio as the physical constants in the Schrödinger equation (2.145). Thus, we require

$$\frac{\hbar^2/\mathrm{m}}{\gamma} = \frac{a^2}{a} = a, \quad \text{or} \quad a = r_0 = \frac{\hbar^2}{\mathrm{m}\,\gamma}. \tag{2.153}$$

The quantity r_0 is called the first Bohr radius of the hydrogenic system.[4] With $a = r_0$, the Schrödinger equation (2.152) becomes

$$\mathrm{i}\, b\, \frac{\partial}{\partial \hat{t}}\, \phi(\hat{\mathbf{x}}, \hat{t}) = \frac{\hbar^2}{\mathrm{m}\, \gamma^2} \Big(-\frac{\hbar^2}{2\mathrm{m}}\, \hat{\Delta} \phi(\hat{\mathbf{x}}, \hat{t}) - \frac{\gamma}{|\hat{\mathbf{x}}|}\, \phi(\hat{\mathbf{x}}, \hat{t}) \Big). \tag{2.154}$$

Now we can choose the time-conversion factor

$$b = s_0 = \frac{\hbar^3}{\mathrm{m}\, \gamma^2} \tag{2.155}$$

and multiply the resulting equation by $\mathrm{m}\, \gamma^2 / \hbar^2$. This converts Eq. (2.154) into the Schrödinger equation in physical units, Eq. (2.145).

You can compute the physical dimensions of the quantities

$$\mathbf{x} = \frac{\hat{\mathbf{x}}}{r_0} = \frac{\mathrm{m}\, \gamma}{\hbar^2}\, \hat{\mathbf{x}}, \qquad t = \frac{\hat{t}}{s_0} = \frac{\mathrm{m}\, \gamma^2}{\hbar^3}\, \hat{t} \tag{2.156}$$

by inserting the dimensions of $\hbar$ (= energy $\times$ time), m (= mass), and γ (= mass length^3 / time^2). You will find that all dimensions cancel out. We say that $\mathbf{x}$ and t are *dimensionless quantities*.

Our result states that you can always replace the Schrödinger equation in physical units by the equation in dimensionless units, thereby getting rid of physical constants. You just have to measure lengths in units of r_0 and times in units of s_0.

The Bohr radius r_0 sets the length scale of atomic phenomena. Note that the radial oscillations of the wave function will take place in a neighborhood of the minimum of the effective radial potential

$$V_{\mathrm{eff}}(r) = \frac{\hbar^2}{2\mathrm{m}}\, \frac{\ell(\ell+1)}{r^2} - \frac{\gamma}{r}. \tag{2.157}$$

A little calculation shows that the minimum of the effective potential is at $r = \ell(\ell+1)\, r_0$.

Let us now consider the hydrogen atom in Bohr's model, for which we have $\mathrm{m} = \mathrm{m}_e$ and $Z = 1$. In this case, we denote the Coulomb coupling constant by γ_0, the Bohr radius by a_0, and the time unit s_0 by t_0, that is

$$\gamma_0 = \frac{e^2}{4\pi\epsilon_0}, \quad a_0 = \frac{\hbar^2}{\mathrm{m}_e\, \gamma_0}, \quad t_0 = \frac{\hbar^3}{\mathrm{m}_e\, \gamma_0^2}. \tag{2.158}$$

The values of these quantities in SI units are given in Table 2.1.

The dimensionless units for the hydrogen atom in the Bohr model are usually called *atomic units*. In atomic units, the mass of the electron is $\mathrm{m}_e = 1$, also, Planck's constant and the Coulomb coupling constant γ_0 have

[4]Here, we have γ proportional to $Z \geq 1$. In case of a hydrogen atom ($Z = 1$), the Bohr radius is denoted by a_0.

the value 1. Hence, Bohr's radius for the hydrogen atom is the atomic length unit and t_0 is the atomic time unit.

The dimension of the Coulomb coupling constant is (mass) × (length)^3 / (time)^2, as you can see from Table 2.1. We can combine γ_0 with the physical constants $\hbar$ and c (the speed of light) into a dimensionless quantity,

$$\alpha = \frac{\gamma_0}{\hbar c} = 1/137.036 \tag{2.159}$$

which is called Sommerfeld's fine structure constant. Because this quantity is dimensionless, it has the same numerical value with respect to all systems of units. From this value, we can therefore compute the value of c in atomic units, where $\hbar = e = 1$. We obtain immediately that

$$c = 137.036 \qquad \text{is the velocity of light in atomic units.} \tag{2.160}$$

The energies and eigenfunctions in SI units will be discussed in Section 2.7.3 below.

EXERCISE 2.17. *Convert the speed of light from the SI to atomic units using a scaling transformation with $a = a_0$ and $b = t_0$ as given in* (2.158)*, and verify the value of c by a direct computation.*

EXERCISE 2.18. *Verify in a similar way that $\hbar$ has the numerical value* 1 *in atomic units.*

EXERCISE 2.19. *Assume that you live in a universe where $\hbar$ has the value $1.0546 \times 10^{-27}\, kg\, m^2\, s^{-1}$ instead of the value given in Table 2.1. How large is a hydrogen atom in this universe (assuming that mass and charge have their usual values)?*

EXERCISE 2.20. *Assume that $\psi(\mathbf{x})$ is an eigenfunction of the Coulomb Hamiltonian in dimensionless units belonging to the eigenvalue E. Perform a scaling transformation $\hat{\mathbf{x}} = \mathbf{x}/a_0$ and show that the eigenvalue parameter of the stationary Schrödinger equation in SI units is $\mathrm{m}\gamma^2 E/\hbar^2$.*

2.7.2. Scaling transformations

In the Hilbert space of wave functions, a change of units induces a linear transformation, called a *scaling transformation* or *dilation*.

Given a wave function ϕ, we define a scaled wave function ψ by

$$\psi(\mathbf{x}) = N_\lambda\, \phi(\lambda \mathbf{x}), \tag{2.161}$$

with a suitable normalization constant N_λ. Conversely, we can express ϕ in terms of ψ by

$$\phi(\hat{\mathbf{x}}) = \frac{1}{N_\lambda}\, \psi\Big(\frac{\hat{\mathbf{x}}}{\lambda}\Big). \tag{2.162}$$

We choose the constant N_λ such that ψ is normalized whenever ϕ is normalized. The norm of ψ is

$$\|\psi\|^2 = \int |\psi(\mathbf{x})|^2 \, d^3x = N_\lambda^2 \int |\phi(\lambda\mathbf{x})|^2 \, d^3x. \tag{2.163}$$

Substituting $\hat{\mathbf{x}} = \lambda\mathbf{x}$ and $d^3\hat{x} = \lambda^3 \, d^3x$, we find

$$\|\psi\|^2 = N_\lambda^2 \int |\phi(\hat{\mathbf{x}})|^2 \, \frac{d^3\hat{x}}{\lambda^3} = \frac{N_\lambda^2}{\lambda^3} \, \|\phi\|^2. \tag{2.164}$$

We see that the linear transformation $\phi \to \psi$ is norm-preserving (that is, unitary) if we define

$$N_\lambda = \lambda^{3/2}. \tag{2.165}$$

Dilations:

The linear operators $U_\lambda : \phi \to \psi$, defined on $L^2(\mathbb{R}^3)$ by

$$\psi(\mathbf{x}) = (U_\lambda \, \phi)(\mathbf{x}) = \lambda^{3/2} \, \phi(\lambda\mathbf{x}) \quad \text{for } \lambda > 0, \tag{2.166}$$

are called *scaling transformations* or *dilations*. They are unitary in the Hilbert space $L^2(\mathbb{R}^3)$. The inverse transformation is

$$U_\lambda^{-1} = U_\lambda^\dagger = U_{1/\lambda}. \tag{2.167}$$

The unitary transformation U_λ induces a transformation of linear operators in the usual way, $A \to U_\lambda^\dagger \, A \, U_\lambda$. Let us now compute the effect of a scaling transformation on the Hamiltonian. First, we compute the action of $-\mathrm{i}\nabla$ (the momentum operator) on a scaled wave function

$$\begin{aligned}
(-\mathrm{i}\nabla)(U_\lambda\phi)(\mathbf{x}) &= \lambda^{3/2}(-\mathrm{i}\nabla)\phi(\hat{\mathbf{x}}) && \text{with } \hat{x} = \lambda\mathbf{x} \\
&= \lambda^{3/2}\lambda(-\mathrm{i}\hat{\nabla}\phi)(\hat{\mathbf{x}}) && \text{by the chain rule} \\
&= \lambda \left(U_\lambda(-\mathrm{i}\nabla\phi)\right)(\mathbf{x}) && \text{by the definition of } U_\lambda.
\end{aligned} \tag{2.168}$$

This is the three-dimensional analog of (2.151). Multiplying everything by the inverse $U_\lambda^{-1} = U_\lambda^\dagger$ from the left gives (on the domain of ∇)

$$U_\lambda^\dagger \, (-\mathrm{i}\nabla) \, U_\lambda = \lambda \, (-\mathrm{i}\nabla). \tag{2.169}$$

Now it is easy to derive the corresponding transformation of the Laplacian operator $-\Delta = (-\mathrm{i}\nabla)^2$ (and hence for the kinetic energy). We obtain immediately

$$U_\lambda^\dagger \, (-\Delta) \, U_\lambda = \lambda^2 \, (-\Delta). \tag{2.170}$$

We may also write this in the form

$$\Delta \, (U_\lambda\phi)(\mathbf{x}) = \lambda^{3/2} \, \lambda^2 \, \hat{\Delta}\phi(\hat{\mathbf{x}}), \tag{2.171}$$

where $\hat{\Delta}$ denotes the Laplacian with respect to $\hat{\mathbf{x}} = \lambda\mathbf{x}$.

The potential energy $1/|\mathbf{x}|$ behaves as follows:

$$\frac{1}{|\mathbf{x}|} U_\lambda \phi(\mathbf{x}) = \lambda^{3/2} \frac{1}{|\mathbf{x}|} \phi(\lambda\mathbf{x}) = \lambda^{3/2}\, \lambda \frac{1}{|\hat{\mathbf{x}}|} \phi(\hat{\mathbf{x}}). \tag{2.172}$$

Hence, we find

$$\frac{1}{|\mathbf{x}|} U_\lambda \phi(\mathbf{x}) = \lambda\, U_\lambda \frac{1}{|\mathbf{x}|} \phi(\mathbf{x}), \tag{2.173}$$

or, after applying U_λ to both sides of the equation,

$$U_\lambda^\dagger \frac{1}{|\mathbf{x}|} U_\lambda = \lambda \frac{1}{|\mathbf{x}|}. \tag{2.174}$$

2.7.3. Energies and eigenfunctions in physical units

The transitions between physical and dimensionless units discussed in Section 2.7.1 are examples for the application of the scaling transformation U_λ. Consider, for example, the stationary Schrödinger equation in dimensionless units,

$$-\frac{1}{2}\Delta\psi(\mathbf{x}) - \frac{1}{|\mathbf{x}|}\psi(\mathbf{x}) = E\,\psi(\mathbf{x}). \tag{2.175}$$

Writing $\psi(\mathbf{x}) = U_\lambda\,\phi(\mathbf{x})$, we find (after dividing the whole equation by $\lambda^{3/2}$)

$$-\frac{\lambda^2}{2}\hat{\Delta}\phi(\hat{\mathbf{x}}) - \frac{\lambda}{|\hat{\mathbf{x}}|}\phi(\hat{\mathbf{x}}) = E\,\phi(\hat{\mathbf{x}}). \tag{2.176}$$

Here, we have used (2.171) and (2.173). With $\lambda = r_0$ as in (2.153), we obtain the stationary Schrödinger equation in physical units

$$-\frac{\hbar^2}{2\mathrm{m}}\hat{\Delta}\phi(\hat{\mathbf{x}}) - \frac{\gamma}{|\hat{\mathbf{x}}|}\phi(\hat{\mathbf{x}}) = \hat{E}\,\phi(\hat{\mathbf{x}}), \tag{2.177}$$

where $\hat{\mathbf{x}} = r_0\,\mathbf{x}$ and

$$\hat{E} = \frac{\mathrm{m}\gamma^2}{\hbar^2} E = \frac{\gamma}{r_0} E. \tag{2.178}$$

Hence, from our knowledge of the Coulomb Hamiltonian in dimensionless units, we can immediately obtain eigenvalues and eigenfunctions of the Coulomb Hamiltonian in physical units. In the same way, we obtain via a scaling transformation the eigenvalues and eigenfunctions in any given system of units, provided we know the conversion factor corresponding to r_0.

The stationary Schrödinger equation in physical units (2.177) has the eigenvalues

$$\hat{E}_n = \frac{\mathrm{m}\gamma^2}{\hbar^2} E_n = \frac{\mathrm{m}\gamma^2}{2\hbar^2 n^2} = Z^2 R_{\mathrm{H}} hc \frac{1}{n^2}, \quad n = 1, 2, 3, \ldots. \tag{2.179}$$

Here, we have introduced the *Rydberg constant for hydrogen* (see also (2.1))

$$R_{\mathrm{H}} = \frac{\mathrm{m}\gamma_0^2}{4\pi\hbar^3 c}. \tag{2.180}$$

Inserting for m the reduced mass $\mathrm{m}_e\mathrm{m}_p/(\mathrm{m}_e + \mathrm{m}_p)$ of the electron-proton system, one obtains the value $R_{\mathrm{H}} \approx 1.09678 \times 10^{-7}\mathrm{m}$, which agrees with the spectroscopically determined value. In the Bohr model of the hydrogen, one sets $\mathrm{m} = \mathrm{m}_e$ and defines the *Rydberg constant*

$$R_\infty = \frac{\mathrm{m}_e\gamma_0^2}{4\pi\hbar^3 c}. \tag{2.181}$$

The value of this constant is given in Table 2.1.

The atomic energy unit is the hartree

$$E_h = \frac{\mathrm{m}_e\gamma_0^2}{\hbar^2}. \tag{2.182}$$

All eigenvalues obtained in dimensionless units have to be multiplied by this factor if we want to know the energies of the hydrogen atom according to Bohr's model (with infinite proton mass). In physical units, the energies of the hydrogen atom according to the Bohr model are

$$E_n = -\frac{1}{2n^2}\,E_h = -\frac{\mathrm{m}_e\gamma_0^2}{2\hbar^2 n^2} = -\frac{\gamma_0}{a_0}\,\frac{1}{2n^2} = -\frac{R_\infty hc}{n^2}. \tag{2.183}$$

Another often used energy unit in atomic physics is the electron volt, $1\,\mathrm{eV} = 1.60218 \times 10^{-19}\,\mathrm{J}$, approximately. With Table 2.1, we compute the ground state energy of the hydrogen as

$$E_1 = -E_h/2 = R_\infty hc = 2.17987 \times 10^{-18}\,\mathrm{J} = 13.60569\,\mathrm{eV} \tag{2.184}$$

Via the scaling transformation U_{r_0} (with r_0 being the Bohr radius), we can immediately find the normalized eigenfunctions of a hydrogenic atom, given the normalized eigenfunctions (2.69) in dimensionless units. Whenever ψ is a solution in dimensionless units, then $\phi = U_{r_0}^\dagger\,\psi$ is a solution of (2.177):

$$\phi_{n,\ell,m}(\hat{\mathbf{x}}) = \Big(\frac{1}{r_0}\Big)^{3/2}\,\psi_{n,\ell,m}\Big(\frac{\hat{\mathbf{x}}}{r_0}\Big). \tag{2.185}$$

Note that the angular part Y_ℓ^m of the eigenfunction is not affected by the scaling transformation.

EXERCISE 2.21. *Show that in n space dimensions, the definition of the unitary scaling transformation has to be replaced by $(U_\lambda\,\phi)(\mathbf{x}) = \lambda^{n/2}\,\phi(\lambda\mathbf{x})$. Show that the equations* (2.169), (2.170), *and* (2.174) *are independent of the space dimension.*

EXERCISE 2.22. *Consider a hydrogenic atom, where the electron has been replaced by a muon. The muon has the mass* $206.8\,\mathrm{m}_e$*, charge* e*, and a lifetime of about* 2.2×10^{-6} *s. Determine the Bohr radius and the energy levels and compare the muonic atom with an electronic atom.*

2.7.4. The generator of dilations and the virial theorem

A composition of two scaling transformations U_λ and U_μ gives again a dilation:

$$U_{\lambda_1}\, U_{\lambda_2}\, \psi(\mathbf{x}) = (\lambda_1\lambda_2)^{3/2}\, \psi(\lambda_1\lambda_2\, \mathbf{x}) = U_{\lambda_1\lambda_2}\, \psi(\mathbf{x}). \tag{2.186}$$

The scaling parameter should always be positive (otherwise the dilation would include a reflection about the origin). Hence, we write $\lambda = \mathrm{e}^\theta$ and denote the dilation operators by

$$U(\theta) = U_{\exp(\theta)}, \quad U(\theta)\phi(\mathbf{x}) = \mathrm{e}^{3\theta/2}\, \psi(\mathrm{e}^\theta \mathbf{x}) \quad \text{with } \theta \in \mathbb{R}. \tag{2.187}$$

Then the composition law (2.186) and the properties of the exponential function imply

$$U(\theta_1)\, U(\theta_2) = U(\theta_1 + \theta_2). \tag{2.188}$$

The operators $U(\theta)$, $\theta \in \mathbb{R}$, form a (strongly continuous) unitary group. The infinitesimal generator can be determined by computing

$$\frac{d}{d\theta}\, U(\theta)\, \psi\Big|_{\theta=0} = D\, \psi \tag{2.189}$$

for differentiable wave functions ψ. We obtain the following result (see Exercise 2.23 below).

The dilation generator:

The generator of the dilation group $U(\theta)$, $\theta \in \mathbb{R}$, is the operator

$$D = \frac{1}{2}(\mathbf{x} \cdot \mathbf{p} + \mathbf{p} \cdot \mathbf{x}). \tag{2.190}$$

It is self-adjoint (on a suitable domain) and

$$U(\theta) = U_{\exp(\theta)} = \exp(-\mathrm{i}D\theta). \tag{2.191}$$

EXERCISE 2.23. *Verify that the operator* D *is the generator of the dilation group.*

With the help of the dilation generator, we can derive a useful result, the *virial theorem*. Whenever ψ is a stationary state, then there exists a number E with $H\psi = E\psi$. Evaluating the commutator of the Hamiltonian

$H = p^2/2 + 1/|\mathbf{x}|$ with the generator of dilations $D = (\mathbf{x}\cdot\mathbf{p} + \mathbf{p}\cdot\mathbf{x})/2$, we find

$$\mathrm{i}\,[H, D] = p^2 - \frac{1}{|\mathbf{x}|} = 2H_0 - \frac{1}{|\mathbf{x}|}. \tag{2.192}$$

Here. H_0 is the operator of kinetic energy. But the expectation value of the commutator in an eigenstate is zero,

$$\begin{aligned}\langle\psi, [H, D]\,\psi\rangle &= \langle H\psi, D\psi\rangle - \langle D\psi, H\psi\rangle \\ &= E(\langle\psi, D\psi\rangle - \langle\psi, D\psi\rangle) = 0.\end{aligned} \tag{2.193}$$

The calculation above is valid for ψ in the intersection of the domains of H and D. (This is always the case for eigenfunctions with negative energy as one can infer from their explicit form given above). Thus

$$2\langle\psi, H_0\,\psi\rangle = \langle\psi, \frac{1}{|\mathbf{x}|}\psi\rangle, \tag{2.194}$$

which means that for bound states in a Coulomb field, the expectation value of the potential energy is minus twice the expectation value of the kinetic energy. Therefore, the total energy satisfies

$$\begin{aligned}E = \langle\psi, H\,\psi\rangle &= \langle\psi, H_0\,\psi\rangle - \langle\psi, \frac{1}{|\mathbf{x}|}\psi\rangle \\ &= -\langle\psi, H_0\,\psi\rangle = -\frac{1}{2}\,\langle\psi, \frac{1}{|\mathbf{x}|}\psi\rangle.\end{aligned} \tag{2.195}$$

Ψ A technical refinement of this argument generalizes it to arbitrary bound states, not just states in the common domain of H and D. In this generalized form, the consideration above can be used to prove that bound state energies of the Coulomb Hamiltonian are always negative. That is, there are no stationary states with positive energy in the Coulomb field. This statement is based on our choice of the zero of energy (which is characterized by $V(\mathbf{x}) \to 0$, as $|\mathbf{x}| \to \infty$). We could always shift the energy scale by adding or subtracting a constant potential, thereby producing bound states with positive energy. Physically, it is therefore more appropriate to say that the Coulomb system has no bound states above the so-called *ionization threshold*.

The result (2.195) for the Coulomb potential can be generalized to arbitrary (differentiable) potentials:

The virial theorem:

Let $H = H_0 + V$ with $H_0 = p^2/2$ and V a differentiable function of $\mathbf{x} \in \mathbb{R}^3$. Suppose that $H\psi = E\psi$. Then

$$2\,\langle\psi, H_0\,\psi\rangle = \langle\psi, \mathbf{x}\cdot\nabla V\,\psi\rangle. \tag{2.196}$$

Equation (2.194) agrees with (2.196), because for $V(\mathbf{x}) = -1/|\mathbf{x}|$ we have $\mathbf{x} \cdot \nabla V(\mathbf{x}) = 1/|\mathbf{x}|$ (for $\mathbf{x} \neq 0$).

EXERCISE 2.24. *Prove the virial theorem by mimicking the corresponding proof for the Coulomb potential.*

EXERCISE 2.25. *Show that the total energy E of a classical particle moving with angular speed ω on a circular orbit in a Coulomb field is*

$$E = \frac{\mathrm{m} r^2 \omega^2}{2} - \frac{\gamma}{r}. \tag{2.197}$$

Eliminate ω from this expression using the fact that the Coulomb force is equal in magnitude to the centrifugal force, and prove the formula

$$2E = -\frac{\gamma}{r}. \tag{2.198}$$

2.8. Special Topic: Dynamics of Rydberg States



The time evolution in quantum mechanics is unitary and hence quite the opposite of a chaotic motion. Nevertheless, the motion of a superposition of several bound states can appear arbitrarily complicated. CD 2.20 shows a superposition of just four eigenstates. A few examples in two dimensions are presented in the movies CD 2.10. The examples in CD 2.11 give a detailed analysis of Gaussian wave packets projected onto the subspace of bound states.

The time evolution of a wave packet belonging to the subspace of bound states,

$$\Psi(r, \vartheta, \varphi, t) = \sum_{n=1}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{\ell} c_{n,\ell,m}\, \phi_{n,\ell,m}(r, \vartheta, \varphi)\, \exp(-\mathrm{i}\, E_n\, t) \tag{2.199}$$

is, in general, very complicated. But in this section, we are going to consider very special states of the form

$$\Psi(r, \vartheta, \varphi, t) = \sum_{\ell=0}^{\infty} c_\ell\, \psi_{\ell+1,\ell,\ell}(r, \vartheta, \varphi)\, \exp(-\mathrm{i}\, E_{\ell+1}\, t). \tag{2.200}$$

These wave functions have the following property: The radial quantum number $n_r = n - \ell - 1$ is always zero. This implies that in each summand, the radial part has no zeros and precisely one maximum near $r = \ell(\ell + 1)$. The magnetic quantum number is maximal. In each summand, the angular momentum is as vertical as possible. That is, the particle is most likely to be found near the polar angle $\vartheta = \pi/2$.

We make one further restriction. We assume that the coefficients c_ℓ are nonzero only in an ℓ-interval $\ell_0 - \Delta \le \ell \le \ell_0 + \Delta$ around some large ℓ_0, and $\ell_0 \gg \Delta$. Hence, Ψ is of the form

$$\Psi(r,\vartheta,\varphi,t) = \sum_{\ell=\ell_0-\Delta}^{\ell_0+\Delta} c_\ell\, \psi_{\ell+1,\ell,\ell}(r,\vartheta,\varphi)\, \exp(-\mathrm{i}\, E_{\ell+1}\, t). \tag{2.201}$$

Coulomb states of this type are often called circular Rydberg states. These states can serve as a simple model for a *Rydberg atom*. An atom with Z electrons can become a Rydberg atom when its outermost electron is excited to a very high energy level. This electron is then in an orbit far outside the core formed by the nucleus and the remaining electrons. To the excited electron, the core appears as a "nucleus" with charge e. As long as the electron does not get too close to the core, it behaves like the electron of a hydrogen atom.

The states $\psi_{\ell+1,\ell,\ell}$ have the following structure:

$$\psi_{\ell+1,\ell,\ell}(r,\vartheta,\varphi) = (-1)^\ell\, h_\ell(r)\, y_\ell(\vartheta)\, \frac{\mathrm{e}^{\mathrm{i}\ell\varphi}}{\sqrt{2\pi}} \tag{2.202}$$

with

$$h_\ell(r) = \frac{2^{\ell+1}}{(\ell+1)^{\ell+2}\,\sqrt{(2\,\ell+1)!}}\, \exp\Big(-\frac{r}{\ell+1}\Big)\, r^\ell, \tag{2.203}$$

$$y_\ell(\vartheta) = \frac{\sqrt{(2\ell+1)!}}{2^\ell\,\ell!\,\sqrt{2}}\, \sin^\ell(\vartheta). \tag{2.204}$$



CD 2.21.1 shows the angular functions y_ℓ for $50 \le \ell \le 300$. All these functions are similar to Gaussians centered at $\vartheta = \pi/2$ (see also Figure 2.11). CD 2.21.2 shows the radial parts h_ℓ for $60 \le \ell \le 140$. These functions also resemble Gaussian functions (centered at $r = \ell(\ell+1)$). CD 2.21.3 illustrates that for ℓ close to ℓ_0, the function h_ℓ becomes similar to h_{ℓ_0}, as ℓ_0 becomes large (see also Figure 2.12).

For ℓ large, the distance between functions y_ℓ and $y_{\ell+1}$ becomes small. Figure 2.11 illustrates this fact. Hence, within the range $\ell_0 - \Delta \le \ell \le \ell_0 + \Delta$, all functions y_ℓ are more or less similar to y_{ℓ_0}. Also, as indicated by numerical computations, the distance between the functions h_ℓ and $h_{\ell+1}$ tends to become small with increasing ℓ (see Fig. 2.12). Thus, for large ℓ_0,

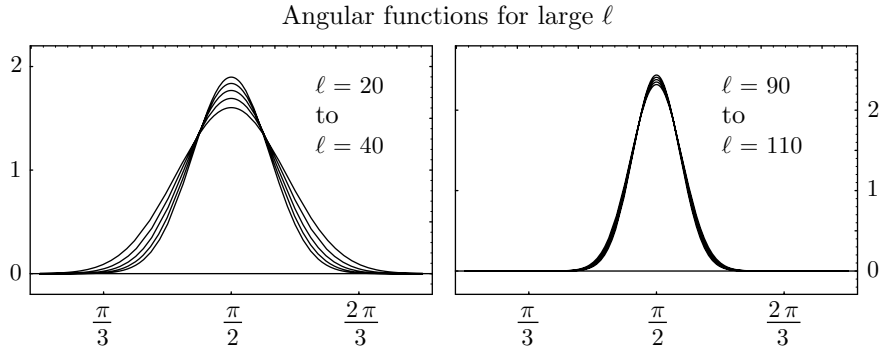


FIGURE 2.11. The angular functions $y_\ell(\vartheta)$ for ℓ in an interval of length 20 around $\ell_0 = 30$ (left image) and $\ell_0 = 100$ (right image). For large ℓ_0, the functions become similar to a Gaussian centered at $\vartheta = \pi/2$.

we make only a small error by rewriting (2.201) as follows:

$$\begin{aligned}\Psi(r,\vartheta,\varphi,t) &= \sum_{\ell=\ell_0-\Delta}^{\ell_0+\Delta} c_\ell\,(-1)^\ell\, h_\ell(r)\, y_\ell(\vartheta)\, \frac{\mathrm{e}^{\mathrm{i}\ell\varphi}}{\sqrt{2\pi}}\, \exp(-\mathrm{i}\, E_{\ell+1}\, t) \\ &\approx h_{\ell_0}(r)\, y_{\ell_0}(\vartheta) \sum_{\ell=\ell_0-\Delta}^{\ell_0+\Delta} c_\ell\,(-1)^\ell\, \frac{\mathrm{e}^{\mathrm{i}\ell\varphi}}{\sqrt{2\pi}}\, \exp(-\mathrm{i}\, E_{\ell+1}\, t) \\ &= h_{\ell_0}(r)\, y_{\ell_0}(\vartheta)\, g(\varphi,t). \end{aligned} \tag{2.205}$$

The radial function h_{ℓ_0} has its single maximum at $r_{\max} = \ell_0(\ell_0+1)$, and the angular function y_{ℓ_0} has its maximum at $\vartheta = \pi/2$. For large ℓ_0, the functions h_{ℓ_0} and y_{ℓ_0} both resemble Gaussian functions, which is also apparent from Figures 2.11 and 2.12. More precisely,

$$h_{\ell_0}(r) \approx N_{\ell_0}\, \exp\Big(-\frac{(r-\ell_0(\ell_0+1))^2}{2\ell_0(\ell_0+1)^2}\Big). \tag{2.206}$$

The approximation is with respect to the distance in $L^2([0,\infty))$. The form of the exponent is obtained by comparing the second derivatives of $h_{\ell_0}(r)$ and $N_{\ell_0}\, \exp\big(-k(r-r_{\max})\big)$ at $r = r_{\max}$. For the angular function we write

$$y_{\ell_0}(\vartheta) \approx M_{\ell_0}\, \exp\Big(-\frac{\ell_0}{2}\,\big(\vartheta - \frac{\pi}{2}\big)^2\Big). \tag{2.207}$$

Hence, our wave packet can be described as follows: It is a product of (approximately) Gaussian wave functions in r and ϑ and a time-dependent function of φ. We are now going to investigate the function g in (2.205).

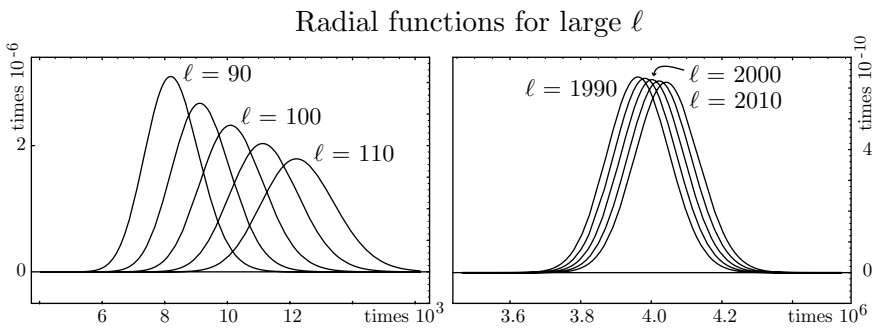


FIGURE 2.12. The radial functions $h_\ell(r)$ for ℓ in an interval of length 20 around $\ell_0 = 100$ (left image) and $\ell_0 = 2000$ (right image). For large ℓ_0, the functions become similar to a Gaussian centered at $r = \ell_0(\ell_0 + 1)$.

Writing $\ell = \ell_0 + k$ and using k instead of ℓ as the summation index, we obtain

$$g(\varphi, t) = \mathrm{e}^{\mathrm{i}\ell_0\varphi} \sum_{k=-\Delta}^{+\Delta} d_k \, \frac{\mathrm{e}^{\mathrm{i}k\varphi}}{\sqrt{2\pi}} \, \exp(-\mathrm{i}\, E_{\ell_0+k+1}\, t). \tag{2.208}$$

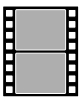
Here, the coefficients d_ℓ determine the shape of the initial function

$$g_0(\varphi) = g(\varphi, 0) = \mathrm{e}^{\mathrm{i}\ell_0\varphi} \, \frac{1}{\sqrt{2\pi}} \sum_{k=-\Delta}^{\Delta} d_k \, \mathrm{e}^{\mathrm{i}k\varphi}. \tag{2.209}$$

This is a simple Fourier sum describing a function that is 2π-periodic in φ (Fourier series are treated in Book One, Chapter 2). Choosing, for example,

$$d_k \approx \exp\Big(-\frac{k^2}{2\Delta}\Big), \tag{2.210}$$

we find that g_0 approximates a Gaussian function centered at $\varphi = 0$.



The movie CD 2.21.4. shows the time evolution of the angular part $g(\varphi, t)$ with a Gaussian distribution of coefficients around $\ell_0 = 90$, and with $\Delta = 10$. You can see the motion of $\Psi(r, \vartheta, \varphi, t)$ in the xy-plane in CD 2.22.1. Neglecting the radial motion for $\ell_0 = 90$ is a rather crude approximation. This approximation is much better for higher angular momenta. This is shown for the wave packet with $\ell_0 = 300$ in CD 2.22.2.

Any finite superposition of eigenstates is periodic in time, because the ratio between any two Coulomb-energies is a rational number. A finite superposition contains the time factors

$$\exp\Big(\mathrm{i}\,\frac{1}{2n_j^2}\,t\Big), \quad n_j = n_{rj} + \ell_j + 1, \quad j = 1, \ldots, k \tag{2.211}$$

(n_j are the principal quantum numbers of the states in the superposition). The interval of periodicity is determined by the first time $t = T_0$, for which all time factors are 1. That is, all exponents must be integer multiples of 2π. Hence, the period T_0 of the wave function is the smallest number t such that $t/(2n_j^2) = 2\pi m_j$ with integers m_j, and

$$T_0 = 4\pi\,\mathrm{LCM}(n_1, n_2, \ldots, n_k), \tag{2.212}$$

where LCM denotes the least common multiple. If there are more than a few states in the superposition, this period is too large to be of any practical importance. For a superposition of the 21 states with $n_r = 0$ and $1990 \leq \ell \leq 2010$, the time-period is

$$T \approx 2.4 \times 10^{114}. \tag{2.213}$$

There are, however, much smaller times at which the system returns to the initial state in an approximate sense. Let us rewrite the energy as

$$E_{\ell_0+k+1} = \frac{1}{2(\ell_0 + k + 1)^2} \approx \frac{1}{2(\ell_0 + k)^2} = \frac{1}{2\ell_0^2\,(1 + k/\ell_0)^2}. \tag{2.214}$$

For $k \in [-\Delta, \Delta]$ and $\ell_0 \gg \Delta$, we find that k/ℓ_0 is small. Hence, we can expand the energy as

$$E_{\ell_0+k+1} \approx \frac{1}{2\ell_0^2} - \frac{k}{\ell_0^3} + \frac{3}{2}\,\frac{k^2}{\ell_0^4} + \mathrm{O}\Big(\frac{1}{\ell_0^5}\Big). \tag{2.215}$$

Assuming that t is not too large (such that $t/\ell_0^5 \ll 1$), we may approximate the time factor in (2.208),

$$\exp(-\mathrm{i}E_{\ell+1}t) \approx \exp\Big(-\mathrm{i}\,\frac{1}{2\ell_0^2}\,t\Big)\,\exp\Big(\mathrm{i}\,\frac{k}{\ell_0^3}\,t\Big)\,\exp\Big(-\mathrm{i}\,\frac{3}{2}\,\frac{k^2}{\ell_0^4}\,t\Big). \tag{2.216}$$

The first exponential is just an overall phase factor. We can ignore it in the evolution of the state. The second factor is periodic with a period

$$T_0 = 2\pi\ell_0^3. \tag{2.217}$$

This is precisely the relation between angular momentum and time period for a circular orbit in the classical Coulomb problem. As long as the third exponential factor in (2.216) may still be neglected, the time evolution of the function g in (2.209) is essentially a multiplication of the coefficients d_ℓ with the exponential factor $\exp(\mathrm{i}\,kt/\ell_0^3)$. This simply amounts to a translation

with respect to φ, that is, $g(\varphi, t) \approx g(\varphi - t/\ell_0^3)$ (up to a phase factor of modulus 1). We see that (in a first approximation for small times) the initial wave packet just rotates around the origin with angular speed $1/\ell_0^3$, thereby retaining its shape.



CD 2.22.1 and 2.22.2 show that an initial function of the type described in this section moves on a circular orbit, revolving around the center of force several times. After a few cycles, the wave packet has spread along its orbit because the influence of the last exponential in (2.216) cannot be neglected any longer. The wave packet behaves in a more quasiclassical way if the average angular momentum is higher (see CD 2.23.1).

The last exponential factor in (2.216) usually cannot be neglected any longer after a few periods T_0. Its presence will distort the shape of the initial function, causing it to spread along its orbit. This exponential becomes equal to 1 again after the time

$$T_1 = \frac{2}{3}\,\ell_0^4\,\pi = \frac{\ell_0}{3}\,T_0. \tag{2.218}$$

For ℓ_0 large enough, the condition $t/\ell_0^5 \ll 1$ is satisfied even for $t \approx T_1$, and we may still neglect higher-order terms in (2.216). As a consequence, the time evolution becomes simple again, and the shape of the initial state is restored at time T_1. This is called a revival of the initial state.



CD 2.22.3 shows the wave packet with $\ell_0 = 300$ at times that are integer multiples of T_0. You can observe the approximate revival of the initial state at $T_1 \approx 100T_0$. In CD 2.23, you can see the angular motion $g(\varphi, t)$ with average angular momentum $\ell_0 \approx 2000$ showing a much "cleaner" revival of the initial state (because the condition $t/\ell_0^5 \ll 1$ is well satisfied). In all these examples, the wave packet stays close to a circular orbit in the xy-plane (with a sharp ϑ-distribution around $\vartheta = \pi/2$. Hence, these states behave as if they were solutions of the two-dimensional Coulomb problem. Examples of the two-dimensional motion are shown in CD 2.24.

For $\ell_0 = 300$, we find that

$$T_1 \approx 1.7 \times 10^{10}. \tag{2.219}$$

This might seem a long time, but in SI units this is still only $4 \times 10^{-7}\,\mathrm{s}$, which is a realistic lifetime for a Rydberg state. Actually, Rydberg states have a very low probability for spontaneous decay. Hence, they can have a surprisingly long lifetime up to $10^{-4}\,\mathrm{s}$. (This is extremely long compared to the typical lifetimes of low-lying energy levels, which is about $10^{-8}\,\mathrm{s}$). We

finally note that the radius of the circular orbit with $\ell_0 = 300$ would be 9×10^5 in atomic units, hence about 4.8×10^{-6} m.

EXERCISE 2.26. *Show that the arguments in this section also apply to the two-dimensional Coulomb problem.*

Chapter 3

Particles with Spin

Chapter summary: This chapter introduces the spin as an important intrinsic property of electrons. We discuss, in particular, some results that are relevant for atomic physics. More results will be presented from a slightly more abstract point of view in Chapter 4.

We start by describing the connection between angular momentum and magnetic moment according to classical electrodynamics. Then we proceed to discuss the Stern-Gerlach experiment, which can only be explained by assuming that the state of an electron is characterized by the eigenvalues of a spin operator $\mathbf{S}$. This spin operator has the properties of an angular momentum, but half-integer eigenvalues.

In Section 3.5, we describe the mathematical consequences of this assumption. We construct a Hilbert space for particles with spin 1/2 and define the operators describing the components of the spin.

In Section 3.6, we define the Pauli operator, that is, the Hamiltonian for a spin-1/2 particle in an external field. We discuss the solutions in a constant, homogeneous magnetic field, thereby generalizing results from Book One in Section 3.7. An important difference from the results without spin is the occurence of bound states with zero energy. This phenomenon also occurs for nonhomogeneous magnetic fields and for certain situations in three dimensions (Section 3.8).

The spin is most important for understanding finer details of the spectrum of hydrogenic atoms. In Section 3.9, we introduce the spin-orbit coupling and describe the spinor eigenfunctions of the hydrogen atom and the structure of the energy spectrum.

3.1. Introduction

The description of an elementary particle by a wave function that is a function of the position (or momentum) alone is often insufficient. In a more realistic model, one has to use wave functions with a more complicated structure. A first experimental evidence that electrons have to be characterized by additional quantum numbers came from spectroscopic observations. The doublet fine structure splitting in the spectra of one-electron atoms led Goudsmit and Uhlenbeck in 1925 to the suggestion that electrons should have an intrinsic angular momentum which was called the spin. The spin is a very important property of electrons and other elementary particles, and

it is crucial for the explanation of many phenomena. For example, together with the Pauli principle, the spin accounts for the structure of the periodic system of elements.

The spin is associated with the magnetic moment of the electrons. Hence, it influences the motion of electrons in an inhomogeneous magnetic field. In 1922, Stern and Gerlach used this effect to measure the magnetic moment (see Section 3.3).

The spin is an observable with the dimension of an angular momentum. The total angular momentum, which is the vector sum of the spin and the orbital angular momentum, is a conserved quantity in rotationally symmetric situations. But despite the fact that the spin is an angular momentum, it cannot be attributed to a rotational motion of the electron. A rotating particle must have some extended structure—a spinning point is meaningless. But if we assume that the electron is a rigid rotator (the quantum mechanical model for a rotating extended structure), then the electron's angular momentum would have properties that are very different from the observed properties of the electron spin. Such a mechanistic model of the spin and the magnetic moment of the electron would fail. According to our present-day knowledge, there is no hint that electrons are something else than point-like[1] particles with mass and charge.

The most striking difference between the spin angular momentum and the orbital angular momentum is the following. Whereas for the orbital angular momentum $\mathbf{L}$ the possible eigenvalues are given by non-negative integer values of the angular-momentum quantum number ℓ, one finds that the quantum number describing the spin of an electron must be $1/2$. This is a nice example of Theorem 1.1, which already stated the possibility of half-integer quantum numbers.

While this section concentrates on the description of the electron, we would like to emphasize that most elementary particles have a nonvanishing spin. For example, the most familiar particles like neutrons, protons, quarks, positrons, muons, and neutrinos all have spin $1/2$ like the electron, whereas some mesons and the photon have spin 1. The spin is indeed an omnipresent phenomenon in elementary particle physics. We also note that particles with half-integral spin are called fermions, and particles with integer spin (including zero) are called bosons. This classification refers to the different behavior of systems composed of several identical particles and will be discussed further in Section 5.9.

[1]The wave function is an extended structure, but it describes the position probability density of a point-like particle.

3.2. Classical Theory of the Magnetic Moment

3.2.1. Magnetic moment of an extended particle

Consider a rigid rotating body in classical mechanics. We denote (for the moment) by $\mathbf{L}_{\text{int}}$ the angular momentum measured with respect to a non-rotating frame of reference that is attached to the center of mass of the body. Let us assume that the dimensions of the body are negligible in comparison to the other dimensions in the experimental setup. In this case, the body is called a particle, and $\mathbf{L}_{\text{int}}$ is called the *internal angular momentum* or *spin* of the particle. The internal angular momentum $\mathbf{L}_{\text{int}}$ is to be distinguished from the orbital angular momentum $\mathbf{L}$ due to the center-of-mass motion. The orbital angular momentum of the body is $\mathbf{L} = \mathbf{x} \times \mathbf{p}$, where $\mathbf{x}$ is the position of the center of mass, and $\mathbf{p}$ is the total momentum of the body. The total angular momentum of the body is given by $\mathbf{J} = \mathbf{L} + \mathbf{L}_{\text{int}}$.

If electric charges are distributed over a spinning body, then the revolution around the center of mass causes circulating currents. According to classical electrodynamics, a circular current is the origin of a magnetic moment $\boldsymbol{\mu}$. A magnetic moment may interact with external electromagnetic fields (as described below) even if the body as a whole is neutral. Hence, the internal angular momentum of a particle may have a notable influence on its center-of-mass motion.

As an example, consider the classical model of the hydrogen atom, where an electron with charge $q = -e$ and with mass m_e moves on a circular orbit with orbital angular momentum $\mathbf{L}_e$ around a heavy[2] nucleus (a proton). As a whole, the atom is a neutral particle, because the nucleus has the charge $+e$. The nucleus is assumed to be at rest in the center-of-mass frame of the atom. Hence, the internal angular momentum $\mathbf{L}_{\text{int}}$ of the atom comes entirely from the motion of the electron, that is, $\mathbf{L}_{\text{int}} = \mathbf{L}_e$.

According to classical electrodynamics (see, for example, Jackson's book [**3**]), a charge q with mass m and angular momentum $\mathbf{L}$ has the magnetic dipole moment

$$\boldsymbol{\mu} = \frac{q}{2\mathrm{m}}\,\mathbf{L} = \frac{q\hbar}{2\mathrm{m}}\,\frac{\mathbf{L}}{\hbar}. \tag{3.1}$$

The factor appearing in front of $\mathbf{L}/\hbar$ is called *magneton*.[3] For an electron,

[2] By "heavy" we mean that it is a good approximation to neglect the influence of the electronic motion on the motion of the nucleus.

[3] This expression depends on the chosen system of units. In the SI used here, the unit of a magneton is $\mathrm{A\,m^2}$ (current × area). In the Gaussian system, a magneton is given by $q\hbar/2mc$ and its unit is $\mathrm{g^{1/2}\,cm^{5/2}\,s^{-1}}$.

we set $q = -e$ and $\mathrm{m} = \mathrm{m}_e$ and obtain

$$\boldsymbol{\mu}_e = -\mu_\mathrm{B} \frac{\mathbf{L}}{\hbar}, \quad \text{with} \quad \mu_\mathrm{B} = \frac{e\hbar}{2\mathrm{m}_e}. \tag{3.2}$$

Here, μ_B is called the *Bohr magneton*. Its value (in SI units) is

$$\mu_\mathrm{B} = 9.274009 \times 10^{-24}\,\mathrm{A\,m^2}. \tag{3.3}$$

In general, it will be necessary to write

$$\boldsymbol{\mu} = g \frac{q\hbar}{2\mathrm{m}} \frac{\mathbf{L}}{\hbar}, \tag{3.4}$$

with a dimensionless factor g, called the *Landé g-factor*. It is undetermined and can be used to fit $\boldsymbol{\mu}$ to its experimentally measured value. In classical physics, the g-factor could be needed to describe the magnetic moment of an extended spinning body in a phenomenological way, because the relation between angular momentum and circular current depends on the details of the distributions of charge and mass.

The Landé g-factor determines the *gyromagnetic ratio* γ, which is defined as the ratio of the magnetic moment to the angular momentum,

$$\gamma = \frac{|\boldsymbol{\mu}|}{|\mathbf{L}|}. \tag{3.5}$$

For the orbital motion of a classical electron we have $g = 1$, and the gyromagnetic ratio is $e/(2\mathrm{m}_e)$.

3.2.2. The influence of an external magnetic field on a magnetic moment

Perhaps the most common realization of a magnetic moment is the needle of a compass. A magnetic field tries to turn the needle in the direction of the field lines. In a more quantitative way, this action of the magnetic field[4] $\mathbf{B}$ on a magnetic moment $\boldsymbol{\mu}$ is described by a torque

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}. \tag{3.6}$$

A particle with magnetic moment also possesses an internal angular momentum $\mathbf{L}_\mathrm{int}$, and hence the particle acts as a gyroscope. The reaction of a gyroscope to a torque is described in the classical mechanics of rigid bodies. It is found that a torque $\boldsymbol{\tau}$ changes the angular momentum $\mathbf{L}_\mathrm{int}$ of a gyroscope according to

$$\frac{d\mathbf{L}_\mathrm{int}}{dt} = \boldsymbol{\tau}. \tag{3.7}$$

[4]What is called here "magnetic field" and denoted by $\mathbf{B}$ is often called *magnetic induction* or *magnetic flux density*. In the SI, $\mathbf{B}$ is measured in units of Tesla, $1\,\mathrm{T} = 1\,\mathrm{kg\,s^{-2}\,A^{-1}}$. The expression "magnetic field" often refers to the quantity $\mathbf{H}$. In vacuum, however, the fields $\mathbf{B}$ and $\mathbf{H}$ are strictly proportional.

Combining Eqs. (3.6) and (3.7) with (3.4), we obtain the following equation for the motion of the magnetic moment

$$\frac{d\boldsymbol{\mu}}{dt} = g\,\frac{q}{2\mathrm{m}}\,\boldsymbol{\mu}\times\mathbf{B}. \tag{3.8}$$

This equation describes a precession about the direction of $\mathbf{B}$. The magnetic moment $\boldsymbol{\mu}$ remains at a fixed angle with respect to $\mathbf{B}$, and the length of the vector $\boldsymbol{\mu}$ remains constant, because the change $d\boldsymbol{\mu}/dt$ is always orthogonal to both $\boldsymbol{\mu}$ and $\mathbf{B}$.

Whereas a torque changes the direction of $\boldsymbol{\mu}$, it does not influence the center-of-mass motion of the particle. But an inhomogeneous magnetic field actually excerts a force on a particle with a magnetic moment. This effect can be used to measure $\boldsymbol{\mu}$, as you will learn in the next section. If the particle as a whole is neutral, the force is just given by[5]

$$\mathbf{F}(\mathbf{x}) = \nabla\big(\boldsymbol{\mu}\cdot\mathbf{B}(\mathbf{x})\big). \tag{3.9}$$

This can be interpreted as the negative gradient of a potential energy,

$$V_{\mathrm{mgn}} = -\boldsymbol{\mu}\cdot\mathbf{B}(\mathbf{x}). \tag{3.10}$$

For convenience, we collect our results in the following box.

Magnetic moment in a magnetic field:

Consider a classical neutral particle with mass m and a magnetic moment $\boldsymbol{\mu}$ that is related to an internal angular momentum according to

$$\boldsymbol{\mu} = g\,\frac{q}{2\mathrm{m}}\,\mathbf{L}. \tag{3.11}$$

Then the motion of the particle in a magnetic field $\mathbf{B}(\mathbf{x})$ is determined by

$$\frac{d}{dt}\,\mathbf{p}(t) = \nabla\big(\boldsymbol{\mu}(t)\cdot\mathbf{B}(\mathbf{x}(t))\big), \qquad \frac{d}{dt}\,\mathbf{x}(t) = \frac{1}{\mathrm{m}}\,\mathbf{p}(t), \tag{3.12}$$

$$\frac{d}{dt}\,\boldsymbol{\mu}(t) = g\,\frac{q}{2\mathrm{m}}\;\boldsymbol{\mu}(t)\times\mathbf{B}(\mathbf{x}(t)). \tag{3.13}$$

Exercise 3.1. *According to classical electrodynamics, the magnitude of the magnetic moment $\boldsymbol{\mu}$ in a closed plane circuit is*

$$\mu = \textit{current}\times\textit{area}, \tag{3.14}$$

[5] A derivation of this formula can be found in textbooks about electrodynamics, see, e.g., [**3**], Chapter 5.

and the direction of $\boldsymbol{\mu}$ *is perpendicular to the plane of the current loop. Assuming that the current is generated by a single charge* q *(with mass* m*) moving on a circle, show that* (3.14) *is equivalent to* $\mu = (q/2\mathrm{m})\, L$.

EXERCISE 3.2. *Can you suggest how the equations of motion* (3.12) *should be modified if the particle is not neutral?*

EXERCISE 3.3. *Using Eq.* (3.13)*, show that* $|\boldsymbol{\mu}|$ *is a constant of motion.*

EXERCISE 3.4. *Consider a neutral particle with magnetic moment* $\boldsymbol{\mu}$*. At time* $t = 0$*, the particle is at the origin with velocity in the positive* x*-direction. Assume that the magnetic field* $\mathbf{B}$ *has only a* z*-component near the* xz*-plane, that is,* $\mathbf{B}(\mathbf{x}) = (0, 0, B(\mathbf{x}))$ *for* $\mathbf{x} = (x, 0, z)$*. For the sake of a simple model, assume that only the third component of* $\mathbf{B}$ *has a nonvanishing gradient, and let* $\nabla B(\mathbf{x}) = (0, 0, f)$ *for* $\mathbf{x} = (x, 0, z)$ *in the* xz*-plane. Find the classical motion of the particle. In particular, show that the trajectory of the particle stays in the* xz*-plane for all times and that the* z*-component of* $\boldsymbol{\mu}$ *remains constant.*



The influence of an inhomogeneous magnetic field on a classical neutral particle with a magnetic moment is shown in CD 3.1. The first movie CD 3.1.1 shows the acceleration in the direction of the field gradient in a simple model case. The situation of Exercise 3.4 is described in CD 3.1.2. The following movie CD 3.1.3 shows a rather chaotic motion in the field $\mathbf{B}(\mathbf{x}) = (0, -y, z)$. For a larger z component, the motion becomes simple again, as shown in CD 3.1.4.

3.3. The Stern-Gerlach Experiment

From the discussion in the previous section, it should be clear that an investigation of the magnetic moment could give some hints about the internal structure of atoms. Already in the year 1921, Otto Stern and Walther Gerlach designed an apparatus for measuring the magnetic moment of atoms and molecules. The experimental arrangement is schematically depicted in Figure 3.1. A beam of neutral particles (atoms or molecules) is sent through the strongly inhomogeneous magnetic field provided by a magnet with a pointed pole tip. If the particles carry a magnetic moment, then we expect an influence of the magnetic field on the motion of the particles.

Now, let us look at this experiment more closely. The design of the pole pieces should guarantee that near the symmetry plane of the magnet (which we take as the xz-plane), the magnetic field points in the vertical direction (the z-direction) and increases strongly with increasing z (see Fig. 3.2(a)). We assume $\mathbf{B}(\mathbf{x}) = (0, 0, B(z))$ (on the xz-plane inside the apparatus), with

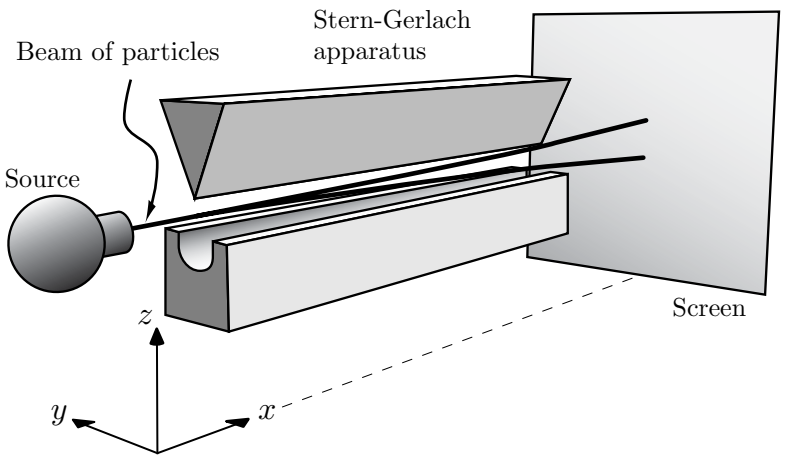


FIGURE 3.1. Schematic setup of a Stern-Gerlach experiment for measuring the spin. Particles in a strongly inhomogeneous magnetic field are deflected according to the component of the magnetic moment in the direction of the gradient of $\mathbf{B}$.

a large positive z-component $B(z)$ that increases with z (that is, $\nabla B(z)$ points in the same direction as $\mathbf{B}$). The force on a neutral particle near the symmetry plane is then approximately given by

$$\mathbf{F}(\mathbf{x}) = \nabla\big(\boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{x})\big) = \nabla\big(\mu_z\, B(z)\big) = \mu_z\, \nabla B(z). \tag{3.15}$$

Ψ According to the Maxwell equations, a magnetic field with the properties required above does not exist. For a magnetic field with a vanishing x-component, Maxwell's equation $\nabla \cdot \mathbf{B} = 0$ requires $\partial B_3/\partial z = -\partial B_2/\partial y$. You can see this in Figure 3.2(a). The horizontal y-component of the magnetic field is positive on the left side and negative on the right side of the center. This gradient gives a contribution to the force on the particle whenever $\mu_y \neq 0$. Depending on the direction of $\boldsymbol{\mu}$, this force is of the same size than the force in the z-direction. But note that the z-component of $\mathbf{B}$ has a large value whereas the y-component is close to 0. Hence, a magnetic moment will perform a rapid precession around the z-axis, which causes μ_y to oscillate around 0. The time average of the force in the y-direction will therefore be very small. The precession of $\boldsymbol{\mu}$ around the y-axis can be neglected, because the y-component of B vanishes near the symmetry plane of the apparatus.

As an example, we consider a beam of atoms entering the apparatus along the x-direction. We assume that all atoms in the beam have the same velocity and the same size $|\boldsymbol{\mu}|$ of the magnetic moment (notice that

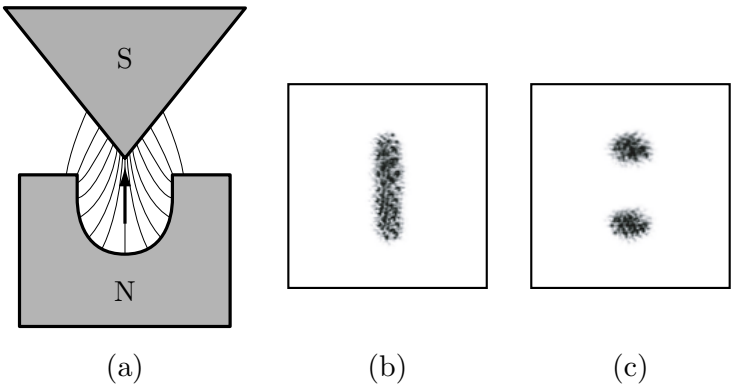


FIGURE 3.2. (a) Qualitative behavior of the magnetic field in a Stern-Gerlach apparatus. The arrow indicates the direction of the magnetic field *and* of its gradient near the symmetry axis. Particles with a magnetic moment entering the device here will only feel a force in the vertical direction. (b) Classically expected result for the distribution of particles on a screen behind the Stern-Gerlach apparatus. (c) Actually observed result.

$|\boldsymbol{\mu}|$ remains constant during the experiment; see Exercise 3.3). In case of hydrogen atoms,[6] the magnetic moment should be determined completely by the angular momentum of the electron as in Eq. (3.1). We know already that the ground state of the hydrogen is characterized by a vanishing orbital angular momentum, and on the basis of (3.1) we would expect no magnetic moment at all. (In fact, the proton forming the nucleus of the hydrogen atom does have a magnetic moment—but it is very small and may be neglected here.)

For atoms with a nonzero magnetic moment $\boldsymbol{\mu}$, the deflecting force (3.15) is proportional to the z-component μ_z and points in the vertical direction

[6]Actually, the experiment was performed first with silver atoms. The electronic configuration of a silver atom in the ground state consists of a closed shell structure and a single valence electron in a state with orbital angular momentum 0. Later, the same result was obtained by T.E. Phillips and J.B. Taylor with a beam of hydrogen atoms. We note that a Stern-Gerlach experiment has never been attempted with charged particles, because the Lorentz force on a moving charge is, in general, much stronger than the weak force (3.15) due to the field gradient. In a method developed by H.G. Dehmelt and others in 1976, one confines charged particles in an electromagnetic trap where they perform oscillations around an equilibrium position. By measuring the frequency of this oscillation, it is possible to determine the magnetic moment with a very high precision.

(the direction of $\nabla B(z)$). When the atoms enter the magnetic field, the torque $\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}$ causes a precession of the vector $\boldsymbol{\mu}$ around the direction of $\mathbf{B}$ (the z-direction). This precession does not change the component μ_z of the magnetic moment (see Exercise 3.4). The amount of the vertical deflection from the straight path is thus proportional to the constant value of μ_z.

The atoms that enter the Stern-Gerlach apparatus usually emerge from a heated gas. If we think in terms of classical physics, we would expect that the magnetic moments in the incident beam have randomly distributed directions. All the possible values $-|\boldsymbol{\mu}| \leq \mu_z \leq |\boldsymbol{\mu}|$ would occur, and the force on the particles would vary over all values between $-|\boldsymbol{\mu}|\,\nabla(|\mathbf{B}(\mathbf{x})|)$ and $+|\boldsymbol{\mu}|\,\nabla(|\mathbf{B}(\mathbf{x})|)$. Hence, classical physics predicts a continuous broadening of the beam in the vertical direction. On the screen behind the apparatus, the particles would leave a trace in the form of a continuous vertical line as in Figure 3.2(b). Instead, one finds a splitting in discrete partial beams, as indicated in Figure 3.2(c).



CD 3.4 schematically shows the Stern-Gerlach experiment for classical neutral particles. All particles have magnetic moments of the same size, but with randomly chosen directions. In the inhomogenous magnetic field of a Stern-Gerlach apparatus, the particles get deflected up or down according to the (random) value of the component μ_z. In CD 3.4.3 we explain that the Stern-Gerlach apparatus could be replaced by an arrangement of two or four parallel wires.

In quantum mechanics, we can try to explain the outcome of the Stern-Gerlach experiment by the quantization of the angular momentum. We assume that the magnetic moment $\boldsymbol{\mu}$ of an atom is related to the angular momentum of the electron as in Eq. (3.1). Because of (3.15), the splitting of the beam into n partial beams then just means that the component of the angular momentum in the vertical direction has precisely n distinct values.

Indeed, this is precisely what we expect from our analysis of the angular momentum in Chapter 1. The vertical component of the orbital angular momentum can only have $2\ell + 1 = 1, 3, 5, \ldots$ different values, corresponding to the angular-momentum quantum numbers $\ell = 0, 1, 2, \ldots$. In any case, we expect a splitting of the beam into an odd number of components. In particular, for hydrogen atoms in their ground state, we would expect no splitting at all, because the orbital angular momentum is zero in that case. But, surprisingly, the Stern-Gerlach experiment with hydrogen atoms in their ground state shows a splitting of the beam in two components, as shown in Figure 3.2(c).

Therefore, hydrogen atoms must have a magnetic moment, even if the electron is in the ground state with orbital angular momentum zero. What

is the origin of that magnetic moment if it is not the orbital angular momentum? As stated above, the nucleus does have a magnetic moment, but it is way too small to acount for the observed splitting.

One concludes that the origin of the atomic magnetic moment must lie in the electron. One assumes that the electron itself has a magnetic moment, whose vertical component can only have two distinct values. With a Stern-Gerlach experiment, one can determine these values quantitatively. One finds that the z-component of the electronic magnetic moment has the values

$$\mu_z = \pm \frac{e\hbar}{2\mathrm{m}_e} = \pm \mu_\mathrm{B}. \tag{3.16}$$

The splitting into an even number of partial beams in a Stern-Gerlach experiment cannot be explained with the quantization of an orbital angular momentum (which would give a splitting into an odd number of partial beams). Hence, the magnetic moment of the electron is not related to its orbital motion or to the angular momentum of an extended charge distribution inside the electron. The electron has no "inner structure" in that sense. The electron's magnetic moment has to be considered simply as an intrinsic property (like the charge or mass).



CD 3.5 shows quantum wave packets in a Stern-Gerlach apparatus. The visualization shows the position probability density only. When passing the inhomogeneous magnetic field, any wave packet would split into precisely two spatially separated parts. Hence, a general wave packet may be interpreted as a superposition of two types of wave packets, both having the same initial position and momentum distribution, but showing an opposite behavior in an inhomogeneous magnetic field: When sent through a Stern-Gerlach apparatus, type 1 gets pushed upwards, and type 2 moves downwards. CD 3.6 shows phase-colored plots of the two types of wave packets.



A variant of the Stern-Gerlach experiment can be used to prepare wave packets of particular type. This is shown in CD 3.7. By putting an obstacle in the lower part of a Stern-Gerlach apparatus, we can make sure that any wave packet leaving the apparatus is of type 1. Such a wave packet would again be deflected upwards in a second Stern-Gerlach apparatus.

The result of the Stern-Gerlach experiment does not depend on the orientation of the inhomogenous magnetic field. One can, for example, rotate the Stern-Gerlach apparatus about the x-axis (direction of the beam) by an arbitrary angle. Then, one observes the same splitting of the beam into

two components. One is led to the (paradoxical) conclusion that the component of the magnetic moment along any direction in space can have just two values $\pm\mu_{\mathrm{B}}$.

3.4. The Spin Operators

3.4.1. Magnetic moment and spin

Every component of the electron's magnetic moment is an observable (it can be measured by a Stern-Gerlach apparatus). In the quantum mechanical formalism, these observables should be represented by self-adjoint operators in the Hilbert space of the particle. Here, we are going to postulate the existence of operators $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ with suitable properties. In analogy to (3.4) we write

$$\boldsymbol{\mu} = -g\,\mu_{\mathrm{B}}\,\frac{\mathbf{S}}{\hbar}, \tag{3.17}$$

with an unknown operator $\mathbf{S} = (S_1, S_2, S_3)$. If g is a number, then $\mathbf{S}$ must correspond to a physical quantity with the dimension of angular momentum. This quantity will be called the *spin angular momentum* or simply the *spin* of the electron. We assume that the components of $\mathbf{S}$ satisfy the angular-momentum commutation relations and call $\mathbf{S}$ the *spin operator* of the particle. We will make no attempt to express $\mathbf{S}$ in terms of position and momentum operators, because we know that the spin cannot be interpreted as an orbital angular momentum.

The splitting of a beam of hydrogen atoms into two components can now be explained by assuming that the electron (which has orbital angular momentum zero) has spin 1/2. More precisely, we mean by this that the square S^2 of the spin operator has the eigenvalue $\hbar^2\, s(s+1)$ with $s = 1/2$. Theorem 1.1 tells us that $s = 1/2$ is indeed among the possible angular-momentum quantum numbers. Moreover, for $s = 1/2$ there are precisely $2s + 1 = 2$ eigenvalues of the vertical component S_3, namely $+\hbar/2$ and $-\hbar/2$. Likewise, because of (3.17), the vertical component μ_3 of $\boldsymbol{\mu}$ can only have two different values.

We are going to discuss the formal consequences of assuming the existence of spin operators in Section 3.5 below. But first, we want to compare our results with the outcome of a Stern-Gerlach experiment.

3.4.2. The g-factor

Taking into account that the possible values of S_3 are $\pm\hbar/2$, we obtain from Eq. (3.17) for the z-component $\mu_z = \mu_3$ of $\boldsymbol{\mu}$ the values

$$\mu_3 = \mp\, g\,\mu_{\mathrm{B}}\,\frac{S_3}{\hbar} = \mp\frac{g}{2}\,\mu_{\mathrm{B}}. \tag{3.18}$$

Comparing this result with the actually measured values (3.16), we find that electrons obviously have the g-factor $g = 2$. We note that the value $g = 2$ is predicted by the relativistic Dirac equation, which is considered one of its big successes.

Because of the negative charge of the electron, the magnetic moment is opposite to its spin. The negative sign in (3.18) corresponds to a positive eigenvalue of S_3. The eigenstates of S_3 belonging to the positive eigenvalue $+\hbar/2$ are called *spin-up eigenstates*. When passing a Stern-Gerlach apparatus with a field gradient in the positive z-direction, a spin-up electron is deflected downwards. The eigenstates of S_3 belonging to the eigenvalue $-\hbar/2$ (the *spin-down eigenstates*) describe the opposite behavior. The two eigenvalues of S_3 just correspond to the two types of behavior of electrons in a Stern-Gerlach experiment.

The magnetic moment of an electron:

Electrons are spin 1/2 particles with g-factor $g = 2$. Any component of its magnetic moment can have only two values. In particular,

$$\mu_3 = \mp\mu_{\mathrm{B}}. \tag{3.19}$$

All components of an angular-momentum operator $\mathbf{S}$ have the same eigenvalues. However, different components of $\mathbf{S}$ do not have simultaneous eigenvectors. We can choose only one component (usually S_3, but this choice is arbitrary) and find simultaneous eigenvectors of that component and of S^2 (see Section 1.5). The arbitrariness of selecting the third component S_3 corresponds to the experimental observation that the choice of the vertical axis as the direction of the field in the Stern-Gerlach experiment is completely irrelevant. The experiment gives the same result with respect to any direction in space.

The g-factor can be measured very precisely, and it turns out that the actual value is slightly different from two. The presently known value[7]. is

$$g = 2.0023193043738 \pm 0.0000000000082. \tag{3.20}$$

The small deviation from the value 2 is called the *anomalous magnetic moment*. The origin of the anomalous magnetic moment can be explained with quantum electrodynamics.

It turns out that the magnetic moment due to the orbital angular momentum has the g-factor 1, as predicted by classical electrodynamics. (In

[7]Groom, D.E., et al. (Particle Data Group), Eur. Phys. J. **C15**, 1 (2000) and 2001 partial update for edition 2002 (URL: http://pdg/lbl.gov)

principle, this can be measured by considering the splitting of a beam of excited hydrogen atoms sent through a Stern-Gerlach apparatus.) An electron with orbital angular momentum $\mathbf{L}$ thus has the magnetic moment

$$\boldsymbol{\mu} = -\mu_{\mathrm{B}} \frac{\mathbf{L} + 2\mathbf{S}}{\hbar} = -\mu_{\mathrm{B}} \frac{\mathbf{J} + \mathbf{S}}{\hbar}, \tag{3.21}$$

where $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is called the *total angular momentum* of the electron. If it is necessary to distinguish between the various types of magnetic moments, one denotes by $\boldsymbol{\mu}_{\mathrm{L}}$ the magnetic moment related to the orbital angular momentum and by $\boldsymbol{\mu}_{\mathrm{S}}$ the magnetic moment due to the spin.

$$\boldsymbol{\mu}_{\mathrm{L}} = -\mu_{\mathrm{B}} \frac{\mathbf{L}}{\hbar}, \qquad \boldsymbol{\mu}_{\mathrm{S}} = -2\,\mu_{\mathrm{B}} \frac{\mathbf{S}}{\hbar}. \tag{3.22}$$

Other particles also have spin 1/2, and the results of this chapter can be easily adapted for these other particles. For example, the proton is a spin 1/2 particle. It has the magnetic moment

$$\boldsymbol{\mu} = g_p\, \mu_{\mathrm{N}} \frac{\mathbf{S}}{\hbar}, \quad \mu_{\mathrm{N}} = \frac{e\hbar}{2\mathrm{m}_p} \tag{3.23}$$

where μ_{N} is called *nuclear magneton*. It is smaller by a factor $\mathrm{m}_e/\mathrm{m}_p$ than the Bohr magneton, $\mu_{\mathrm{N}} \approx 5.05 \times 10^{-27}\,\mathrm{A\,m^2}$. The g-factor of the proton is rather large, $g_p \approx 5.59$. The neutron is also a spin 1/2 particle. Note that the neutron has no charge at all. But it does have a magnetic moment due to its spin, which is given in terms of the nuclear magneton by

$$\boldsymbol{\mu} = -g_n\, \mu_{\mathrm{N}} \frac{\mathbf{S}}{\hbar}, \quad \mu_{\mathrm{N}} = \frac{e\hbar}{2\mathrm{m}_p}, \quad \text{with } g_n \approx 3.83. \tag{3.24}$$

3.4.3. Electron in a constant magnetic field

A charged particle in a magnetic field is subject to the Lorentz force. The Lorentz force is always orthogonal to the magnetic field $\mathbf{B}$ and to the velocity $\mathbf{v}$ of the particle. Hence, the Lorentz force changes only the direction but not the magnitude of the velocity vector. For a classical electron with charge $-e$ and mass m_e, the Lorentz force causes the acceleration

$$\frac{d}{dt}\,\mathbf{v}(t) = -\frac{e}{\mathrm{m}_e}\,\mathbf{v}(t) \times \mathbf{B}. \tag{3.25}$$

Here, it is assumed that the magnetic field is homogeneous ($\mathbf{B}$ is independent of $\mathbf{x}$) because in an inhomogeneous magnetic field, the magnetic moment $\boldsymbol{\mu}$ would cause an additional acceleration, see Eq. (3.12).

Let us compare this with the classical equation of motion for the magnetic moment (3.13) with g-factor 2:

$$\frac{d}{dt}\,\boldsymbol{\mu}(t) = -\frac{e}{\mathrm{m}_e}\,\boldsymbol{\mu}(t) \times \mathbf{B}. \tag{3.26}$$

This equation is mathematically identical with the Lorentz-force equation. As a consequence, the velocity $\mathbf{v}(t)$ and the magnetic moment $\boldsymbol{\mu}(t)$ in a constant field are strictly "in phase." From (3.25) and (3.26) we obtain the following result.

Classical motion of electrons with magnetic moment:

For a classical particle with charge e and and g-factor 2 in a homogeneous magnetic field, the velocity $\mathbf{v}(t)$ and the magnetic moment $\boldsymbol{\mu}(t)$ both remain constant in magnitude and precess about the direction of $\mathbf{B}$ with a constant angular speed

$$\omega = \frac{e}{\mathrm{m}_e} |\mathbf{B}| = 2 \frac{\mu_\mathrm{B}}{\hbar} |\mathbf{B}| = 2\omega_\mathrm{L}. \tag{3.27}$$

The quantity ω_L is called *Larmor frequency*.

The anomalous magnetic moment of real electrons, however, causes a slight asynchronism of the orbital motion and the motion of the magnetic moment.



For suitable initial conditions, the classical motion of a charged particle in a homogeneous field takes place on a circle. In CD 3.2, we show the behavior of particles with various g-factors. If the initial velocity has a component in the direction of the field, the orbit is a helix, as shown in CD 3.3.

3.4.4. Properties of the spin operator

In order to describe the two different types of behavior of electrons in an inhomogenous magnetic field, we have to postulate that the magnetic moment of the particles is related to a spin operator $\mathbf{S}$ with the quantum number $s = 1/2$. The components of $\mathbf{S}$ are to be defined as self-adjoint operators in an appropriate Hilbert space. They should satisfy commutation relations characteristic of angular-momentum operators, so that the results of Section 1.5 can be applied. In the absence of magnetic fields, this operator should not disturb the "normal behavior" of electrons, in order to be compatible with our previously obtained results. In particular, we assume that the spin operator commutes with all components of position and momentum. Let us collect our assumptions on the spin operator in the following box.

The spin postulate:

The spin is represented by three self-adjoint operators $\mathbf{S} = (S_1, S_2, S_3)$ satisfying the angular-momentum commutation relations

$$[S_i\,,\,S_j] = \sum_{k=1}^{3} \mathrm{i}\hbar\,\epsilon_{ijk}\,S_k, \tag{3.28}$$

and commuting with position and momentum operators. The possible states of electrons have to be described by eigenstates of the operator S^2 belonging to the quantum number $s = 1/2$. In the Hilbert space of an electron, the operator S^2 must be a multiple of the identity operator,

$$S^2 = S_1^2 + S_2^2 + S_3^2 = \hbar^2\, s(s+1)\,\mathbf{1} = \hbar^2\,\frac{3}{4}\,\mathbf{1}. \tag{3.29}$$

It follows that every component S_k has precisely two eigenvalues $\pm\hbar/2$.

Ψ Because of (3.29), all components of the spin are bounded operators that are defined everywhere in the Hilbert space.

We can choose one component, which is usually S_3, and decompose the Hilbert space into the corresponding eigenspaces,

$$\mathfrak{H} = \mathfrak{H}_1 \oplus \mathfrak{H}_2 \quad \text{(orthogonal direct sum)}. \tag{3.30}$$

Here, $\mathfrak{H}_1$ is the eigenspace of S_3 belonging to the eigenvalue $+\hbar/2$. We say that the states in this subspace have *spin-up*. Similarly, the eigenspace $\mathfrak{H}_2$ belonging to the eigenvalue $-\hbar/2$ contains the states with *spin-down*.

In the next section we are going to construct a suitable Hilbert space that allows us to define spin operators with the desired properties.

3.5. Spinor-Wave Functions

3.5.1. A Hilbert space for a spin 1/2 particle

In the sole presence of an electrostatic field or a homogeneous magnetic field, the state of the spin has no influence on the spatial motion of an electron.[8] In these situations, the electron's space-time behavior is still described by a solution of the ordinary Schrödinger equation, no matter whether the electron has spin-up or spin-down. Hence, we can describe the spin state simply by attaching an index to the wave function. We write $\psi_1(\mathbf{x}, t)$ for a wave function of a particle with spin-up and $\psi_2(\mathbf{x}, t)$ if we want to indicate that

[8] This is true, at least, within a reasonable approximation. See the discussion in Section 3.9.3 later in this chapter.

the particle has spin-down. The two types of electron wave functions should behave in the same way unless there is an inhomogeneous magnetic field.

Thus, we construct a Hilbert space of a particle with spin 1/2 in such a way that $L^2(\mathbb{R}^3)$, the Hilbert space of "ordinary" wave functions, can be identified with both eigenspaces of S_3. Because of (3.30), we have to form the direct sum of two copies of $L^2(\mathbb{R}^3)$.

The state space for particles with spin 1/2:

The Hilbert space for a particle with spin 1/2 is isomorphic to the direct sum of two copies of $L^2(\mathbb{R}^3)$,

$$\mathfrak{H} \cong L^2(\mathbb{R}^3)^2 = L^2(\mathbb{R}^3) \oplus L^2(\mathbb{R}^3). \tag{3.31}$$

The elements of $L^2(\mathbb{R}^3) \oplus L^2(\mathbb{R}^3)$ can be obtained by combining two square-integrable wave functions ψ_1 and ψ_2 into a column vector with two components. Hence, the wave function of a particle with spin is assumed to be of the form

$$\psi(\mathbf{x}) = \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix}, \quad \text{with } \psi_i \in L^2(\mathbb{R}^3), \text{ for } i = 1, 2. \tag{3.32}$$

Wave functions of this type are called *spinor-wave functions* or *spinors*.

When we consider dynamical processes, ψ will also depend on time. We write

$$\psi(\mathbf{x}, t) = \begin{pmatrix} \psi_1(\mathbf{x}, t) \\ \psi_2(\mathbf{x}, t) \end{pmatrix}, \tag{3.33}$$

and this means that for each time t, the two components of ψ are square-integrable functions of $\mathbf{x}$.

Ψ A spinor-wave function (3.32) can also be interpreted as a function ψ of $\mathbf{x}$ with values in $\mathbb{C}^2$, the two-dimensional complex linear space whose vectors are pairs of complex numbers,

$$\psi : \mathbb{R}^3 \to \mathbb{C}^2. \tag{3.34}$$

By this we emphasize that at each point of space, the electron has two internal degrees of freedom. The Hilbert space of $\mathbb{C}^2$-valued functions is denoted by $L^2(\mathbb{R}^3, \mathbb{C}^2)$, and it is isomorphic to $L^2(\mathbb{R}^3)^2$. Another way of writing the Hilbert space of spinor-wave functions is the tensor product

$$L^2(\mathbb{R}^3) \otimes \mathbb{C}^2. \tag{3.35}$$

We are going to define the tensor product of Hilbert spaces in Chapter 5 (see, in particular, Section 5.2.5).

We have pointed out already that the discrimination of S_3 was an arbitrary decision. We could have started with any other component of the spin and ended up with a different (but equivalent) description. This will be discussed further in Section 3.5.3 below. The description obtained here will be called the *standard representation*. It is characterized by the choice of a particular Cartesian coordinate system in space (that is, a fixed z-axis) and by the choice of the operator S_3 and its eigenspaces in the construction of a Hilbert space.

3.5.2. Spin operators in the standard representation

In the standard representation, spinor-wave functions of the type

$$\psi_{\text{up}}(\mathbf{x}) = \begin{pmatrix} \psi_1(\mathbf{x}) \\ 0 \end{pmatrix}, \qquad \psi_{\text{down}}(\mathbf{x}) = \begin{pmatrix} 0 \\ \psi_2(\mathbf{x}) \end{pmatrix}, \tag{3.36}$$

correspond to particles with spin-up or spin-down, respectively. These are eigenvectors of the operator S_3 corresponding to the two possible eigenvalues $\pm\hbar/2$. Hence, in the Hilbert space $L^2(\mathbb{R}^3)^2$, the operator S_3 must have the form

$$S_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.37}$$

This is a self-adjoint operator, defined everywhere in $L^2(\mathbb{R}^3)^2$. Its action on a spinor-wave function consists in changing the sign of the lower component and multiplying everything with $\hbar/2$. The observable corresponding to this operator is interpreted as the component of the spin into the z-direction of a fixed Cartesian coordinate system.

It is now rather straightforward to guess the operators that describe the other components S_1 and S_2 of the spin. In Exercise 1.22, we introduced the three Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.38}$$

These matrices are often formally combined into a vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. It is easy to verify (Exercise 1.22) that the three components of

$$\mathbf{S} = \frac{\hbar}{2}\,\boldsymbol{\sigma} \tag{3.39}$$

satisfy the angular-momentum commutation relations (3.28). Hence, the spin observable can be represented by the three matrices $S_j = \sigma_j/2$, $j = 1, 2, 3$. You can also verify that

$$S^2 = S_1^2 + S_2^2 + S_3^2 = \hbar^2\,\frac{3}{4}\,\mathbf{1}_2 = \hbar^2\, s(s+1)\,\mathbf{1}_2 \quad \text{with} \quad s = \frac{1}{2}. \tag{3.40}$$

Hence, all wave functions in the spinor Hilbert space $\mathfrak{H}$ are indeed eigenvectors of S^2 with spin quantum number $1/2$.

The spin operators (standard representation):

In the standard representation, the spin $\mathbf{S} = (S_1, S_2, S_3)$ is defined in terms of the Pauli σ-matrices by $\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}$. All spinors in the Hilbert space $L^2(\mathbb{R}^3)^2$ are eigenvectors of

$$S^2 = \mathbf{S} \cdot \mathbf{S} = \hbar^2 \frac{3}{4} \mathbf{1}_2. \tag{3.41}$$

The kinetic energy operator, the momentum and position operators are assumed to act component-wise on the spinors in $L^2(\mathbb{R}^3)^2$. For example,

$$-\frac{\hbar^2}{2\mathrm{m}} \Delta\psi(\mathbf{x}) = -\frac{\hbar^2}{2\mathrm{m}} \begin{pmatrix} \Delta\psi_1(\mathbf{x}) \\ \Delta\psi_2(\mathbf{x}) \end{pmatrix}. \tag{3.42}$$

It follows immediately that all these operators commute with all components of the spin operator.

EXERCISE 3.5. *Describe the spinor-wave functions in $\mathfrak{H}$ that are eigenvectors of the spin component S_1.*

EXERCISE 3.6. *Verify that in the Hilbert space $L^2(\mathbb{R}^3)^2$, the operators of kinetic energy, position, momentum, and orbital angular momentum all commute with S_1, S_2, and S_3.*

3.5.3. Changing representations

The notion of "spin-up" and "spin-down" refers to the arbitrarily chosen direction of the z-axis in $\mathbb{R}^3$. For example, one can choose the direction defined by the inhomogeneous magnetic field in the Stern-Gerlach apparatus as the z-direction in space. Equivalently, we can assume a magnetic field in the y-direction (just rotate the Stern-Gerlach apparatus in Figure 3.1 by 90 degrees about the x-axis) and start our construction of the spinor-Hilbert space with the eigenspaces of S_2. The operator S_2 also has the two eigenvalues $\pm\hbar/2$, and the construction described in Section 3.5.1 again leads to the Hilbert space $L^2(\mathbb{R}^3)^2$. But now the interpretation of the spinors (3.36) is different, because "spin-up" and "spin-down" now refer to the y-direction (looking in the direction of the beam in Figure 3.1, we could call this "spin-left" and "spin-right"). Moreover, in this Hilbert space, it is the operator S_2 that is represented by the diagonal matrix (3.37).

How is this representation related to the standard representation where S_3 is diagonal? It is clear that physical predictions should not depend on

which axis in space is chosen to define spin-up and spin-down or—put in a mathematical language—which component S_k of the spin is chosen to be diagonal. Hence, this relation must be a symmetry transformation. In the Hilbert space of the system, it must be possible to implement this symmetry transformation by a unitary or antiunitary operator.

In the standard representation, the eigenvectors of S_2 are given by

$$\psi_l = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}, \qquad \psi_r = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix}. \tag{3.43}$$

The unitary matrix U that maps the eigenvectors of S_3 onto the eigenvectors of S_2 is given by

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix}. \tag{3.44}$$

This matrix defines a unitary operator in $L^2(\mathbb{R}^3)^2$. We can apply it to an arbitrary spinor-wave function,

$$U \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix} = \psi_1(\mathbf{x})\, \psi_l + \psi_2(\mathbf{x})\, \psi_r. \tag{3.45}$$

The result is spinor whose components in the C^2-basis $\{\psi_l, \psi_r\}$ are just ψ_1 and ψ_2. With respect to this basis, S_2 is diagonal. Hence, the operator U just maps the standard representation to the representation where S_2 is diagonal.

Ψ The operator U actually implements a rotation through an angle $-\pi/2$ about the x-axis. Arbitrary rotations and unitary transformations in the Hilbert space of spinors will be discussed in Chapter 4 (see Section 4.4.2).

3.5.4. Interpretation of spinor-wave functions

The scalar product in $\mathfrak{H} = L^2(\mathbb{R}^3)^2$ is

$$\langle \psi, \phi \rangle = \int_{\mathbb{R}^3} \left(\overline{\psi_1(\mathbf{x})}\, \phi_1(\mathbf{x}) + \overline{\psi_2(\mathbf{x})}\, \phi_2(\mathbf{x}) \right) d^3x \tag{3.46}$$

and the associated norm $\| \cdot \|$ is given by

$$\|\psi\|^2 = \int_{\mathbb{R}^3} \left(|\psi_1(\mathbf{x})|^2 + |\psi_2(\mathbf{x})|^2 \right) d^3x. \tag{3.47}$$

Let us introduce the following abbreviation for the scalar product of vectors ψ and ϕ in $\mathbb{C}^2$,

$$\langle \psi, \phi \rangle_2 = \overline{\psi_1}\, \phi_1 + \overline{\psi_2}\, \phi_2, \tag{3.48}$$

and denote the norm in $\mathbb{C}^2$ by

$$\left| \psi \right|_2^2 = |\psi_1|^2 + |\psi_2|^2 = \langle \psi, \psi \rangle_2. \tag{3.49}$$

Then the scalar product (3.46) and the norm (3.47) can be simply written as

$$\langle \psi, \phi \rangle = \int_{\mathbb{R}^3} \big\langle \psi(\mathbf{x}), \phi(\mathbf{x}) \big\rangle_2 \, d^3x, \quad \|\psi\|^2 = \int_{\mathbb{R}^3} \big|\psi(\mathbf{x})\big|_2^2 \, d^3x. \tag{3.50}$$

For $B \subset \mathbb{R}^3$ we interpret the expression

$$p(B) = \int_B \big|\psi(\mathbf{x})\big|_2^2 \, d^3x \tag{3.51}$$

as the probability of finding the particle in the region B. Correspondingly, we define the partial probabilities

$$p_i(B) = \int_B |\psi_i(\mathbf{x})|^2 \, d^3x. \tag{3.52}$$

The interpretation is as follows: p_1 is the probability of finding the particle in the region B with spin-up, and p_2 is the probability of finding the particle in B with spin-down. We have $p(B) = p_1(B) + p_2(B)$.

3.5.5. Visualization of spinor-wave functions

Visualization by vectors. A spinor-wave function associates two complex numbers with each point $\mathbf{x} \in \mathbb{R}^3$. Hence, it consists of four independent real-valued functions, the real and imaginary parts of the components ψ_1 and ψ_2. As described already in Book One, Chapter 1, we can introduce the following vector field:

$$\mathfrak{v}(\mathbf{x}) = \begin{pmatrix} \big\langle \psi(\mathbf{x}), \sigma_1 \, \psi(\mathbf{x}) \big\rangle_2 \\ \big\langle \psi(\mathbf{x}), \sigma_2 \, \psi(\mathbf{x}) \big\rangle_2 \\ \big\langle \psi(\mathbf{x}), \sigma_3 \, \psi(\mathbf{x}) \big\rangle_2 \end{pmatrix}, \quad \text{where } \psi(\mathbf{x}) = \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix} \in \mathbb{C}^2. \tag{3.53}$$

Here, we used the $\mathbb{C}^2$ scalar product introduced in Eq. (3.48). The length of the vector $\mathfrak{v}$ is given by the position probability density

$$\sqrt{\mathfrak{v}_1(\mathbf{x})^2 + \mathfrak{v}_2(\mathbf{x})^2 + \mathfrak{v}_3(\mathbf{x})^2} = \big|\psi(\mathbf{x})\big|_2^2. \tag{3.54}$$

The vector field $\mathfrak{v}(\mathbf{x})$ can be visualized by arrows attached to a grid of $\mathbf{x}$-values. Our visualizations, however, often show the vector field $\vec{v}(\mathbf{x}) = \mathfrak{v}(\mathbf{x})/|\psi(\mathbf{x})|_2$ in order to reduce the contrast between the smallest and the largest values. Note that

$$\int_{\mathbb{R}^3} \mathfrak{v}_i(\mathbf{x}) \, d^3x = \int_{\mathbb{R}^3} \big\langle \psi(x), \sigma_i \, \psi(x) \big\rangle_2 \, d^3x = \langle \psi, \sigma_i \, \psi \rangle = \langle \sigma_i \rangle_\psi \tag{3.55}$$

is just the expectation value of the ith component of $\boldsymbol{\sigma}$. The vector field $\mathfrak{v}$ describes the density of the expectation value of $\boldsymbol{\sigma}$,

$$\int_{\mathbb{R}^3} \mathfrak{v}(\mathbf{x}) \, d^3x = \langle \boldsymbol{\sigma} \rangle_\psi = \frac{2}{\hbar} \, \langle \mathbf{S} \rangle_\psi. \tag{3.56}$$

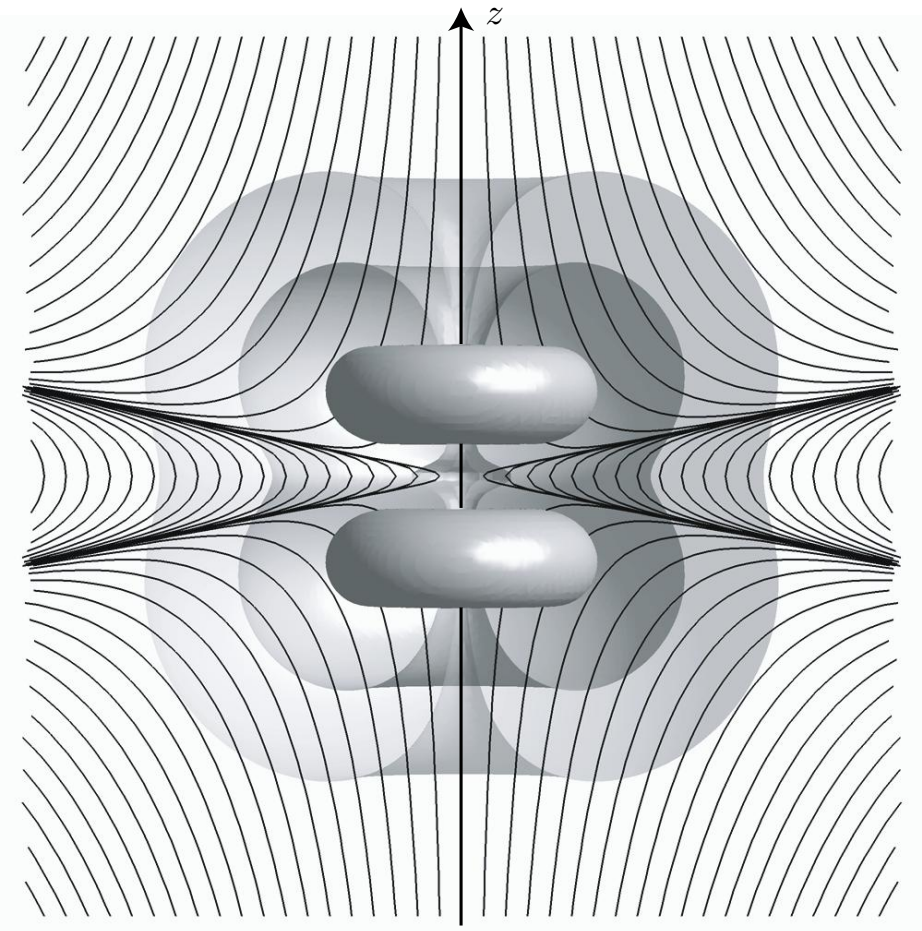


FIGURE 3.3. A spinor-wave function of the hydrogen atom. The gray isosurface indicate three levels of the modulus of the wave function. The lines are the flux lines of the vector field generated by the spin. The figure is symmetric with respect to rotations about the z-axis. The quantum numbers for this state are $n = 3$, $\kappa = 3$, and $m_j = 3/2$ (see Section 3.9.4).

The direction of $\mathfrak{v}(\mathbf{x})$ at a point $\mathbf{x}$ gives the local spin-up direction of the spinor-wave function. An example is given in Figure 3.3.

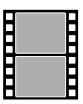
The mapping between spinors $\psi(\mathbf{x})$ and vectors $\mathfrak{v}(\mathbf{x})$ will be discussed further in Section 4.4.1. Here, we just note that this correspondence is not one-to-one. Multiplication of the spinor $\psi(x)$ with a phase factor leaves the functions $\mathfrak{v}_i(\mathbf{x})$ unchanged.



By plotting arrows on a regular grid of space points, CD 3.8.1 visualizes the spin-vector field $\vec{v}(\mathbf{x})$ associated with a spinor-wave function in two dimensions. A spinor-wave packet splits into parts with spin-up and spin-down due to the influence of an inhomogeneous magnetic field. In CD 3.8.2 and CD 3.9.1, we use small "magnetic needles" whose two poles are colored according to the complex values of the two components of the spinor. Whereas this method is useful for a single spinor in $\mathbb{C}^2$ (see CD 4.8), it is not capable of showing finer details of a spinor field. In CD 3.8.3 and CD 3.9.2, we use enlarged pixels, each containing the two colors derived from the two spinor components via the standard color map. This method shows the information about the phases of the two components in great detail (in particular, in regions where the parts with spin-up and spin-down are well separated), but is not useful to depict the local spin-direction.

Vizualization by colors. In order to display finer spatial details of a spinor field, arrays of vectors are not very useful. In these cases we use a color map. We can use the coloring of the unit sphere in the HLS color system (as described in Book One) to describe a direction in $\mathbb{R}^3$. According to this method, the spin-up vector $\mathfrak{v} = (0, 0, 1)$ is represented by white, and spin-down is black.

In addition, the absolute value can be symbolized by the saturation of the color. If the wave function is very small or zero, this will be represented by a saturation close to zero, that is, by the color gray. As the (three-dimensional) color information is derived from the vector $\mathfrak{v}(\mathbf{x})$, this method of visualization again discards the phase information in the spinor $\psi(x)$.



CD 3.10 describes the color map for vectors in three dimensions. We use this color map to describe the local expectation value of the spinor field, that is, the vector field defined in (3.53). Examples for the use of this color map can be found in CD 3.11 and CD 3.12.

3.6. The Pauli Equation

3.6.1. The Pauli operator

In the following sections, we are going to describe the quantum dynamics of an electron in a magnetic field. We start with a pure magnetic field, that is, we assume that there are no other external forces. The Hamiltonian of an electron in a pure magnetic field is obtained by adding the potential energy (3.10) of the magnetic moment interaction to the usual expression for the kinetic energy in the magnetic field. This gives (writing m_e for the mass and

$-e$ for the charge)

$$H_{\text{Pauli}} = \frac{1}{2\mathrm{m}_e}\big(\mathbf{p} + e\,\mathbf{A}(\mathbf{x})\big)^2 - \boldsymbol{\mu}\cdot\mathbf{B}(\mathbf{x}), \tag{3.57}$$

with $\mathbf{p} = -\mathrm{i}\hbar\nabla$, acting component-wise on spinor-wave functions. The operator H_{Pauli} will be called the *Pauli operator* for an electron in a pure magnetic field. Here, $\mathbf{A}$ is the vector potential, and $\mathbf{B} = \nabla\times\mathbf{A}$ is the magnetic field.

It is useful to write the expression (3.57) in the standard representation, where S_3 is diagonal. We assume, for simplicity, that the g-factor is 2. Then

$$\boldsymbol{\mu} = -\frac{e\hbar}{2\mathrm{m}_e}\,\boldsymbol{\sigma} = -\mu_{\mathrm{B}}\,\boldsymbol{\sigma}. \tag{3.58}$$

For the negatively charged electrons, the magnetic moment vector $\boldsymbol{\mu}$ is antiparallel to the spin. The matrix for the potential energy in the magnetic field becomes

$$-\boldsymbol{\mu}\cdot\mathbf{B}(\mathbf{x}) = \mu_{\mathrm{B}}\begin{pmatrix} B_3(\mathbf{x}) & B_1(\mathbf{x}) - \mathrm{i}B_2(\mathbf{x}) \\ B_1(\mathbf{x}) + \mathrm{i}B_2(\mathbf{x}) & -B_3(\mathbf{x}) \end{pmatrix}. \tag{3.59}$$

The Pauli operator in the standard representation acts on two-component wave functions and is, in fact, a matrix-differential operator

$$H_{\text{Pauli}} = \begin{pmatrix} \frac{1}{2\mathrm{m}_e}\big(-\mathrm{i}\hbar\nabla + e\,\mathbf{A}\big)^2 - \mu_{\mathrm{B}}\,B_3 & -\mu_{\mathrm{B}}(B_1 - \mathrm{i}B_2) \\ -\mu_{\mathrm{B}}(B_1 + \mathrm{i}B_2) & \frac{1}{2\mathrm{m}_e}\big(-\mathrm{i}\hbar\nabla + e\,\mathbf{A}\big)^2 + \mu_{\mathrm{B}}\,B_3 \end{pmatrix}. \tag{3.60}$$

If necessary, we may add to this operator an electrostatic potential energy $V(\mathbf{x})\,\mathbf{1}_2$. This is a diagonal matrix multiplication operator. That is, the function $V(\mathbf{x})$ has to be added in the main diagonal of the matrix H_{Pauli}.

The evolution equation with the Pauli operator

$$\mathrm{i}\,\hbar\,\frac{d}{dt}\,\psi(t) = H_{\text{Pauli}}\,\psi(t) \tag{3.61}$$

is called the *Pauli equation*. It describes the time evolution of a particle with spin 1/2 in an external magnetic field.

The matrix-form of the Pauli operator in the standard representation shows that in the absence of a magnetic field, the Pauli equation decouples into two completely unrelated Schrödinger equation equations. Only the presence of a magnetic field with nonzero B_1 or B_2 introduces a coupling between the upper and the lower component of the spinor-wave function.[9]

Having introduced the Pauli operator with all physical constants in the international system of units, we now switch to atomic units, where $\mathrm{m}_e =$

[9]This is only true in an approximate sense, because relativistic effects are neglected here. See the discussion in Section 3.9.3.

$e = \hbar = 1$, and $\mu_{\mathrm{B}} = 1/2$. A simple scaling transformation will bring you back, as described in Section 2.7.

3.6.2. Magnetic fields with constant direction

A special case is given by the magnetic fields $\mathbf{B}(\mathbf{x})$ with constant direction. We choose this direction as the x_3-direction and conclude from $\nabla \cdot \mathbf{B} = 0$ that $\mathbf{B}$ does not depend on x_3 at all (see Book One, Section 8.3). Hence, a magnetic field with constant direction is of the form $\mathbf{B}(\mathbf{x}) = (0, 0, B(x_1, x_2))$. We remind the reader that the third component B of the magnetic field is just given by the third component of $\nabla \times \mathbf{A}$,

$$B(x_1, x_2) = \frac{\partial}{\partial x_1} A_2(x_1, x_2) - \frac{\partial}{\partial x_2} A_1(x_1, x_2). \tag{3.62}$$

Moreover, we can assume that $A_3(x_1, x_2) = 0$.

In atomic units, the Pauli operator in the standard representation is the diagonal matrix

$$H_{\mathrm{Pauli}} = \frac{1}{2} \sum_{i=1}^{3} \Big(-\mathrm{i} \frac{\partial}{\partial x_i} + A_i(x_1, x_2) \Big)^2 \mathbf{1}_2 + \frac{1}{2} B(x_1, x_2)\, \sigma_3. \tag{3.63}$$

($\mathbf{1}_2$ is the two-dimensional unit matrix). We note that the sign in front of the vector potential and the magnetic field reflects the negative charge of an electron (-1 in atomic units).

The Pauli equation can now be separated into a part that describes the free motion in the x_3-direction and a part that describes the planar motion in a magnetic field. The Pauli equation for the motion in the x_1x_2-plane consists of two independent equations, because H_{Pauli} is a diagonal matrix. With $\mathbf{x} = (x_1, x_2)$, $\nabla = (\partial_1, \partial_2)$, and $\mathbf{A} = (A_1, A_2)$, we write

$$\begin{aligned} \mathrm{i} \frac{d}{dt} \psi_1 &= \frac{1}{2} \big(-\mathrm{i}\nabla + \mathbf{A}(\mathbf{x}) \big)^2 \psi_1 + \frac{1}{2} B(\mathbf{x})\, \psi_1, \\ \mathrm{i} \frac{d}{dt} \psi_2 &= \frac{1}{2} \big(-\mathrm{i}\nabla + \mathbf{A}(\mathbf{x}) \big)^2 \psi_2 - \frac{1}{2} B(\mathbf{x})\, \psi_2. \end{aligned} \tag{3.64}$$

Hence, a particle that starts with spin-up (that is, $\psi_2 = 0$) will remain spin-up all the time. The third component of the spin is a constant of motion. Indeed, we see immediately from (3.63) that H_{Pauli} commutes with $S_3 = \sigma_3/2$.

EXERCISE 3.7. *A neutron is a neutral particle with an anomalous magnetic moment $\boldsymbol{\mu}$. Set up the Pauli equation for a neutron in a magnetic field in two dimensions. As an example, consider the magnetic field $B(x, y) = y$. Use the Avron-Herbst formula (Book One) to give a solution of the corresponding initial-value problem.*

3.6.3. Supersymmetric structure

It is an important observation that the Pauli operator with a purely magnetic field is the square of another operator. Consider the operator

$$\mathcal{D} = \boldsymbol{\sigma} \cdot \big(-\mathrm{i}\nabla + \mathbf{A}(\mathbf{x})\big) = \sum_{i=1}^{3} \sigma_i \Big(-\mathrm{i}\frac{\partial}{\partial x_i} + A_i(\mathbf{x})\Big). \tag{3.65}$$

A little calculation using the formula

$$(\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} + \mathrm{i}\boldsymbol{\sigma} \cdot (\mathbf{A} \times \mathbf{B}) \tag{3.66}$$

gives (with $\boldsymbol{\mu} = -\boldsymbol{\sigma}/2$ in atomic units)

$$\begin{aligned} \frac{1}{2}\,\mathcal{D}^2 &= \frac{1}{2}\Big(\boldsymbol{\sigma} \cdot \big(\mathbf{p} + \mathbf{A}(\mathbf{x})\big)\Big)^2 \\ &= \frac{1}{2}\big(\mathbf{p} + \mathbf{A}(\mathbf{x})\big)^2 + \mathrm{i}\frac{1}{2}\boldsymbol{\sigma} \cdot \big(-\mathrm{i}\nabla \times \mathbf{A}(\mathbf{x})\big) \\ &= \frac{1}{2}\big(\mathbf{p} + \mathbf{A}(\mathbf{x})\big)^2 - \boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{x}) = H_{\text{Pauli}}. \end{aligned} \tag{3.67}$$

As the square of another operator, the Pauli operator is always nonnegative. There are no states with a negative total energy in a pure magnetic field.

The two-dimensional Pauli equation (3.64) can also be factored. Define the operator

$$D = -\mathrm{i}\frac{\partial}{\partial x_1} + A_1(\mathbf{x}) + \frac{\partial}{\partial x_2} + \mathrm{i}\,A_2(\mathbf{x}). \tag{3.68}$$

This operator is not self-adjoint. Its adjoint is given by

$$D^\dagger = -\mathrm{i}\frac{\partial}{\partial x_1} + A_1(\mathbf{x}) - \frac{\partial}{\partial x_2} - \mathrm{i}\,A_2(\mathbf{x}). \tag{3.69}$$

Nevertheless, the operator D is useful because we have

$$\begin{aligned} D^\dagger D &= \big(-\mathrm{i}\nabla + \mathbf{A}(\mathbf{x})\big)^2 + B(\mathbf{x}), \\ DD^\dagger &= \big(-\mathrm{i}\nabla + \mathbf{A}(\mathbf{x})\big)^2 - B(\mathbf{x}), \end{aligned} \tag{3.70}$$

with $B(\mathbf{x})$ as in (3.62), and hence

$$H_{\text{Pauli}} = \frac{1}{2}\begin{pmatrix} D^\dagger D & 0 \\ 0 & DD^\dagger \end{pmatrix}. \tag{3.71}$$

We remind the reader that the nonzero eigenvalues of the self-adjoint operators $D^\dagger D$ and $DD^\dagger$ are closely related. Whenever $D^\dagger D$ has an eigenvector ψ belonging to a nonzero eigenvalue λ, then $DD^\dagger$ has the same eigenvalue with corresponding eigenvector $D\psi$. Whenever $DD^\dagger$ has an eigenvalue with eigenvector ϕ, then $D^\dagger D$ has the same eigenvalue with eigenvector $D^\dagger\phi$ (you can check this by a short calculation; see also Exercise 2.5 and Book One, Section 7.6). This symmetric relation between the spectra of $D^\dagger D$ and $DD^\dagger$

is a very simple example of a so-called supersymmetry. We also note that the operators $D^\dagger D$ and $DD^\dagger$ are nonnegative, that is, all nonzero eigenvalues are positive.

3.7. Solution in a Homogeneous Magnetic Field

3.7.1. The g-factor of orbital motion

For a constant, homogeneous magnetic field, we may choose the coordinate system such that $\mathbf{B}(\mathbf{x}) = (0, 0, B)$ (for all $\mathbf{x}$). We write the vector potential in the Poincaré gauge (see Book One, Section 8.4),

$$\mathbf{A}(\mathbf{x}) = \frac{B}{2}\,(-y, x, 0), \quad \text{for all } \mathbf{x} = (x, y, z). \tag{3.72}$$

In atomic units and with the g-factor 2 the spin magnetic moment $\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathrm{S}}$ becomes

$$\boldsymbol{\mu}_{\mathrm{S}} = -\mathbf{S} \tag{3.73}$$

The Pauli operator now reads

$$\begin{aligned} H_{\mathrm{Pauli}} &= \frac{1}{2}\big(\mathbf{p} + \mathbf{A}(\mathbf{x})\big)^2 - \boldsymbol{\mu}_{\mathrm{S}} \cdot \mathbf{B}(\mathbf{x}) \\ &= \frac{1}{2}\left(\mathbf{p}^2 + \mathbf{p}\cdot\mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\cdot\mathbf{p} + \mathbf{A}(\mathbf{x})^2\right) + B\,S_3 \\ &= \frac{1}{2}\left(\mathbf{p}^2 + B\,(xp_2 - yp_1) + \mathbf{A}(\mathbf{x})^2\right) + B\,S_3 \\ &= \frac{\mathbf{p}^2}{2} + \frac{B}{2}\,L_3 + B\,S_3 + \frac{B^2}{8}\,(x^2 + y^2). \end{aligned} \tag{3.74}$$

The second term is obviously the magnetic interaction energy $-\boldsymbol{\mu}_{\mathrm{L}} \cdot \mathbf{B}$ of the magnetic moment $\boldsymbol{\mu}_{\mathrm{L}}$ that is caused by the orbital angular momentum,

$$\boldsymbol{\mu}_{\mathrm{L}} = -\frac{1}{2}\,\mathbf{L}. \tag{3.75}$$

Comparison with (3.73) shows that the g-factor belonging to the magnetic moment $\boldsymbol{\mu}_{\mathrm{S}}$ is twice as large as the g-factor belonging to $\boldsymbol{\mu}_{\mathrm{L}}$. The magnetic moment caused by the orbital motion of a charged particle has the g-factor 1.

3.7.2. Solutions with zero energy

The constant magnetic field is a two-dimensional situation with $B(\mathbf{x}) = B$. Now each of the equations (3.64) becomes equivalent to a Schrödinger equation in a constant magnetic field. The considerations of Book One also apply to this case (see Book One, Chapter 8). Here, we present a method of finding the ground-state solutions which will be useful also in the next section.

We assume that the constant magnetic field is $B > 0$. A vector potential for this field is

$$\mathbf{A}(\mathbf{x}) = \frac{B}{2}\,(-y, x). \tag{3.76}$$

We also define the auxiliary function

$$\phi(\mathbf{x}) = \frac{B}{4}\,(x^2 + y^2) = \frac{B}{4}\,|\mathbf{x}|^2, \tag{3.77}$$

with the property $\Delta\phi(\mathbf{x}) = B$. Next, we consider the operator $D^\dagger$ defined in (3.69). Assume we can find a solution ψ_0 of the equation $D^\dagger\psi = 0$. This function would also satisfy the equation $DD^\dagger\psi_0 = 0$. Hence, in view of (3.71), the spinor

$$\begin{pmatrix}\psi_1(\mathbf{x})\\ \psi_2(\mathbf{x})\end{pmatrix} = \begin{pmatrix}0\\ \psi_0(\mathbf{x})\end{pmatrix} \tag{3.78}$$

would be an eigenvector of the Pauli operator (3.74) for a constant field in two dimensions, belonging to the eigenvalue $E = 0$.

In order to find a solution ψ_0, let us first describe the properties that any solution of $D^\dagger\psi = 0$ must have. First of all, we require ψ_0 to be square-integrable. Next, we write

$$\psi_0(\mathbf{x}) = \mathrm{e}^{-\phi(\mathbf{x})}\,\omega(\mathbf{x}) \tag{3.79}$$

and find that

$$D^\dagger\psi_0 = -\mathrm{i}\mathrm{e}^{-\phi}\Big(\frac{\partial}{\partial x} - \mathrm{i}\,\frac{\partial}{\partial x_2}\Big)\omega. \tag{3.80}$$

Hence, the equation $D^\dagger\psi_0 = 0$ is obviously equivalent to the equation

$$\frac{\partial\omega}{\partial x_1} - \mathrm{i}\,\frac{\partial\omega}{\partial x_2} = 0. \tag{3.81}$$

If you write $\omega(x, x) = u(x, y) + \mathrm{i}v(x, y)$, you will see that (3.81) is in turn equivalent to the famous Cauchy-Riemann equations characterizing the real and imaginary parts of an analytic function. More precisely, it turns out that the function ω has to be an entire analytic function of $x - \mathrm{i}y$! Taking into account that $\psi_0 = \mathrm{e}^{-\phi}\omega$ should be square-integrable, we have to exclude those analytic functions ω that increase too fast in some direction of the complex plane. But this still leaves, for example, all polynomials in $x - \mathrm{i}y$. Choosing $\omega(\mathbf{x}) = (x - \mathrm{i}y)^\ell$, where ℓ is a non-negative integer, we find that

$$\psi_0(\mathbf{x}) = \mathrm{e}^{-B|\mathbf{x}|^2/4}(x - \mathrm{i}y)^\ell \tag{3.82}$$

satifsfies $D^\dagger\psi_0 = 0$. As explained above, we conclude

$$H_{\mathrm{Pauli}}\begin{pmatrix}0\\ \psi_0(\mathbf{x})\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}. \tag{3.83}$$

We note that *all* square-integrable functions of the form

$$\psi_0(\mathbf{x}) = \mathrm{e}^{-B|\mathbf{x}|^2/4}\,\omega(x - \mathrm{i}y) \tag{3.84}$$

where ω is analytic, lead to a zero-energy eigenvector of the two-dimensional H_{Pauli} operator for the homogeneous field. We have singled out the solutions of the type (3.82), because they are simultaneous eigenfunctions of the angular-momentum operator $L_3 = xp_2 - yp_1$. This can be seen as follows: We write the complex number $x - \mathrm{i}y$ in polar coordinates as $r\,\mathrm{e}^{-\mathrm{i}\varphi}$. Then, we obtain $(x-\mathrm{i}y)^\ell = r^\ell\,\mathrm{e}^{-\mathrm{i}\ell\varphi}$. In spherical coordinates, $L_3 = -\mathrm{i}\partial/\partial\varphi$ and hence $L_3\,\psi_0^{(\ell)} = -\ell\,\psi_0^{(\ell)}$ follows immediately.

Eigenvectors belonging to different eigenvalues of a self-adjoint operator are orthogonal, therefore $\{\psi_0^{(\ell)} \mid \ell = 0, 1, 2, \ldots\}$ is an orthogonal set of simultaneous eigenstates of H_{Pauli} and L. This set is even a basis in the zero-energy eigenspace, because every zero-energy solution can be written as a superposition of the $\psi_0^{(\ell)}$. This can be seen as follows: Any zero-energy solution must be of the form

$$\psi(\mathbf{x}) = \mathrm{e}^{-B|\mathbf{x}|^2/4}\begin{pmatrix} 0 \\ \omega(z) \end{pmatrix} \quad (\text{with } z = x - \mathrm{i}y), \tag{3.85}$$

where $\omega(z)$ is an entire analytic function of z. Writing ω as a power series in z,

$$\omega(z) = \sum_{\ell=0}^{\infty} a_\ell\, z^\ell \quad (\text{for all } z), \tag{3.86}$$

we can combine (3.85) and (3.88) to see that ψ has the expansion

$$\psi(\mathbf{x}) = \sum_{\ell=0}^{\infty} a_\ell\, \psi_0^{(\ell)}(\mathbf{x}). \tag{3.87}$$

Ground states in a homogeneous magnetic field:

The Pauli equation for an electron in a constant magnetic field $B > 0$ in two dimensions has infinitely many zero-energy solutions with spin-down. The solutions

$$\psi_0^{(\ell)}(\mathbf{x}) = \mathrm{e}^{-B|\mathbf{x}|^2/4}\begin{pmatrix} 0 \\ (x - \mathrm{i}y)^\ell \end{pmatrix}, \quad \ell = 0, 1, 2, 3, \ldots \tag{3.88}$$

are also eigenfunctions of the orbital angular-momentum operator $L = xp_2 - yp_1$,

$$H_{\text{Pauli}}\,\psi_0^{(\ell)} = 0, \qquad L_3\,\psi_0^{(\ell)} = -\ell\,\psi_0^{(\ell)}, \qquad S_3\,\psi_0^{(\ell)} = -\frac{1}{2}\,\psi_0^{(\ell)}. \tag{3.89}$$

The set $\{\psi_0^{(\ell)} \mid \ell = 0, 1, 2, \ldots\}$ is a basis in the zero-energy eigenspace.

Zero-energy eigenfunctions of the Pauli operator are time-independent solutions of the time-dependent Pauli equation (3.64).

The operator (3.74) without the spin term BS_3 is equivalent to a Schrödinger operator in a constant magnetic field. We know that the Schrödinger operator has no zero-energy eigenvalues. Instead, the ground-state energy is $|B|/2$ (Book One, Section 8.5). Obviously, the spin term in (3.74) just compensates the "spinless" ground-state energy.

For the Schrödinger operator as well as for the Pauli operator in a homogeneous magnetic field, the ground-state energy has an infinite degree of degeneracy. This degeneracy is caused by the translational symmetry discussed in Book One, Section 8.6.

3.7.3. The spectrum of eigenvalues

In this section, we are going to use the operators D and $D^\dagger$ as ladder operators to generate all eigenvalues and eigenfunctions of the Pauli operator, just as we did with the harmonic oscillator (see Book One, Section 7.7). In the last section, we obtained an infinite number of solutions for $D\psi = 0$, where D is the operator defined in (3.68) in case of a constant magnetic field,

$$D = -\mathrm{i}\frac{\partial}{\partial x_1} - \frac{B}{2}\,x_2 + \frac{\partial}{\partial x_2} + \mathrm{i}\,\frac{B}{2}\,x_1. \tag{3.90}$$

This operator has the property

$$D^\dagger D = DD^\dagger + 2B. \tag{3.91}$$

Given a function ψ_0 with $D^\dagger\psi_0 = 0$, we immediately find

$$D^\dagger D\,\psi_0 = 2B\,\psi_0. \tag{3.92}$$

Moreover, equation (3.91) tells us that $2B$ is the smallest eigenvalue of $D^\dagger D$ because the operator $DD^\dagger$ is non-negative (and B is assumed to be positive).

We can use ψ_0 to form an eigenvector with spin-up of the Pauli operator (3.71):

$$H_{\text{Pauli}}\begin{pmatrix}\psi_0\\0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}D^\dagger D & 0\\0 & DD^\dagger\end{pmatrix}\begin{pmatrix}\psi_0\\0\end{pmatrix} = B\begin{pmatrix}\psi_0\\0\end{pmatrix}. \tag{3.93}$$

Hence, we found that H_{Pauli} has the eigenvalue B. There is also a spin-down eigenvector belonging to the same eigenvalue. This is so because, whenever $D^\dagger D$ has an eigenvalue with eigenvector ψ_0, then $DD^\dagger$ has the same eigenvalue with eigenvector $D\psi_0$. Hence,

$$H_{\text{Pauli}}\begin{pmatrix}0\\\psi_1\end{pmatrix} = B\begin{pmatrix}0\\\psi_1\end{pmatrix} \quad \text{with } \psi_1 = D\psi_0. \tag{3.94}$$

We can continue in this way. Because of (3.91), the function ψ_1 is also an eigenfunction of $D^\dagger D$:

$$D^\dagger D\,\psi_1 = DD^\dagger\,\psi_1 + 2B\,\psi_1 = 4B\,\psi_1. \tag{3.95}$$

Also, $\psi_2 = D\psi_1$ is an eigenfunction of $DD^\dagger$ belonging to the eigenvalue $4B$. This leads to the eigenvalue $2B$ of the Pauli operator,

$$H_{\text{Pauli}} \begin{pmatrix} \psi_1 \\ 0 \end{pmatrix} = 2B \begin{pmatrix} \psi_1 \\ 0 \end{pmatrix}, \tag{3.96}$$

$$H_{\text{Pauli}} \begin{pmatrix} 0 \\ \psi_2 \end{pmatrix} = 2B \begin{pmatrix} 0 \\ \psi_2 \end{pmatrix} \quad \text{with } \psi_2 = D\psi_1. \tag{3.97}$$

If we proceed in this way, we obtain a whole sequence of eigenfunctions and eigenvalues of the Pauli operator. We conclude

Energy eigenstates in a homogeneous magnetic field:
The two-dimensional Pauli operator in a constant magnetic field $B > 0$ has the eigenvalues

$$E_n = nB \qquad (n = 0, 1, 2, 3, \ldots). \tag{3.98}$$

Let ψ_0 be a nonzero function with $D^\dagger\psi_0 = 0$, and define

$$\psi_n = D\,\psi_{n-1} = D^n\,\psi_0. \tag{3.99}$$

Eigenvectors of the Pauli operator belonging to the energies E_n are

$$\begin{pmatrix} 0 \\ \psi_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \psi_{n-1} \\ 0 \end{pmatrix} \qquad (n = 1, 2, 3\ldots) \tag{3.100}$$

with spin-down and spin-up, respectively. There are no spin-up eigenvectors for zero energy.

It has to be emphasized that such a sequence of eigenvectors is obtained for each of the zero-energy eigenstates $\psi_0^{(\ell)}$ defined in (3.88).

3.8. Special Topic: Magnetic Ground States

Because of the negative charge of an electron, spin-up corresponds to magnetic moment down and hence the potential energy in a field that points in the positive z-direction is negative. Obviously, this energy precisely compensates the ground state energy of a charged particle in a homogeneous magnetic field (see Book One, Chapter 8).

A generalization of the argument that leads to the ground state in a homogeneous field in Section 3.7.2 can be used to show that the Pauli operator in two dimensions has zero-energy bound states even if the magnetic field has

a finite flux. The number of zero-energy bound states of H_{Pauli} provides a lower bound for the number of negative-energy bound states of the operator $H = H_{\text{Pauli}} + V(\mathbf{x})$, with the same magnetic field and a negative electrostatic potential. If ψ is any eigenstate of H_{Pauli} belonging to the eigenvalue 0, then we find for the expectation value of the total energy

$$\langle \psi, H\psi \rangle = \langle \psi, H_{\text{Pauli}}\psi \rangle + \langle \psi, V(\mathbf{x})\psi \rangle = \langle \psi, V(\mathbf{x})\psi \rangle < 0. \tag{3.101}$$

For this result it is only necessary that $V(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ and $V(\mathbf{x}) < 0$ in some open region, because it turns out that the eigenstates of H_{Pauli} cannot vanish on open regions. Whenever $\langle \psi_j, H\psi_j \rangle < 0$ for linearly independent states $\psi_1, \ldots, \psi_n$, one can prove that the dimension of the subspace with negative energy is at least n.

It is interesting that the same problem turns out to be much more difficult in three dimensions. But at least we can show that there are magnetic fields giving rise to zero-energy bound states also in three dimensions. This result depends on the spin in a crucial way.

3.8.1. Two dimensions

The following theorem by Aharonov and Casher[10] describes the zero-energy states of the Pauli equation with a magnetic field in two dimensions:

THEOREM 3.1. *Let $B(\mathbf{x})$ be a magnetic field that vanishes outside a bounded region in two dimensions and let the flux*

$$F = \frac{1}{2\pi} \int_{\mathbb{R}^2} B(\mathbf{x}) d^2x \tag{3.102}$$

be finite. For $|F| > 1$, the Pauli operator has an eigenvalue with zero energy. Assume $F = n + r$, where n is a positive integer, and $0 < r < 1$. Then, the Pauli operator for an electron has precisely n spin-down eigenstates with energy zero. If $F = -n - r$, then there are precisely n spin-up eigenstates belonging to the eigenvalue zero. For $r = 0$, the multiplicity of the eigenvalue is $n - 1$ in both cases.

PROOF. The proof uses the idea that a function ϕ with $\Delta\phi(\mathbf{x}) = B(\mathbf{x})$ can also be defined for inhomogeneous fields. The Green function of Δ in two dimensions is $\frac{1}{2\pi} \ln |\mathbf{x} - \mathbf{y}|$. Therefore, the function

$$\phi(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \ln |\mathbf{x} - \mathbf{y}| \, B(\mathbf{y}) \, d^2y \tag{3.103}$$

satisfies $\Delta\phi(\mathbf{x}) = B(\mathbf{x})$. Moreover, it can be shown that

$$\phi(\mathbf{x}) - F \ln |\mathbf{x}| = O\big(\frac{1}{|\mathbf{x}|}\big), \quad \text{as } |\mathbf{x}| \to \infty. \tag{3.104}$$

[10] Aharonov, Y., and Casher, A., Phys. Rev. **A19**, 2461–2462 (1979).

We choose the vector potential

$$\mathbf{A}(\mathbf{x}) = \Big(-\frac{\partial}{\partial y}\,\phi(\mathbf{x}), \frac{\partial}{\partial x}\,\phi(\mathbf{x})\Big) \tag{3.105}$$

and look for a solution of

$$\boldsymbol{\sigma}\cdot(\mathbf{p}+\mathbf{A})\psi = 0, \qquad \boldsymbol{\sigma} = (\sigma_1,\sigma_2). \tag{3.106}$$

Writing

$$\omega(\mathbf{x}) = \mathrm{e}^{-\sigma_3\phi(\mathbf{x})}\,\psi(\mathbf{x}) \tag{3.107}$$

we find that (3.106) is equivalent to

$$\boldsymbol{\sigma}\cdot\mathbf{p}\,\omega = 0 \quad \text{or} \quad \begin{cases} \big(\frac{\partial}{\partial x} + \mathrm{i}\,\frac{\partial}{\partial y}\big)\omega_1(\mathbf{x}) = 0, \\ \big(\frac{\partial}{\partial x} - \mathrm{i}\,\frac{\partial}{\partial y}\big)\omega_2(\mathbf{x}) = 0. \end{cases} \tag{3.108}$$

These equations are equivalent to the Cauchy-Riemann equations. Hence, ω_1 has to be an entire analytic function in the variable $z = x + \mathrm{i}\,y$ whereas ω_2 must be analytic in $\overline{z} = x - \mathrm{i}\,y$. It follows from (3.104) and (3.107) that for large $|z| = |\mathbf{x}|$, these functions behave as

$$\omega_1(\mathbf{x}) \approx \mathrm{e}^{-F\ln|\mathbf{x}|}\psi_1(\mathbf{x}) = |\mathbf{x}|^{-F}\psi_1(\mathbf{x}), \tag{3.109}$$

$$\omega_2(\mathbf{x}) \approx \mathrm{e}^{+F\ln|\mathbf{x}|}\psi_2(\mathbf{x}) = |\mathbf{x}|^{+F}\psi_2(\mathbf{x}). \tag{3.110}$$

For positive F we find that ω_1 is square integrable at infinity and hence zero, because an analytic function cannot vanish in all directions, as $|z| \to \infty$. This shows that $\psi_1 = 0$, and therefore any solution of (3.106) must have spin-down. But for this we have to fulfill the condition

$$\psi_2 = \mathrm{e}^{-\phi}\omega_1 \in L^2(\mathbb{R}^2), \tag{3.111}$$

which requires that ω_2 should not increase faster than $|\mathbf{x}|^{F-1-\delta}$, for some (arbitrarily small) $\delta > 0$. Here, it is assumed that $F = n + r$ with n a positive integer and $0 \le r < 1$. Because ω_2 is an entire function of $\overline{z}$ that should not increase faster than $|z|^{n-1+r-\delta}$, it must be a polynomial in $x - \mathrm{i}\,y$ of degree $\le n-1$ if $r > 0$, or of degree $\le n-2$ if $r = 0$. Hence, for $r > 0$ we may choose the n linearly independent polynomials

$$\omega_2(\mathbf{x}) = (x - \mathrm{i}y)^\ell \quad \text{with} \quad \ell = 0, 1, 2, \ldots n-1, \tag{3.112}$$

and for $r = 0$ we have the condition $\ell \le n-2$. Correspondingly, the n (resp. $n-1$) solutions of (3.106) are given by

$$\mathrm{e}^{\sigma_3\phi(\mathbf{x})}\begin{pmatrix} 0 \\ (x-\mathrm{i}y)^\ell \end{pmatrix} = \begin{pmatrix} 0 \\ \mathrm{e}^{-\phi(\mathbf{x})}\,(x-\mathrm{i}y)^\ell \end{pmatrix} \tag{3.113}$$

and these are the zero-energy solutions of the Pauli operator

$$H_{\text{Pauli}} = \frac{1}{2}\,\big(\boldsymbol{\sigma}(\mathbf{p}+\mathbf{A})\big)^2. \tag{3.114}$$

An analogous reasoning applies to the case $F < 0$. □

The eigenfunctions obtained here are, in general, not eigenfunctions of L_3, because $B(\mathbf{x})$ (and hence $\phi(\mathbf{x})$) need not be cylindrically symmetric. The theorem relates the number of zero-energy eigenstates of the Pauli operator with a global property (flux) of the magnetic field. The theorem may be regarded as an elementary example of the famous Atiyah-Singer index theorem in global analysis.

3.8.2. Three dimensions

There is no analog of the previous theorem in three dimensions. We do not know general conditions on magnetic fields that lead to zero-energy bound states, but we do know some examples. Here is a particularly simple one due to Loss and Yau.[11]

We want to find a three-dimensional vector potential $\mathbf{A}(\mathbf{x})$ and a spinor $\psi(\mathbf{x})$ with

$$\boldsymbol{\sigma} \cdot \big(\mathbf{p} + \mathbf{A}(\mathbf{x})\big)\, \psi(\mathbf{x}) = 0. \tag{3.115}$$

By Eq. (3.67), this would give a magnetic field such that the stationary Pauli equation $H_{\text{Pauli}}\psi = E\psi$ has a zero-energy solution.

Assume we are given a spinor with $\psi(\mathbf{x}) \neq 0$ for all $\mathbf{x}$. Then, we can form the vector field

$$\mathbf{v}(\mathbf{x}) = \frac{\big\langle \psi(\mathbf{x}), \boldsymbol{\sigma}\, \psi(\mathbf{x}) \big\rangle_2}{\big|\psi(\mathbf{x})\big|_2^2}. \tag{3.116}$$

According to (3.54), $\mathbf{v}(\mathbf{x})$ is a unit vector in $\mathbb{R}^3$. It may be verified that

$$\boldsymbol{\sigma} \cdot \mathbf{v}(\mathbf{x})\, \psi(\mathbf{x}) = \psi(\mathbf{x}). \tag{3.117}$$

(See also Section 4.4.1 below). Define

$$\mathbf{A}(\mathbf{x}) = -\lambda(\mathbf{x})\, \mathbf{v}(\mathbf{x}) \tag{3.118}$$

with some real-valued function λ and you can see that

$$\boldsymbol{\sigma} \cdot \mathbf{A}(\mathbf{x})\, \psi(\mathbf{x}) = -\lambda(\mathbf{x})\, \psi(\mathbf{x}). \tag{3.119}$$

It remains to find a real-valued λ and an everywhere nonzero spinor ψ such that the equation

$$\boldsymbol{\sigma} \cdot \mathbf{p}\, \psi(\mathbf{x}) = \lambda(\mathbf{x})\, \psi(\mathbf{x}) \tag{3.120}$$

holds. For example, let

$$\psi(\mathbf{x}) = \frac{\mathbf{1}_2 + \mathrm{i}\, \boldsymbol{\sigma} \cdot \mathbf{x}}{(1 + x^2)^{3/2}}\, \phi_0, \tag{3.121}$$

[11] Loss, M., and Yau, H.T., Commun. Math. Phys. **104**, 283–290 (1986).

where ϕ_0 is any unit vector in $\mathbb{C}^2$. We obtain

$$\boldsymbol{\sigma} \cdot \mathbf{p}\, \psi(\mathbf{x}) = \frac{3}{1+x^2}\, \psi(\mathbf{x}). \tag{3.122}$$

Now combine Eqs. (3.119) and (3.120) with $\lambda(\mathbf{x}) = 3/(1+x^2)$ to obtain an example for (3.115). The solution ψ produces the vector field

$$\mathbf{v}(\mathbf{x}) = \frac{1}{1+x^2}\left((1-x^2)\mathbf{w} + 2(\mathbf{w}\cdot\mathbf{x})\,\mathbf{x} + 2\mathbf{w}\times\mathbf{x}\right), \tag{3.123}$$

which is everywhere nonzero. Here, $\mathbf{w} = \langle \phi_0, \boldsymbol{\sigma}\, \phi_0 \rangle_2$ is a unit vector in $\mathbb{R}^3$. The magnetic field belonging to the vector potential $\mathbf{A}(\mathbf{x}) = \lambda(\mathbf{x})\, \mathbf{v}(\mathbf{x})$ is given by

$$\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}) = \frac{12}{(1+x^2)^2}\mathbf{v}(\mathbf{x}). \tag{3.124}$$

3.9. The Coulomb Problem with Spin

3.9.1. Coulomb-Hamiltonian for a particle with spin

Due to the interaction term $-\boldsymbol{\mu}\cdot\mathbf{B}$ in the Pauli Hamiltonian, the motion of the particle in space in general depends on the behavior of its magnetic moment, and vice versa. But this is not the case for a purely electrostatic field when $\mathbf{B} = 0$. For electrons in a purely electrostatic field, the Hamiltonian has the familiar form $H = p^2/(2\mathrm{m}_e) + V(\mathbf{x})$ acting component-wise on spinor-wave functions. Hence, the time evolution of one spinor-component is completely independent from the behavior of the other. The spin is a constant of motion, and the initial spin-state has no influence on the behavior of the spinor-wave functions in space.

As an example, we consider a hydrogenic atom. In the standard representation, the Hamiltonian is (cf. Eq. (2.21))

$$H = \left(-\frac{\hbar^2}{2\mathrm{m}_e}\,\Delta - \frac{\gamma}{|\mathbf{x}|}\right)\mathbf{1}_2 = \begin{pmatrix} H_{\mathrm{coul}} & 0 \\ 0 & H_{\mathrm{coul}} \end{pmatrix}. \tag{3.125}$$

Here, H_{coul} is the familiar Coulomb Hamiltonian as defined in (2.21). Here, it acts component-wise on spinor wave functions.

Ψ The Hamiltonian H in (3.125) is the direct sum of two Coulomb Hamiltonians,

$$H = H_{\mathrm{coul}} \oplus H_{\mathrm{coul}}. \tag{3.126}$$

This operator is defined in the direct sum $L^2(\mathbb{R}^3) \oplus L^2(\mathbb{R}^3)$ of "spinless" Hilbert spaces. Equivalently, we may write the Hilbert space as $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ and define H as

$$H = H_{\mathrm{coul}} \otimes \mathbf{1}_2, \tag{3.127}$$

where $\mathbf{1}_2$ is the identity operator in $\mathbb{C}^2$. In that way, we can see that H acts like the ordinary Coulomb Hamiltonian on the "space part" of a spinor and like the identity on the "spin part."

The Coulomb Hamiltonian (3.125) commutes with the operator S_3,

$$[H, S_3] = 0 \quad \text{(on the domain of } H\text{)}. \tag{3.128}$$

A system of simultaneous eigenfunctions is, for example, given by

$$\psi_{n,\ell,m,+}(\mathbf{x}) = \begin{pmatrix} \psi_{n,\ell,m}(\mathbf{x}) \\ 0 \end{pmatrix}, \qquad \psi_{n,\ell,m,-}(\mathbf{x}) = \begin{pmatrix} 0 \\ \psi_{n,\ell,m}(\mathbf{x}) \end{pmatrix}, \tag{3.129}$$

with the Coulomb eigenfunctions described in Section 2.5.2, see Eq. (2.106). Every eigenfunction of the Coulomb problem appears once with spin-up and once with spin-down. These spinors are simultaneous eigenvectors of the set of commuting operators H, L^2, L_3, and S_3. The spin does not change the eigenvalues E_n of the Coulomb Hamiltonian. But, accounting for the spin, the degree of degeneracy of each eigenvalue E_n is twice the degree of degeneracy of that eigenvalue in the problem without spin.

3.9.2. Complete set of observables

For the Coulomb problem without spin, it was necessary to specify the eigenvalues of the operators H, L^2, and L_3 in order to specify a unique eigenstate. For the Coulomb problem with spin, there is an additional degree of freedom. Given eigenvalues of H, L^2, and L_3, we are still left with two orthogonal eigenstates. We need the eigenvalue of the spin-operator S_3 to specify a unique state.

A useful set of quantum numbers for the Coulomb problem with spin therefore consists of

$n = 1, 2, \ldots$	the principal quantum number labeling the eigenvalues of H
$\ell = 0, 1, \ldots, n-1$	the angular-momentum quantum number defining the eigenvalues of L^2
$m_\ell = -\ell, -\ell+1, \ldots, \ell$	the magnetic quantum number belonging to the orbital angular momentum L_3,
$m_s = -1/2, +1/2$	the magnetic quantum number due to the spin S_3.

Each quadrupel of values (n, ℓ, m_ℓ, m_s) specifies a simultaneous eigenvector of (H, L^2, L_3, S_3). This set of quantum numbers is complete in the sense that the corresponding eigenvector is unique up to a scalar factor (that is, the corresponding quantum state is unique). A set of commuting observables whose eigenvalues specify a unique simultaneous eigenstate is called a *complete set of commuting observables*.

Ψ More generally, let us consider n self-adjoint operators A_i, $i = 1, \ldots, n$, all having a "pure point spectrum." That is, we assume that the spectrum contains only eigenvalues and no continuous part. (It is technically much more demanding to define the completeness when observables with a continuous spectrum are involved.) The assumption of a pure point spectrum is fulfilled for the operators L^2, L_3, S_3, and for H in the subspace of bound states (which is also a Hilbert space).

DEFINITION 3.1. A set of self-adjoint operators $\{A_1, A_2, \ldots, A_n\}$ with pure point spectrum in a Hilbert space $\mathfrak{H}$ is called a *complete set of observables* if the following two conditions are satisfied:

(1) For all $(\lambda_1, \ldots, \lambda_n)$, where λ_i is an eigenvalue of A_i, there is a normalized simultaneous eigenvector $\Psi_{\lambda_1,\ldots,\lambda_n}$ of all A_i that is unique up to multiplication with a phase factor.
(2) The set of all $\Psi_{\lambda_1,\ldots,\lambda_n}$ is an orthonormal basis of the Hilbert space.

By definition, the common eigenvectors $\Psi_{\lambda_1,\ldots,\lambda_n}$ belong to the common domain $\mathfrak{D}$ of the operators, that is, to the intersection

$$\mathfrak{D} = \mathfrak{D}(A_1) \cap \mathfrak{D}(A_2) \cap \cdots \cap \mathfrak{D}(A_n). \tag{3.130}$$

It follows from the definition that $\mathfrak{D}$ is a dense subspace of $\mathfrak{H}$, because it contains all vectors of an orthonormal basis. Moreover, the operators A_i commute with each other

$$[A_i, A_j] = 0 \quad \text{on } \mathfrak{D}. \tag{3.131}$$

It also follows that the operators A_i commute in the sense that the corresponding unitary groups commute (see Book One, Section 6.11).

3.9.3. The spin-orbit interaction

Due to the following relativistic effect, the spin $\mathbf{S}$ of electrons is not completely independent from its orbital motion, even if there is only an electrostatic field. Consider a classical electron that moves at a certain time t with velocity $\mathbf{v}$ in an electrostatic field $\mathbf{E}$. We go into a coordinate system that moves with velocity $\mathbf{v}$. This coordinate system is at time t the rest frame of the electron. But while the electron is at rest, the charges generating the electric field move with velocity $-\mathbf{v}$. Moving charges constitute a current, and currents are the origin of a magnetic field. In its rest frame, the electron still has a magnetic moment $\boldsymbol{\mu}_S$ due to its spin. Hence, we expect a contribution to the energy of the electron from the interaction of the spin magnetic moment with the magnetic field $\mathbf{B}'$ in the moving frame.

The strength of the magnetic field $\mathbf{B}'$ in the electron's rest frame is given by

$$\mathbf{B}' = -\frac{1}{c^2}\,\mathbf{v}\times\mathbf{E}. \tag{3.132}$$

This formula can be obtained from the relativistic transformation law for electromagnetic fields. It is an approximation that is valid if v is much smaller than c, the speed of light. Within that approximation, we expect that the Hamiltonian contains an additional interaction term of the form (3.10), that is,

$$-\boldsymbol{\mu}_{\mathrm{S}}\cdot\mathbf{B}' = \frac{1}{c^2}\,\boldsymbol{\mu}\cdot\mathbf{v}\times\mathbf{E}. \tag{3.133}$$

Next, we consider the classical model of a hydrogenic atom. The atomic nucleus (with Z protons) causes the electric field

$$\mathbf{E} = \frac{Ze}{4\pi\epsilon_0}\,\frac{\mathbf{x}}{|\mathbf{x}|^3} = \frac{\gamma}{e}\,\frac{\mathbf{x}}{|\mathbf{x}|^3}. \tag{3.134}$$

In the rest frame of an electron at time t, the proton moves with velocity $-\mathbf{v}$ and causes an additional magnetic field

$$\mathbf{B}' = -\frac{\gamma}{ec^2}\,\frac{\mathbf{v}\times\mathbf{x}}{|\mathbf{x}|^3}. \tag{3.135}$$

This is just the Biot-Savart law for the magnetic field of a moving charge. Introducing the angular momentum $\mathbf{L} = \mathrm{m}\mathbf{x}\times\mathbf{v}$, we can write this as

$$\mathbf{B}' = \frac{\gamma}{e\mathrm{m}_e c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{L}. \tag{3.136}$$

Using the expression (3.17) for the magnetic moment of the electron due to its spin, the interaction term becomes

$$-\boldsymbol{\mu}_{\mathrm{S}}\cdot\mathbf{B}' = \frac{g\gamma\mu_{\mathrm{B}}}{\hbar e\mathrm{m}_e c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{S}\cdot\mathbf{L} = \frac{g\gamma}{2\mathrm{m}_e^2 c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{S}\cdot\mathbf{L}. \tag{3.137}$$

This expression involves the spin and the orbital angular momentum of the electron and is therefore called the *spin-orbit interaction*. A more careful consideration of relativistic effects, in particular, of the Thomas precession, shows that the actual size of the spin-orbit energy is only half as large. The interested reader will find the details of this calculation, for example, in [**3**]. The final result for the spin-orbit interaction is therefore

$$E_{\text{spin-orbit}} = \frac{g\gamma}{4\mathrm{m}_e^2 c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{S}\cdot\mathbf{L}. \tag{3.138}$$

In quantum mechanics, this gives a contribution to the Hamiltonian of the hydrogen atom. Usually, this contribution is treated by perturbation theory. This term and other contributions from relativistic corrections of the kinetic energy are automatically accounted for by the Dirac equation.

Let us now consider the Hamiltonian of a hydrogenic atom if the spin-orbit energy is taken into account. By adding the operator corresponding to (3.138) to the Coulomb-Hamiltonian (3.125) for a particle with spin, we obtain

$$H = \Big(-\frac{\hbar^2}{2\mathrm{m}}\,\Delta - \frac{\gamma}{|\mathbf{x}|}\Big)\,\mathbf{1}_2 + \frac{g\gamma}{4\mathrm{m}_e^2 c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{S}\cdot\mathbf{L} \tag{3.139}$$

For the hydrogen atom ($Z = 1$) in atomic units, we may set $\gamma = 1$, $\mathrm{m}_e = 1$, $c \approx 137$, $g \approx 2$, and the Hamiltonian simplifies to

$$H = \Big(-\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|}\Big)\,\mathbf{1}_2 + \frac{1}{2c^2}\,\frac{1}{|\mathbf{x}|^3}\,\mathbf{S}\cdot\mathbf{L} \tag{3.140}$$

and we see that the constant in front of the spin-orbit term has a very small numerical value ($\approx 2.7 \times 10^{-5}$). Hence, we expect that the spin-orbit term has only a small influence on the Coulomb eigenfunctions.



The spin-orbit term is usually treated by first-order perturbation theory, because the radial problem corresponding to (3.139) cannot be solved exactly. This is a very delicate problem in mathematical perturbation theory. The spin-orbit term dominates the Coulomb potential at small distances from the origin. Hence, there are wave functions in the Hilbert space for which the spin-orbit term contributes much more to the energy than the Coulomb potential. Fortunately, this is not the case for the Coulomb eigenfunctions, which are indeed only slightly perturbed by the presence of the spin-orbit interaction. We note that the spin-orbit term is implicitly contained in the relativistic Dirac equation for the hydrogen atom. In that framework it is not necessary to employ any perturbation theory, because the solution is known explicitly. In the following, we investigate the angular part of the wave function according to the Pauli equation with a spin-orbit term.

3.9.4. Eigenfunctions of the spin-orbit operator

In a more realistic model of the hydrogen atom that takes into account the spin-orbit term $f(|\mathbf{x}|)\,\mathbf{S}\cdot\mathbf{L}$, the spin $\mathbf{S}$ and the orbital angular momentum $\mathbf{L}$ are no longer constants of motion. Neither L_3 nor S_3 commutes with the spin-orbit term. Therefore, these operators also do not commute with the Hamiltonian (3.139). But the system is still spherically symmetric, because the scalar product $\mathbf{S}\cdot\mathbf{L}$ is invariant under rotations. Hence, the generator of rotations, the total angular-momentum operator

$$\mathbf{J} = \mathbf{L} + \mathbf{S}, \tag{3.141}$$

must be a constant of motion. Indeed, one finds

$$[H\,,\,J_k] = H\,J_k - J_k\,H = 0, \qquad k = 1, 2, 3. \tag{3.142}$$

We may choose the operators J^2, J_3, $\mathbf{S}\cdot\mathbf{L}$ as a set of observables that commute with each other and with the Hamiltonian (3.140). The operator $S^2 = (3/4)\mathbf{1}$ (sticking with atomic units, $\hbar = 1$) is just a multiple of the identity operator, hence it trivially commutes with all other operators.

It is useful to introduce the *spin-orbit operator*

$$K = 2\,\mathbf{S}\cdot\mathbf{L} + \mathbf{1}. \tag{3.143}$$

We can express the operators J^2 and L^2 in terms of K:

$$J^2 = K^2 - \frac{1}{4}, \qquad L^2 = K^2 - K. \tag{3.144}$$

It is clear that K commutes with the Hamiltonian (3.140), because K commutes with the spin-orbit term and both $\mathbf{S}$ and $\mathbf{L}$ commute with the scalar Coulomb Hamiltonian. Hence, K is a conserved quantity and so are the observables J^2 and L^2, which can be expressed in terms of K.

Exercise 3.8. *With the help of*

$$\boldsymbol{\sigma}\cdot\mathbf{L} = \begin{pmatrix} L_3 & L_1 - \mathrm{i}L_2 \\ L_1 + \mathrm{i}L_2 & -L_3 \end{pmatrix}, \tag{3.145}$$

write K as a two-by-two matrix operator and verify (3.144)*, using the angular-momentum commutation relations.*

Equation (3.144) implies that any eigenvector of K is also eigenvector for the operators J^2 and L^2. The eigenvalue of J^2, which is usually written $j(j+1)$ with a non-negative j, is determined by the eigenvalue κ of K. By solving $\kappa^2 = j(j+1) + 1/4$ for $j > 0$, we find immediately

$$j = |\kappa| - \frac{1}{2}. \tag{3.146}$$

We know that the eigenvalue of L^2 is $\ell(\ell+1)$, where the orbital angular-momentum quantum number ℓ is a non-negative integer. Because of (3.144), the eigenvalues of K must fulfill the relation

$$\ell(\ell+1) = \kappa^2 - \kappa. \tag{3.147}$$

Solving for κ, we obtain the two solutions $\kappa = \ell + 1$ and $\kappa = -\ell$. Note that the value $\kappa = -\ell = 0$ has to be excluded, because the relation $K^2 = J^2 + 1/4$ implies $\kappa^2 > 0$. Hence, the only possible values for κ are

$$\kappa = \pm 1,\ \pm 2,\ \pm 3,\ \ldots. \tag{3.148}$$

This in turn shows that the possible values of the quantum number j are

$$j = \frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots. \tag{3.149}$$

The eigenvalue of J_3 will be denoted by m_j. As we know from Theorem 1.1, the only possible values for m_j are

$$m_j = -j,\, -j+1,\, \ldots\, , +j. \tag{3.150}$$

The simultaneous eigenvectors of the operators K and J_3 are the *spinor harmonics*. They can be defined in terms of the spherical harmonics Y_ℓ^m as follows

$$\mathcal{Y}_{\kappa,m_j} = \frac{1}{\sqrt{2\kappa-1}} \begin{pmatrix} \sqrt{\kappa - \frac{1}{2} + m_j}\; Y_{\kappa-1}^{m_j-1/2} \\ \sqrt{\kappa - \frac{1}{2} - m_j}\; Y_{\kappa-1}^{m_j+1/2} \end{pmatrix}, \quad \kappa \geq 1, \tag{3.151}$$

$$\mathcal{Y}_{\kappa,m_j} = \frac{1}{\sqrt{1-2\kappa}} \begin{pmatrix} -\sqrt{\frac{1}{2} - \kappa - m_j}\; Y_{-\kappa}^{m_j-1/2} \\ \sqrt{\frac{1}{2} - \kappa + m_j}\; Y_{-\kappa}^{m_j+1/2} \end{pmatrix}, \quad \kappa \leq -1. \tag{3.152}$$



CD 3.13 presents a gallery of visualization of the spinor harmonics $\mathcal{Y}_{\kappa,m_j}$ with angular momentum $j \leq 7/2$. You may select the function by specifying the quantum numbers j, m_j, and the sign of $\mathbf{S}\cdot\mathbf{L}$ (which is related to the sign of κ, see (3.159) below). Because these functions depend only on the angles ϑ and φ, they may be regarded as functions on the surface of the unit sphere. CD 3.13.2 shows the upper and lower components as separate plots, CD 3.13.3 visualizes the spin directions via colors, as described in Section 3.5.5. The associated spin-vector field is symmetric with respect to rotations around the z-axis (see CD 3.13.4 and CD 3.13.5).

The spinor harmonics are also eigenfunctions of L^2 and S^2, but not of L_3, or S_3. We have the following result.

Spin-orbit and angular-momentum eigenvalues:

$$K\,\mathcal{Y}_{\kappa,m_j} = \kappa\,\mathcal{Y}_{\kappa,m_j} \qquad \kappa = \pm 1, \pm 2, \pm 3, \ldots, \tag{3.153}$$

$$J^2\,\mathcal{Y}_{\kappa,m_j} = j(j+1)\,\mathcal{Y}_{\kappa,m_j} \qquad j = |\kappa| - \frac{1}{2} = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots, \tag{3.154}$$

$$J_3\,\mathcal{Y}_{\kappa,m_j} = m_j\,\mathcal{Y}_{\kappa,m_j} \qquad m_j = -j, -j+1, \ldots, j. \tag{3.155}$$

It is not difficult to verify the eigenvalue equation for $J_3 = L_3 + S_3$ using

$$L_3\, Y_\ell^m(\vartheta,\varphi) = m\, Y_\ell^m(\vartheta,\varphi). \tag{3.156}$$

In order to compute the eigenvalue of the matrix operator

$$\mathbf{S}\cdot\mathbf{L} = \frac{1}{2}\begin{pmatrix} L_3 & L_1 - \mathrm{i}L_2 \\ L_1 + \mathrm{i}L_2 & -L_3 \end{pmatrix}, \tag{3.157}$$

we need that $L_1 \pm \mathrm{i}\,L_2$ are ladder operators with respect to m. We have

$$(L_1 \pm \mathrm{i}\,L_2)\,Y_\ell^m = \sqrt{(\ell \pm m + 1)(\ell \mp m)}\,Y_\ell^{m\pm 1}. \tag{3.158}$$

A little calculation now shows

$$\mathbf{S}\cdot\mathbf{L}\,\mathcal{Y}_{\kappa,m_j} = \frac{1}{2}(\kappa - 1)\,\mathcal{Y}_{\kappa,m_j} \tag{3.159}$$

for both signs of κ. This proves (3.153). The spinor harmonics are also eigenfunctions of L^2.

$$L^2\,\mathcal{Y}_{\kappa,m_j} = \ell(\ell+1)\,\mathcal{Y}_{\kappa,m_j} \quad \text{with} \quad \ell = \begin{cases} \kappa - 1 & \text{for } \kappa > 0, \\ -\kappa & \text{for } \kappa < 0. \end{cases} \tag{3.160}$$

The eigenvalues of $J^2 = L^2 + 2\mathbf{S}\cdot\mathbf{L} + S^2$ follow by a straightforward computation.

The sign of the eigenvalue of K also determines the sign of $\mathbf{S}\cdot\mathbf{L}$, that is, whether the spin and the orbital angular momentum are more parallel or rather antiparallel. (Neither $\mathbf{L}$ nor $\mathbf{S}$ have a well-defined direction, because these operators have non-commuting components.) If the eigenvalue of K is positive, then we find $j = \kappa - 1/2$ and $\ell = \kappa - 1$ and hence $j = \ell + 1/2$. The total angular momentum j is obtained by adding the spin $1/2$ to the orbital angular momentum ℓ. In a state belonging to a negative eigenvalue of K, the operator $\mathbf{S}\cdot\mathbf{L}$ is strictly negative ($\mathbf{S}$ and $\mathbf{L}$ have opposite directions). In this case, we have $j = -\kappa - 1/2$, $\ell = -\kappa$, and the total angular momentum $j = \ell - 1/2$ is obtained by subtracting the spin from the orbital angular momentum.

Without proof, we note the formula

$$|\mathcal{Y}_{\kappa,m_j}(\vartheta,\varphi)|^2 = |\mathcal{Y}_{-\kappa,m_j}(\vartheta,\varphi)|^2. \tag{3.161}$$

Here, the modulus is defined as the $\mathbb{C}^2$-norm of the spinor, that is,

$$|\mathcal{Y}_{\kappa,m_j}(\vartheta,\varphi)|^2 = \big\langle\, \mathcal{Y}_{\kappa,m_j}\,,\, \mathcal{Y}_{\kappa,m_j}\,\big\rangle_2. \tag{3.162}$$

This quantity is symmetric with respect to rotations about the z-axis, hence it actually does not depend on m_j.

It is interesting to compare the spinor harmonics with the spherical harmonics. In Figure 3.4, we show "polar plots" of $|Y_4^1|$ and $|\mathcal{Y}_{\pm 5,3/2}|$ (similar to Figure 1.8). These are curves in the xz-plane where at each angle ϑ, the distance from the origin is the absolute value of the function. The spherical harmonics and the corresponding spinor harmonics have maxima in roughly the same directions. But the absolute value of the spinor harmonic has no

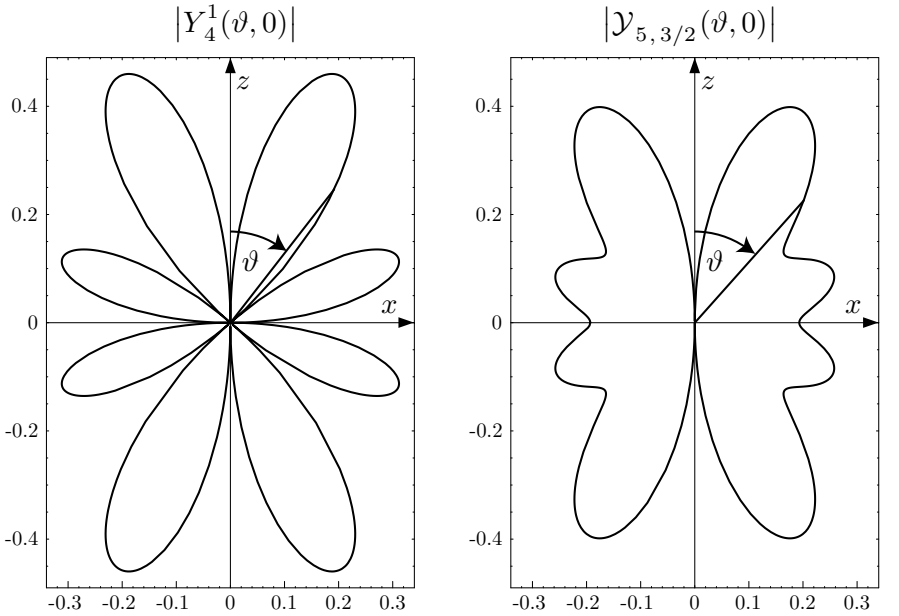
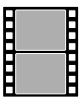


FIGURE 3.4. Polar plots of the absolute values of the spherical harmonic Y_4^1 and of the corresponding spinor harmonics $\mathcal{Y}_{\pm 5,3/2}$.

zeros for $0 < \vartheta < \pi$, because the two components of the spinor do not vanish simultaneously for ϑ inside this interval.

Some visualizations of the simultaneous eigenfunctions of the Coulomb Hamiltonian, J^2, and K are depicted in Figure 3.5.



CD 3.14 is a gallery of eigenfunctions of the hydrogen atom with spin (see also Fig. 3.5). Two navigation palettes either show the quantum numbers n, κ, and m_j, or the quantum numbers j, ℓ, n_r, and m_j. The visualizations show flux lines of the spin-vector field. In addition, the flux lines are colored according the direction of that vector field. (This helps to distinguish the sign of m_j.) In these visualizations, the influence of the spin-orbit term is neglected. Note, however, that the spin-orbit term would change only the radial part of the eigenfunction. Moreover, this distortion would affect the radial wave function only in a small neighborhood of $r = 0$, which could not be seen in these images.

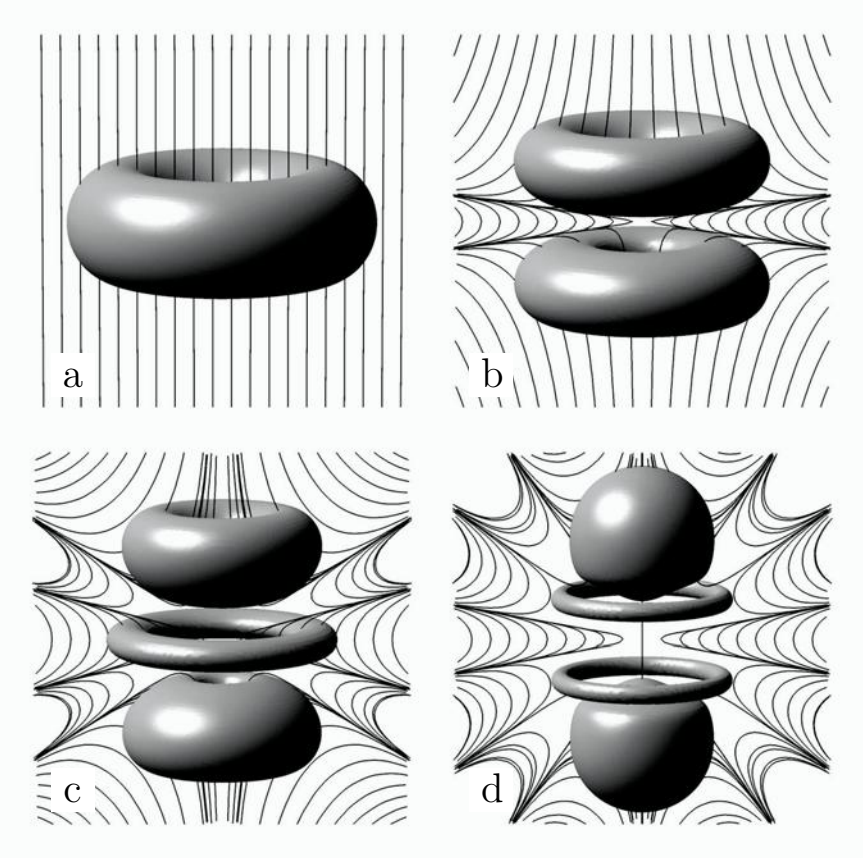


FIGURE 3.5. Some hydrogen eigenfunctions with $n = 4$, $\kappa = 4$, and (a) $m_j = 7/2$, (b) $m_j = 5/2$, (c) $m_j = 3/2$, and (d) $m_j = 1/2$. The images show isosurfaces of the absolute value of $|\psi_{n,\kappa,m_j}|$ and some flux lines representing the spin-vector field (which has a rotational symmetry about the z-axis).

3.9.5. The radial equation

In spherical coordinates, the Coulomb Hamiltonian with the spin-orbit coupling (3.140) becomes (using atomic units)

$$H = -\frac{1}{2}\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{2}\frac{L^2}{r^2} - \frac{1}{r} + \frac{1}{2c^2}\frac{\mathbf{S}\cdot\mathbf{L}}{r^3}\,. \tag{3.163}$$

Here, the part containing the radial derivatives commutes with the part containing angular derivatives, and hence we can find a solution of the eigenvalue equation in the form of a product of a radial part and an angular part

(separation of variables)

$$\psi(r,\vartheta,\varphi) = \frac{1}{r}\, f_\kappa(r)\, \mathcal{Y}_{\kappa,m_j}(\vartheta,\varphi). \tag{3.164}$$

The angular part is a simultaneous eigenfunction of L^2 and of $\mathbf{S}\cdot\mathbf{L}$. Inserting ψ into $H\psi = E\psi$ therefore leads to the radial equation for f_κ,

$$\Big(-\frac{1}{2}\,\frac{\partial^2}{\partial r^2} + \frac{1}{2}\,\frac{\kappa(\kappa-1)}{r^2} - \frac{\gamma}{r} + \frac{1}{4c^2}\,\frac{\kappa-1}{r^3}\Big)\, f_\kappa(r) = E\, f_\kappa(r). \tag{3.165}$$

Chapter 4

Qubits

Chapter summary: A qubit (quantum bit) is a quantum-mechanical two-state system. Any quantum system that can have two different states can also assume an arbitrary superposition of these states. Compared to a classical bit, a qubit has a significantly higher complexity (a continuum of possible states versus two distinct states 0 and 1). But nevertheless, because the state space is just two-dimensional, the measurement of any observable can produce at most two different results.

A canonical example of a qubit is provided by the spin of a spin-1/2 particle. Many ideas in this chapter are formulated with this system in mind. But there are other realizations of qubits, for example, the polarization states of photons, which we are going to discuss in Section 4.5.

Because of the relative simplicity of a qubit system, it is worthwhile to review some peculiarities of quantum mechanics. In Section 4.2, we use a typical Stern-Gerlach experiment to illustrate the projection postulate, the state preparation by single-particle measurements, and the state verification by ensemble measurements (Section 4.3). We ask whether it is meaningful to talk about the state of a single qubit, and we describe how one can determine (or rather estimate) an unknown quantum state. Moreover, we discuss the impossibility of "classical" teleportation in quantum mechanics.

In Section 4.4, we associate a unique "spin-up direction" with every qubit state. We describe the implementation of rotations as unitary transformations in the qubit's Hilbert space and compute the transition probabilities between different qubit states.

The strange topic of single-particle interference is presented in Section 4.6. We introduce interferometers and discuss the problems of acquiring the "which-way" information. We describe a variant of the double-slit experiment and discuss what it means to rotate a qubit through an angle of 2π. Interaction-free measurement (the detection of a bomb without actually looking at it) is presented in Section 4.6.4 as an example illustrating the meaning of the interference of probability amplitudes.

Section 4.7 deals with quantum cryptography. We present an example of a secure key distribution protocol that allows one to establish a secure communication via the classical one-time pad. The security of the method depends on the fact that quantum mechanics indeed gives a complete description of the state of a qubit. In a hidden variable theory, one assumes that the state of a qubit can be described by some additional parameters whose knowledge would enable us to make more

accurate predictions. Section 4.8 presents an example of a hidden variable theory and discusses its implications.

We conclude this chapter with a section about the spin in a time-dependent magnetic field (Section 4.9). In particular, we discuss the time evolution in a periodically time-dependent magnetic field and the phenomenon of spin resonance, or magnetic resonance. The results are relevant for technological applications like nuclear spin tomography.

4.1. States and Observables

4.1.1. The Hilbert space of a qubit

Many investigations of quantum systems do not require a "complete" description of the state. For example, one often neglects the position and momentum of a particle when one is only interested in the "inner degrees of freedom" related to the spin. This simplifies the description considerably, because the Hilbert space describing the spin of a particle with spin 1/2 is just the two-dimensional complex vector space $\mathbb{C}^2$.

Definition:

A quantum system with a two-dimensional Hilbert space is called a *two-state system* or a *qubit* (quantum bit). The vectors in the Hilbert space of a qubit are often called *spinors*.

The qubit is the simplest nontrivial quantum system, but it has many typical properties of larger quantum systems. Moreover, the qubit is at the basis of quantum information theory.

The classical analog of a qubit is a *bit*—the unit of information that can only have one of two possible values (usually denoted by "0" and "1"). In a classical computer, bits are usually represented by the voltage in a capacitor. If miniaturization is brought to an extreme, one might think of representing a bit by two different states of an elementary particle. But then quantum mechanics takes over, and the properties of a qubit are very different from the properties of a classical bit.

The crucial difference between a bit and a qubit is that the two states of a qubit can form quantum-mechanical superpositions and hence an infinite number of further states. It is still appropriate to speak of a two-state system, because all possible states of a qubit can be described as superpositions of just two basis states. Moreover, any measurement of a qubit can produce at most two different results, which may be called "0" and "1". But the physical meaning of "0" and "1" depends on the observable one chooses to

measure. For example, the two eigenstates of the spin observable S_1 are physically distinct from the eigenstates of S_3.

It should be stressed that particles with spin 1/2 are not the only examples of qubits. Other physically important qubit systems are realized, for example, by the polarization states of photons (Section 4.5.1), or by two oscillating states of trapped ions. As a mathematical example, the qubit obtained by restricting the harmonic oscillator to the subspace spanned by the first two eigenstates will be discussed in Section 4.5.3.

4.1.2. States of a qubit

Let us first choose an orthonormal basis in the two-dimensional Hilbert space of a qubit, that is, we choose two vectors ψ_+ and ψ_- with

$$\langle \psi_+, \psi_- \rangle = 0, \qquad \|\psi_+\| = \|\psi_-\| = 1. \tag{4.1}$$

The physical interpretation of the basis vectors depends on the physical system under consideration. Thinking of particles with spin 1/2, we may assume that the vectors ψ_+ and ψ_- describe the states with spin-up and spin-down in the z-direction of a chosen Cartesian coordinate system in $\mathbb{R}^3$. Thinking of photons, we may take ψ_+ to describe the state of vertical polarization, and ψ_- to describe the state of horizontal polarization (see Section 4.5.1).

With respect to this basis, vectors are represented by column vectors in $\mathbb{C}^2$, and linear operators are represented by two-by-two matrices. For example, the basis vectors become

$$\psi_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \psi_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.2}$$

This representation (again called the *standard representation*) will be used throughout this chapter, and the identification between states and column vectors, observables and two-by-two matrices is generally made without further notice.

A general state of a qubit is an arbitrary superposition of the two basis states,

$$\psi = c_+ \, \psi_+ + c_- \, \psi_- = \begin{pmatrix} c_+ \\ c_- \end{pmatrix}, \quad \text{with } c_\pm \in \mathbb{C}. \tag{4.3}$$

The norm of ψ and the scalar product with $\phi = d_+ \, \psi_+ + d_- \, \psi_-$ are given by

$$\|\psi\|^2 = |c_+|^2 + |c_-|^2, \quad \langle \psi, \phi \rangle = \overline{c_+} \, d_+ + \overline{c_-} \, d_-. \tag{4.4}$$

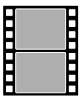
We omit the index 2 used in Section 3.5 to denote the $\mathbb{C}^2$-scalar product, as long as no confusion with the norm and scalar product of spinor-wave

functions can arise. Usually, it is assumed that the states are normalized, that is, $\|\psi\| = 1$.

Very often, the two base vectors are denoted by

$$\psi_+ = |\uparrow\rangle, \qquad \psi_- = |\downarrow\rangle, \tag{4.5}$$

with the arrows indicating the physical interpretation as spin-up and spin-down. The *ket-symbol* $|\cdot\rangle$ is used in Dirac's notation, which is predominant in the literature about qubit systems. Here, it comes in handy to convert a graphical symbol into a Hilbert-space vector. In Appendix A.8, we give a short introduction to Dirac's notation. Other commonly used ket-symbols for the $\mathbb{C}^2$-basis vectors are $|0\rangle$ and $|1\rangle$, or $|+\rangle$ and $|-\rangle$.



We can visualize a qubit state $\psi = c_+\,\psi_+ + c_-\,\psi_-$ by a bar diagram showing the complex values of coefficients c_+ and c_-. A few examples are given in CD 4.1. Other methods of visualizing qubit states will be discussed later.

4.1.3. Qubit observables

Any observable has to be represented by a self-adjoint operator. With respect to a chosen orthonormal basis in the Hilbert space of a qubit, observables are thus represented by Hermitian[1] two-by-two matrices. The three Pauli matrices $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ defined in (3.38) are examples of particular importance. The basis vector ψ_+ is an eigenvector of σ_3 belonging to the eigenvalue $+1$, and ψ_- belongs to the eigenvalue -1 of σ_3. The physical meaning of the Pauli matrices thus depends on the interpretation of $\psi_\pm$. If the qubit is realized as a particle with spin 1/2, then the Pauli matrices may be interpreted as the standard representation of the spin observables S_1, S_2, and S_3 according to Eq. (3.39).

It is usually assumed that, in turn, every self-adjoint operator corresponds to an observable. Hence, for any Hermitian matrix Q, it is, at least in principle, possible to build a measurement device that produces the eigenvalues of Q as possible results. The measurement of a qubit-observable can produce at most two different results, because a two-by-two matrix Q can have at most two different eigenvalues.

The three Pauli matrices together with the two-dimensional unit matrix $\mathbf{1}_2$ form a basis in the four-dimensional real vector space of all Hermitian two-by-two matrices (see Exercise 4.2 below). Hence, we can make the following observation:

[1] By definition, a square-matrix A is *Hermitian* if A equals the adjoint matrix $A^\dagger = \overline{A}^\top$.

Pauli matrices and qubit observables:

With respect to an orthonormal basis in $\mathbb{C}^2$, any qubit observable Q is represented by a linear combination of Pauli matrices,

$$Q = \frac{1}{2}\left(a_0 \mathbf{1}_2 + \sum_{k=1}^{3} a_k \sigma_k\right) = \frac{1}{2}\begin{pmatrix} a_0 + a_3 & a_1 - \mathrm{i}a_2 \\ a_1 + \mathrm{i}a_2 & a_0 - a_3 \end{pmatrix}, \tag{4.6}$$

with real coefficients $a_0, \ldots, a_3$.

Finally, we note the following useful formulas for Pauli matrices,

$$\sigma_j \, \sigma_k + \sigma_k \, \sigma_j = 2 \, \delta_{jk} \, \mathbf{1}_2, \tag{4.7}$$

$$\sigma_j \, \sigma_k - \sigma_k \, \sigma_j = 2 \sum_{m=1}^{3} \epsilon_{jkm} \, \sigma_m. \tag{4.8}$$

The verification is left as an exercise (see Exercise 4.4 below).

EXERCISE 4.1. *Show that all the Pauli matrices have the eigenvalues* $+1$ *and* -1. *Find a basis in* $\mathbb{C}^2$ *that consists of eigenstates of the matrix* σ_1.

EXERCISE 4.2. *Show that any Hermitian two-by-two matrix* Q *can be written in the form* (4.6). *Show that for a given matrix* Q, *the coefficients* a_0 *and* $\mathbf{a} = (a_1, a_2, a_3)$ *can be determined from*

$$a_0 = \operatorname{Tr} Q, \qquad a_k = \operatorname{Tr}(Q\sigma_k), \tag{4.9}$$

where Tr *denotes the trace of the matrix* (*the sum of the diagonal elements*).

EXERCISE 4.3. *Verify the anticommutation relation* (4.7) *and the commutation relation* (4.8) *for Pauli matrices.*

EXERCISE 4.4. *Use the explicit definition* (3.38) *of the Pauli matrices to verify the following formulas*

$$\sigma_j \, \sigma_k = \delta_{jk} \, \mathbf{1}_2 + \sum_{m=1}^{3} \epsilon_{jkm} \, \sigma_m \quad \textit{for all } j, \, k = 1, 2, 3. \tag{4.10}$$

$$\sigma_j \, \sigma_k = \sigma_m \quad \textit{if } (j, k, m) \textit{ is a cyclic permutation of } (1, 2, 3). \tag{4.11}$$

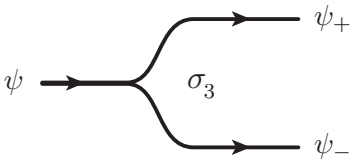


FIGURE 4.1. Feynman's schematic diagram of a measurement of the spin-component σ_3. The result is determined by acquiring the "which-way information."

4.2. Measurement and Preparation

The procedures of state preparation (preparatory measurement) and state verification (determinative experiment) can be illustrated clearly with spin measurements. This gives us an occasion to discuss once again the subtleties of the quantum measurement process.

4.2.1. Stern-Gerlach experiment

Given qubits in a state ψ, we want to measure the observable σ_3. To that purpose we send the qubits through a vertically oriented Stern-Gerlach apparatus schematically depicted in Figure 4.1. Each particle that passes through the device is either deflected up or down. The measurement is done by looking which way the particle takes ("*which-way information*"). Every single measurement gives a result that can be described by one of the eigenvalues of σ_3, and never something else.

What does quantum mechanics predict about the measurement of σ_3? We write the initial state as a linear combination $\psi = c_+\,\psi_+ + c_-\,\psi_-$ of the eigenvectors $\psi_\pm$ of σ_3. The expressions

$$p_{\psi\to\psi_\pm} = \big|\langle\psi_\pm, \psi\rangle\big|^2 = |c_\pm|^2 \tag{4.12}$$

are the probabilities that a particle in the state ψ is actually found to be in the state ψ_+ or ψ_-. Measuring σ_3 in the state ψ will thus give $+1$ ("spin-up") with probability $|c_+|^2$ and -1 ("spin-down") with probability $|c_-|^2$.

One has to admit that the word "measurement" is badly chosen, as it suggests that observables have some value that is revealed by the measurement procedure. But, this is a classical preconception. Actually, the quantum mechanical formalism makes no statement about the value of an observable before the measurement, or whether the observable actually has any value. In quantum mechanics, observables are represented by operators, and operators are not values. Except for the special situation that the system is in an eigenstate of the observable, the observable does not have a value prior to, and independent of, the measurement.

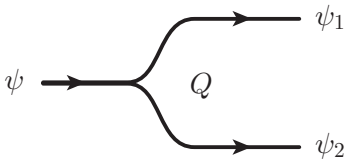
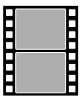


FIGURE 4.2. Diagram for the measurement of a qubit observable Q. The device has one input channel and two output channels, corresponding to the two possible results of the measurement. ψ_1 and ψ_2 denote the eigenvectors belonging to the two real eigenvalues of the Hermitian two-by-two matrix Q.

Quantum mechanics gives only probabilities for the various possible results of a measurement. We cannot predict the outcome of an *elementary measurement* (a measurement that is performed only once on a single quantum system), unless we know that the system is in an eigenstate of the observable being measured.

Figure 4.1 is an abstract symbol for the measurement of the observable σ_3. It may be realized by a Stern-Gerlach apparatus for a particle with spin 1/2, or by a birefringent (doubly refracting) crystal for measuring the polarization states of a photon (see Section 4.5.1 below). We are going to use the same symbol for an arbitrary qubit observable Q, even if the practical realization of that measurement bears no similarity with a Stern-Gerlach experiment (see Fig. 4.2).



Stern-Gerlach experiments are often done with atoms effusing from a heated gas. The particles thus enter the apparatus in a random state, and the spin measurement produces just a random bit. CD 4.2 simulates a Stern-Gerlach experiment with qubits in a random state. As soon as the which-way information becomes available, the qubit gets projected into the corresponding spin-eigenstate.

The measurement of a quantum observable is a typical *random experiment*. In probability theory, a random experiment is an experiment whose outcome is uncertain, until the experiment is actually run. The set of possible outcomes forms the *sample space* of the experiment. It is usually assumed that a random experiment can be repeated indefinitely under identical conditions. An elementary experiment is just a single run of the random experiment. A *random variable* is a function defined on the sample space.

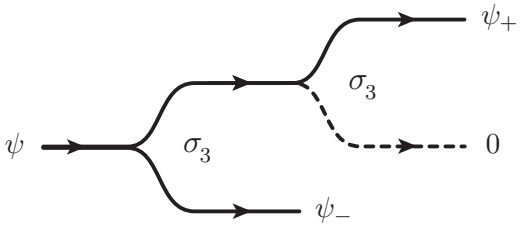


FIGURE 4.3. A qubit emerging with spin-up from a Stern-Gerlach apparatus is in the state described by ψ_+. A second measurement of σ_3 gives spin-up with certainty.

In each run of the experiment, the random variable thus assumes a value that depends on the random outcome.

In the case of an experiment with qubits, the sample space consists of only two elements. For the Stern-Gerlach experiment measuring σ_3 the sample space {"upper path", "lower path"}. We can define a random variable S with the values $+1$ and -1 for the two possible outcomes. For each run of the experiment, the random variable S describes the value of the qubit observable σ_3. Note that a random variable is only associated with the observable being measured. In the measurement of σ_3, no random variable is associated with σ_1 or σ_2, because no values are obtained for these observables.

4.2.2. Projection postulate

After measuring σ_3 of an individual qubit with a Stern-Gerlach apparatus, the qubit is in the state specified by the outcome of the measurement. If, for example, we find the result $+1$, then—right after the measurement—the system is in the state described by the eigenvector ψ_+ of σ_3. This can be verified by sending the particle through a second Stern-Gerlach apparatus whose inhomogeneous field has the same orientation (see Fig. 4.3). One finds that whenever a particle emerges from the first apparatus with spin-up, it will pass the second apparatus also with spin-up. The first measurement obviously changes the initial state vector ψ to a state vector $\phi = c_+\psi_+ + c_-\psi_-$ that satisfies $|c_+|^2 = 1$ and $|c_-|^2 = 0$, because the second measurement of σ_3 gives the result "spin up" with certainty. We conclude that $\phi = \psi_+$, up to an irrelevant phase factor.

The effect of the measurement of σ_3 is well summarized by the projection postulate (see Book One, Section 4.8): Define the projection operators

$$P = |\psi_+\rangle\langle\psi_+|, \qquad 1 - P = |\psi_-\rangle\langle\psi_-| \tag{4.13}$$

onto the eigenspaces of σ_3. Assume that the particle is in a state ψ before the measurement of σ_3. Then, after the measurement, the particle is either

in the state $P\psi$ or in the state $(1-P)\psi$, depending on the result of the measurement.

> The measurement of an observable projects the state vector into the subspace that corresponds to the measured result. This changes the state, except if the state already belongs to that subspace.

For wave packets, the projection postulate claims that a measurement causes an instantaneous "collapse." Let us discuss this with the help of an example. Consider a spinor-wave packet that splits into two spatially separated parts when entering a Stern-Gerlach apparatus.



CD 3.5.3 and 3.12.1 show the splitting of a quantum wave packet in an inhomogenous magnetic field (Stern-Gerlach apparatus) without showing the measurement of the state.

Measuring the deflection of the particle (the which-way information) tells us in which part of the wave packet we are going to find the particle. This amounts to a measurement of the position. The projection erases the "wrong part" of the wave packet and keeps only the part that agrees with the result of the measurement. This instantaneous collapse to the "right part" appears dramatic only if the wave function is regarded as a physical field. I prefer to see the wave function as a container of physically available information about the particle at each moment. Then, the collapse just reflects the fact that new information about the position of the particle has become available.[2]

In any case, the projection postulate is a rather simple model of the measuring process. Any real measurement involves several physical systems (the object, a measuring apparatus, and an observer), and any description involving only the states of the measured object can only be a crude simplification.

4.2.3. Stern-Gerlach filter and state preparation

We measure a qubit observable Q by sending the qubit into an appropriate measurement device (see Fig. 4.2) and by determining through which channel it leaves the apparatus. One radical method to obtain this which-way information is to block one of the paths in the apparatus, say the path corresponding to the eigenstate ψ_2. Then, we can be sure that all particles emerging from the apparatus are in the other eigenstate ψ_1. The modified

[2]This does not mean that there has to be a conscious observer to register that information. With "becoming available" I mean that in the combined system consisting of qubit and measuring apparatus the information is physically realized in a way that it has influence upon the time evolution in the future.

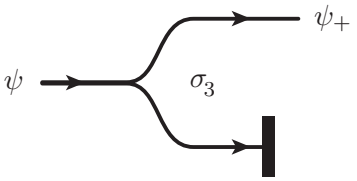


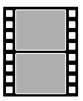
FIGURE 4.4. A Stern-Gerlach filter is a Stern-Gerlach apparatus where one of the partial beams is blocked off. All particles emerging from this device have spin-up (with respect to the positive z-direction). Hence, the Stern-Gerlach filter is a device for preparing the state ψ_+.

Stern-Gerlach apparatus with one blocked channel will be called a *Stern-Gerlach filter* (see Fig. 4.4). A filter is a device that passes only particles with a certain value of the observable the filter measures, that is, particles with a certain *property*. In the formalism, properties are represented by projection operators. For example, the Stern-Gerlach filter for the state ψ_+ shown in Figure 4.4 is the physical equivalent of the projection operator $P = |\psi_+\rangle\langle\psi_+|$.

A filter serves to prepare a system in a certain state. For example, if we need a qubit in the state ψ_+, we can send it into the filter shown in Figure 4.4. When the qubit (whose initial state may be unknown) leaves the filter, then it is in the desired quantum state ψ_+. When it gets absorbed or reflected (which, in principle, can be detected), then our attempt to prepare ψ_+ has been unsuccessful, and we have to repeat the preparation procedure with another qubit.

A state preparation requires only a single measurement. Hence, the state ψ that is obtained in that way characterizes a single quantum system.

In Section 4.4.1 below, we are going to prove that any state of a qubit can be prepared with an appropriately oriented Stern-Gerlach filter.



The act of projecting a spinor-wave packet onto an eigenstate of the spin by blocking one path in the Stern-Gerlach apparatus is shown in CD 3.7, CD 3.12.2, and CD 3.12.3. In CD 4.3, we simulate an experiment where qubits emerging from a random source are first prepared in the state ψ_+, and then the observable σ_3 is measured.

4.3. Ensemble Measurements

4.3.1. State verification

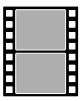
The state ψ describes how a qubit was prepared. Hence, it is meaningful to speak of the state of an individual qubit. But what is the meaning of the probabilities that quantum mechanics allows us to compute for this state? Can we associate probabilities with single events? Indeed, there are approaches to probability theory that attempt to do just that.[3] But no matter how you think about this, a quantitative verification (or falsification) of the predicted probabilities can only be done by repeating an experiment very often and by counting the various outcomes. Therefore, we make a clear distinction between *single measurements* (which we also called *elementary experiments*) and *ensemble measurements*.

> An ensemble measurement consists of repeated elementary experiments performed on a very large number of identically prepared systems.

An ensemble measurement of σ_3 in a Stern-Gerlach apparatus divides an ensemble $\mathcal{E}$ of incoming particles into two sub-ensembles $\mathcal{E}_+$ and $\mathcal{E}_-$ according to the which-way information obtained in the course of the experiment. One then counts the numbers n_+ and n_- of particles in the sub-ensembles and divides by the total number of particles n. If n is large enough, the fractions n_+/n and n_-/n are approximations to the probabilities for spin-up and spin-down, respectively. That is, for qubits with initial state ψ, we obtain for the probabilities (4.12):

$$p_{\psi\to\psi_\pm} \approx \frac{n_\pm}{n} \quad \text{for } n \text{ large.} \tag{4.14}$$

> A probability distribution can only be verified by an ensemble measurement, where we count the fraction of elementary measurements with a specified outcome.



CD 4.4 simulates a measurement of the observable σ_1 on qubits prepared in an eigenstate of σ_3. In CD 4.5, you can determine the probability that a qubit in the state ψ_+ is found with spin-up in a Stern-Gerlach apparatus that is rotated through an arbitrary angle α about the direction of the beam.

[3] According to the Bayesian point of view, the numerical value of the probability is a measure of the plausibility one gives to a hypothesis on the basis of available information.

When experiments are actually performed, one often works with a beam of particles, and the probabilities are found by measuring the intensities of the partial beams emerging from the Stern-Gerlach apparatus. Thus, ensemble measurements are carried out automatically. Note that we do not consider interference or interaction between different members of the same ensemble. In order to make sure that the particles of the ensemble are indeed independent of each other, one can make the intensity of the particle beam so low that one particle is already in the detector before the next particle leaves the source. It is an experimentally verified fact that all predictions of quantum mechanics remain valid for this situation.

4.3.2. Determining an unknown state

Ensemble measurements are also necessary to determine an unknown state ψ (determinative measurement). In order to determine an unknown state $\psi = c_+\psi_+ + c_-\psi_-$ experimentally, we have to measure, among other things, the probability $|c_+|^2 = 1 - |c_-|^2$. Counting a finite number of spin-up events of course cannot give the exact numerical value.[4] An accurate state determination is impossible with a finite number of elementary measurements.

An ensemble measurement of σ_3 in the state $\psi = c_+\psi_+ + c_-\psi_-$ can reveal the values of $|c_\pm|^2$ (with a fidelity depending on the number of elementary experiments). But even the knowledge of $|c_\pm|$ is not sufficient to determine the state (unless the state is one of the eigenstates of σ_3). Because $c_\pm$ are complex numbers, we can write $c_\pm = |c_\pm| \exp(\mathrm{i}\phi_\pm)$ and

$$\psi = \mathrm{e}^{\mathrm{i}\phi_+} \Big(|c_+|\, \psi_+ + |c_-|\, \mathrm{e}^{\mathrm{i}(\phi_- - \phi_+)}\, \psi_- \Big). \tag{4.15}$$

The overall phase ϕ_+ has no physical significance, because all vectors in the one-dimensional subspace spanned by ψ represent the same state. But the relative phase $\phi_r = \phi_- - \phi_+$ has a physical meaning. (For example, the linear combination $\psi_+ + \mathrm{e}^{\mathrm{i}\pi}\,\psi_-$ with relative phase π is orthogonal to the linear combination $\psi_+ + \psi_-$ with relative phase 1.) In order to determine ϕ_r, one also has to measure the probabilities for spin-up and down with respect to other directions (see the exercises below).

[4] A notable exception occurs if we know the orientation of the Stern-Gerlach apparatus that has been used to prepare the unknown state. Then a single measurement with a Stern-Gerlach apparatus in the same orientation can tell you whether the state was prepared as spin-up or spin-down.

In order to measure an unknown state exactly, one would need an infinite ensemble of identically prepared systems. Moreover, it is not sufficient to measure only σ_3. In general, one needs also information about the other components of the spin.

It is believed that a qubit can work as a basic information processing unit in a quantum computer. Quantum computers have a large potential, because the superposition states of a qubit hold much more information than a classical bit. The main problem with quantum computers is that this information is not directly accessible. A quantum computer has no advantage if it is necessary to repeat the measurement many times in order to determine the final state of the qubit holding the result.

EXERCISE 4.5. *Assume that a qubit is prepared in one of two known non-orthogonal states ψ_1 and ψ_2. Can you decide with a single measurement which of the two states was actually prepared?*

EXERCISE 4.6. *With $p \in [0,1]$ and $\phi \in [0, 2\pi)$, consider the $\mathbb{C}^2$-vector*

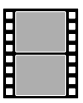
$$\psi = \sqrt{p}\,\psi_+ + \sqrt{1-p}\,\mathrm{e}^{\mathrm{i}\phi}\,\psi_-. \tag{4.16}$$

Can we reasonably assume that any qubit state is represented by a vector like this? What is the physical meaning of p? Describe a procedure for measuring p. Can you measure p in a single elementary measurement, or is it necessary to perform an ensemble measurement?

EXERCISE 4.7. *Given ψ as in the previous exercise, express the probability q for finding spin-up in the x-direction in terms of p and ϕ. Proceed to show that some information about the relative phase ϕ can be obtained from a measurement of σ_1. More precisely, show that*

$$\cos\phi = \frac{q - 1/2}{\sqrt{p(1-p)}}. \tag{4.17}$$

EXERCISE 4.8. *The probabilities p and q from the previous exercises still do not determine ψ uniquely. What is the remaining ambiguity? Would it help to measure σ_2?*



If we perform a sufficiently large number of measurements on qubits in a state ψ, we can get a good approximation to the true state, because the measured relative frequencies approximate the true probabilities. This process is called *state estimation*. It is illustrated by CD 4.6. This simulation uses the results of Exercises 4.6–4.8.

4.3.3. Classical teleportation is impossible

Our observations about preparing a state (of individual qubits) versus determining the state (of an ensemble) are often described in a pregnant way as follows:

> It is impossible to measure the state of single qubit in such a way that the acquired information is sufficient to prepare another qubit in the same state.

Consider the following setup (see Fig. 4.5). A physicist called Alice receives a qubit in a quantum state ψ. It is assumed that the preparation procedure and hence the state of that qubit is unknown to Alice. She measures the state of the qubit with some measuring apparatus M. Then she tells the result to another physicist called Bob who resides in a distant laboratory. This information transfer is done by classical methods, for example, via telephone. Bob uses that information to choose a certain preparation procedure and to prepare a qubit in some quantum state ψ'. Can one arrange things in such a way that for every possible input ψ, Bob's output ψ' is identical with ψ? This would be called *classical teleportation*,[5] because the information transfer uses classical bits and bytes.

We can test the reliability of any proposed teleportation scheme as follows. We send to Alice a sequence of qubits in randomly chosen states only known to us. Then we measure some observable on the qubits produced by Bob. As we know the states of the input qubits, we can predict the statistical distribution of values measured on these qubits. A successful teleportation would mean that any measurement of Bob's output gives the same distribution as if the measurement was performed directly on Alice's input. This is impossible, because the measurement done by Alice on an individual qubit gives insufficient information to infer the preparation procedure used for that qubit. Hence, Bob has to guess the preparation procedure for each qubit, and our statistical test will reveal his errors. The failure to teleport qubit states in that way is remarkable, because the teleportation of classical bits obviously presents no problems.

The impossibility of the classical teleportation is related to the fact that a quantum state cannot be copied: There is no physical process that uses a qubit in some (unknown) state ψ in order to prepare two qubits in the same state ψ. (This "no-cloning theorem" will be proved in Section 6.8.4 below). Otherwise, Alice could use a quantum copier to produce many copies of the

[5]The commonly used expression "teleportation" is badly chosen. One does not attempt to transfer the qubit itself ("only" its state), and the transfer is by no means instantaneous.

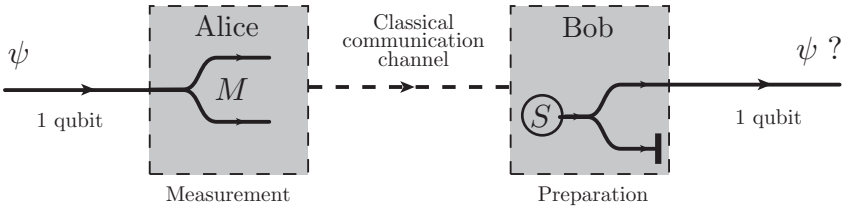


FIGURE 4.5. The (impossible) classical teleportation protocol. Alice receives a qubit in a state ψ and performs some measurement. The result is sent to Bob via a classical information channel. On the basis of this information, Bob tries to prepare a qubit in the same state.

input and estimate the state via an ensemble measurement as described in the previous section. Alice can then advise Bob to produce a more or less accurate copy of the input. (It is still impossible to reconstruct the state *exactly* on the basis of a finite number of measurements.) In Section 6.6.2 below, we are going to discuss the possibility of *quantum teleportation*, where the classical communication between Alice and Bob is assisted by quantum-mechanical correlations.

4.4. Qubit Manipulations

It is very natural to ask the following question. If ψ is not an eigenstate of σ_3, perhaps it is a spin-up eigenstate with respect to some other direction in $\mathbb{R}^3$? The answer is given in this section.

4.4.1. All states are "spin-up" in some direction

The component of a vector $\mathbf{a} \in \mathbb{R}^3$ in the direction of a unit vector $\mathbf{n}$ is given by the scalar product $\mathbf{a} \cdot \mathbf{n}$. Similarly, the Hermitian matrix

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \begin{pmatrix} n_3 & n_1 - \mathrm{i}n_2 \\ n_1 + \mathrm{i}n_2 & -n_3 \end{pmatrix} \tag{4.18}$$

defines the component of the spin in the direction of $\mathbf{n}$. We can measure the observable $\boldsymbol{\sigma} \cdot \mathbf{n}$ with a Stern-Gerlach apparatus whose inhomogeneous magnetic field is oriented in the direction of $\mathbf{n}$.[6] Hence, we can also prepare states that are spin-up in the direction of $\mathbf{n}$ by blocking one of the paths in

[6]The usual Stern-Gerlach method has to be changed in order to measure, say, σ_2 for particles moving in the y-direction. Nowadays, spin-measurements with respect to arbitrary directions can be performed on laser-cooled atoms released from a magneto-optical trap.

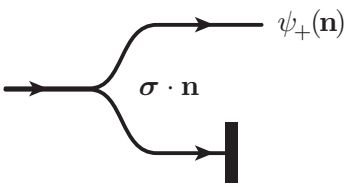


FIGURE 4.6. A Stern-Gerlach filter with magnetic field pointing in the direction $\mathbf{n}$. It prepares the spin-up eigenstate of $\sigma\cdot\mathbf{n}$. Every possible state of a qubit has a unique "spin-up direction" and can hence be realized in this way.

the Stern-Gerlach apparatus. The corresponding Feynman symbol is shown in Figure 4.6.

The Hermitian two-by-two matrix $\boldsymbol{\sigma}\cdot\mathbf{n}$ has the eigenvalues $+1$ and -1 and normalized eigenvectors

$$\psi_\pm(\mathbf{n}) = \frac{1}{\sqrt{2(1\pm n_3)}}\begin{pmatrix} n_3\pm 1 \\ n_1+\mathrm{i}n_2\end{pmatrix}. \tag{4.19}$$

On the north pole $\mathbf{n}=\mathbf{e}_z$ and on the south pole $-\mathbf{e}_z$ of the unit sphere, we have to complete the definition (4.19). We choose

$$\psi_+(-\mathbf{e}_z) = \begin{pmatrix}0\\1\end{pmatrix} = \psi_-(\mathbf{e}_z). \tag{4.20}$$

$\psi_+(\mathbf{n})$ defines the state with spin-up in the direction of $\mathbf{n}$, $\psi_-(\mathbf{n})$ has spin-down. The two eigenvalues of $\boldsymbol{\sigma}\cdot\mathbf{n}$ are the only possible results of measuring the spin $\boldsymbol{\sigma}$ in the direction $\mathbf{n}$. It is indeed one of the very counterintuitive features of the spin that its component along any axis shows just two possible values. The spin cannot be understood as a classical vector attached to the particle (see Section 4.8.1 below).

Now, let ψ be any normalized state in $\mathbb{C}^2$. Then

$$\mathbf{n}(\psi) = \begin{pmatrix}\langle\sigma_1\rangle_\psi\\ \langle\sigma_2\rangle_\psi\\ \langle\sigma_3\rangle_\psi\end{pmatrix}, \quad \text{with } \langle\sigma_i\rangle_\psi = \langle\psi,\sigma_i\psi\rangle, \tag{4.21}$$

is a unit vector in $\mathbb{R}^3$. By a little calculation you can verify that ψ is an eigenvector of $\boldsymbol{\sigma}\cdot\mathbf{n}(\psi)$ belonging to the eigenvalue $+1$,

$$\big(\boldsymbol{\sigma}\cdot\mathbf{n}(\psi)\big)\,\psi = \psi. \tag{4.22}$$

The vector $\psi_+\big(\mathbf{n}(\psi)\big)$ defined as in (4.19) is another eigenvector of $\boldsymbol{\sigma}\cdot\mathbf{n}(\psi)$ belonging to the eigenvalue $+1$. Normalized eigenvectors are unique up to a phase, hence

$$\psi = \mathrm{e}^{\mathrm{i}\lambda}\psi_+\big(\mathbf{n}(\psi)\big), \tag{4.23}$$

with some suitable phase factor $\exp(\mathrm{i}\lambda)$, that is, ψ and $\psi_+\big(\mathbf{n}(\psi)\big)$ describe the same quantum state. The phase λ is just the argument of the upper component of ψ, because the upper component of $\psi_+(\mathbf{n})$ is defined to be real-valued for all $\mathbf{n}$. We call $\mathbf{n}(\psi)$ the *spin-up direction* of ψ.

Conversely, if $\hat{\mathbf{n}}$ is some unit vector in $\mathbb{R}^3$, then $\psi_+(\hat{\mathbf{n}})$ is the corresponding spin-up eigenstate and

$$\mathbf{n}\big(\psi_+(\hat{\mathbf{n}})\big) = \hat{\mathbf{n}}. \tag{4.24}$$

Any vector $\psi \in \mathbb{C}^2$ with the property $\mathbf{n}\big(\psi\big) = \hat{\mathbf{n}}$ can differ from $\psi_+(\hat{\mathbf{n}})$ only by a phase.

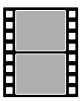
Hence, any vector $\psi \in \mathbb{C}^2$ can be interpreted as a spin-up state with respect to a unique direction $\mathbf{n}(\psi)$. The spin-up direction in turn determines the vector ψ uniquely up to a phase factor.

There is a one-to-one correspondence between directions $\mathbf{n} \in \mathbb{R}^3$ and the states of a qubit. The correspondence is given by Eqs. (4.19) and (4.21)

$$\mathbf{n} \mapsto \psi_+(\mathbf{n}), \qquad \psi \mapsto \mathbf{n}(\psi). \tag{4.25}$$

For a given $\mathbf{n}$, the spinor $\psi_+(\mathbf{n})$ is the normalized spin-up eigenvector of $\boldsymbol{\sigma} \cdot \mathbf{n}$ (it is unique up to a phase). For a given ψ, the components of $\mathbf{n}(\psi)$ are the expectation values of the Pauli matrices in the state ψ. Any state of the qubit can be prepared as a spin-up eigenstate with an appropriately oriented Stern-Gerlach filter.

Ψ The mapping $\psi_+ : \mathbf{n} \mapsto \psi_+(\mathbf{n})$ from S^2 (the unit sphere in $\mathbb{R}^3$) into the set of unit vectors in $\mathbb{C}^2$ is discontinuous. The lower component has a phase-discontinuity at the south pole of the unit sphere. The normalized eigenvectors $\psi_+(\mathbf{n})$ are only unique up to a phase factor $\exp\big(\mathrm{i}\lambda(\mathbf{n})\big)$. In Eq. (4.19), we defined the phase by requiring that the first component of $\psi_\pm$ be real-valued. There is no choice of the phase factor that would lead to a continuous mapping ψ_+.



The association of qubit states with directions in $\mathbb{R}^3$ is visualized in CD 4.7. In CD 4.8, we represent states by the magnetic needles already familiar from the visualization of a spinor field in CD 3.8.

EXERCISE 4.9. *Verify* $\boldsymbol{\sigma} \cdot \mathbf{n}(\psi)\, \psi = \psi$ *with* $\mathbf{n}(\psi)$ *as in* (4.21).

EXERCISE 4.10. *In spherical coordinates, the axis* $\mathbf{n}$ *is described by a polar angle* ϑ *and an azimuthal angle* φ. *Show that the matrix* $\boldsymbol{\sigma} \cdot \mathbf{n}$ *has the*

following representation in spherical coordinates

$$\boldsymbol{\sigma}\cdot\mathbf{n} = \begin{pmatrix} \cos\vartheta & \mathrm{e}^{-\mathrm{i}\varphi}\sin\vartheta \\ \mathrm{e}^{\mathrm{i}\varphi}\sin\vartheta & -\cos\vartheta \end{pmatrix} \tag{4.26}$$

EXERCISE 4.11. *Show that for any unit vector* $\mathbf{n}$ *in* $\mathbb{R}^3$*, the operator* $P_{\mathbf{n}} = \frac{1}{2}(\mathbf{1}_2 + \boldsymbol{\sigma}\cdot\mathbf{n})$ *is a projection operator that projects onto the eigenstate with spin up in the direction of* $\mathbf{n}$*. Take an arbitrary state* $\psi = c_+\psi_+ + c_-\psi_- \neq \psi_-$ *and assume that* $c_+ > 0$ *(which can always be achieved by multiplying* ψ *with a phase factor). Show that you can choose* $\mathbf{n}$ *such that* $P_{\mathbf{n}}\psi_+$ *is equal to* ψ *after renormalization (that is,* $P_{\mathbf{n}}\psi_+ = k\psi$ *with some constant* k*). Hence, it is possible to prepare any state except* ψ_- *by applying a Stern-Gerlach filter to particles that are originally in the state* ψ_+*.*

EXERCISE 4.12. *Compute* $\mathbf{n}(\psi)$ *for* $\psi = \sqrt{p}\,\psi_+ + \sqrt{1-p}\,\mathrm{e}^{\mathrm{i}\phi}\,\psi_-$ *and for* $\overline{\psi}$*.*

4.4.2. Rotations of a qubit

Just as rotations of wave functions are generated by the orbital angular-momentum operator $\mathbf{L}$, the rotations of qubits are generated by the spin operator $\mathbf{S}$. The rotations around the axis given by a unit vector $\mathbf{n}$ are generated by the component of the spin with respect to this axis, that is, by $\mathbf{S}\cdot\mathbf{n}$.

Spinor rotation:

The rotation of a particle with spin 1/2 through an angle α about an axis $\mathbf{n}$ is given by the unitary operator

$$U(\boldsymbol{\alpha}) = \exp\Big(-\frac{\mathrm{i}}{\hbar}\,\mathbf{S}\cdot\boldsymbol{\alpha}\Big), \quad \text{with } \boldsymbol{\alpha} = \alpha\mathbf{n}. \tag{4.27}$$

In the standard representation,

$$U(\boldsymbol{\alpha}) = \exp\Big(-\frac{\mathrm{i}}{2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\alpha}\Big) = (\cos\frac{\alpha}{2})\,\mathbf{1}_2 - \mathrm{i}(\sin\frac{\alpha}{2})\,\boldsymbol{\sigma}\cdot\mathbf{n}. \tag{4.28}$$

The exponential function of the matrix $\boldsymbol{\sigma}\cdot\mathbf{n}$ can be calculated by the power series

$$\exp\Big(-\mathrm{i}\,\frac{\alpha}{2}\,\boldsymbol{\sigma}\cdot\mathbf{n}\Big) = \sum_{k=0}^{\infty}\frac{1}{k!}\left(-\mathrm{i}\,\frac{\alpha}{2}\right)^k(\boldsymbol{\sigma}\cdot\mathbf{n})^k. \tag{4.29}$$

The result (4.28) is easily obtained if we use the property $(\boldsymbol{\sigma}\cdot\mathbf{n})^2 = 1$, and compare the resulting power series with the well-known Taylor series for the sine and cosine functions.

EXERCISE 4.13. *Verify that the infinite sum* (4.29) *indeed gives* (4.28).

The operator $U(\boldsymbol{\alpha})$ is unitary, $U(\boldsymbol{\alpha})^\dagger = U(\boldsymbol{\alpha})^{-1} = U(-\boldsymbol{\alpha})$, and it has the astonishing property that

$$U(2\pi\mathbf{n}) = -\mathbf{1}_2. \tag{4.30}$$

This means that a rotation through an angle 2π turns a state vector into its negative. We can accept this, because a vector ψ and its negative $-\psi$ describe the same state (they are elements of the same ray in the Hilbert space).



CD 4.8 visualizes rotations of a qubit around the z- and y-axes. A rotation through an angle 2π multiplies the state vector with -1.

Now it is time for a little consistency check. On the one hand, rotations in $\mathbb{R}^3$ are given by orthogonal 3×3 matrices $\mathbf{R}(\boldsymbol{\alpha})$ with determinant $+1$. On the other hand, we have claimed that a unitary two-by-two matrix $U(\boldsymbol{\alpha})$ describes a rotation of spinors. This is only meaningful if a rotation by $U(\boldsymbol{\alpha})$ turns the basis state ψ_+ into the spin-up eigenstate with respect to an axis that is obtained from the z-axis by the rotation $\mathbf{R}(\boldsymbol{\alpha})$. More generally, for any unit vector $\hat{\mathbf{n}} \in \mathbb{R}^3$,

$$U(\boldsymbol{\alpha})\,\psi_+(\hat{\mathbf{n}}) = \psi_+\big(\mathbf{R}(\boldsymbol{\alpha})\hat{\mathbf{n}}\big) \quad \text{(up to a phase)}. \tag{4.31}$$

It is easy to verify this statement in special cases; a proof of the general case is left to Section 4.4.4. For example, consider a rotation around the y-axis, which maps the z-direction into the positive x-direction. With $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ denoting the unit vectors in $\mathbb{R}^3$,

$$\mathbf{R}(\pi\mathbf{e}_y/2)\,\mathbf{e}_z = \mathbf{e}_x. \tag{4.32}$$

It is easy to see that the corresponding unitary two-by-two rotation matrix is

$$U(\pi\mathbf{e}_y/2) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \tag{4.33}$$

Hence, the basis vector ψ_+ (spin-up in z-direction) is transformed into

$$U(\pi\mathbf{e}_y/2)\,\psi_+ = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(\psi_+ + \psi_-), \tag{4.34}$$

and it is easy to verify that this is an eigenstate of the matrix σ_1 belonging to the eigenvalue $+1$ (spin-up in x-direction).

We note that any unitary two-by-two matrix U can be written as $\exp(\mathrm{i}Q)$ with some Hermitian matrix Q. Because any Hermitian two-by-two matrix

can be written as $Q = \theta\,\mathbf{1}_2 - \boldsymbol{\alpha}\cdot\boldsymbol{\sigma}/2$ (see Eq. (4.6)) we find that every unitary matrix is of the form

$$U = \mathrm{e}^{\mathrm{i}\theta}\,\mathrm{e}^{-\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}/2}. \tag{4.35}$$

Characterization of unitary qubit transformations:

Up to a phase factor, any two-by-two unitary matrix U can be interpreted as a rotation matrix, that is,

$$U = \mathrm{e}^{\mathrm{i}\theta}\,U(\boldsymbol{\alpha}), \tag{4.36}$$

with suitable parameters θ and $\boldsymbol{\alpha}$. The unitary rotation matrices $U(\boldsymbol{\alpha})$ are distinguished by the property that their determinant is $+1$,

$$\mathrm{Det}\,U(\boldsymbol{\alpha}) = 1. \tag{4.37}$$

For example, the matrix

$$U_{\mathrm{H}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = \frac{1}{\sqrt{2}}\,(\sigma_1+\sigma_3) \tag{4.38}$$

can be written as

$$U_{\mathrm{H}} = \mathrm{i}\,U(\pi\mathbf{n}), \quad \text{with } \mathbf{n} = \frac{1}{\sqrt{2}}\,(1,0,1). \tag{4.39}$$

The matrix U_{H} is called the *Hadamard transformation*. It is unitary *and* Hermitian. Therefore, it also satisfies $U_{\mathrm{H}}^2 = \mathbf{1}_2$. We find $U_{\mathrm{H}}\,\sigma_1\,U_{\mathrm{H}} = \sigma_3$ and $U_{\mathrm{H}}\,\sigma_3\,U_{\mathrm{H}} = \sigma_1$. Hence, U_{H} maps the eigenstates of σ_3 onto the eigenstates of σ_1. This is intuitively clear, because the rotation in (4.39) exchanges the z-axis with the x-axis.

Exercise 4.14. *Show that with* $\mathbf{n}(\psi)$ *as in Eq.* (4.21),

$$U\big(2\alpha\,\mathbf{n}(\psi)\big)\,\psi = \mathrm{e}^{-\mathrm{i}\alpha}\,\psi. \tag{4.40}$$

4.4.3. Time evolution of the spin in a magnetic field

Qubits realized as particles with a magnetic moment can be rotated with the help of a homogeneous magnetic field. Neglecting the kinetic energy, the energy of a particle with spin in a magnetic field $\mathbf{B}$ is

$$H = -\boldsymbol{\mu}\cdot\mathbf{B}. \tag{4.41}$$

As an example, we consider an electron. Using Eq. (3.58), we insert $\boldsymbol{\mu} = -\mu_{\mathrm{B}}\,\boldsymbol{\sigma}$ and obtain the energy operator (in the standard representation)

$$H = \mu_{\mathrm{B}}\,\boldsymbol{\sigma}\cdot\mathbf{B}. \tag{4.42}$$

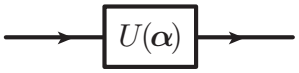


FIGURE 4.7. Diagram representing the unitary transformation $U(\boldsymbol{\alpha})$ of a qubit.

The unitary operator of time evolution generated by H is thus a spinor rotation

$$\exp\Big(-\frac{\mathrm{i}}{\hbar}Ht\Big) = \exp\Big(-\mathrm{i}\,\frac{t}{2}\,\boldsymbol{\omega}\cdot\boldsymbol{\sigma}\Big) \qquad \text{with} \qquad \boldsymbol{\omega} = 2\,\frac{\mu_{\mathrm{B}}}{\hbar}\,\mathbf{B}. \tag{4.43}$$

The time evolution of a particle with spin 1/2 in the presence of a magnetic field $\mathbf{B}$ is described by a rotation $U(t\boldsymbol{\omega})$ where the axis of the rotation is defined by the direction of $\mathbf{B}$. The rotation has the uniform angular speed (using $\mu_{\mathrm{B}} = e\hbar/2\mathrm{m}_e$)

$$\omega = \frac{e}{\mathrm{m}_e}\,|\mathbf{B}|, \tag{4.44}$$

in agreement with the classical result (3.27). Hence, we can always perform an arbitrary rotation of the state just by putting the particle into a suitable magnetic field for a certain time. At least in principle, we have a method to realize an arbitrary unitary transformation (up to a phase) in an experiment.

It is a crucial assumption in quantum information theory that any unitary transformation can be applied to a qubit. The Feynman symbol for a unitary transformation U is shown in Figure 4.7.

If the qubit is not realized by a particle with spin, but by two states of some other system (for example, by the polarization states of photons; see below), then the realization of unitary transformations depends on physics of that system and the experimental means available for the manipulation of the system.

4.4.4. Special topic: Spinor rotations

There is a one-to-one correspondence between vectors $\mathbf{a} \in \mathbb{R}^3$ and Hermitian two-by-two matrices with trace zero. This correspondence is given by the mapping

$$\mathbf{a} \longrightarrow \mathbf{a}\cdot\boldsymbol{\sigma} = \begin{pmatrix} a_3 & a_1 - \mathrm{i}a_2 \\ a_1 + \mathrm{i}a_2 & -a_3 \end{pmatrix} \tag{4.45}$$

In Exercise 4.2, it was shown that the components of the vector $\mathbf{a}$ can be obtained from a given Hermitian matrix Q with trace zero by the formulas

$$a_k = \frac{1}{2}\,\mathrm{Tr}\,(Q\sigma_k). \tag{4.46}$$

Furthermore, a little calculation shows that the determinant of $\mathbf{a}\cdot\boldsymbol{\sigma}$ is related to the length a of the vector $\mathbf{a}$ by

$$\operatorname{Det}\mathbf{a}\cdot\boldsymbol{\sigma} = -\mathbf{a}\cdot\mathbf{a} = -a^2. \tag{4.47}$$

Now an arbitrary unitary transformation U transforms a Hermitian matrix $Q = \mathbf{a}\cdot\boldsymbol{\sigma}$ into a matrix $Q' = UQU^\dagger$. A similarity transformation does not change the trace of a matrix, therefore $\operatorname{Tr} Q' = 0$. Moreover, $Q'^\dagger = (UQU^\dagger)^\dagger = (U^\dagger)^\dagger Q^\dagger U^\dagger = UQU^\dagger = Q'$, hence Q' is again Hermitian. We may write

$$U\,\mathbf{a}\cdot\boldsymbol{\sigma}\,U^\dagger = \mathbf{b}\cdot\boldsymbol{\sigma}. \tag{4.48}$$

This shows that the similarity transformation with a unitary matrix induces a unique mapping between vectors in $\mathbb{R}^3$. The vector $\mathbf{b}$ can be determined from our trace formula Eq. (4.46),

$$b_k = \frac{1}{2}\operatorname{Tr}(\sigma_k\,\mathbf{b}\cdot\boldsymbol{\sigma}) = \frac{1}{2}\operatorname{Tr}(\sigma_k\,U\,\mathbf{a}\cdot\boldsymbol{\sigma}\,U^\dagger) = \frac{1}{2}\sum_{j=1}^{3}\operatorname{Tr}(\sigma_k U\,\sigma_j\,U^\dagger)\,a_j \tag{4.49}$$

Hence, the transformation $\mathbf{a}\to\mathbf{b}$ is obviously described by a 3×3 matrix with the matrix elements

$$\mathbf{R}_{kj} = \frac{1}{2}\operatorname{Tr}(\sigma_k U\,\sigma_j\,U^\dagger). \tag{4.50}$$

Because the determinant of a complex two-by-two matrix is not changed under a unitary transformation, we find that

$$-a^2 = \operatorname{Det}\mathbf{a}\cdot\boldsymbol{\sigma} = \operatorname{Det}U\,\mathbf{a}\cdot\boldsymbol{\sigma}\,U^{-1} = \operatorname{Det}\mathbf{b}\cdot\boldsymbol{\sigma} = -b^2. \tag{4.51}$$

Hence, $\mathbf{b}$ has the same length as $\mathbf{a}$, which means that the transformation is orthogonal. In fact, one can show that the matrix $\mathbf{R}_{kj}$ is just the 3×3 rotation matrix corresponding to U. The matrix U can be written as $U = \mathrm{e}^{\mathrm{i}\theta}\,U(\boldsymbol{\alpha})$ with a uniquely given rotation vector $\boldsymbol{\alpha}$. The phase factor $\mathrm{e}^{\mathrm{i}\theta}$ drops out of the formula (4.50), and

$$\mathbf{R}_{kj} = \frac{1}{2}\operatorname{Tr}\big(\sigma_k U(\boldsymbol{\alpha})\,\sigma_j\,U^\dagger(\boldsymbol{\alpha})\big) = \mathbf{R}(\boldsymbol{\alpha})_{kj} \tag{4.52}$$

are just the matrix elements of the 3×3 rotation matrix defined in Eq. (1.11).

Exercise 4.15. *Combine the equations above to show that* (4.48) *can be written as*

$$U(\boldsymbol{\alpha})\,\mathbf{a}\cdot\boldsymbol{\sigma}\,U(\boldsymbol{\alpha})^{-1} = \big(\mathbf{R}(\boldsymbol{\alpha})\mathbf{a}\big)\cdot\boldsymbol{\sigma}. \tag{4.53}$$

Use this to show the relation

$$U(\boldsymbol{\alpha})^{-1}\,\sigma_k\,U(\boldsymbol{\alpha}) = \sum_{j=1}^{3}\mathbf{R}(\boldsymbol{\alpha})_{kj}\,\sigma_j, \quad k = 1,2,3, \tag{4.54}$$

which is usually abbreviated to

$$U(\boldsymbol{\alpha})^{-1}\,\boldsymbol{\sigma}\,U(\boldsymbol{\alpha}) = \mathbf{R}(\boldsymbol{\alpha})\,\boldsymbol{\sigma}. \tag{4.55}$$

EXERCISE 4.16. *With* $\mathbf{n}(\psi)$ *defined as in Eq.* (4.21)*, show that*

$$\mathbf{n}\big(U(\boldsymbol{\alpha})\psi\big) = \mathbf{R}(\boldsymbol{\alpha})\,\mathbf{n}(\psi). \tag{4.56}$$

EXERCISE 4.17. *Prove the relation*

$$U(\boldsymbol{\alpha})\psi = \mathrm{e}^{-\mathrm{i}\lambda}\,\psi_+\big(\mathbf{R}(\boldsymbol{\alpha})\,\mathbf{n}(\psi)\big), \tag{4.57}$$

where λ *is the argument of the first component of the spinor* $U(\boldsymbol{\alpha})\psi$*.*

The correspondence between unitary rotation matrices $U(\boldsymbol{\alpha})$ and orthogonal matrices $\mathbf{R}(\boldsymbol{\alpha})$ is not one-to-one. As noted earlier, $U(2\pi\mathbf{n}) = -\mathbf{1}_2$, hence

$$U((\alpha+2\pi)\mathbf{n}) = U(\alpha\mathbf{n})\,U(2\pi\mathbf{n}) = U(\alpha\mathbf{n})\,(-\mathbf{1}_2) = -U(\alpha\mathbf{n}), \tag{4.58}$$

and the matrices $\pm U(\alpha\mathbf{n})$ are both mapped to the same matrix $\mathbf{R}(\alpha\mathbf{n})$ by (4.50).

Recall that a symmetry transformation is defined as a ray transformation (see Section 1.1). Unitary transformations U_1 and U_2 that differ only by a phase factor, $U_1 = \mathrm{e}^{\mathrm{i}\lambda}U_2$, define the same symmetry transformation $\hat{U}$. Hence, the mapping

$$\mathbf{R}(\boldsymbol{\alpha}) \to \hat{U}(\boldsymbol{\alpha}) \tag{4.59}$$

is one-to-one from the group of rotations into the set of symmetry transformations. It is called a *ray representation* or *projective representation* of the rotation group.

4.4.5. Transition probabilities between qubit states

It is a strange result of quantum mechanics that a qubit prepared with spin up in the z-direction can be found with spin up with respect to some other direction $\mathbf{n}$ (except in the negative z-direction). Here, we are going to determine the probability for this to happen. In particular, we are going to prove the following important result for the transition probability:

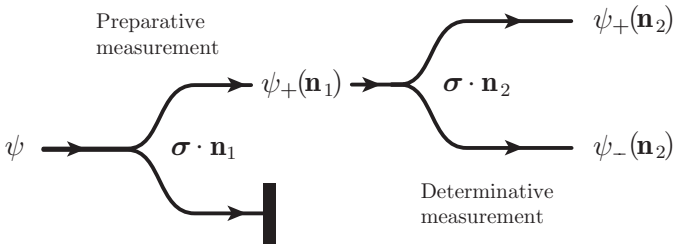


FIGURE 4.8. Experimental setup for determining a transition probability. All particles are prepared in the spin-up eigenstate in the direction of $\mathbf{n}_1$. In the second Stern-Gerlach apparatus, the magnetic field points in the direction defined by $\mathbf{n}_2$.

Qubit transitions:

Let ψ and ϕ be any two qubit states (unit vectors in $\mathbb{C}^2$). Then, the transition probability between ψ and ϕ is

$$|\langle \psi, \phi \rangle|^2 = \left(\cos \frac{\alpha}{2} \right)^2, \tag{4.60}$$

where α is the angle between the spin-up directions of ψ and ϕ, that is

$$\cos \alpha = \mathbf{n}(\psi) \cdot \mathbf{n}(\phi) \tag{4.61}$$

with $\mathbf{n}(\cdot)$ defined as in Eq. (4.21).

You can see again that only the states with spins pointing in opposite directions ($\alpha = \pi$) are orthogonal.

In order to prove this result, we note that any state is "spin-up" with respect to some axis. We denote the spin-up direction of ψ by $\mathbf{n}_1$ and the spin-up direction of ϕ by $\mathbf{n}_2$. Hence, $\psi = \psi_+(\mathbf{n}_1)$ and $\phi = \psi_+(\mathbf{n}_2)$, at least up to a phase factor. We can ignore the phase factors, because they play no role for the transition probability anyway. The experimental arrangement for measuring the transition probability from $\psi_+(\mathbf{n}_1)$ to $\psi_+(\mathbf{n}_2)$ is depicted in Figure 4.8.

Next, we perform a suitable rotation such that the vector $\mathbf{n}_1$ becomes the unit vector $\mathbf{e}_z$ in z-direction and such that $\mathbf{n}_2$ becomes a unit vector $\mathbf{m}$ that lies in the xz-plane. This rotation is characterized by some rotation vector $\boldsymbol{\alpha}_0$. We can apply the unitary matrix $U(\boldsymbol{\alpha}_0)$ to both spinors ψ and ϕ without changing the scalar product (this is just the definition of unitarity). Then

$$|\langle \psi, \phi \rangle|^2 = |\langle \psi_+(\mathbf{n}_1), \psi_+(\mathbf{n}_2) \rangle|^2 = |\langle \psi_+(\mathbf{e}_z), \psi_+(\mathbf{m}) \rangle|^2. \tag{4.62}$$

We can rotate $\mathbf{e}_z$ into the vector $\mathbf{m}$ by a rotation through an angle α around the y-axis. The angle α is the same as the angle between the vectors $\mathbf{n}_1$ and $\mathbf{n}_2$. The rotation matrix for rotations around the y-axis is

$$U(\alpha \mathbf{e}_y) = \begin{pmatrix} \cos(\alpha/2) & -\sin(\alpha/2) \\ \sin(\alpha/2) & \cos(\alpha/2) \end{pmatrix}, \tag{4.63}$$

and we have $U(\alpha \mathbf{e}_y)\psi_+(\mathbf{e}_z) = \psi_+(\mathbf{m})$ (up to a phase). Now, because $\psi_+(\mathbf{e}_z) = \psi_+$, the transition probability becomes simply

$$|\langle \psi, \phi \rangle|^2 = |\langle \psi_+, U(\alpha \mathbf{e}_y)\psi_+ \rangle|^2 = \left(\cos \frac{\alpha}{2}\right)^2. \tag{4.64}$$



You can verify the formula (4.64) with the screen experiment CD 4.5.

EXERCISE 4.18. *Use Equation* (4.53) *to show the following. Assume that a qubit is in the state* ψ*, with* $\|\psi\| = 1$*. Given an arbitrary unit vector* $\mathbf{n} \in \mathbb{R}^3$*, write* $\mathbf{m} = \mathbf{R}(\boldsymbol{\alpha})\mathbf{n}$*, where* $\mathbf{R}(\boldsymbol{\alpha})$ *is the* 3×3*-rotation matrix that transforms the spin-up direction* $\mathbf{n}(\psi)$ *into* $\mathbf{e}_z$ *(the unit vector in the positive z-direction). Then, the expectation value of the spin-component in the direction of* $\mathbf{n}$ *is given by the third component of* $\mathbf{m}$*, that is,*

$$\langle \boldsymbol{\sigma} \cdot \mathbf{n} \rangle_\psi = \langle \psi, \boldsymbol{\sigma} \cdot \mathbf{n}\, \psi \rangle = m_3. \tag{4.65}$$

4.5. Other Qubit Systems

4.5.1. Photon polarizations

In many books, two state-systems are discussed with photons as the primary example. Indeed, many qubit experiments discussed so far are much easier to accomplish with photons and polarization filters than with spin-1/2 particles and Stern-Gerlach devices.

Photons are massless particles with spin 1. They are quantum mechanical objects, and the electromagnetic wave may be regarded as their wave function. A classical electromagnetic wave is a transverse wave where the electric and the magnetic field vectors $\mathbf{E}$ and $\mathbf{B}$ are perpendicular to each other and to the direction of the propagation. The wave is said to be *linearly polarized* if $\mathbf{E}$ is always in a plane. If $\mathbf{E}$ rotates on a circle, the wave is *circularly polarized*. The square of the amplitude of the electromagnetic wave, that is, the intensity of the light beam, is proportional to the number of photons in the beam. Photons do not interact among themselves because

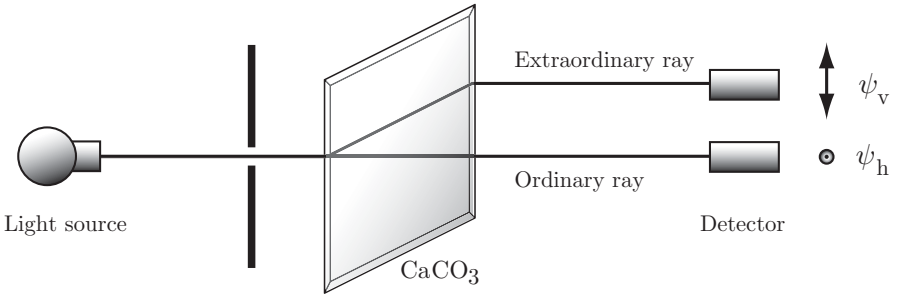


FIGURE 4.9. A clacit crystal is a natural Stern-Gerlach apparatus for photons.

they are chargeless.[7] They are bosons, which means that an arbitrary number can occupy the same state, and hence it is easy to perform experiments with a large ensemble of particles. For light beams with low intensity, the number of photons in the beam can be counted with a photomultiplier.

But photons are unusual because they are massless, and they move with the velocity of light. The wave equation that describes the propagation through space and time is relativistically invariant and hence completely different from the Schrödinger equation. But here we are only interested in the internal degrees of freedom that are related to the spin of the photons.

We distinguish two polarization states of photons, $\psi_{\rm v}$ and $\psi_{\rm h}$. These two states can be distinguished very easily using a birefringent crystal (calcit), which is the analog of a Stern-Gerlach apparatus (see Fig. 4.9). We obtain a polarization filter if one of the beams is prevented from passing the crystal (nicol prism). Such a filter can also be realized by a thin sheet of transparent plastic with embedded microcrystals. Hence, unlike a Stern-Gerlach filter, a polarization filter is an everyday device. Probably you own polarization filters in form of sun glasses.

Assume that the photons move in the positive y-direction. Then, we make the following identification: The basis vector $\psi_+ \in \mathbb{C}^2$ describes the state $\psi_{\rm v}$ of photons that pass a vertically oriented polarization filter. The vector $\psi_- \in \mathbb{C}^2$ is the state $\psi_{\rm h}$ of those photons that pass a horizontally oriented polarization filter. The orthogonality of these two states is confirmed experimentally, because none of the particles emerging from a vertical polarizer can pass the horizontal polarizer and vice versa.

By measuring the fraction of particles that can pass a second polarization filter rotated through an angle α against the first, we can determine the behavior under rotations experimentally. Denote by ϕ_+ (ϕ_-) the state

[7] A *very* small attractive interaction between photons has been predicted due to nonlinear effects in quantum electrodynamics.

of photons that pass through a polarization filter obtained from the vertically (horizontally) oriented filter by a rotation through the angle α. The new states are also orthogonal, $\langle \phi_+, \phi_- \rangle = 0$, and hence they form another orthogonal basis in $\mathbb{C}^2$. Hence, we assume that the new basis $\{\phi_+, \phi_-\}$ is related to the reference basis $\{\psi_{\rm v}, \psi_{\rm h}\}$ via a unitary transformation:

$$\phi_+ = V(\alpha)\psi_{\rm v}, \quad \phi_- = V(\alpha)\psi_{\rm h}. \tag{4.66}$$

EXERCISE 4.19. *Try to determine the matrix elements of the two-by-two matrix $V(\alpha)$ on the basis of the following observations. By measuring intensities, one obtains the transition probabilities*

$$\begin{aligned} |\langle \psi_{\rm v}, \phi_+ \rangle|^2 &= |\langle \psi_{\rm h}, \phi_- \rangle|^2 = \cos^2 \alpha, \\ |\langle \psi_{\rm v}, \phi_- \rangle|^2 &= |\langle \psi_{\rm h}, \phi_+ \rangle|^2 = \sin^2 \alpha. \end{aligned} \tag{4.67}$$

Moreover, we have (verify this!)

$$\langle \phi_+, \phi_- \rangle = \langle \phi_+, \psi_+ \rangle \langle \psi_+, \phi_- \rangle + \langle \phi_-, \psi_{\rm v} \rangle \langle \psi_{\rm h}, \phi_- \rangle = 0. \tag{4.68}$$

What can you learn from all this about the matrix $V(\alpha)$?

It turns out that the rotation matrix can be chosen as

$$V(\alpha) = \exp(-{\rm i}\,\sigma_2 \alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = U(2\alpha \mathbf{e}_y). \tag{4.69}$$

Hence, the eigenstates of σ_2 remain invariant under rotations. These states are interpreted as circularly polarized states,

$$\begin{aligned} \psi_R &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ {\rm i} \end{pmatrix} = \frac{1}{\sqrt{2}} (\psi_{\rm v} + {\rm i}\psi_{\rm h}), \\ \psi_L &= \frac{1}{\sqrt{2}} \begin{pmatrix} {\rm i} \\ 1 \end{pmatrix} = \frac{\rm i}{\sqrt{2}} (\psi_{\rm v} - {\rm i}\psi_{\rm h}). \end{aligned} \tag{4.70}$$

Note that the generator of rotations in the plane orthogonal to the direction of motion is σ_2, and that it has the eigenvalues ± 1 ("spin 1"). The component σ_2 of the angular momentum in the direction of motion is also called the helicity. Hence, the circularly polarized states are the helicity eigenstates of the photons.

The eigenstates of σ_1 are obtained from $\psi_{\rm v}$ and $\psi_{\rm h}$ by a rotation through $\pi/4$. We have

$$V(\pi/4) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{4.71}$$

and hence

$$\begin{aligned}\phi_+ &= V(\pi/4)\,\psi_{\mathrm{v}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{\sqrt{2}}(\psi_{\mathrm{v}} + \psi_{\mathrm{h}}),\\ \phi_- &= V(\pi/4)\,\psi_{\mathrm{h}} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix} = \frac{1}{\sqrt{2}}(-\psi_{\mathrm{v}} + \psi_{\mathrm{h}}).\end{aligned} \tag{4.72}$$

The vector ϕ_+ describes the state of linearly polarized photons that can pass a polarization filter rotated through an angle of 45° against the vertical axis. ϕ_- corresponds to a linear polarization with an angle of 135°.

You can see that the polarization states of the photon are realized mathematically in exactly the same way as the spin 1/2 states of massive particles, but the interpretation of the state vectors and operators is quite different. Whereas $\exp(\mathrm{i}\sigma_2\alpha)$ can be interpreted as a rotation in space (about the direction of the light ray), the unitary operators $\exp(\mathrm{i}\sigma_1\alpha)$ and $\exp(\mathrm{i}\sigma_3\alpha)$ have no such interpretation.

At first sight, it also appears strange that photons should be a two-state system, because spin 1 lets us expect a three-state system. But for a particle moving at the velocity of light, the internal (rotational) degrees of freedom are reduced. For the sake of a heuristic argument, let us consider the photons as a limit case of particles with $m > 0$, $v < c$. In the limit $v \to c$, the Lorentz contraction reduces the rest frame of the particles to a plane transverse to the direction of motion. Hence, the rotational degrees of freedom are reduced to the rotations around the axis defined by the direction of motion. The rotational motion can be clockwise and counterclockwise, resulting in the two helicity eigenstates. The linearly polarized states are just superpositions of the helicity states ψ_R and ψ_L.

4.5.2. Spatial states of photons

The spatial part of a photon's wave function can also be used to realize a qubit. Consider the Mach-Zehnder interferometer depicted in Figure 4.10. In this arrangement, the photons can only occur in two states: moving up ($\nearrow$) or moving down ($\searrow$). We denote these states by ψ_{u} and ψ_{d}, respectively. These two states form an orthonormal basis in the Hilbert space of the system. In this description of photons, we ignore the polarization and consider only the two alternative directions of motion. A general state of the qubit is a superposition of up-moving and down-moving photon states. Such a state can be prepared by sending a photon through a beam splitter. Let us describe the action of the beam splitters and mirrors in the Mach-Zehnder interferometer with repect to the reference basis $\{\psi_{\mathrm{u}}, \psi_{\mathrm{d}}\}$.

Assume that a photon enters the first beam splitter from below. This photon is initially in the state ψ_{u}. The beam splitter splits the photon's

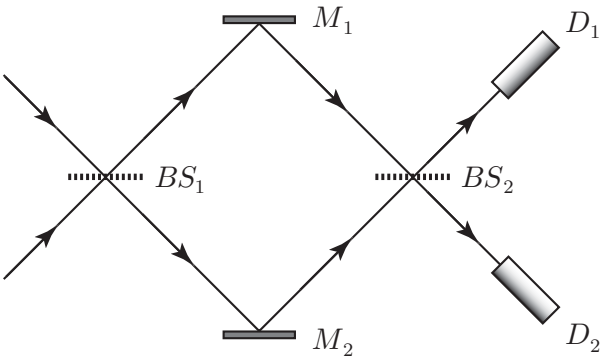


FIGURE 4.10. The Mach-Zehnder interferometer. It consists of two symmetric 50/50 beam splitters BS_1 and BS_2 (e.g., half-silvered mirrors) and two mirrors M_1 and M_2. In this setup (ignoring polarizations), the photons are "spatial qubits." The two possible states are described as "moving up" and "moving down." These basis states can be measured by two detectors D_1 or D_2 on the possible paths of the photons.

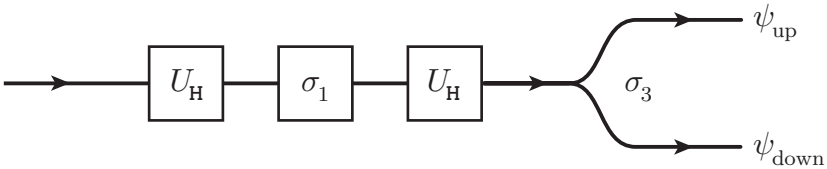


FIGURE 4.11. The Mach-Zehnder interferometer of Figure 4.10 as a qubit diagram. The beam splitters can be described by Hadamard transformations and the two mirrors by σ_1. The pair of detectors D_1 and D_2 is an analyzer measuring the observable that is represented by the matrix σ_3 with respect to the reference basis $\{\psi_{\mathrm{u}}, \psi_{\mathrm{d}}\}$.

state into two parts of equal size, one moving up, the other moving down. Therefore, after the beam splitter, the photon is in an equal superposition of ψ_{u} and ψ_{d}. The action of the beam splitter on the state ψ_{u} may thus be described by

$$\psi_{\mathrm{u}} \longrightarrow \frac{1}{\sqrt{2}} (\psi_{\mathrm{u}} + \psi_{\mathrm{d}}) \tag{4.73}$$

Similarly, a photon incident from above in the state ψ_{d} is converted into another equal superposition of ψ_{u} and ψ_{d}. If the beam splitter is to be

described by a unitary transformation, then orthogonal initial states have to be converted into orthogonal states after the beam splitter. Hence, we assume

$$\psi_{\mathrm{u}} \longrightarrow \frac{1}{\sqrt{2}}\,(\psi_{\mathrm{u}} - \psi_{\mathrm{d}}), \tag{4.74}$$

which is orthogonal to the final state in (4.73). The unitary transformation describing this transformation is the Hadamard transformation U_{H} defined in (4.38).

Now, let us consider the mirrors. Obviously, in the given arrangement, the mirrors just convert the state ψ_{u} into ψ_{d} and vice versa. This is conveniently done by the (unitary) matrix σ_1. Hence, the combined action of the Mach-Zehnder interferometer is described by the composition of the three unitary transformations U_{H}, σ_1, and U_{H}, as depicted in Figure 4.11:

$$U_{\mathrm{MZ}} = U_{\mathrm{H}}\,\sigma_1\,U_{\mathrm{H}} = \sigma_3. \tag{4.75}$$

Another important operation that can be performed with spatial photon states is to put a phase shifter (a waveplate shifting the phase of an electromagnetic wave) in the path of one photon. With respect to the basis $\{\psi_{\mathrm{u}}, \psi_{\mathrm{d}}\}$, the phase shifter placed in the path of an up-moving photon is represented by the matrix $\Phi(\alpha)$ and in the path of a down-moving photon by $\Psi(\alpha)$, where

$$\Phi(\alpha) = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\alpha} & 0 \\ 0 & 1 \end{pmatrix}, \quad \Psi(\alpha) = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\alpha} \end{pmatrix}. \tag{4.76}$$

Figure 4.12 shows a Mach-Zehnder interferometer with additional phase shifters and the corresponding qubit diagram. Up to a phase factor, the operator $\Psi(\gamma)\,U_{\mathrm{H}}\,\Phi(\beta)\,\sigma_1\,U_{\mathrm{H}}\,\Phi(\alpha)$ can represent any unitary operator in the Hilbert space of the qubit.

Exercise 4.20. *Verify that*

$$\mathbf{1} = \Psi(\pi)\,U_{\mathrm{H}}\,\sigma_1\,U_{\mathrm{H}} \tag{4.77}$$

$$\sigma_2 = -\mathrm{i}U_{\mathrm{H}}\,\Phi(\pi)\,\sigma_1\,U_{\mathrm{H}}. \tag{4.78}$$

If you send a photon from below into a modified Mach-Zehnder interferometer according to Figure 4.12 with $\alpha = 0$, $\beta = \pi$, $\gamma = 0$, in which direction will the photon finally leave the interferometer?

Exercise 4.21. *There is a certain freedom in describing the action of the beam splitters and mirrors. By definition, a 50/50 beam splitter is unitary and converts each state of the reference basis into a superposition where both basis states occur with equal probability $1/2$. What is the remaining freedom?*

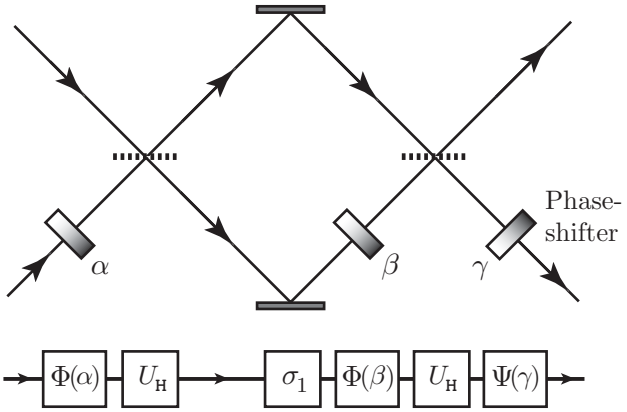


FIGURE 4.12. The Mach-Zehnder interferometer with additional phase shifters is a universal unitary gate. With a suitable choice of the angles α, β, and γ, it can perform an arbitrary unitary transformation with the qubit (up to a phase factor).

Can we also use the matrix

$$U_{\mathrm{BS}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \tag{4.79}$$

to describe the action of the beam splitter? Discuss also the freedom to describe the action U_{M} of the two mirrors on the qubit's state. Can you define U_{BS} and U_{M} in such a way that $U_{\mathrm{BS}}\, U_{\mathrm{M}}\, U_{\mathrm{BS}} = \mathbf{1}$?

EXERCISE 4.22. *Show that*

$$U(\alpha \mathbf{e}_x) = \mathrm{e}^{-\mathrm{i}\alpha/2}\, \Psi(\pi)\, U_{\mathrm{H}}\, \Phi(\alpha)\, \sigma_1\, U_{\mathrm{H}}, \tag{4.80}$$

$$U(\alpha \mathbf{e}_y) = -\mathrm{i}\mathrm{e}^{-\mathrm{i}\alpha/2}\, \Psi(-\pi/2)\, U_{\mathrm{H}}\, \Phi(\alpha)\, \sigma_1\, U_{\mathrm{H}}\, \Phi(\pi/2), \tag{4.81}$$

$$U(\alpha \mathbf{e}_z) = -\mathrm{e}^{-\mathrm{i}\alpha/2}\, \Psi(\alpha)\, U_{\mathrm{H}}\, \sigma_1\, U_{\mathrm{H}}\, \Phi(\pi). \tag{4.82}$$

Hence, the qubit rotations about the x-, y-, and z-axes can be represented by a Mach-Zehnder interferometer equipped with additional phase shifters as in Figure 4.12.

4.5.3. Two states of a harmonic oscillator

Consider a one-dimensional harmonic oscillator. More precisely, we consider only the subspace spanned by the ground state ϕ_0 and the first excited state ϕ_1,

$$\phi_0(x) = \Big(\frac{1}{\pi}\Big)^{1/4} \exp\Big(-\frac{x^2}{2}\Big), \quad \phi_1(x) = \sqrt{2}\, x\, \phi_0(x). \tag{4.83}$$

These two eigenfunctions span a two-dimensional Hilbert space, and hence they realize a qubit. This simple mathematical model of a qubit will provide us with additional ways to visualize two-qubit systems in Chapter 5.

The Hilbert space spanned by ϕ_0 and ϕ_1 is $P_{01}L^2(\mathbb{R})$ where P_{01} is the projection operator

$$P_{01} = |\phi_0\rangle\langle\phi_0| + |\phi_1\rangle\langle\phi_1|. \tag{4.84}$$

Of course, we have to be careful with all measurements and manipulations in order not to excite the oscillator to energies beyond the first level.

We consider the Hamiltonian operator

$$H = \frac{1}{2}(p^2 + x^2) - 1, \tag{4.85}$$

which is the harmonic oscillator Hamiltonian (in dimensionless units) shifted by a constant potential. In the subspace spanned by ϕ_0 and ϕ_1, the operator H has the eigenvalues $-1/2$ and $1/2$ with eigenfunctions ϕ_0 and ϕ_1. The two-dimensional Hilbert space is isomorphic to $\mathbb{C}^2$. We can identify ϕ_0 with $\psi_+ \in \mathbb{C}^2$, and ϕ_1 with ψ_-. The two-by-two matrix $S_3 = (1/2)\sigma_3$ can be identified with the operator $-H$, because

$$(-H)\phi_0 = \frac{1}{2}\,\phi_0, \quad (-H)\phi_1 = -\frac{1}{2}\phi_1. \tag{4.86}$$

The position and momentum operators do not leave the subspace $P_{01}L^2(\mathbb{R})$ invariant. You can see this if you write x and p in terms of the ladder operators $A^\dagger$ and A introduced in Book One.

$$x = \frac{1}{\sqrt{2}}(A^\dagger + A), \quad p = \frac{\mathrm{i}}{\sqrt{2}}(A^\dagger - A) \tag{4.87}$$

Then, for example, $x\phi_1 = \phi_2 + \phi_0/\sqrt{2}$, which is not in $P_{01}L^2(\mathbb{R})$. But we may consider the restrictions of x and of p to the range of P_{01}, that is, the operators

$$\hat{x} = P_{01}xP_{01}, \quad \hat{p} = P_{01}pP_{01}. \tag{4.88}$$

It is easy to see that

$$\begin{aligned} \hat{x}\phi_0 &= \frac{1}{\sqrt{2}}\phi_1 \qquad & \hat{p}\phi_0 &= \frac{\mathrm{i}}{\sqrt{2}}\phi_1 \\ \hat{x}\phi_1 &= \frac{1}{\sqrt{2}}\phi_0 \qquad & \hat{p}\phi_1 &= -\frac{\mathrm{i}}{\sqrt{2}}\phi_1 \end{aligned} \tag{4.89}$$

and we have the following interpretation of the qubit observables,

$$\begin{aligned} \sigma_1 &\quad\longleftrightarrow\quad \sqrt{2}\,\hat{x}, \\ \sigma_2 &\quad\longleftrightarrow\quad \sqrt{2}\,\hat{p}, \\ \sigma_3 &\quad\longleftrightarrow\quad -2H. \end{aligned} \tag{4.90}$$

The time evolution is hence a qubit rotation around the z-axis:

$$\exp\Big(\mathrm{i}\frac{t}{2}\sigma_3\Big) \quad \longleftrightarrow \quad e^{-\mathrm{i}Ht}. \tag{4.91}$$

4.6. Single-Particle Interference

4.6.1. Interferometer

Consider the following situation. We send a particle in a certain spin state ψ into a Stern-Gerlach apparatus. Ignoring the spatial distribution of the spinor-wave packet, we describe the spin state (a qubit) by the vector

$$\psi = \begin{pmatrix} c_+ \\ c_- \end{pmatrix}. \tag{4.92}$$

The two components are deflected into different directions, and the spinor-wave packet splits into two spatially separated parts,

$$\psi \to \begin{pmatrix} c_+ \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ c_- \end{pmatrix}. \tag{4.93}$$

Now, assume that we do not make the slightest attempt to determine where the particle actually is. That is, we make no measurement that forces the qubit to give away any information about its spin-state ("which-way information"). This would certainly change the state of the qubit. Instead, by a clever arrangement of inhomogeneous magnetic fields, we bring the two parts of the spinor-wave packet together again, without changing the spin-state of either part. After the reunion, the initial situation is restored completely. The final state of the qubit equals the initial state.

Schematically, this experiment is depicted by the diagram in Figure 4.13. A device that temporarily splits the state of the qubit into separated parts is called an *interferometer*. A beam of qubits sent through the interferometer will split into two spatially separated beams. The two partial beams are refocused again, and a single beam leaves the interferometer. Input and output as well as the qubits inside the interferometer are described by the same qubit state. As an operator acting on qubit states, the interferometer is just the identity. The importance of such a device lies in the possibility to manipulate the partial beams individually and then observe the effect by measuring the output state. Examples will be presented in the following sections.

It is certainly difficult to build an interferometer on the basis of Stern-Gerlach devices. But in quantum optics where qubits are realized by photons, all kinds of interferometers are standard devices. In principle, one could realize a Stern-Gerlach interferometer for photons by arranging two

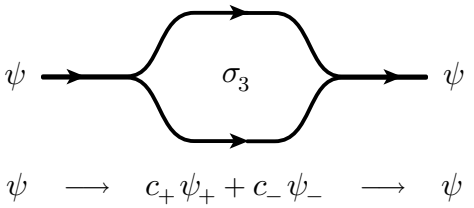


FIGURE 4.13. Stern-Gerlach interferometer: The two partial beams are refocused again (for example, with the help of another inhomogeneous magnetic field with opposite polarity), and the initial state is restored. The device has no measurable effect on the state. It just visualizes the expansion into the basis of eigenvectors of σ_3, as indicated below the graph.

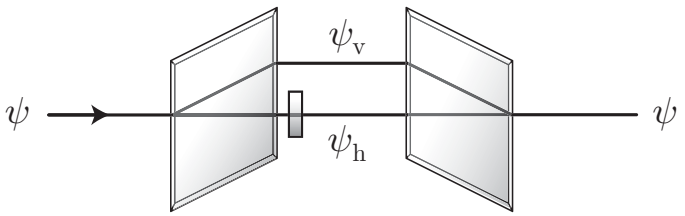


FIGURE 4.14. Two birefringent crystals realizing a Stern-Gerlach interferometer for photons. An additional phase shifter (retarder plate) in one of the beams can compensate for eventual phase differences between the partial beams.

birefringent calcit crystals as shown in Figure 4.14. Note, however, that the Mach-Zehnder interferometer described in in Section 4.5.2 operates in a different way, which is obvious from the qubit diagram in Figure 4.11. We also mention neutron interferometers, which play an important role in experiments illustrating the quantum mechanics of qubits. We refer to the literature for details (see, for example, the book [**8**]).

Observing a qubit inside the interferometer changes the physically available information about the qubit. This amounts to a measurement and changes the state either to "spin-up" or to "spin-down." An interferometer with an observer acts like the Stern-Gerlach analyzer in Figure 4.1: In an ensemble measurement, the which-way information allows an observer to sort the qubits into two ensembles realizing the states ψ_+ and ψ_-. On the other hand, if no measurement is made, no projection takes place, and the state remains unchanged during the qubit's passage through the interferometer. Indeed, the spatial separation of the two parts cannot be described

at all in the two-dimensional Hilbert space of a qubit. Hence, the diagram in Figure 4.13 is just a visualization of the fact that any qubit state ψ is a superposition of the two orthogonal states ψ_+ and ψ_-.



CD 4.9 shows an interferometer operating with or without an observer. In the presence of an observer, the interferometer changes the state of the qubit into a mixture of qubits with spin-up and spin-down.

Interferometers are important measurement devices. The spatial separation gives us the experimental possibility to manipulate one partial beam without affecting the other component. The change of the state can be observed via the interference that takes place when the two parts are recombined.

One cannot help asking what happens if a single qubit is sent through the interferometer. Will the particle take just one of the two possible ways, and which? We know for sure that a single particle does not split in two. If we place detectors along the two paths we always find the entire particle on just one of the paths (in accordance with the observation that the measurement of σ_3 always gives a definite result). Nevertheless, it appears as if the unobserved particle takes both paths at once, because both components ψ_+ and ψ_- are present in the final state ψ. If we do something to one of the components, the final state will be changed. This phenomenon is called *single-particle interference*.

In order to illustrate this strange behavior, we consider several experiments. Each experiment features an interferometer of the type shown in Figure 4.13. When the experiment is actually performed, the apparatus temporarily splits the wave packet into two parts that are separated by a macroscopic distance (see also the discussion about Schrödinger cat states in Book One). Per se, the splitting of the wave packet into spatially separated parts has no effect on the qubit state and is not reflected by the formalism, because the motion in space is completely ignored. If no further manipulations are performed, the qubit remains in the same state before entering, inside, and after leaving the interferometer.

The quintessence of all the experiments is the following: During its passage through the apparatus, even a single particle can gather information about the conditions on both paths. Something that happens on one of the possible paths can change the final outcome. We can put this in a more abstract way: Whenever there are two possible alternatives for the behavior of a quantum system, it is wrong to assume that one of these possibilities is actually realized, unless one performs a measurement.

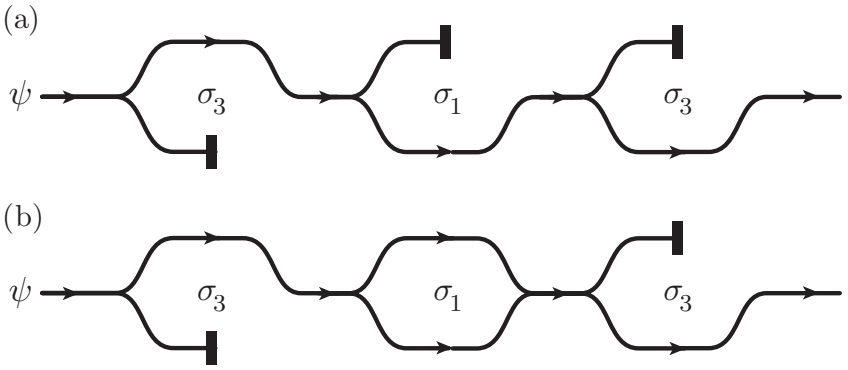


FIGURE 4.15. Double-slit experiment with Stern-Gerlach interferometers and filters.

4.6.2. A double-slit experiment

The two arrangements of Stern-Gerlach devices shown in Figure 4.15 serve as a first example exhibiting the interference phenomenon. In both cases, the first apparatus (a filter) just serves to prepare a qubit in the state ψ_+ (spin-up in the z-direction). Consider first Figure 4.15(b). The qubit prepared by the first apparatus is sent through an interferometer oriented in the positive x-direction. Because

$$\psi_+ = \frac{1}{\sqrt{2}}\big(\psi_+(\mathbf{e}_x) - \psi_-(\mathbf{e}_x)\big), \tag{4.94}$$

the probability amplitudes for both paths in the interferometer are the same. In a space-time picture, the wave function splits into two orthogonal parts inside this apparatus (Schrödinger cat state).

The output of the interferometer is collected by a Stern-Gerlach filter that projects onto the state ψ_-. As explained in the previous section, the interferometer has no influence on the state ψ_+ of the prepared particles. But as ψ_+ is orthogonal to the state ψ_-, none of the prepared particles can pass the final Stern-Gerlach filter.

Now, consider Figure 4.15(a). Here, we have blocked one of the paths in the interferometer (no matter which one). As a consequence, the interferometer now acts as a filter. Only one-half of the prepared particles can pass, because $|\langle\psi_\pm(\mathbf{e}_x), \psi_+\rangle|^2 = 1/2$. From the remaining particles another half is filtered by the final Stern-Gerlach apparatus, because $|\langle\psi_-, \psi_\pm(\mathbf{e}_x)\rangle|^2 = 1/2$. Hence, a quarter of the successfully prepared particles finally survive the experiment.

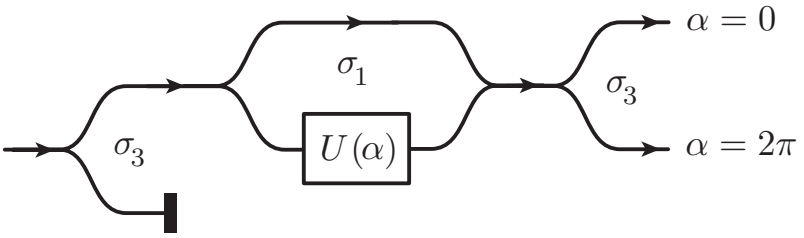


FIGURE 4.16. Experiment verifying the effect of a rotation through an angle 2π. An interferometer with a unitary transformation in one channel represents, in general, a non-unitary transformation (except for $\alpha = 0, 2\pi, 4\pi$).



In CD 4.10.1, you can perform the double-slit experiment as described in Figure 4.15. You may verify that the probability to end in the state ψ_- is 1/4 if one of the paths in the interferometer is blocked. The probability for the result ψ_- is zero (destructive interference) if there is no obstacle.

This experiment is the two-state analog of the double-slit experiment described in Book One. The result is highly paradoxical: If we increase the number of ways to reach the detector, the probability of getting there actually decreases. This cancellation of probability amplitudes (if both "slits" of the splitter are left open) is called *destructive interference*.

It is strange that the interference also happens if single particles are sent through the arrangement one after another. The probabilities are thus accumulated from single-particle events (*single-particle interference*).

4.6.3. A rotation through 2π

The unitary rotation matrix has the property $U(2\pi\mathbf{n}) = -\mathbf{1}_2$. A rotation through an angle 2π around any axis turns the state vector into its negative. The vector $-\psi$ is physically not distinguishable from ψ. If, however, the state is split into two orthogonal components, then a rotation of only one of the components will influence the state in a measurable way. Consider two orthogonal states ψ_1, ψ_2, and the superposition $\psi = \psi_1 + \psi_2$. If you multiply ψ_2 by a phase factor, then the state described by $\psi_1 + \mathrm{e}^{\mathrm{i}\lambda}\psi_2$ is, in general, different from $\psi_1 + \psi_2$, although $\mathrm{e}^{\mathrm{i}\lambda}\psi_2$ and ψ_2 represent the same state. Consider the following experiment, depicted in Figure 4.16.

A system is prepared in the state ψ_+ and then enters an interferometer that splits the beam according to its components with respect to the x-direction. Because of (4.94), the amplitudes of the two components are the same. Now a rotation through an angle $\alpha = 0$ or $\alpha = 2\pi$ about an arbitrary

axis is applied to the lower partial beam in the interferometer. If $\alpha = 0$ nothing happens, because $U(0) = \mathbf{1}$. The state ψ_+ is restored when it leaves the interferometer and the particle is found in the upper channel of the final analyzer. If $\alpha = 2\pi$, the component $\psi_-(\mathbf{e}_x)$ is multiplied by -1. Hence, the superposition (4.94) is changed into

$$\psi_- = \frac{1}{\sqrt{2}}\big(\psi_+(\mathbf{e}_x) + \psi_-(\mathbf{e}_x)\big), \tag{4.95}$$

and the particle is finally detected in the lower channel of the final Stern-Gerlach apparatus.



CD 4.10.2 simulates an experiment similar to the one shown in Figure 4.16. You may rotate the qubit in one path of the interferometer through an arbitrary angle α. However, the transformation is not unitary unless α is an integer multiple of 2π (see Exercise 4.24).

EXERCISE 4.23. *Let P_1 and P_2 be orthogonal projection operators, such that $P_1 + P_2 = \mathbf{1}$. An interferometer like the one depicted in Figure 4.16 corresponds to the transformation $T = P_1 + UP_2$ with a unitary operator U. Under what condition is T unitary?*

EXERCISE 4.24. *Assume that in $P_1 = |\psi_+(\mathbf{e}_x)\rangle\langle\psi_+(\mathbf{e}_x)|$, $P_2 = \mathbf{1} - P_1$, and that $U(\alpha)$ is a rotation about the z-axis. Show that $T(\alpha) = P_1 + U(\alpha)P_2$ is not unitary (except for $\alpha = 0, 2\pi, 4\pi, \ldots$). Show that $T(\pi)\psi_+(\mathbf{e}_y) = 0$. That is, the apparatus in Figure 4.16 with $\alpha = \pi$ acts as a filter annihilating qubits in the state $\psi_+(\mathbf{e}_y)$.*

EXERCISE 4.25. *Define $T(\alpha)$ as in the previous exercise. Compute the transition probabilities*

$$|\langle\psi_+, T(\alpha)\,\psi_+\rangle|^2 \quad \textit{and} \quad |\langle\psi_-, T(\alpha)\,\psi_+\rangle|^2. \tag{4.96}$$

4.6.4. Interaction-free measurement

The two situations depicted in Figure 4.15 are often described in terms of an *interaction-free measurement* and packed into the following story[8]. Some malicious person has threatened to put an atomic bomb into a dark room. It is your task to find out whether this person has spoken the truth and the bomb is actually there. But there is a difficulty that prevents you from just looking: The trigger is so sensitive that already the absorption of a single elementary particle would detonate the bomb. Fortunately, the dark room contains a perfect vacuum and the light is turned off, so that no gas molecule

[8]Elitzur, A.C., and Vaidman, L., Foundations of Physics **2**, 987 (1993).

and no photon inadvertently triggers the explosion. For obvious reasons you must be very careful. You are not allowed to open the door and turn the light on. Is it still possible to detect the presence of the bomb?

Despite the total darkness, you have a certain chance to detect the bomb without catastrophe. You just have to build an experimental setup according to Figure 4.17 and arrange it around the dark room in such a way that the bomb, if it is actually there, would block one path of the interferometer. Next, you send a single particle through the arrangement and had better look for somewhere to hide.

The idea of the procedure can be explained with the help of the double-slit experiment in Section 4.6.2. Consider Figure 4.15 and suppose that we do not know whether one of the paths in the interferometer is actually blocked. That is, we do not know if the experimental setup is described by (a) or by (b). We can just send in a particle and see whether it comes out again. Now, suppose that we detect the particle after the analyzer. Then we know for sure that one of the paths in the second Stern-Gerlach apparatus is blocked by an obstacle. The bomb is there! Because the particle has passed all devices, it has not been absorbed by the bomb. In this single event, the particle did not interact with the obstacle. The mere fact that it finally arrives in the detector proves that the bomb is there.

We note that in the presence of a bomb, there is a 25% chance to detect the bomb without actually touching it. This chance can be improved considerably by the refinement[9] of the double-slit experiment depicted in Figure 4.17.



CD 4.11 lets you play around with the apparatus depicted in Figure 4.17. The simulation contains a detailed step-by-step explanation of the method.

The filter to the left prepares a particle in the state ψ_+. The second Stern-Gerlach apparatus may or may not contain a bomb that explodes as soon as it is touched by the particle. If there is a bomb, then it has been placed such that it blocks the spin-down path of the apparatus. Because the particle enters the apparatus with spin-up, we are sure that it takes the upper path. Hence, the bomb is left untouched and the particle leaves the apparatus, still in the state ψ_+. Now the arrangement contains a switch S that diverts the particle and sends it through a device that performs a rotation $U_n = U(\frac{\pi}{n}\,\mathbf{e}_y)$ around the y-axis through the angle π/n, where n is some positive integer. After that, the particle is sent again through

[9]This experiment was actually done with photons, see Kwiat, P.G., Weinfurter, H., Herzog, T., Zeilinger, A., and Kasevich, M.A., Phys. Rev. Lett. **74**, 4763 (1995).

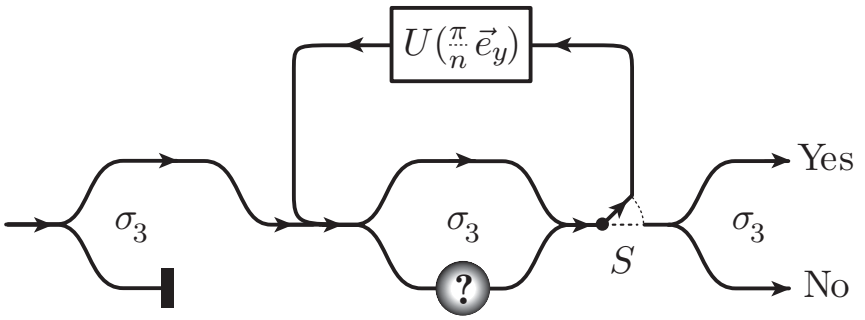


FIGURE 4.17. Schematic arrangement for an interaction free measurement. It is the goal to determine whether an object (a bomb) blocks the "spin-down path" of a Stern-Gerlach magnet. With a probability close to one, the experiment gives a unique answer. The presence of the bomb is detected with the help of a particle that apparently never gets in contact with the bomb. The particle is sent n-times through a circuit containing a device

the apparatus that possibly contains a bomb. It might seem imprudent to tempt fate twice, but by choosing n large, we can make the probability of a catastrophe very small. The probability that the particle takes the upper path (thus leaving the bomb untouched) is given by

$$p_{\rm up} = |\langle \psi_+, U_n \psi_+ \rangle|^2 = \left(\cos \frac{\pi}{2n}\right)^2, \tag{4.97}$$

which is close to 1 if n is large. When the particle (and everybody else) survives the passage through the apparatus, the process is repeated. The switch contains a counter, and after n cycles it sends the particle through an analyzer. Note that the probability that we are still alive is

$$p_{\rm up}^n = \left(\cos \frac{\pi}{2n}\right)^{2n} \underset{n\to\infty}{\longrightarrow} 1, \tag{4.98}$$

which is close to one if n is large. Let us now consider the two cases "bomb" and "no bomb" separately:

(A) A bomb blocks the "spin-down path": The Stern-Gerlach apparatus containing the bomb acts as a projection operator onto ψ_+. After each cycle, the state ψ_+ is restored with probability $p_{\rm up}$. If the particle finally leaves the circuit after n cycles (with probability $p_{\rm up}^n$), the state is still ψ_+ and it will be detected in the spin-up channel of the analyzer.

(B) There is no bomb: If the Stern-Gerlach magnet contains no bomb, it does nothing to the state of the particle. Hence, in each cycle, the state

of the particle is changed by the rotation U_n applied to it. After n cycles, the state is

$$\psi_{\text{final}} = U_n^n \, \psi_+ = U(\pi \, \mathbf{e}_y) \, \psi_+ = \psi_-, \tag{4.99}$$

because a rotation through the angle π turns spin-up into spin-down. Hence, the state will be detected in the spin-down channel of the analyzer.

The experiment—although performed with a single qubit—gives a clear answer: Finding the particle in the state ψ_+ means "yes, the bomb is there," and finding the particle in the state ψ_- means, "no, there is no bomb." And, at least in principle, the probability of obtaining an answer (and thus surviving the experiment) can be made arbitrarily close to 1.

4.7. Quantum Cryptography

4.7.1. One-time pad

Another application of the quantum mechanics of qubits is the secure key distribution protocol of quantum cryptography. We consider the following situation. Alice wants to send secret information to Bob, but an eavesdropper Eve might intercept the message. In order to prevent this, Alice has to encrypt her message. Presently used encryption methods are only computationally secure, which means that on the basis of current technology, it is highly improbable (but not impossible) to crack the encryption within a reasonable time. But there is one (and only one) method that has been proved to be absolutely secure. It is called the *one-time pad* or *Vernam cipher*.[10] The one-time pad uses classical communication to transmit classical bits of information. It is assumed that Alice and Bob share a secret key. This key is a string of randomly chosen bits $\{k_1, k_2, \ldots, k_n\}$, with $k_i \in \{0, 1\}$. The message is another string consisting of n bits $\{m_1, m_2, \ldots, m_n\}$. The key-string has to be as long as the message-string. Alice encrypts the message by adding the key-bits to the message-bits. Then, the cryptogram is a string consisting of the bits

$$c_i = m_i + k_i \pmod 2, \quad i = 1, 2, \ldots, n. \tag{4.100}$$

Alice sends the cryptogram $\{c_1, c_2, \ldots c_n\}$ to Bob. Bob subtracts the key to recover the message,

$$m_i = c_i - k_i \pmod 2, \quad i = 1, 2, \ldots, n. \tag{4.101}$$

Although the encryption and decryption algorithms and the encrypted message are publicly known, it is impossible for Eve to decrypt the message. Trying every possible sequence of key-bits k_i on the message just produces

[10] It was invented by Gilbert Vernam from AT&T in 1917, see Vernam, G.S., J. AIEE **45**, 109 (1926).

all possible n-bit messages and Eve has no way to decide which is the right one. To her, any decryption is as likely as any other. In fact, the transmitted string is just a sequence of random bits that contains no information about the message. This information about the message is contained in the correlation between the cryptogram and the key. Hence, the code is unbreakable and the method is secure, provided the key material is secret, truly random, and used only once.[11]

The critical feature of the one-time pad is the exchange of the key information between Alice and Bob, because the security of the protocol depends on the secrecy of the key. If Alice and Bob communicate over a longer period of time, they are bound to run out of keys, because the keys have to be as long as the messages and can be used only once. If the necessity arises to distribute new keys between Alice and Bob, how can they make sure that information about the key cannot be acquired by Eve? This is the problem of secure key distribution for which quantum mechanics provides a solution.

4.7.2. Quantum key distribution

Presently, there are several known quantum key distribution schemes. Here, we are going to describe the B92 protocol.[12] This protocol assumes that Alice and Bob work together (by exchanging classical information) to generate a new key in a secure way. They use the setup depicted in Figure 4.18. The method works as follows.

(1) Alice and Bob both generate a sufficiently long list of true random bits, S^A and S^B. From these two sequences, the key will be "distilled."
(2) For each bit of the list S^A, Alice prepares a qubit. If the bit is 0, she prepares the state $|\uparrow\rangle = \psi_+(\mathbf{e}_z)$ (spin-up in z-direction). If the bit is 1, she prepares $|\rightarrow\rangle = \psi_+(\mathbf{e}_x)$ (spin-up in x-direction).
(3) Alice sends her qubits one after another to Bob. Bob filters the qubits according to his bit list S^B. If the ith bit S^B_i is 0, he projects the ith qubit onto the state $|\leftarrow\rangle = \psi_-(\mathbf{e}_x)$ (spin-down in x-direction), and if the bit is 1, he projects onto $|\downarrow\rangle = \psi_-(\mathbf{e}_z)$ (spin-down in z-direction).
(4) Bob records the outcomes of his measurements. For each qubit, the outcome is either 1 (the qubit passes the filter), or 0 (the qubit gets absorbed). In that way, Bob will generate another list R of random bits. Note that Bob can obtain the result $R_i = 1$ for qubit i only

[11] Shannon, C.E., Bell Syst. Tech. J. **28**, 657 (1949).

[12] Bennett, C.H., Phys. Rev. Lett. **68**, 3121 (1992). See Chapter 2 of [**2**] for more information about quantum cryptography.

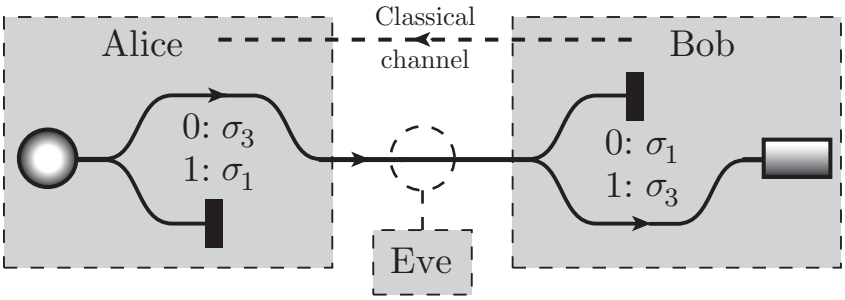


FIGURE 4.18. Setup for the B92 quantum key distribution protocol. Alice and Bob generate an encryption key under the nose of an eavesdropper Eve. Alice can prepare qubits in two non-orthogonal states, for example spin-up in the z- or x-direction. Bob can project onto the spin-down state either in the x- or z-direction.

if $S_i^A = S_i^B$, because otherwise Bob projects onto a state that is orthogonal to the state prepared by Alice.

(5) Via classical communication, Bob informs Alice about his results. He sends only the list R. This list contains no information about the settings chosen for Bob's measurement device. The classical communication may be public, but it is assumed that it cannot be disturbed by Eve.

(6) Both Alice and Bob now examine their lists S^A and S^B and generate new lists K^A and K^B by keeping only those bits, for which Bob recorded 1 in his measurement. This procedure results in two identical lists $K^A = K^B = K$ for Alice and Bob, because $R_i = 1$ can only happen for $S_i^A = S_i^B$. The list K is the key. It is a subset of S^A or S^B. Its length is about one-quarter of the length of the original lists.

Figure 4.19 shows an example illustrating the generation of the key as a common subset of the random lists S^A and S^B. Both lists S^A and S^B remain with Alice and Bob, and it is assumed that Eve has no knowledge about the contents of these lists. The message R selects the elements of S^A and S^B that constitute the common key. This message can be made public, because it contains no information about the values of the selected bits.

It can be shown that the B92 key generation protocol is secure. In this introduction, however, we are not going to analyze all possible eavesdropping attacks. In one scenario, Eve knows the two non-orthogonal states used for the protocol, and it is assumed that she can intercept and replace the qubits being sent to Bob. A single measurement, however, does not give

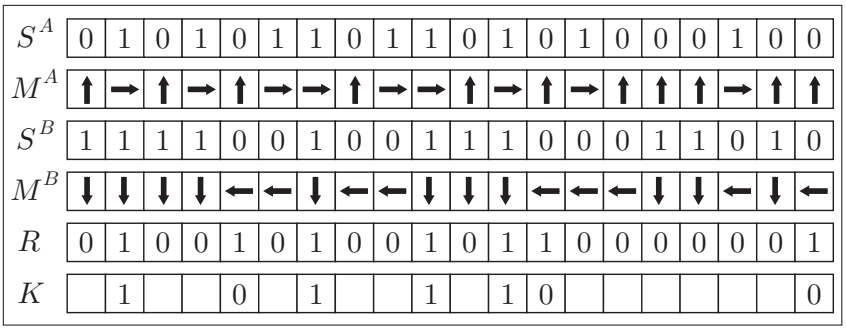


FIGURE 4.19. Example for the generation of a key using the B92 protocol. S^A: list of random bits generated by Alice. M^A: settings of the device for the state preparation. S^B: Bob's random list. M^B: settings chosen for Bob's filter. R: list with Bob's results, showing 1 if and only if the qubit passes through the filter. K: the resulting key.

sufficient information about the state of the intercepted qubit (Section 4.3.3), and copying the qubit is forbidden by the no-cloning theorem (Section 6.8.4 below). Quite generally, it is impossible to distinguish between two non-orthogonal states without disturbing them. The fake qubits received by Bob would thus introduce errors into the key. This can be detected if Alice and Bob sacrifice a portion of their key and compare it via the classical channel.

4.8. Hidden Variables

4.8.1. Failure of classical picture

In quantum mechanics, the three components of the spin are described by operators. Values are obtained only in measurements and usually cannot be predicted, except in some cases. For example, the observable "length of $\boldsymbol{\sigma}$" is represented by the matrix $|\boldsymbol{\sigma}| = \sqrt{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}} = \sqrt{3}\,\mathbf{1}_2$. Any measurement of that observable will produce the value $\sqrt{3}$ with certainty. This is conveniently but imprecisely formulated as "the observable $|\boldsymbol{\sigma}|$ has the value $\sqrt{3}$." Moreover, we can prepare a state where, for example, σ_3 has a definite value (either $+1$ or -1) simultaneously with $|\boldsymbol{\sigma}|$. But then the other components of the spin are totally uncertain; a measurement randomly produces $+1$ or -1. Can we assume that for an individual qubit these values already exist prior to the measurement?

The hypothesis of "hidden variables" assumes that physical quantities actually do have values, although these values might be inaccessible. Of

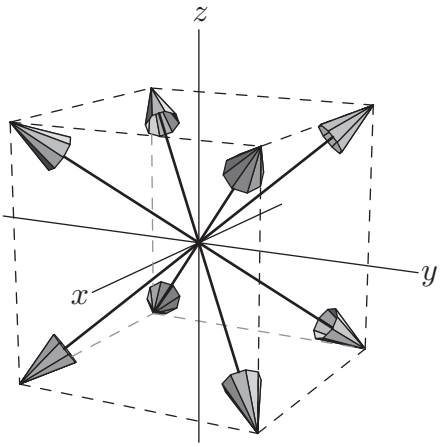


FIGURE 4.20. A classical vector, for which the x-, y-, and z-components can only have the values $\pm 1/2$, can only point to the corners of a centered unit-cube. Quantum mechanically, this property holds for any orientation of the coordinate system.

course, the spin is not a vector whose components have pre-existing values equal to those measured in Stern-Gerlach experiments: Such a vector would have the length $\sqrt{3}$, and every component would have either the value $+1$ or -1 (the only values ever obtained in any measurement of these components). This is impossible. If all components were ± 1, this would only leave eight possible spin vectors (see Fig. 4.20). But the orientation of the coordinate system is completely arbitrary. We could draw the same picture with respect to a rotated coordinate system and thus obtain eight other possible spin vectors in contradiction to the first result.

A hidden variable theory therefore usually assumes that the spin components have a continous range of possible values and that the experimental results ± 1 are an artifact of the method. From that point of view the quantum-mechanical results just reflect our incomplete knowledge of the system and of its interactions with our measurement device. Let us pursue this idea further in the following section.

4.8.2. Hidden-variable interpretation

Assume, for the moment, that a single qubit is characterized not only by its quantum state ψ (descibing the preparation procedure) but also by a parameter vector $\boldsymbol{\lambda} \in \Omega \subset \mathbb{R}^n$ that determines its behavior during a measurement. The components of $\boldsymbol{\lambda}$ are called *hidden variables*. An improved

version of quantum theory could perhaps one day provide us with a better understanding, but according to our present limitations in both experiment and theory, we are completely ignorant of the hidden variables. Hence, we expect that the available preparation procedures cannot yield reproducible values of $\boldsymbol{\lambda}$. Instead, when we prepare an ensemble of particles, then the values of $\boldsymbol{\lambda}$ will be distributed randomly according to some probability distribution. Naturally, this would lead to a probability distribution for the values of the spin components. But these probabilities would be related to insufficient knowledge and need not be regarded as a matter of principle. Perhaps Einstein had something like this in mind when an "inner voice" told him that God "does not play dice."

We denote the probability distribution of the hidden variables in a state ψ by $\rho_\psi(\boldsymbol{\lambda})$. Here, the subscript indicates that the distribution possibly depends on the preparation procedure, that is, on ψ. As a probability distribution, ρ_ψ must have the property

$$\rho_\psi(\boldsymbol{\lambda}) \geq 0, \quad \int_{\mathbb{R}^n} \rho_\psi(\boldsymbol{\lambda})\, d^n\lambda = 1. \tag{4.102}$$

A measurement of the spin in the direction $\mathbf{n}$ on a system in the state ψ gives either $+1$ or -1. The hidden variable theory assumes that the result depends on the actual value of $\boldsymbol{\lambda}$ for the individual system. The mechanism of this dependence is still unknown, but here we assume for simplicity that it is deterministic. The dependence of the result on $\boldsymbol{\lambda}$ is a function $s_{\mathbf{n},\psi}(\boldsymbol{\lambda})$ that can only have the values $+1$ and -1. According to the classical theory of probability, the probability for spin-up in the direction of $\mathbf{n}$ can be expressed in terms of the probability distribution $\rho_\psi(\boldsymbol{\lambda})$ by

$$p^{\text{up}}_{\mathbf{n},\psi} = \int_{B^{\text{up}}_{\mathbf{n},\psi}} \rho_\psi(\boldsymbol{\lambda})\, d^n\lambda \quad \text{with} \quad B^{\text{up}}_{\mathbf{n},\psi} = \{\boldsymbol{\lambda} \mid s_{\mathbf{n},\psi}(\boldsymbol{\lambda}) = +1\}. \tag{4.103}$$

Similarly, the expectation value of the spin-component $\boldsymbol{\sigma}\cdot\mathbf{n}$ in the state ψ is given by

$$\langle s_{\mathbf{n},\psi}\rangle = \int_{R^n} \rho_\psi(\boldsymbol{\lambda})\, s_{\mathbf{n},\psi}(\boldsymbol{\lambda})\, d^n\lambda. \tag{4.104}$$

Exercise 4.27 below gives an example of a hidden variable theory that can reproduce all expectation values of all spin components in all states of a single qubit.

A possible generalization could take into account another source of randomness that originates in our imprecise knowledge of the interaction between the particle and the measurement device in the determinative experiment. Most likely, the device is not built to react precisely to the actual value of $\boldsymbol{\lambda}$, because the hidden variables influence the outcome in some presently unknown way. Therefore, one assumes that for a given value of $\boldsymbol{\lambda}$, the result

$+1$ in the direction $\mathbf{n}$ only appears with a certain probability $p^{\mathrm{up}}_{\mathbf{n},\psi}(\boldsymbol{\lambda})$. The result -1 would thus appear with probability $p^{\mathrm{down}}_{\mathbf{n},\psi}(\boldsymbol{\lambda}) = 1 - p^{\mathrm{up}}_{\mathbf{n},\psi}(\boldsymbol{\lambda})$. The probability for spin-up in direction of $\mathbf{n}$ would then be given by

$$p^{\mathrm{up}}_{\mathbf{n},\psi} = \int \rho_\psi(\boldsymbol{\lambda})\, p^{\mathrm{up}}_{\mathbf{n},\psi}(\boldsymbol{\lambda})\, d^n\lambda. \tag{4.105}$$

A hidden variable theory is considered successful if it makes for all presently known measurements the same statistical predictions as quantum mechanics. It is indeed possible to invent such theories (see Exercise 4.27 for an example). The opinion shared by a majority of physicists is that it is not reasonable to explain the statistical properties of quantum systems on the basis of hidden variables. Any such assumption would lead into serious troubles with well established physical principles (in particular, locality) as soon as one considers systems consisting of more than one qubit. We shall return to this question in Section 6.5.

EXERCISE 4.26. *Show that the expectation value of the spin in the direction of* $\mathbf{n}$ *is always given by*

$$\langle \boldsymbol{\sigma} \cdot \mathbf{n} \rangle_\psi = 2\, p^{up}_{\mathbf{n},\psi} - 1. \tag{4.106}$$

EXERCISE 4.27. *Investigate Bell's example of a hidden variable theory: In this example, each qubit is characterized by its quantum state* ψ *and by a single hidden parameter* $\lambda \in \mathbb{R}$. *The probability distribution of* λ *is given by*

$$\rho_\psi(\lambda) = \begin{cases} 1/2 & \text{for } -1 \le \lambda \le 1, \\ 0 & \text{else.} \end{cases} \tag{4.107}$$

(*Here,* ρ_ψ *is in fact independent of* ψ). *Assume that the hidden parameter* λ *determines the value of the spin-component* $\boldsymbol{\sigma} \cdot \mathbf{n}$ *in the direction of an arbitrary unit vector* $\mathbf{n}$ *by the following formula:*

$$s_{\mathbf{n},\psi}(\lambda) = \mathrm{sgn}(\lambda + |m_3|)\,\mathrm{sgn}(m_3), \quad \mathrm{sgn}(x) = \begin{cases} 1, & x \ge 0, \\ -1, & x < 0. \end{cases} \tag{4.108}$$

Here, m_3 *is the third component of the vector* $\mathbf{m} = \mathbf{R}(\boldsymbol{\alpha})\,\mathbf{n}$. *The rotation matrix* $\mathbf{R}(\boldsymbol{\alpha})$ *is determined by the requirement that it should transform the spin-up direction* $\mathbf{n}(\psi)$ *of the state* ψ *into the positive* z*-direction. Show that all quantum-mechanical expectation values of the spin are described correctly by this hidden variable theory. To this purpose, show that the relation*

$$\langle \psi, \boldsymbol{\sigma} \cdot \mathbf{n}\, \psi \rangle = \int_{-\infty}^{\infty} \rho_\psi(\lambda)\, s_{\mathbf{n},\psi}(\lambda)\, d\lambda \tag{4.109}$$

holds for all directions $\mathbf{n}$ *and all spinors* ψ (*see also Exercise 4.18*).

4.9. Special Topic: Qubit Dynamics

4.9.1. Time-dependent Hamiltonian

A general time-independent qubit Hamiltonian has the form

$$H = \omega_0 \, \mathbf{1}_2 + \boldsymbol{\omega} \cdot \boldsymbol{\sigma} \tag{4.110}$$

and the time evolution generated by this Hamiltonian is essentially a rotation:

$$\exp(-\mathrm{i}Ht) = \mathrm{e}^{-\mathrm{i}\omega_0 t} \, U(\boldsymbol{\omega} t/2), \tag{4.111}$$

(assuming $\hbar = 1$) with U as defined in (4.28) (see also Section 4.4.3).



CD 4.12.1 and 2 show the precession of the spin vector about the direction of the magnetic field, that is, the time evolution generated by a Hamiltonian like (4.110). In CD 4.12.3, the magnetic field vector $\boldsymbol{\omega}$ slowly rotates about the z-axis.

Here, we consider a Hamiltonian that depends explicitly on time. It describes, for example, a particle with spin in a time-dependent magnetic field. In this case, the interaction energy is

$$H(t) = -\boldsymbol{\mu} \cdot \mathbf{B}(t). \tag{4.112}$$

Now the time evolution is not simply the exponential function of the Hamiltonian.

In the simplest case, the Hamiltonians for different times commute,

$$[H(t), H(s)] = 0, \quad \text{for all } t \text{ and } s. \tag{4.113}$$

In this case, the solution of the Schrödinger equation

$$\mathrm{i} \frac{d}{dt} \, \psi(t) = H(t) \, \psi(t) \tag{4.114}$$

can be obtained by

$$\psi(t) = \exp\Big(-\mathrm{i} \int_{t_0}^{t} H(s) \, ds\Big) \, \psi(t_0). \tag{4.115}$$

Note that the time evolution operator not only depends on the time t, but also on the initial time t_0. The time evolution is thus given by a two-parameter family of unitary operators, $(t, s) \to U(t, s)$. Quite generally, it has the following properties:

$$\begin{aligned} &U(t_0, t_0) = \mathbf{1}, \\ &U(t, s) \, U(s, t_0) = U(t, t_0), \\ &U(t, s)^\dagger = U(s, t) = U(t, s)^{-1}. \end{aligned} \tag{4.116}$$

In general, we cannot assume that $H(t)$ commutes with $H(s)$ for $t \neq s$. In this case, we have to be particularly careful. We can expect that the solution

is still described by a unitary operator depending on two parameters with the properties (4.116). We have to solve the Schrödinger equation

$$\mathrm{i}\,\frac{d}{dt}\,U(t,t_0)\,\psi = H(t)\,U(t,t_0)\,\psi \tag{4.117}$$

for all initial states ψ. Integrating over time from t_0 to t, we obtain the integral equation

$$\psi(t) = \psi(t_0) + (-\mathrm{i})\int_{t_0}^{t} H(s)\,\psi(s)\,dt, \tag{4.118}$$

which can be solved by iteration,

$$\psi(t) = \psi(t_0) + (-\mathrm{i})\int_{t_0}^{t} H(t)\,\psi(t_0)\,dt \tag{4.119}$$

$$+\,(-\mathrm{i})^2\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\,H(t_1)\,H(t_2)\,\psi(t_0) + \dots \tag{4.120}$$

$$= \sum_{n=0}^{\infty} (-\mathrm{i})^n \int_{t_0}^{t} dt_1 \dots \int_{t_0}^{t_{n-1}} dt_n\,H(t_1)\dots H(t_n)\,\psi(t_0). \tag{4.121}$$

This is called the *Dyson expansion*. Notice that it is not possible to exchange the order of the factors in the integrand, because the operators $H(t_j)$ and $H(t_k)$ do not commute. If the Hamiltonian $H(t)$ is a finite-dimensional matrix, the series always converges in the norm of bounded operators.

EXERCISE 4.28. *In view of the applications, it is useful to consider a Hamiltonian of the form*

$$H(t) = H_0 + H_1(t), \tag{4.122}$$

where H_0 is independent of t. Find the equation of motion for $\tilde{\psi}(t) = \exp(\mathrm{i}H_0 t)\,\psi(t)$ and solve it using the Dyson expansion.

4.9.2. Time dependence generated by unitary operators

The Schrödinger equation can be solved exactly if the Hamiltonians at different times are connected by unitary transformations,

$$H(t+s) = \mathrm{e}^{\mathrm{i}At}\,H(s)\,\mathrm{e}^{-\mathrm{i}At}. \tag{4.123}$$

Let $U(t,s)$ be the propagator defined by $H(t)$, that is,

$$\mathrm{i}\,\frac{d}{dt}\,U(t,s) = H(t)U(t,s), \tag{4.124}$$

then, for any fixed $s \in \mathbb{R}$,

$$\mathrm{i}\,\frac{d}{dt}\,\mathrm{e}^{-\mathrm{i}As}\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As} = \mathrm{e}^{-\mathrm{i}As}\left(\mathrm{i}\,\frac{d}{dt}\,U(t+s,s)\right)\mathrm{e}^{\mathrm{i}As} \tag{4.125}$$

$$= \mathrm{e}^{-\mathrm{i}As}\,H(t+s)\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As} \tag{4.126}$$

$$= \mathrm{e}^{-\mathrm{i}As}\,H(t+s)\,\mathrm{e}^{\mathrm{i}As}\,\mathrm{e}^{-\mathrm{i}As}\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As} \tag{4.127}$$

$$= H(t)\,\mathrm{e}^{-\mathrm{i}As}\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As}. \tag{4.128}$$

Hence, $V(t) = \mathrm{e}^{-\mathrm{i}As}\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As}$ is the solution of (4.124) that fulfills the initial condition $V(0) = \mathbf{1}$. This shows that

$$V(t) = \mathrm{e}^{-\mathrm{i}As}\,U(t+s,s)\,\mathrm{e}^{\mathrm{i}As} = U(t,0), \tag{4.129}$$

because $U(t,0)$ is the unique solution of (4.124) with $U(0,0) = \mathbf{1}$. Multiplying from the left by $\mathrm{e}^{-\mathrm{i}At}$ and from the right by $\mathrm{e}^{-\mathrm{i}As}\,U(s,0)$,

$$\mathrm{e}^{-\mathrm{i}A(t+s)}\,U(t+s,0) = \mathrm{e}^{-\mathrm{i}At}\,U(t,0)\,\mathrm{e}^{-\mathrm{i}As}\,U(s,0). \tag{4.130}$$

This shows that the operators

$$W(t) = \mathrm{e}^{-\mathrm{i}At}\,U(t,0) \tag{4.131}$$

form a one-parameter unitary group, because (4.130) is just the group property $W(t+s) = W(t)W(s)$. The generator of the unitary group W is

$$\frac{d}{dt}W(t)\Big|_{t=0} = H(0) + A, \quad \text{hence} \quad W(t) = \mathrm{e}^{-\mathrm{i}(H(0)+A)t}. \tag{4.132}$$

Thus, we find that the solution of the initial value problem is given by

$$\psi(t) = U(t,0)\,\psi_0 = \mathrm{e}^{\mathrm{i}At}\,W(t)\,\psi_0 = \mathrm{e}^{\mathrm{i}At}\,\mathrm{e}^{-\mathrm{i}\,(H(0)+A)t}\,\psi_0. \tag{4.133}$$

4.9.3. Magnetic resonance

We consider the Hamiltonian

$$H(t) = \frac{1}{2}\,\omega_0\,\sigma_3 + \frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{b}(t) \tag{4.134}$$

with

$$\mathbf{b}(t) = \lambda\,\omega_0\,(\cos\omega t, \sin\omega t, 0). \tag{4.135}$$

The time-independent part of the Hamiltonian has the two eigenstates ψ_+ and ψ_-. The energy difference between these states is just given by ω_0. The second summand is a time-dependent perturbation. The constant $\lambda \geq 0$ describes the strength of this perturbation. The vector $\mathbf{b}(t)$ rotates with constant angular speed ω in the xy-plane. Physically, the perturbation can be interpreted as a time-dependent magnetic field that is homogeneous in space, but whose field vector $\mathbf{B}$ rotates with a constant angular speed in the xy-plane. Hence, the time-dependence of the Hamiltonian $H(t)$ is generated

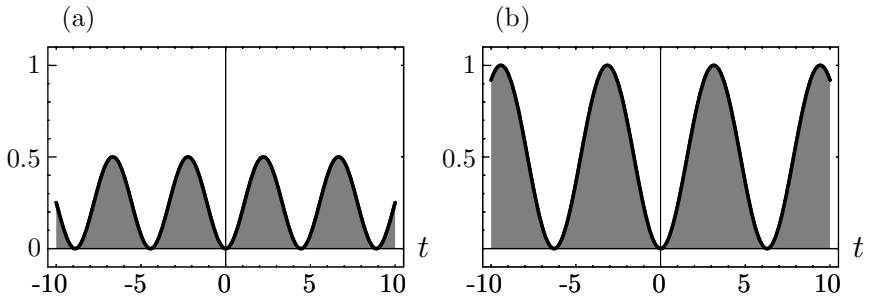


FIGURE 4.21. The transition probability $p(t) = |\langle\psi(t), \psi_-\rangle|^2$ for the initial state $\psi(0) = \psi_+$ as a function of time. We have $\omega_0 = 10$, $\lambda = 0.2$, and (a) $\omega = 9$, (b) $\omega = 10$.

by a rotation around the z-axis. Indeed, using $\mathbf{R}(\omega t \mathbf{e}_z)\mathbf{b}(0) = \mathbf{b}(t)$, we find from Eq. (4.53) that

$$\begin{aligned} H(t) &= \mathrm{e}^{-\mathrm{i}\omega t\sigma_3/2}\, H(0)\, \mathrm{e}^{+\mathrm{i}\omega t\sigma_3/2} \\ &= \frac{1}{2}\,\omega_0\,\sigma_3 + \mathrm{e}^{-\mathrm{i}\omega t\sigma_3/2}\,\frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{b}(0)\,\mathrm{e}^{+\mathrm{i}\omega t\sigma_3/2}. \end{aligned} \tag{4.136}$$

This is of the form (4.123) with $A = -\omega\sigma_3/2$, and the solution is therefore given by (4.133),

$$\begin{aligned} \psi(t) &= \exp\Big(-\mathrm{i}t\,\frac{\omega}{2}\,\sigma_3\Big)\,\exp\Big(-\mathrm{i}t\frac{\omega_0-\omega}{2}\,\sigma_3 - \mathrm{i}t\frac{1}{2}\,\lambda\,\omega_0\,\sigma_1\Big)\,\psi(0) \\ &= \exp\Big(-\mathrm{i}t\,\frac{\omega}{2}\,\sigma_3\Big)\,\exp\Big(-\mathrm{i}t\frac{\Omega(\lambda,\omega_0,\omega)}{2}\,\boldsymbol{\sigma}\cdot\hat{\mathbf{b}}\Big)\,\psi(0) \end{aligned} \tag{4.137}$$

with the unit vector

$$\begin{aligned} \hat{\mathbf{b}} &= \frac{1}{\Omega(\lambda,\omega_0,\omega)}\,(\lambda\omega_0, 0, \omega_0-\omega), \quad \text{where} \\ \Omega(\lambda,\omega_0,\omega) &= \lambda\omega_0\,\sqrt{1+\frac{1}{\lambda^2}\,\Big(1-\frac{\omega}{\omega_0}\Big)^2}. \end{aligned} \tag{4.138}$$

Now it is easy to determine, for example, the time evolution of $\psi(0) = \psi_+$. It is interesting to consider the (time-dependent) transition probability that $\psi(t)$ is found to be in the state ψ_-. This probability is given by

$$p(t) = |\langle\psi(t), \psi_-\rangle|^2 = \frac{\lambda^2\omega_0^2}{\Omega(\lambda,\omega_0,\omega)^2}\,\sin^2\Big(\frac{t\,\Omega(\lambda,\omega_0,\omega)}{2}\Big) \tag{4.139}$$



Explore the parameter dependence of the function $p(t)$ in CD 4.13.1. The animations CD 4.13.2 show the maximal value $p_{\max}(\lambda,\omega_0,\omega)$ of $p(t)$, and CD 4.13.3 investigates the period $T(\lambda,\omega_0,\omega)$ of $p(t)$.

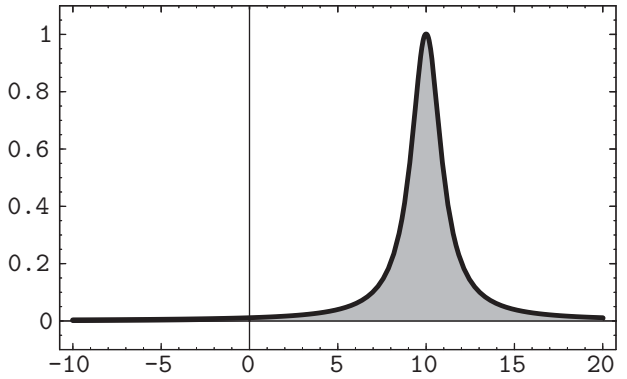


FIGURE 4.22. Resonance curve ($p_{\max}$ as a function of ω) for $\lambda = 0.1$ and $\omega_0 = 10$.

We plot the function $p(t)$ in Figure 4.21. We see that $p(t)$ is periodic in time. At the zeros of $p(t)$, the state is orthogonal to ψ_-, that is, the system is again in the initial state. For these times, the solution $\psi(t)$ must be equal to ψ_+ (up to a phase factor). In particular, we have

$$\psi(T) = \mathrm{e}^{\mathrm{i}\theta}\,\psi_+ \tag{4.140}$$

(where $\mathrm{e}^{\mathrm{i}\theta}$ is a suitable phase factor) at the time

$$T = T(\lambda, \omega_0, \omega) = \frac{2\pi}{\Omega(\lambda, \omega_0, \omega)}. \tag{4.141}$$

The transition probability $p(t)$ oscillates in time between 0 and a maximal value given by

$$p_{\max}(\lambda, \omega_0, \omega) = \frac{\lambda^2\omega_0^2}{\Omega(\lambda, \omega_0, \omega)^2} = \left(1 + \frac{1}{\lambda^2}\Big(1 - \frac{\omega}{\omega_0}\Big)^2\right)^{-1}. \tag{4.142}$$

This maximal value describes the strength of the system's response to the perturbation $\mathbf{b}(t)$. As a function of ω (at fixed values of λ and ω_0), $p_{\max}$ has the shape of a resonance curve (see Fig. 4.22). Notice the following points:

(1) The maximum is at the resonance frequency $\omega = \omega_0$, where the maximal value is 1, irrespective of the value of λ.
(2) The width of the resonance depends on $\lambda\omega_0$.
(3) There is no resonance for negative ω, that is, if the external magnetic field rotates in the opposite direction.



The movies CD 4.14–CD 4.17 show the behavior of a qubit under various time-dependent perturbations. Resonant and non-resonant behavior of the magnetic moment vector is shown for circularly and linearly oscillating magnetic fields. In CD 4.15, the perturbation is large and the resonance curve is wide. In CD 4.16, the perturbation is small and the resonance curve is a sharp peak. At the resonance frequency, it is always possible to turn an initial state ψ_+ into the orthogonal state ψ_-, irrespective of the strength of the perturbation.

Atomic nuclei with an odd number of nucleons have a nonzero spin (for example, the nucleus of a hydrogen atom). The resonant behaviour of the nuclear spin in magnetic fields can be used to measure the distribution of these atoms in a sample of material. This technique, known as nuclear magnetic resonance (NMR), has important applications to medical imaging.

Chapter 5

Composite Systems

Chapter summary: Up to now, we have only considered single-particle systems. Real physics starts where at least two particles are involved. As a first example of a two-particle system we consider the Schrödinger equation for two free particles. This shows us how to construct two-particle states as products of one-particle states (Section 5.1). An abstract formulation of this method is given by the tensor product of Hilbert spaces. The Hilbert space of a composite system contains not only product states, but also their linear combinations (Section 5.2). In general, these states are entangled, that is, they cannot be written as simple products. Entanglement cannot be created by local measurements or manipulations of the subsystems, but usually, an interaction between the subsystems immediately leads to entanglement (Sections 5.3 and 5.4).

The theory presented here has applications not only to atomic physics but also to quantum information theory. An entangled state of a composite system encodes information about the system as a whole that cannot be measured locally (that is, by measurements on the subsystems alone). This information describes correlations between the subsystems. In general, the state of a subsystem cannot be described by a state vector in the Hilbert space of the subsystem. If the bipartite system is in an entangled state, then any subsystem is in a statistical mixture of states. We describe this new situation by a density operator ρ (Section 5.5). The density operator generalizes the orthogonal projection operator onto the subspace spanned by a "pure state" ψ. In Section 5.6, we compare pure and mixed states and discuss the ensemble interpretation of mixed states. We describe the ambiguities in the preparation of mixed states in Section 5.7 and some mathematical aspects of bipartite systems in Section 5.8 (the normal form of the state vector, maximally entangled states, purification, and the projection postulate for mixed states).

Section 5.9 describes the physics of composite systems whose parts are indistinguishable. A major new ingredient is the symmetrization postulate, which says that all states of the composite system are either symmetric or antisymmetic under exchange of the subsystems. Hence, one may conclude that there are two families of elementary particles, bosons and fermions. Fermions obey Pauli's exclusion principle, which states that two fermions cannot occupy the same state. Section 5.10 describes multiparticle systems with spin, thus forming the theoretical foundation of the physics of atoms, molecules, and matter. Finally, in Section 5.11 we present the addition of angular momenta as a problem where two systems are combined into a larger system.

5.1. States of Two-Particle Systems

5.1.1. The Hamiltonian for two free particles

In order to get a feeling for the wave functions of a two-particle system, we neglect all interactions and consider a system of two free particles. For reasons that will become clear later, we assume here that the two particles are not identical, but can be distinguished by some physical property (for example, we may assume that the particles have different masses $\mathrm{m}_1 \neq \mathrm{m}_2$). Moreover, we assume that the Hilbert space of either particle is $L^2(\mathbb{R}^n)$. We are going to follow our usual procedure to set up a quantum Hamiltonian by correspondence with the classical mechanical system. From the solutions of the Schrödinger equation with this Hamiltonian, we will learn about the typical structure of two-particle wave functions.

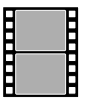
Assuming that each particle moves in an n-dimensional space, the classical configuration space of the two-particle system is $\mathbb{R}^{2n}$. The points $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ in this space describe the configuration of the system, that is, the positions $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ of the individual particles. The Hamiltonian operator for this system is obtained from the classical expression for the energy with the usual translation rule. According to this rule, the operator for the momentum $\mathbf{p}^{(j)}$ of the particle j is the differential operator $-\mathrm{i}\hbar\nabla^{(j)}$ (gradient with respect to the coordinates $\mathbf{x}^{(j)}$). The Hamiltonian is the sum of the kinetic energies of the two particles, that is,

$$H = \frac{(\mathbf{p}^{(1)})^2}{2\mathrm{m}_1} + \frac{(\mathbf{p}^{(2)})^2}{2\mathrm{m}_2} = -\frac{\hbar^2}{2\mathrm{m}_1}\,\Delta^{(1)} - \frac{\hbar^2}{2\mathrm{m}_2}\,\Delta^{(2)}, \tag{5.1}$$

with m_1 and m_2 denoting the masses of the two particles, respectively. Here, it is assumed that the two particles move freely. In particular, there is no potential energy term that would describe an interaction between the two particles.

The linear operator H acts on (differentiable) functions $\psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ of the $2n$ position variables, and $\Delta^{(j)}$ is the Laplace operator that differentiates with respect to the variables $\mathbf{x}^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$,

$$\Delta^{(j)} = \frac{\partial^2}{\partial (x_1^{(j)})^2} + \cdots + \frac{\partial^2}{\partial (x_n^{(j)})^2}. \tag{5.2}$$



CD 5.1.1 shows the independent (interaction-free) motion of two free particles on a line. The particles have different mass, move in opposite directions, and both are described by Gaussian wave functions. The second particle's wave function is plotted upside-down. CD 5.1.2 shows two independent oscillating particles, that is, two particles moving without interaction in a harmonic oscillator potential.

5.1.2. The Schrödinger equation of a two-particle system

We consider the time-dependent Schrödinger equation with the two-particle Hamiltonian above. Scaling the units of length, we can get rid of the constant $\hbar$ and write the Schrödinger equation in the form

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)},t) = -\Big(\frac{1}{2\mathrm{m}_1}\,\Delta^{(1)} + \frac{1}{2\mathrm{m}_2}\,\Delta^{(2)}\Big)\,\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)},t). \tag{5.3}$$

We note that the Hamiltonian is a sum of two terms that commute with one another. Hence, we can find a special set of solutions by a separation of variables, that is, by writing

$$\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)},t) = \psi_1(\mathbf{x}^{(1)},t)\,\psi_2(\mathbf{x}^{(2)},t). \tag{5.4}$$

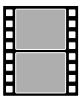
This is a solution of the two-particle Schrödinger equation (5.3) whenever ψ_j, $j = 1, 2$, are solutions of the one-particle Schrödinger equations for particles with mass m_j.

EXERCISE 5.1. *Insert* (5.4) *into the Schrödinger equation* (5.3) *and show that* $\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)},t)$ *is a solution if and only if* $\psi_j(\mathbf{x},t)$ *is a solution of the equation*

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi_j(\mathbf{x},t) = -\frac{1}{2\mathrm{m}_j}\,\Delta\,\psi_j(\mathbf{x},t). \tag{5.5}$$

We conclude that the Schrödinger equation for two free particles can be separated completely into two independent one-particle problems. This should not be surprising, because a system of two interaction-free particles consists of completely independent one-particle subsystems. Each of the particles moves as if the other was not there.

If ψ and ϕ are solutions of the two-particle Schrödinger equation, then any linear combination $a\,\psi + b\,\phi$ is again a solution (superposition principle, linearity of the Schrödinger equation). One soon realizes that even if the two solutions ψ and ϕ are in product form, it is in general impossible to write the linear combination $a\,\psi + b\,\phi$ as a product of one-particle solutions. Thus, the two-particle Schrödinger equation has many solutions that are not just products of one-particle wave functions.



CD 5.2.1 combines two one-particle solutions into a solution of the two-particle Schrödinger equation. This is a wave function of the form (5.4), defined on the configuration space $\mathbb{R}^2$ of the two-particle system. CD 5.2.2 shows the wave function in configuration space that is obtained by forming the product of the two independent oscillators in CD 5.1.2. The Hamiltonian of this system is a sum of two harmonic-oscillator Hamiltonians without interaction term.

5.1.3. Two-particle Hilbert space

What is the most general form of the wave function at a given time t? Assume that the set $\{\psi_j \mid j = 1, 2, \ldots\}$ is an orthonormal basis in the single-particle Hilbert space $L^2(\mathbb{R}^n)$. Among the possible states of the two-particle system are all the products

$$\psi_j(\mathbf{x}^{(1)})\,\psi_k(\mathbf{x}^{(2)}), \quad j, k = 1, 2, 3 \ldots \tag{5.6}$$

and because the set of all possible states is a linear space (superposition principle), we can form arbitrary linear combinations from these products. The most general two-particle wave function (at a given time t) thus has the form

$$\psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{jk}\,\psi_j(\mathbf{x}^{(1)})\,\psi_k(\mathbf{x}^{(2)}). \tag{5.7}$$

This is a square-integrable function of the $2n$ variables $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\}$ whenever the double-sum converges with respect to the norm in $L^2(\mathbb{R}^{2n})$. For this it is necessary and sufficient that the coefficients c_{jk} be square-summable,

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |c_{jk}|^2 < \infty. \tag{5.8}$$

In this case, the two-particle wave function ψ is square-integrable over $\mathbb{R}^{2n}$,

$$\|\psi\|^2 = \int_{\mathbb{R}^{2n}} |\psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})|^2\, d^{2n}x = \sum_{j,k} |c_{jk}|^2 < \infty. \tag{5.9}$$

In general, it is not possible to write a linear combination of product functions again as a product of two wave functions.

Hilbert space of two particles:

If the state space of a single particle is $L^2(\mathbb{R}^n)$, then the Hilbert space of two-particle wave functions is $L^2(\mathbb{R}^{2n})$. If $\{\psi_k \mid k = 1, 2, 3, \ldots\}$ is an orthonormal basis in $L^2(\mathbb{R}^n)$, then the products of the basis states

$$\{\psi_{jk} \mid \psi_{jk}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \psi_j(\mathbf{x}^{(1)})\,\psi_k(\mathbf{x}^{(2)}),\ j, k = 1, 2, 3, \ldots\} \tag{5.10}$$

form an orthonormal basis of the two-particle Hilbert space.

At any given time t, the most general wave function is thus a (possibly infinite) linear combination of products of one-particle wave functions.

EXERCISE 5.2. *Show that the states of the two-particle basis* (5.10) *are orthonormal if the one-particle states ψ_j are pairwise orthonormal* (*that is, if $\langle \psi_j, \psi_k \rangle = \delta_{ik}$*). *Prove Eq.* (5.9).



In CD 5.3.1, we show a wave function of the two-particle system that cannot be written as a product of one-particle wave functions. A state with this property is called *entangled*. Entangled states will be discussed later in this section and in Chapter 6. CD 5.3.2 shows wave function that is a product of one-particle wave functions (this is called a *separable state* of the two-particle system).

5.1.4. The interpretation of two-particle wave functions

Analogous to the interpretation of a one-particle wave function, we say that

$$\int_{B\subset\mathbb{R}^{2n}} |\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})|^2\, d^{2n}x \tag{5.11}$$

is the probability of finding the configuration $\mathbf{x} = (\mathbf{x}^{(1)},\mathbf{x}^{(2)})$ in $B \subset \mathbb{R}^{2n}$. A natural choice for the region B is a "rectangle" $B_1 \times B_2$, with $B_j \subset \mathbb{R}^n$. The above expression then gives the probability of finding the first particle in the region B_1 and the second particle in B_2. For the special case that the wave function is a product, this probability also becomes a product,

$$\int_{B\subset\mathbb{R}^{2n}} |\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})|^2\, d^{2n}x = \int_{B_1} |\psi_1(\mathbf{x})|^2\, d^nx \int_{B_2} |\psi_2(\mathbf{x})|^2\, d^nx. \tag{5.12}$$

This is the joint probability of the independent events E_1 = "particle 1 in B_1" and E_2 = "particle 2 in B_2." If the wave function is not a product, the joint probability will not factorize, which means that the events E_1 and E_2 are not independent.

Because of the high dimension of the configuration space of two particles, it is difficult to visualize a two-particle wave function (except in the case that the space dimension is $n = 1$). We can, however, visualize the quantity

$$\rho_1(\mathbf{x}) = \int_{\mathbb{R}^n} |\psi(\mathbf{x},\mathbf{y})|^2\, d^ny, \tag{5.13}$$

which is called the *one-particle density function*. We interpret $\rho_1(\mathbf{x})$ as the probability density for finding particle 1 at $\mathbf{x}$ and particle 2 somewhere. Thus, $\int_B \rho_1(\mathbf{x})\, d^nx$ is the probability that the position of particle 1 is in B_1, irrespective of the position of particle 2. Similarly, we define

$$\rho_2(\mathbf{x}) = \int_{\mathbb{R}^n} |\psi(\mathbf{y},\mathbf{x})|^2\, d^ny \tag{5.14}$$

as the position probability density of particle 2 irrespective of the position of particle 1. Whenever ψ is a product of normalized one-particle states, ρ_1 and ρ_2 are just the corresponding one-particle densities:

$$\rho_1(\mathbf{x}) = |\psi_1(\mathbf{x})|^2, \quad \rho_2(\mathbf{x}) = |\psi_2(\mathbf{x})|^2, \quad \text{if } \psi(\mathbf{x},\mathbf{y}) = \psi_1(\mathbf{x})\psi_2(\mathbf{y}). \tag{5.15}$$



CD 5.5 visualizes the time evolution of a two-particle system in terms of the one-particle density functions $\rho_1(\mathbf{x})$ and $\rho_2(\mathbf{x})$. The system consists of two particles interacting via a harmonic-oscillator force.

5.2. Hilbert Space of a Bipartite System

Let us consider quite generally a quantum system that is made up of two different parts A and B (this is called a *bipartite system*). How do we construct the formalism of the combined system if we know how to describe the constituent parts? This precisely was the problem we had to face when we introduced the wave functions of a system consisting of two free particles. We solved this problem by introducing a larger Hilbert space, which contains products of single-particle wave functions. The same ideas also works for other quantum systems (e.g., for qubits).

5.2.1. Construction of the tensor product

We consider two physical systems A and B. The Hilbert spaces of the individual systems are denoted by $\mathfrak{H}^A$ and $\mathfrak{H}^B$, respectively. When we join the two parts, we have to find a larger Hilbert space that is capable of describing the states of the compound system. The form of the two-particle wave functions (5.6) suggests that the compound states should be products of individual states. Mathematically, the *tensor product* provides a nice method to construct a Hilbert space containing all possible products of elements from two given Hilbert spaces $\mathfrak{H}^A$ and $\mathfrak{H}^B$.

Let the system A be in a state $\psi^A \in \mathfrak{H}^A$ and B be in a state $\psi^B \in \mathfrak{H}^B$. Without further assumptions, a multiplication of ψ^A with ψ^B need not make any sense, because the two Hilbert spaces could be completely different. Hence, we introduce a new symbol $\otimes$ and denote the state of the combined system AB by

$$\psi^A \otimes \psi^B. \tag{5.16}$$

This formal product just represents the ordered pair of the states ψ^A and ψ^B. In case of ordinary wave functions, we may replace $\otimes$ by the usual product of two functions. Next, we introduce the scalar product

$$\langle\, \psi^A \otimes \psi^B \,,\, \phi^A \otimes \phi^B \,\rangle = \langle \psi^A \,, \phi^A \,\rangle \, \langle \psi^B \,, \phi^B \,\rangle. \tag{5.17}$$

The expressions on the right side of this definition are well defined, because they involve only scalar products of vectors in the same Hilbert space.

The product states of the form $\psi^A \otimes \psi^B$ cannot be the only states of the compound system. According to the superposition principle, we have to include arbitrary linear combinations of product states in the state space of the system AB. We extend the set of product states in such a way that the

formal product $\otimes$ is linear in both factors. To that purpose, we introduce the following rule

$$\sum_j c_j \psi_j^A \otimes \sum_k d_k \psi_k^B = \sum_{j,k} c_j \, d_k \, (\psi_j^A \otimes \psi_k^B). \tag{5.18}$$

For the scalar product (5.17) we assume, as usual, linearity in the second and antilinearity in the first factor. In that way, the definition of the scalar product "automatically" carries over to all vectors of the form (5.18).

The vector space that consists of all elements of the form (5.18) defines[1] the *tensor product* of the Hilbert spaces $\mathfrak{H}^A$ and $\mathfrak{H}^B$. We denote this vector space by

$$\mathfrak{H}^{AB} = \mathfrak{H}^A \otimes \mathfrak{H}^B. \tag{5.19}$$

The introduction of the tensor product for the description of compound systems is a basic new ingredient in the quantum mechanical theory. Ultimately, it can only be justified by its success.

Basic assumption about compound systems:

The Hilbert space of a quantum mechanical system consisting of two parts A and B is the tensor product of the Hilbert spaces $\mathfrak{H}^A$ and $\mathfrak{H}^B$ associated to the subsystems.

It is straightforward to generalize the considerations above to systems composed of several subsystems. The tensor product is associative. Up to isomorphism, we may, for example, identify the following tensor products

$$\mathfrak{H}^A \otimes \mathfrak{H}^B \otimes \mathfrak{H}^C = (\mathfrak{H}^A \otimes \mathfrak{H}^B) \otimes \mathfrak{H}^C = \mathfrak{H}^A \otimes (\mathfrak{H}^B \otimes \mathfrak{H}^C). \tag{5.20}$$

5.2.2. Orthonormal basis of the tensor product space

Assume that $\{\psi_j^A\}$ is an orthonormal basis in $\mathfrak{H}^A$ and $\{\psi_k^B\}$ is an orthonormal basis in $\mathfrak{H}^B$. Then, one can show that the set formed by the vectors

$$\psi_{jk} = \psi_j^A \otimes \psi_k^B \tag{5.21}$$

forms an orthonormal basis of the tensor product space. Hence, every vector ψ in $\mathfrak{H}^{AB}$ has a unique representation as

$$\psi = \sum_{j,k} c_{jk} \, \psi_j^A \otimes \psi_k^B \quad \text{with} \quad c_{jk} = \langle \psi_j^A \otimes \psi_k^B \, , \, \psi \rangle. \tag{5.22}$$

[1] In the infinite dimensional case, the definition of the tensor product includes the topological closure of the set of finite linear combinations of product states. We refer to the mathematical literature for a more precise definition of the tensor product of Hilbert spaces.

The sum above might be finite or infinite, depending on the dimensions of the constituent Hilbert spaces. In the infinite-dimensional case, one has the additional condition

$$\|\psi\|^2 = \sum_{j,k} |c_{jk}|^2 < \infty. \tag{5.23}$$

EXERCISE 5.3. *Show that the basis states ψ_{jk} defined in* (5.21) *form an orthonormal set, that is,*

$$\langle \psi_{jk} \, , \, \psi_{lm} \rangle = \delta_{jl} \, \delta_{km}. \tag{5.24}$$

EXERCISE 5.4. *Describe the Hilbert space of a two-qubit system. What is its dimension?*

5.2.3. Entangled states

Whenever a bipartite system is in a product state, that is,

$$\psi = \psi^A \otimes \psi^B, \tag{5.25}$$

then we can say that system A is in the state ψ^A and system B is in the state ψ^B. The state $\psi^A \otimes \psi^B$ is obtained if the two systems are separately prepared in the states ψ^A and ψ^B, respectively, and juxtaposed without interaction.

But in the Hilbert space of the compound system, most states cannot be written as a single tensor product. For example, the linear combination

$$\psi = \frac{1}{\sqrt{2}} \, (\psi_0^A \otimes \psi_0^B + \psi_1^A \otimes \psi_1^B) \tag{5.26}$$

is not a product state, at least if we assume that both sets $\{\psi_0^A, \psi_1^A\}$ and $\{\psi_0^B, \psi_1^B\}$ are linearly independent.

Entangled and separable states:

A state of a compound system is called *entangled* if it cannot be written as a single tensor product of subsystem states. A state in the product form (5.25) is called *unentangled* or *separable*.

Even if the initial state of a compound system is separable, it will, in general, become entangled during the time evolution if there is some coupling between the subsystems. Whenever the state of a compound system is entangled, it is meaningless to speak of the state of a subsystem in the sense used so far (that is, as something that can be described by a single vector in the Hilbert space of the system). We will be forced to generalize our notion of a state.

EXERCISE 5.5. *Given two orthogonal vectors ϕ_1^A and ϕ_2^A in $\mathfrak{H}^A$ and two orthogonal vectors ϕ_1^B and ϕ_2^B in $\mathfrak{H}^B$, show that the state*

$$\psi = \frac{1}{2}\big(\phi_1^A \otimes \phi_1^B + \phi_1^A \otimes \phi_2^B + \phi_2^A \otimes \phi_1^B + \phi_2^A \otimes \phi_2^B\big) \tag{5.27}$$

is separable whereas the state

$$\psi = \frac{1}{\sqrt{2}}\big(\phi_1^A \otimes \phi_1^B + \phi_2^A \otimes \phi_2^B\big) \tag{5.28}$$

is entangled.

5.2.4. Example: Two-qubit system

As an example, we consider a bipartite system where the parts are simple qubits. The Hilbert space of two qubits A and B is $\mathbb{C}^2 \otimes \mathbb{C}^2$. In each one-qubit subspace, we choose a basis $\{\psi_+, \psi_-\}$. The basis states could be realized as spin-up and spin-down eigenstates with respect to some common reference direction (for example, the z-direction in some inertial frame used in the description of both particles). For the two-qubit system, we obtain the following basis of product states,

$$\{\, \psi_+ \otimes \psi_+ \,,\, \psi_+ \otimes \psi_- \,,\, \psi_- \otimes \psi_+ \,,\, \psi_- \otimes \psi_- \,\}. \tag{5.29}$$

These four basis vectors span a four-dimensional complex vector space. This shows that the tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2$ is isomorphic to $\mathbb{C}^4$.

In the Hilbert space of the two-qubit system, we can choose another basis that consists entirely of entangled states. Define

$$\begin{aligned} \psi_{\mathrm{e}}^{\pm} &= \frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_+ \pm \psi_- \otimes \psi_-\big), \\ \psi_{\mathrm{o}}^{\pm} &= \frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_- \pm \psi_- \otimes \psi_+\big). \end{aligned} \tag{5.30}$$

The states $\psi_{\mathrm{e}}^{\pm}$ are superpositions of product states in which the spins are parallel. They are sometimes called *even parity states*. The states $\psi_{\mathrm{o}}^{\pm}$, where the two summands have antiparallel spins, are called the *odd parity states*. The basis formed by these states is called the *Bell basis*.

The experimental realizability of entangled states is of particular importance for possible applications of quantum computers. A number of methods for creating two-qubit systems in an entangled state are described in [**2**].

EXERCISE 5.6. *Show that the Bell basis*

$$\{\psi_{\mathrm{e}}^{+}, \psi_{\mathrm{e}}^{-}, \psi_{\mathrm{o}}^{+}, \psi_{\mathrm{o}}^{-}\} \tag{5.31}$$

is an orthonormal basis in the two-qubit Hilbert space.



CD 5.7 explores several methods to visualize two-qubit states: (a) As a bar diagram that shows the four coefficients of the expansion with respect to the basis (5.29). This could be called the "standard representation" of the two-qubit states. (b) In case of a separable state, we can show the states of the subsystem in a Bloch sphere as described in Section 4.4.1. (c) We can visualize qubit states by the first two eigenstates of a harmonic oscillator as described in Section 4.5.3. Two-qubit states may thus be visualized as states of the harmonic oscillator in two dimensions.

5.2.5. Example: Particle with spin

Another example is the combination of a particle in $\mathbb{R}^3$ whose state space is $L^2(\mathbb{R}^3)$ and of a qubit whose state space is $\mathbb{C}^2$. The result is simply a particle with spin in $\mathbb{R}^3$. The tensor product $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ is spanned by the linear combination of the product states

$$\psi(\mathbf{x}) \otimes \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad \psi \in L^2(\mathbb{R}^3), \quad c_i \in \mathbb{C}, \tag{5.32}$$

We can omit the $\otimes$ in (5.32), because it makes sense to multiply a function with a vector. Thus, we identify

$$\psi(\mathbf{x}) \otimes \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \equiv \psi(\mathbf{x}) \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1\, \psi(\mathbf{x}) \\ c_2\, \psi(\mathbf{x}) \end{pmatrix}. \tag{5.33}$$

The Hilbert space $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ therefore consists of two-component wave functions. Every two-component wave function is a linear combination of two vectors of the type (5.33),

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \psi_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.34}$$

It is interesting that this state is entangled whenever ψ_1 and ψ_2 are linearly independent.

In Section 3.5.1, we introduced the Hilbert space $L^2(\mathbb{R}^3)^2$ of spinor-wave functions as a direct sum of two copies of $L^2(\mathbb{R}^3)$. Obviously, this Hilbert space is the same as (isomorphic to) the tensor product of $L^2(\mathbb{R}^3)$ and $\mathbb{C}^2$:

$$L^2(\mathbb{R}^3)^2 = L^2(\mathbb{R}^3) \oplus L^2(\mathbb{R}^3) \cong L^2(\mathbb{R}^3) \otimes \mathbb{C}^2. \tag{5.35}$$

In certain experimental setups, the spatial states of photons also realize a qubit (see Section 4.5.2). We can combine these states with the polarization states into a two-qubit system. For example, a single photon that moves upwards and is vertically polarized is a two-qubit system in the state

$$\psi_{\mathrm{uv}} = \psi_{\mathrm{u}} \otimes \psi_{\mathrm{v}}. \tag{5.36}$$

In a similar way, we may define the other basis states $\psi_{\rm uh}$, $\psi_{\rm dv}$, and $\psi_{\rm uh}$. A Bell state of the photon is, for example, $\psi_{\rm e}^{+} = 2^{(-1/2)}\,(\psi_{\rm uv} + \psi_{\rm dh})$.

5.3. Interacting Particles

5.3.1. Two-particle interactions

It will turn out that entangled states have some strange properties. Therefore, one must ask whether entangled states really exist in nature. For systems of non-interacting particles, we can avoid entangled states by choosing appropriate initial conditions. A state that is initially separable will be separable for all times. But, for interacting particles this is no longer true, and entanglement cannot be avoided.

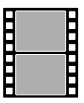
The Hamiltonian for an interacting two-particle systems is again obtained from its analog in classical mechanics. Thus, it could be of the form (assuming $\hbar = 1$)

$$H = -\frac{1}{2\mathrm{m}_1}\,\Delta^{(1)} - \frac{1}{2\mathrm{m}_2}\,\Delta^{(2)} + V(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \tag{5.37}$$

where again $\Delta^{(j)}$ is the Laplace acting on the coordinates of particle (j) (see Eq. (5.2)). Consider, for example, a helium atom. In the approximation of infinite nuclear mass, the atomic nucleus is a fixed, point-like center of a Coulomb force (with charge $2e$). The helium atom has two electrons that interact by Coulomb repulsion. The electrostatic potential energy of the two electrons can thus be described by the expression

$$V(\mathbf{x}_1, \mathbf{x}_2) = -\frac{2\gamma}{|\mathbf{x}^{(1)}|} - \frac{2\gamma}{|\mathbf{x}^{(2)}|} + \frac{\gamma}{|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}|}, \tag{5.38}$$

where $\gamma = e^2/4\pi\epsilon_0$, as usual. The first two summands describe the electrostatic potential energy of the electrons in the field of the nucleus, and the last term is the Coulomb repulsion between the two electrons. The presence of the last term makes it impossible to obtain a solution of the Schrödinger equation in form of a product of one-particle wave functions (that is, in form of a separable state). Even if the initial state is separable, the time evolution will put the system into an entangled state.



CD 5.5 shows the time evolution of a two-particle wave function in configuration space. It describes two particles moving in one dimension and interacting via a harmonic oscillator force. Although the initial state is separable, it soon becomes obvious that the state at time $t > 0$ is entangled.

5.3.2. Separation of the center-of-mass motion

A special case of (5.37) is a Hamiltonian of the form

$$H = -\frac{1}{2\mathrm{m}_1}\,\Delta^{(1)} - \frac{1}{2\mathrm{m}_2}\,\Delta^{(2)} + V(\mathbf{x}^{(1)} - \mathbf{x}^{(2)}). \tag{5.39}$$

Here, the potential energy depends only on the relative position of the two particles. This Hamiltonian would arise, for example, in a description of a hydrogen atom, where the proton is not just treated as a fixed force center, but as a quantum-mechanical particle. The hydrogen atom consists of a proton with mass m_p and an electron with mass m_e. These particles interact by an attractive Coulomb force that depends only on the distance between the proton and the electron. Hence, the Hamiltonian of the hydrogen atom becomes

$$H = -\frac{1}{2\mathrm{m}_p}\,\Delta^{(1)} - \frac{1}{2\mathrm{m}_e}\,\Delta^{(2)} - \frac{\gamma}{|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}|}. \tag{5.40}$$

Again, the time evolution does not preserve the separability of wave functions. But we can find another set of coordinates, where a separation into a product is still possible.

Starting with the Cartesian coordinates $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ of the two-particle system, we introduce *center-of-mass coordinates* $\mathbf{X}$ and *relative coordinates* $\mathbf{x}$ by

$$\mathbf{X} = \frac{\mathrm{m}_1\mathbf{x}^{(1)} + \mathrm{m}_2\mathbf{x}^{(2)}}{\mathrm{m}_1 + \mathrm{m}_2}, \qquad \mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1. \tag{5.41}$$

After the coordinate transformation $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \to (\mathbf{X}, \mathbf{x})$, the Hamiltonian (5.39) decomposes into a sum of a part that depends only on $\mathbf{X}$ and another part that depends only on $\mathbf{x}$,

$$H = -\frac{1}{2M}\,\Delta_{\mathbf{X}} - \frac{1}{2\mu}\,\Delta_{\mathbf{x}} + V(\mathbf{x}), \tag{5.42}$$

where $M = \mathrm{m}_1 + \mathrm{m}_2$ is the total mass and μ is the reduced mass

$$\mu = \frac{\mathrm{m}_1\mathrm{m}_2}{\mathrm{m}_1 + \mathrm{m}_2}. \tag{5.43}$$

The symbols $\Delta_{\mathbf{X}}$ and $\Delta_{\mathbf{x}}$ denote the Laplace operators with respect to the indicated coordinates. Thus, we found a new way to decompose the two-particle system into two subsystems. One of these subsystems describes the free motion of the center of mass. It is given by the Hamiltonian

$$H_{\mathrm{CM}} = -\frac{1}{2M}\,\Delta_{\mathbf{X}}. \tag{5.44}$$

The other subsystem describes the relative motion. This subsystem is described by the Hamiltonian

$$H_{\mathrm{REL}} = -\frac{1}{2\mu}\,\Delta_{\mathbf{x}} + V(\mathbf{x}), \tag{5.45}$$

which is the same as the Hamiltonian for a particle with mass μ in an external potential energy $V(\mathbf{x})$. The Hamiltonian of the two-particle system is a sum

$$H = H_{\mathrm{CM}} + H_{\mathrm{REL}}. \tag{5.46}$$

The two summands commute with each other, hence the time evolution of the combined system is

$$\mathrm{e}^{-\mathrm{i}Ht} = \mathrm{e}^{-\mathrm{i}H_{\mathrm{CM}}t}\,\mathrm{e}^{-\mathrm{i}H_{\mathrm{REL}}t}. \tag{5.47}$$

We see that the two subsystems have been chosen in such a way that a separable initial state remains separable for all times.

Thus, we can solve the two-particle Schrödinger equation with a product

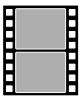
$$\psi(\mathbf{X},\mathbf{x},t) = \Psi(\mathbf{X},t)\,\psi(\mathbf{x},t), \tag{5.48}$$

where Ψ describes a free motion and ψ describes the relative motion of the two particles,

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(\mathbf{x},t) = H_{\mathrm{REL}}\,\psi(\mathbf{x},t). \tag{5.49}$$

The linearity of the Schrödinger equation implies that arbitrary linear combinations of the solutions (5.48) are again solutions.

It is clear that the bound-state energies are given by the eigenvalues of the Hamiltonian describing the relative motion.



The movies CD 5.4 and CD 5.6 show states of the two-particle oscillator and indicate the coordinate axes of the center-of-mass and relative coordinates. A blue line (the "$\mathbf{x}$-axis") shows those configurations of the two-particle system, for which the center of mass is at the origin ($\mathbf{X} = 0$). The green line (the "$\mathbf{X}$-axis") is the coordinate axis with relative coordinates $\mathbf{x} = 0$. The wave packet in configuration space is always symmetric about the blue line. It oscillates in the $\mathbf{x}$-direction and spreads in the $\mathbf{X}$-direction. The corresponding classical system (represented as a white point at the center of the wave packet) always moves on the blue line.

5.4. Observables of a Bipartite System

5.4.1. Tensor product of operators

Given a linear operator S in the Hilbert space $\mathfrak{H}^A$ and another linear operator T in $\mathfrak{H}^B$, we can define a linear operator in the Hilbert space of the compound system. This operator will be called the tensor product of S and T and

denoted by $S \otimes T$. Whenever S and T correspond to observables of the subsystems (that is, self-adjoint operators), the tensor product $S \otimes T$ is an observable of the compound system. The measurable values of $S \otimes T$ are the products of the eigenvalues of S and T.

Let us describe briefly the construction of $S \otimes T$. For simplicity, we consider the case where S and T are bounded (continuous) linear operators. First, we define the tensor product on separable states as

$$(S \otimes T)\, \psi^A \otimes \psi^B = S\psi^A \otimes T\psi^B. \tag{5.50}$$

We extend this definition by linearity to the set of all finite superpositions of separable states:

$$\begin{aligned}(S \otimes T) \sum_{j,k} c_{jk}\, \psi_j^A \otimes \psi_k^B &= \sum_{j,k} c_{jk}\, (S \otimes T)\, \psi_j^A \otimes \psi_k^B \\ &= \sum_{j,k} c_{jk}\, S\psi_j^A \otimes T\psi_k^B. \end{aligned} \tag{5.51}$$

The result is a bounded and densely defined operator that can be extended (by continuity) to all of $\mathfrak{H}^A \otimes \mathfrak{H}^B$ (see Book One, Section 2.5). We have

$$\|S \otimes T\| = \|S\|\, \|T\|. \tag{5.52}$$

A separable state $\psi^A \otimes \psi^B$ consisting of eigenvectors,

$$S\psi^A = s\psi^A, \quad T\psi^B = t\psi^B, \tag{5.53}$$

is an eigenvector of the product observable,

$$(S \otimes T)\, \psi^A \otimes \psi^B = st\, \psi^A \otimes \psi^B, \tag{5.54}$$

and the eigenvalue is the product of the individual eigenvalues. This follows immediately from the definition (5.50). Moreover, *all* eigenvalues of the tensor product can be written as products of the eigenvalues of the factors.

Eigenvalues of the tensor product:

Suppose that S and T are operators with a discrete spectrum of eigenvalues. Then, the eigenvalues of $S \otimes T$ are given by the products st, where s is an eigenvalue of S and t is an eigenvalue of T.

Ψ For unbounded operators (like position and momentum operators) domain questions have to be taken into account. Equation (5.50) makes sense only for $\psi^A \in \mathfrak{D}(S)$ and $\psi^B \in \mathfrak{D}(T)$. Usually, the domains $\mathfrak{D}(S)$ and $\mathfrak{D}(T)$ are dense sets, and hence the set of finite linear combinations of separable states $\psi^A \otimes \psi^B$ is dense in $\mathfrak{H}^A \otimes \mathfrak{H}^B$. Therefore, Eq. (5.51) gives us linear operator on a dense domain in $\mathfrak{H}^A \otimes \mathfrak{H}^B$. If possible, one takes the closure of this densely defined operator to complete the definition of the

tensor product $S \otimes T$. In that way, the tensor product of self-adjoint operators becomes self-adjoint. We refer to the literature for mathematical details concerning this procedure. According to a general mathematical theorem, the spectrum of the tensor product is the (closure of the) product of the individual spectra. (This statement includes the continuous spectra of the operators.)

EXERCISE 5.7. *Show that the composition of* (*bounded*) *tensor-product operators is given by*

$$(S_1 \otimes T_1)(S_2 \otimes T_2) = (S_1 S_2) \otimes (T_1 T_2). \tag{5.55}$$

EXERCISE 5.8. *Given unitary groups* $V(t)$ *and* $W(t)$*, show that the tensor product* $U(t) = V(t) \otimes W(t)$ *is also a unitary group, that is,* $U(0) = \mathbf{1}$*,* $U(s)U(t) = U(s+t)$*.*

There are, of course, many observables of the bipartite system that are not just tensor-products of subsystem operators. An example is the potential energy $V(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ describing the interaction of a two-particle system in (5.37), except if $V(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = V_1(\mathbf{x}^{(1)})\, V_2(\mathbf{x}^{(2)})$. Another interesting example is the *Kronecker sum*. It combines two subsystem operators S and T into the operator $S \otimes \mathbf{1} + \mathbf{1} \otimes T$. The generator of a tensor product of unitary groups is the Kronecker sum of the generators,

$$\mathrm{e}^{-\mathrm{i}tS} \otimes \mathrm{e}^{-\mathrm{i}tT} = \mathrm{e}^{-\mathrm{i}t\,(S \otimes \mathbf{1} + \mathbf{1} \otimes T)}. \tag{5.56}$$

This can be seen as follows. If $U(t)$ is a tensor product of unitary groups,

$$U(t) = V(t) \otimes W(t) = \mathrm{e}^{-\mathrm{i}tS} \otimes \mathrm{e}^{-\mathrm{i}tT}, \tag{5.57}$$

then the generator of $U(t)$ is obtained, as usual, by differentiating $U(t)$ at $t = 0$. The product rule also applies to the tensor product,

$$\mathrm{i}\,\frac{d}{dt}\,U(t) = \Big(\mathrm{i}\,\frac{d}{dt}\,V(t)\Big) \otimes W(t) + V(t) \otimes \Big(\mathrm{i}\,\frac{d}{dt}\,W(t)\Big), \tag{5.58}$$

and we find

$$\mathrm{i}\,\frac{d}{dt}\,U(t)\Big|_{t=0} = S \otimes \mathbf{1} + \mathbf{1} \otimes T. \tag{5.59}$$

5.4.2. Local manipulations

Sometimes one is interested in measuring an observable defined only for one of the subsystems. A measurement performed on subsystem A without perturbation of B is called a *local measurement* of the subsystem A. This is possible only if the two systems can be isolated from each other, which is certainly easier if, as it is sometimes the case, they are already separated by some spatial distance.

In the Hilbert space of a composite system, an observable S of the subsystem A is represented by the operator

$$S \otimes \mathbf{1}. \tag{5.60}$$

This operator acts in a nontrivial way only on the states of system A. It has the same eigenvalues as the operator S, and whenever S is self-adjoint, so is $S \otimes \mathbf{1}$. Similar considerations apply, of course, to the observables $\mathbf{1} \otimes T$ of the subsystem B.

Subsystem observables:

Observables of the type $S \otimes \mathbf{1}$ or $\mathbf{1} \otimes T$ are called *local observables* or *subsystem observables*. Subsystem observables belonging to different subsystems commute with each other,

$$[S \otimes \mathbf{1}\,,\, \mathbf{1} \otimes T] = 0. \tag{5.61}$$

The unitary group generated by $S \otimes \mathbf{1}$ is given by

$$\mathrm{e}^{-\mathrm{i}t(S\otimes\mathbf{1})} = \mathrm{e}^{-\mathrm{i}tS} \otimes \mathbf{1}, \tag{5.62}$$

and similarly for $\mathbf{1} \otimes T$.

Equation (5.61) follows immediately from (5.55). Equation (5.62) follows from (5.56) by setting $T = 0$. It implies that the corresponding unitary groups also commute,

$$[\mathrm{e}^{-\mathrm{i}t(S\otimes\mathbf{1})}\,,\, \mathrm{e}^{-\mathrm{i}t(\mathbf{1}\otimes T)}] = 0. \tag{5.63}$$

A unitary transformation of the form $V \otimes \mathbf{1}$, where V is unitary in the Hilbert space of subsystem A, is called a *subsystem transformation*. A subsystem transformation can be performed by applying V to subsystem A (that is, to the vectors in Hilbert space $\mathfrak{H}^A$) and by doing nothing with subsystem B.

Then, the tensor product of unitary transformations can be written as a product of subsystem transformations,

$$V \otimes W = (V \otimes \mathbf{1})\,(\mathbf{1} \otimes W). \tag{5.64}$$

Unitary transformations of this type are called *local unitary transformations*. Physically, a local unitary transformation can be performed while the subsystems are isolated from each other.

A *property* P of subsystem A is described by an orthogonal projection operator. Remember that an orthogonal projection operator is a bounded self-adjoint operator P with $P^2 = P$. Its only eigenvalues are 0 and 1. The set of states having the property P is the range of P (for a projection operator, $\mathrm{Ran}\, P =$ the eigenspace belonging to the eigenvalue 1). It turns out that

$P \otimes \mathbf{1}$ is an orthogonal projection operator whenever P is (see Exercise 5.9). We conclude that properties of a subsystem are also properties of the compound system. The observable $P \otimes \mathbf{1}$ measures whether the subsystem A has the property P without considering the subsystem B.

As an example, let us consider a two-qubit system. A local measurement of σ_3 on qubit A is a measurement of $\sigma_3 \otimes \mathbf{1}$. If the result is "spin up," then qubit A is in the state ψ_+ afterwards. According to the projection postulate, we have performed the projection $P_+ \otimes \mathbf{1}$, where $P_+ = |\psi_+\rangle\langle\psi_+|$ is an orthogonal projection operator in the Hilbert space of qubit A.

EXERCISE 5.9. *Show that $P \otimes \mathbf{1}$ is an orthogonal projection operator onto the subspace* $\operatorname{Ran} P \otimes \mathfrak{H}^B$ *in* $\mathfrak{H}^A \otimes \mathfrak{H}^B$ *whenever P is an orthogonal projection operator onto* $\operatorname{Ran} P$ *in* $\mathfrak{H}^A$.

EXERCISE 5.10. *Compute the action of the projection operator*

$$P = |\psi_+\rangle\langle\psi_+| \otimes \mathbf{1} \tag{5.65}$$

on the state

$$\frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_+ + \psi_- \otimes \psi_-\big). \tag{5.66}$$

What is the state of qubit B after having found spin-up for qubit A?

EXERCISE 5.11. *Show that a separable state remains separable under a local unitary transformation. Similarly, an entangled state cannot be turned into a separable state by means of a local unitary transformation.*

5.5. The Density Operator

If the bipartite system is in a separable state $\psi^A \otimes \psi^B$, then we can say that the subsystem A is in the state ψ^A. But how do we characterize the subsystem if the composite system is in an entangled state?

5.5.1. What does entanglement mean for the subsystems?

Let us consider a two-qubit system in the Bell state

$$\psi_{\mathrm{e}}^+ = \frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_+ + \psi_- \otimes \psi_-\big). \tag{5.67}$$

What do we know about the state of the qubits A and B? What information about the individual qubits is contained in the entangled state of the compound system?

The state of ψ_{e}^+ is a superposition of two orthogonal product states. In this state, the probability of finding both qubits with spin up is $1/2$, the same as that of finding both with spin down. Hence, qubit A is in the state ψ_+ or ψ_- with equal probability $1/2$. It is important to understand that

the state of qubit A is not a superposition of the states ψ_+ and ψ_-, but a statistical mixture.

Statistical mixture:

A *statistical mixture* of two quantum states ψ_1 and ψ_2 is an ensemble of systems, where each individual system is prepared with probability p in the state ψ_1 and with probability $1-p$ in the state ψ_2. Here, the states ψ_1 and ψ_2 need not be orthogonal.

Let us now consider three different ensembles of qubits. The ensembles are distinguished by physically different preparation procedures:

(I) An ensemble of qubits that are part of a two-qubit system in a Bell state.

(II) An ensemble of qubits that are prepared with equal probability 1/2 either in the state ψ_+ or in the state ψ_- (a statistical mixture).

(III) An ensemble of qubits in the state $\psi_+(\mathbf{e}_x)$. Here, $\mathbf{e}_x$ is the unit vector in the positive x-direction, that is, all qubits are prepared with spin-up in that direction (see (4.19)).

The ensemble (II) is prepared by randomly choosing a preparation procedure for each qubit (that is, without the help of an entangled two-qubit system). This is called a *statistical mixture* of the states ψ_+ and ψ_-. Apart from an obviously different preparation procedure, the two ensembles (I) and (II) behave in an identical way.

There is no statistical test that can distinguish between the ensembles (I) and (II).

The ensemble (III) is a superposition of the states ψ_+ and ψ_-:

$$\psi_+(\mathbf{e}_x) = \frac{1}{\sqrt{2}}\,(\psi_+ + \psi_-). \tag{5.68}$$

This also describes a situation where the probability of finding spin-up or -down in the z direction is 1/2. Hence, all three ensembles cannot be distinguished by measurements of σ_3. But the superposition differs drastically from the statistical mixture with respect to measurements of σ_1. A measurement of σ_1 in the state $\psi_+(\mathbf{e}_x)$ gives the eigenvalue $+1$ with certainty (and hence the expectation value of σ_1 is $+1$). In the statistical mixture, we find ψ_+ and ψ_- with equal probability, and in both cases there is just a 50% chance to find spin-up in the x-direction. Hence, a measurement of σ_1 produces $+1$ and -1 with equal probability (the expectation value of σ_1 is 0).

The expectation values of the components of the spin are

$$\langle\sigma_1\rangle = \langle\sigma_2\rangle = \langle\sigma_3\rangle = 0 \qquad \text{for ensemble (II)}, \tag{5.69}$$

$$\langle\sigma_1\rangle = 1, \quad \langle\sigma_2\rangle = \langle\sigma_3\rangle = 0 \qquad \text{for ensemble (III)}. \tag{5.70}$$

EXERCISE 5.12. *For a two-qubit system in the state* ψ_{o}^{-}*, show that*

$$\langle\psi_{\text{o}}^{-}\,, (\sigma_k \otimes \mathbf{1})\,\psi_{\text{o}}^{-}\rangle = 0, \quad k = 1, 2, 3. \tag{5.71}$$

Hence, the expectation value of the spin of qubit A vanishes along any axis. Show that the same result holds for the other Bell states. This proves (5.69) *for ensemble* (I)*.*

Quite generally, given an ensemble of qubit pairs AB in an entangled state, the ensemble of qubits A is not described by a state vector in the usual sense, but rather by a statistical mixture of state vectors. A description of subsystems in terms of state vectors is no longer adequate, and the theory has to be generalized to accommodate for mixed states. This will be done in the next sections.

An entangled state describes much more than just a statistical mixture of subsystem states. This additional information concerns the correlations between the subsystems. Chapter 6 is, among other things, devoted to a deeper look into these questions.

5.5.2. Expectation values of subsystem observables

In order to develop a theory describing the states of a subsystem, we want to figure out how to extract information about subsystems from a given state of the compound system. To that purpose, we consider an observable $S \otimes \mathbf{1}$ of the subsystem A. For simplicity, we assume that S is a bounded operator in the Hilbert space of system A. Let us compute the expectation value for this observable. We need some orthonormal bases of vectors ψ_i^A in $\mathfrak{H}^A$ and ψ_j^B in $\mathfrak{H}^B$. The product states $\psi_i^A \otimes \psi_j^B$ form an orthonormal basis in $\mathfrak{H}^A \otimes \mathfrak{H}^B$. Hence, the state ψ of the compound system can be written as a linear combination of these product states:

$$\psi = \sum_{i,j} c_{ij}\, \psi_i^A \otimes \psi_j^B. \tag{5.72}$$

Before we compute probabilities, the vector ψ has to be normalized. Hence, we assume

$$\|\psi\|^2 = \sum_{i,j} |c_{ij}|^2 = 1. \tag{5.73}$$

The expectation value of $S \otimes \mathbf{1}$ in the state ψ is obtained by a little calculation:

$$\begin{aligned}
&\langle \psi, S \otimes \mathbf{1}\, \psi \rangle \\
&\quad = \Big\langle \sum_{i,j} c_{ij}\, \psi_i^A \otimes \psi_j^B \,,\, \sum_{k,l} c_{kl}\, S\psi_k^A \otimes \psi_l^B \Big\rangle && \text{by (5.51) and (5.72)} \\
&\quad = \sum_{i,j,k,l} \overline{c_{ij}}\, c_{kl}\, \langle \psi_i^A \otimes \psi_j^B \,,\, S\psi_k^A \otimes \psi_l^B \rangle && \text{(anti-) linearity of } \langle \cdot, \cdot \rangle \\
&\quad = \sum_{i,j,k,l} \overline{c_{ij}}\, c_{kl}\, \langle \psi_i^A, S\, \psi_k^A \rangle\, \langle \psi_j^B, \psi_l^B \rangle && \text{by (5.17)} \\
&\quad = \sum_{i,j,k,l} \overline{c_{ij}}\, c_{kl}\, \langle \psi_i^A, S\, \psi_k^A \rangle\, \delta_{jl} && \text{orthonormality of } \{\psi_j^B\}.
\end{aligned}$$

Finally, using the definition (1.12) of the Kronecker symbol δ_{ij}, we obtain the result

$$\langle \psi, S \otimes \mathbf{1}\, \psi \rangle = \sum_{i,k,l} \overline{c_{il}}\, c_{kl}\, \langle \psi_i^A, S\, \psi_k^A \rangle. \tag{5.74}$$

This involves the "matrix elements" of S between many different states of the subsystem. Usually, the expectation value of an observable S in a state ψ is just $\langle \psi, S\psi \rangle$. Equation (5.74) shows that the expectation value of a subsystem observable is much more complicated. This suggests that for a given state of the compound system, the state of the subsystem cannot be described just by a single vector in the Hilbert space of that subsystem.

EXERCISE 5.13. *Show that for $\psi = \psi^A \otimes \psi^B$, the expectation value of the operator $S \otimes \mathbf{1}$ becomes*

$$\langle S \otimes \mathbf{1} \rangle_\psi = \langle \psi, S \otimes \mathbf{1} \psi \rangle = \langle \psi^A, S\, \psi^A \rangle = \langle S \rangle_{\psi^A}. \tag{5.75}$$

Next, we are going to verify that the result (5.74) can also be written in another way. For this we need to define the trace of an operator.

5.5.3. Trace-class operators

DEFINITION 5.1. Assume that for a linear operator S, the series

$$\operatorname{Tr} S = \sum_j \langle \psi_j,\, S\, \psi_j \rangle \tag{5.76}$$

converges and has the same value in any orthonormal basis $\{\psi_j\}$ of the Hilbert space. Then, $\operatorname{Tr} S$ is called the *trace* of S and the operator S is said to be of the *trace class*.

Ψ The value of the series (5.76) is automatically independent of the chosen orthonormal basis whenever the Hilbert space is finite dimensional, or whenever the operator S is non-negative (meaning $\langle \psi, S\psi \rangle \geq 0$ for all ψ). This is not true in the general case, and therefore we need the assumption in the definition above. Moreover, we note that a trace-class operator is always bounded, but not every bounded operator is of the trace class. For a non-negative trace-class operator we have $\|S\| \leq \operatorname{Tr} S$.

The following definition describes an important subclass of the trace class.

DEFINITION 5.2. A bounded (hence everywhere defined) linear operator ρ is called a *density operator* if it has the following properties:

(1) ρ is non-negative: $\langle \psi, \rho\, \psi \rangle \geq 0$ for all ψ.
(2) ρ is of the trace class and $\operatorname{Tr} \rho = 1$.

As the main characterization of trace-class operators, we quote the following mathematical result.

Canonical form of trace class operators:

A self-adjoint linear operator S in a Hilbert space $\mathfrak{H}$ is of the trace class if and only if there is an orthonormal basis $\{\psi_j\}$ in $\mathfrak{H}$ such that

$$S = \sum_j \lambda_j\, \psi_j\, \langle \psi_j, \cdot \rangle = \sum_j \lambda_j\, |\psi_j\rangle\langle\psi_j| \tag{5.77}$$

with real numbers λ_j satisfying $\sum_j |\lambda_j| < \infty$. The last expression in (5.77) uses Dirac's bra-ket notation (see Appendix A.8).
The trace of S is then given by

$$\operatorname{Tr} S = \sum_j \lambda_j. \tag{5.78}$$

An operator with the representation (5.77) has the eigenvalues λ_j and the corresponding eigenvectors ψ_j. Each nonzero eigenvalue has at most a finite degree of degeneracy.

This result states that a trace-class operator is a sum of one-dimensional orthogonal projection operators. If ρ is a density operator, then its eigenvalues satisfy

$$\lambda_j \geq 0, \qquad \sum_j \lambda_j = 1. \tag{5.79}$$

EXERCISE 5.14. *Show that if ρ_1 and ρ_2 are density operators, then for any λ with $0 \leq \lambda \leq 1$, the operator*

$$\rho(\lambda) = \lambda\,\rho_1 + (1-\lambda)\,\rho_2 \tag{5.80}$$

is also a density operator.

5.5.4. Density operator of a subsystem

We return to our task of rewriting (5.74). Given a normalized state of the compound system

$$\psi = \sum_{i,j} c_{ij}\,\psi_i^A \otimes \psi_j^B \tag{5.81}$$

(where $\{\psi_i^A\}$ and $\{\psi_j^B\}$ are orthonormal bases), we define the operator

$$\rho^A = \sum_{i,k,l} \overline{c_{il}}\,c_{kl}\,|\psi_k^A\rangle\langle\psi_i^A|. \tag{5.82}$$

It can be shown that this operator has the following properties:

(1) ρ^A is bounded, self-adjoint, and non-negative.
(2) ρ^A is trace class with $\operatorname{Tr}\rho^A = 1$.

These results follow by a direct computation that is left as an exercise. Hence, the operator ρ^A is a density operator. It has only non-negative eigenvalues that sum to one. Because ρ^A is trace class and S is bounded, it follows from the mathematical theory that $S\rho^A$ also belongs to the trace class. Let us compute the trace

$$\begin{aligned}
\operatorname{Tr} S\,\rho^A &= \sum_j \langle\psi_j^A, S\,\rho^A\,\psi_j^A\rangle && \text{definition of trace}\\
&= \sum_{j,i,k,l} \overline{c_{il}}\,c_{kl}\,\langle\psi_j^A, S\,\psi_k^A\rangle\langle\psi_i^A, \psi_j^A\rangle && \text{by (5.82)}\\
&= \sum_{j,i,k,l} \overline{c_{il}}\,c_{kl}\,\langle\psi_j^A, S\,\psi_k^A\rangle\,\delta_{ij} && \text{orthonormality of } \{\psi_j^A\}\\
&= \sum_{i,k,l} \overline{c_{il}}\,c_{kl}\,\langle\psi_i^A, S\,\psi_k^A\rangle. && (5.83)
\end{aligned}$$

If we compare this result with Eq. (5.74), we see that $\operatorname{Tr} S\,\rho^A$ is just another way of writing the expectation value of $S \otimes \mathbf{1}$ in the state ψ.

Expectation value for a subsystem:

Let ψ be a normalized state of a bipartite system in $\mathfrak{H}^A \otimes \mathfrak{H}^B$. Let c_{ij} be the expansion coefficients of ψ with respect to an orthonormal basis $\{\psi_i^A \otimes \psi_j^B\}$. Then, the operator

$$\rho^A = \sum_{i,k,l} \overline{c_{il}}\, c_{kl}\, |\psi_k^A\rangle\langle\psi_i^A| \tag{5.84}$$

is a density operator. If S is a bounded operator defined on $\mathfrak{H}^A$, then

$$\langle S \otimes \mathbf{1}\rangle_\psi = \langle \psi\,,\, S \otimes \mathbf{1}\, \psi\rangle = \operatorname{Tr} S\rho^A. \tag{5.85}$$

A similar result holds for observables of the system B.

EXERCISE 5.15. *Assuming that $\mathfrak{H}^A$ and $\mathfrak{H}^B$ have finite dimension, show that the operator ρ^A defined in* (5.84) *is nonnegative, that is, $\langle \phi, \rho^A \phi\rangle \geq 0$ for all $\phi \in \mathfrak{H}^A$. Moreover, show that $\operatorname{Tr} \rho^A = \|\psi\|^2$, where ψ is the corresponding state of the compound system, defined as in* (5.81).

EXERCISE 5.16. *By the same reasoning as above, show that the expectation value of a subsystem observable $1 \otimes S$ of the system B in the state* (5.81) *is given by*

$$\operatorname{Tr} S\rho^B = \sum_k \langle \psi_k^B, S\, \rho^B\, \psi_k^B\rangle \tag{5.86}$$

with

$$\rho^B = \sum_{i,j,l} \overline{c_{ij}}\, c_{il}\, |\psi_l^B\rangle\langle\psi_j^B|. \tag{5.87}$$

5.6. Pure and Mixed States

5.6.1. State of a subsystem

All that can be known about a quantum system is contained in the expectation values of the observables of the system. For a subsystem, we can compute these expectation values according to (5.85) as soon as we know the density operator. Thus, the density operator describes the state of the subsystem.

The state of a subsystem:

If a bipartite system is in a state ψ, then the *state of the subsystem A* is characterized by the density operator ρ^A defined in (5.84), and the state of the subsystem B is given by a similar expression ρ^B.

Earlier in this book, we characterized states as one-dimensional subspaces of the Hilbert space (Book One, Section 4.1). For convenience, we used a normalized vector ψ of the subspace $[\psi] = \{c\psi \mid c \in \mathbb{C}\}$ to represent the state. Equivalently, we may describe $[\psi]$ by the operator $\rho = |\psi\rangle\langle\psi|$, which is the unique orthogonal projection operator with range $[\psi]$. You may check that ρ is a density operator (see Exercise 5.17). A density operator is thus a direct generalization of our earlier concept.

Pure and mixed states:

In general, the state of a quantum system is described by a density operator ρ. A one-dimensional projection operator

$$\rho = |\psi\rangle\langle\psi| \quad \text{with } \|\psi\| = 1 \tag{5.88}$$

is called a *pure state*. It is usually represented by the state vector ψ. Otherwise, ρ is called a *mixed state*.

Let the bipartite system be in a separable state $\psi = \psi^A \otimes \psi^B$. We may assume that ψ^A belongs to some orthonormal basis. Using this basis in (5.84), we see immediately that the density matrix describes a pure state:

$$\rho^A = |\psi^A\rangle\langle\psi^A|. \tag{5.89}$$

Likewise, the expectation value of $S \otimes \mathbf{1}$ is reduced to the familiar expression

$$\operatorname{Tr} S\rho^A = \langle\psi^A, S\psi^A\rangle \tag{5.90}$$

(see also Exercise 5.13).

Whenever a pure state of the compound system is separable, then the subsystem is in a pure state.

A stronger version of this statement is given in Section 5.8.1 below.

EXERCISE 5.17. *Show that any one-dimensional projection operator is a positive self-adjoint trace-class operator with unit trace and hence a density operator.*

5.6.2. Canonical form of the density operator

For a given density operator ρ, we can find an orthonormal basis $\{\psi_j\}$ such that ρ has the canonical form (see Section 5.5.3)

$$\rho = \sum_j p_j \, |\psi_j\rangle\langle\psi_j|. \tag{5.91}$$

Here, the numbers p_j are the eigenvalues of ρ, and the vectors ψ_j are the corresponding eigenvectors. Because ρ is non-negative with $\operatorname{Tr}\rho = 1$, the eigenvalues fulfill the conditions

$$p_j \geq 0 \quad (\text{all } j), \qquad \sum_j p_j = 1. \tag{5.92}$$

The projection operator $P_k = |\psi_k\rangle\langle\psi_k|$ is an observable describing the property of "being in the state ψ_k." The probability that a system in the state ρ is actually found in the state ψ_k is, as usual, given by the expectation value of P_k, that is, by

$$\begin{aligned} \operatorname{Tr} P_k\,\rho &= \sum_l \langle \psi_l, P_k\,\rho\,\psi_l\rangle \\ &= \sum_{l,j} p_j\,\langle\psi_l,\psi_k\rangle\,\langle\psi_k,\psi_j\rangle\,\langle\psi_j,\psi_l\rangle = p_k. \end{aligned} \tag{5.93}$$

Hence, the eigenvalue p_k of ρ is the probability that a system in the mixed state ρ is found to be in the pure state ψ_k. The conditions (5.92) show that the interpretation of the numbers p_j as probabilities is consistent.

The meaning of the density operator:

A mixed state ρ of a physical system is a statistical ensemble of pure states $\{\psi_1, \psi_2, \ldots\}$ (forming an orthonormal set), where each pure state ψ_j occurs with a probability p_j. The probabilities p_j are just the eigenvalues of the density operator ρ, and the pure states ψ_j are the corresponding eigenvectors.

Whenever one of the eigenvalues of ρ is degenerate, then the canonical form of ρ is not unique. Consider, for example, a qubit in a mixed state described by the density matrix

$$\rho = \frac{1}{2}\,|\psi_+\rangle\langle\psi_+| + \frac{1}{2}\,|\psi_-\rangle\langle\psi_-| = \frac{1}{2}\,\mathbf{1}. \tag{5.94}$$

Here, the eigenvalue $p_1 = p_2 = 1/2$ is degenerate with degree 2. But any orthonormal basis $\{\psi_1, \psi_2\}$ in $\mathbb{C}^2$ has the completeness property

$$|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2| = \mathbf{1}, \tag{5.95}$$

and hence the density matrix (5.94) can be represented in a multitude of ways.

5.6.3. Ensemble interpretation of mixed states

Next, we consider an arbitrary set $\{\phi_1, \phi_2, \ldots\}$ of normalized Hilbert space vectors. These vectors need not be orthogonal to each other. It can be shown that the operator

$$\rho = \sum_j q_j \, |\phi_j\rangle\langle\phi_j| \quad \text{with } q_j \geq 0 \text{ for all } j \text{ and } \quad \sum_j q_j = 1 \tag{5.96}$$

is a density operator (see Exercise 5.19). (In general, if the set $\{\phi_j\}$ is not orthonormal, the q_j are not the eigenvalues of ρ.)

The interpretation of the state (5.96) is as follows:

Interpretation of the density operator:

The density operator ρ in (5.96) describes an ensemble of quantum systems that have been prepared in the state ϕ_k with probability q_k.

For an *orthonormal* set $\{\phi_j\}$, the probability q_k of being *prepared* in the state ϕ_k will coincide with the probability p_k of being *found* in the state ϕ_k. But if the states are not orthonormal, the probability p_k is larger than q_k, because a system prepared in the state ϕ_j with $j \neq k$ also has a certain chance to be found in the state ϕ_k. The probability p_k is the expectation value of $P_k = |\phi_k\rangle\langle\phi_k|$ in the state ρ:

$$p_k = \mathrm{Tr}\left(|\phi_k\rangle\langle\phi_k| \, \rho\right) \geq q_k. \tag{5.97}$$

Example 5.1. Consider a Stern-Gerlach filter that serves to prepare a qubit with spin-up in the direction of the inhomogeneous magnetic field. We assume that the filter can be rotated about the x-axis (the direction of the beam). Let us consider three orientations, given by the rotation angles $\alpha = 0$, $\alpha = \pi/3$, and $\alpha = 2\pi/3$. Hence, we measure the spin with respect to three directions given by the unit vectors (shown in Fig. 5.1)

$$\mathbf{n}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{n}_2 = \frac{1}{2} \begin{pmatrix} 0 \\ -\sqrt{3} \\ 1 \end{pmatrix}, \quad \mathbf{n}_3 = \frac{1}{2} \begin{pmatrix} 0 \\ -\sqrt{3} \\ -1 \end{pmatrix}. \tag{5.98}$$

In the standard representation, the spinors with spin up in these directions are $\phi_j = \psi_+(\mathbf{n}_j)$, or

$$\phi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \phi_2 = \frac{1}{2} \begin{pmatrix} \sqrt{3} \\ -\mathrm{i} \end{pmatrix}, \quad \phi_3 = \frac{1}{2} \begin{pmatrix} 1 \\ -\mathrm{i}\sqrt{3} \end{pmatrix}. \tag{5.99}$$

These vectors represent the states that can be prepared by our Stern-Gerlach filter. Assume we throw a dice and if the result is 1 or 2, we prepare a qubit

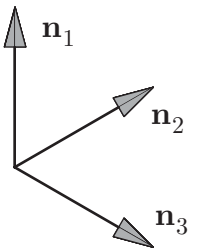


FIGURE 5.1. The three possible orientations of the Stern-Gerlach apparatus in Example 5.1, looking in the direction of the beam.

in the state $\psi_+(\mathbf{n}_1)$ by turning the Stern-Gerlach apparatus (filter) into the upright position. If the result is 3 or 4, we prepare $\psi_+(\mathbf{n}_2)$, and if we get 5 or 6, we prepare $\psi_+(\mathbf{n}_3)$. Hence, we prepare an ensemble of qubits where each of the states (5.99) occurs with probability 1/3. The density operator describing this ensemble is

$$\rho = \sum_{j=1}^{3} \frac{1}{3} \, |\phi_j\rangle\langle\phi_j|. \tag{5.100}$$

In the standard representation we obtain the matrix

$$\rho = \begin{pmatrix} \frac{2}{3} & \frac{\mathrm{i}}{2\sqrt{3}} \\ -\frac{\mathrm{i}}{2\sqrt{3}} & \frac{1}{3} \end{pmatrix}. \tag{5.101}$$

With the explicit form of ρ, it is easy to determine the expectation values of qubit observables. For example, the probability that a measurement of σ_3 on the ensemble gives $+1$ is

$$\mathrm{Tr}\left(|\psi_+\rangle\langle\psi_+| \, \rho \right) = \frac{2}{3}. \tag{5.102}$$

The expectation value of the components of $\boldsymbol{\sigma}$ is $\langle\sigma_j\rangle = \mathrm{Tr}\, \sigma_j\rho$, and a little computation gives

$$\langle\sigma_1\rangle = 0, \quad \langle\sigma_2\rangle = -\frac{1}{\sqrt{3}}, \quad \langle\sigma_3\rangle = \frac{1}{3}. \tag{5.103}$$

EXERCISE 5.18. *Show that* (5.84) *can be written in the form* (5.96) *by setting*

$$q_j = \sum_k |c_{kj}|^2, \quad \phi_j = \frac{1}{\sqrt{q_j}} \sum_k c_{kj} \, \psi_j^A, \tag{5.104}$$

and verify the conditions (5.92).

EXERCISE 5.19. *Show that the operator ρ defined by* (5.96) *is indeed a density operator. Hint: Expand ϕ_j in an orthonormal basis $\{\psi_j^A\}$ and show that* (5.96) *can be written in the form of* (5.84)*. Use the result of Exercise 5.15.*

EXERCISE 5.20. *Repeat the computations of Example 5.1 with the directions for the Stern-Gerlach measurements as specified by the angles $\alpha = 0$, $\alpha = 2\pi/3$, and $\alpha = 4\pi/3$. Determine ρ and compute the expectation value of $\boldsymbol{\sigma} \cdot \mathbf{n}$, where $\mathbf{n}$ is some unit vector in $\mathbb{R}^3$. What is the probability that the qubit is found with spin up with respect to the direction given by $\mathbf{n}$?.*

5.7. Preparation of Mixed States

5.7.1. Preparing an ensemble in a mixed state

The interpretation of the density operator tells us how we can prepare an ensemble of quantum systems in a mixed state ρ if we know how to prepare the pure states ϕ_j in (5.96). We need the help of a (classical) random number generator that produces the number j with probability p_j. Whenever the random number generator gives a number j, we prepare a copy of the system in the state ϕ_j. Repeating this procedure generates the statistical ensemble of pure quantum states described by the density matrix ρ as in (5.96).

Any quantum system can be in a mixed state. We even expect that mixed states are a more realistic description of quantum systems than pure states. Whenever we prepare an ensemble of quantum systems, there are experimental inaccuracies and unavoidable fluctuations that will cause each member of the ensemble to be prepared in a slightly different state, and we are going to end up with a statistical mixture of the form (5.96). In that case, the appearance of statistical mixtures of pure states is related to an imprecise knowledge or inaccuracy of the preparation process.

Suppose we want to prepare an ensemble of hydrogen atoms in the ground state and let them evaporate from a bottle of liquid hydrogen. As this preparation method ignores the electronic spin, we expect that all directions of the spin will occur with equal probability. Hence, we expect a statistical mixture of the states (3.129) with spin up and spin down,

$$\rho = \frac{1}{2} |\psi_{1,0,0,+}\rangle\langle\psi_{1,0,0,+}| + \frac{1}{2} |\psi_{1,0,0,-}\rangle\langle\psi_{1,0,0,-}|, \tag{5.105}$$

rather than a pure state (unless some unknown influence during the preparation prefers a particular state). We see that the mixed state ρ expresses our ignorance about the details of the preparation process. A pure state can only be obtained by a simultaneous preparatory measurement of a complete set of commuting observables. Whenever the experiment leaves us ignorant

of the eigenvalue of at least one of these observables, we have to describe the state by a density operator.

Ignorance or inaccuracy as a reason for the appearance of mixed states is not specific to quantum mechanics. In classical physics, the methods of statistical mechanics have been developed for the very same reason. But the mixed states describing the subsystems of a compound system appear on a more fundamental level. We cannot avoid statistical mixtures by attempting to control the state preparation more carefully (that is by trying to be more accurate and less ignorant). Even if the compound system is prepared in the best possible way (that is, in a pure state), it is inevitable that the subsystem is in a mixed state whenever the compound state is entangled. It has to be stressed, however, that no statistical test can distinguish between an ensemble that is in a mixed state because of an inaccurate preparation procedure and an ensemble that is in a mixed state, because its members are part of a larger quantum system.

As discussed in Section 5.5.1, there is a difference between a mixed state and a superposition. Given an orthonormal set $\{\psi_1, \psi_2, \ldots\}$, consider

$$\rho = \sum_j p_j \, |\psi_j\rangle\langle\psi_j| \qquad \text{versus} \qquad \psi = \sum_j \sqrt{p_j}\, \psi_j. \tag{5.106}$$

To both ρ and ψ, the pure states ψ_j contribute with probability p_j. It is nevertheless possible to find a statistical test that distinguishes between ρ and ψ (for example, by measuring the observable $|\psi\rangle\langle\psi|$). The mathematical difference between ρ and ψ is the following: The pure state ψ contains information about the complex phases of the ψ_j; the mixed state ρ does not. If you replace one of the states ψ_j by $\exp(\mathrm{i}\lambda)\,\psi_j$, this would change ψ but not ρ. The "relative phases" of the ψ_j have observable consequences for the state ψ (interference). The mixed state ρ is totally insensitive against phase differences of its constituents. Sometimes, the pure state ψ is called a *coherent superposition*, and the mixed state ρ is called an *incoherent superposition* of the pure states ψ_j.

EXERCISE 5.21. *Show that $\rho^2 \neq \rho$ for $\rho = p_1 \, |\psi_1\rangle\langle\psi_1| + p_2 \, |\psi_2\rangle\langle\psi_2|$, with $p_i > 0$, $p_1 + p_2 = 1$. Hence, ρ is not a projection operator. Extend the proof to the case where ρ is a mixture of n orthogonal states.*

EXERCISE 5.22. *Show that a density operator ρ describes a pure state if and only if $\rho^2 = \rho$.*

EXERCISE 5.23. *Show that in the Hilbert space of a qubit, any density operator is of the form*

$$\rho = \begin{pmatrix} a & b \\ \overline{b} & 1-a \end{pmatrix} \qquad \textit{with } 0 \leq a \leq 1 \textit{ and suitable } b \in \mathbb{C}. \tag{5.107}$$

EXERCISE 5.24. *Determine the unitary transformation U that diagonalizes the operator ρ in* (5.107),

$$\rho_{\text{diag}} = U^{-1}\rho U = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix} \qquad \text{with } p \geq 1-p, \tag{5.108}$$

and show that the positivity of ρ is equivalent to the condition

$$|b|^2 \leq a - a^2. \tag{5.109}$$

5.7.2. Ambiguity of realizing mixed states

The procedure for preparing a mixed state ρ is not unique. This can be seen already in the simplest case of a qubit. Consider, for example, a qubit in a mixed state described by the density matrix

$$\rho = \frac{1}{2}\,|\psi_+\rangle\langle\psi_+| + \frac{1}{2}\,|\psi_-\rangle\langle\psi_-| = \frac{1}{2}\,\mathbf{1}. \tag{5.110}$$

We can prepare this state by creating an ensemble of qubits, where a randomly chosen 50% of the particles is prepared in the state ψ_+ (spin-up in the z-direction), and the rest is prepared in the state ψ_-. But we could also create an ensemble containing the two eigenstates of, say, σ_1 with equal probability. This would produce the mixed state

$$\frac{1}{2}\,|\psi_+(\mathbf{e}_x)\rangle\langle\psi_+(\mathbf{e}_x)| + \frac{1}{2}\,|\psi_-(\mathbf{e}_x)\rangle\langle\psi_-(\mathbf{e}_x)| = \frac{1}{2}\,\mathbf{1} = \rho. \tag{5.111}$$

It is intuitively clear that the two representations (5.110) and (5.111) are equivalent, because each of the pure states $\psi_\pm(\mathbf{e}_x)$ again has a 50% chance to be in one of the σ_3-eigenstates ψ_+ and ψ_-. In the ensemble described by (5.111), we would find ψ_+ and ψ_- with equal probability, and hence this state could equally well be described by the mixture (5.110). There is no chance to determine experimentally which of the two (physically different) preparation procedures was used to create the given ensemble. Any measurement of any spin component would just produce a random bit.

Ambiguity of the ensemble preparation:

In general, there are many physically different ways to prepare an ensemble of systems in the state ρ. All preparation procedures are completely equivalent with respect to determinative measurements. No statistical test on the ensemble can distinguish between different methods to prepare one and the same state ρ.

Additional information about the preparation procedure, however, can be used to obtain pure states from a mixed state. If, for example, the

preparer makes a record of how the qubits were prepared ("first qubit spin-up, second qubit spin-down, ..."), this information would allow us to choose a subensemble in the pure state ψ_+. This information is therefore physically equivalent to a Stern-Gerlach filter (represented by the projection operator $P_+ = |\psi_+\rangle\langle\psi_+|$), because this filter would select the same subensemble.

EXERCISE 5.25. *Show that the qubit density operator ρ in* (5.107) *can be written as*

$$\rho = (2p-1)|\psi\rangle\langle\psi| + (2-2p)\rho_{\rm m} \quad \textit{with } \psi = U\psi_+ \textit{ and } \rho_{\rm m} = \frac{1}{2}\mathbf{1}. \quad (5.112)$$

The eigenvalue parameter $1/2 \le p \le 1$ and the unitary matrix U have been introduced in Exercise 5.24. For $p > 1/2$, ρ describes a "noisy" qubit state ψ: A measurement of the corresponding component of the spin gives ψ with probability p and the orthogonal state with the probability $1-p$. The case $p = 1$ corresponds to the pure state (ψ is found with certainty), and $p = 1/2$ describes the "maximally mixed" state $\rho = (1/2)\mathbf{1}$.

5.7.3. Example: Mixed qubit states

From (4.6) we know that any density operator ρ in the Hilbert space of a qubit can be written as

$$\rho = \frac{1}{2}\Big(a_0\mathbf{1}_2 + \sum_{k=1}^{3} a_k\sigma_k\Big) \quad (5.113)$$

where, according to (4.9), the coefficients a_k, $k = 1, 2, 3$, are the expectation values of the observables σ_k, that is

$$a_k = \mathrm{Tr}\,(\sigma_k\rho) = \langle\sigma_k\rangle_\rho. \quad (5.114)$$

Moreover, $a_0 = \mathrm{Tr}\,\rho = 1$. The vector $\mathbf{a} = (a_1, a_2, a_3)$ satisfies $0 \le |\mathbf{a}| \le 1$. It can be used to visualize the mixed state ρ.

Visualization of density operators for qubits:

Any qubit density operator ρ corresponds to a unique vector $\mathbf{a} \in \mathbb{R}^3$ with $0 \le |\mathbf{a}| \le 1$ such that $\rho = \rho(\mathbf{a})$ with

$$\rho(\mathbf{a}) = \frac{1}{2}\big(\mathbf{1}_2 + \mathbf{a}\cdot\boldsymbol{\sigma}\big). \quad (5.115)$$

The components of $\mathbf{a}$ describe the expectation values of the components of $\boldsymbol{\sigma}$.

In this connection, the unit sphere $\{\mathbf{a} \in \mathbb{R}^3 \mid |\mathbf{a}| \le 1\}$ is called the *Bloch sphere*. A qubit density operator corresponds to a point inside the Bloch

sphere. The maximally mixed state $\rho = \frac{1}{2}\mathbf{1}$ corresponds to $\mathbf{a} = 0$ (the center of the Bloch sphere).

For $|\mathbf{a}| = 1$ (the surface of the Bloch sphere), the operator $\rho(\mathbf{a})$ is an orthogonal projection operator. The vector $\psi_+(\mathbf{a})$ defined in (4.19) is the eigenvector of $\rho(\mathbf{a})$ belonging to the eigenvalue 1 (see Exercise 4.11). Hence, $\rho(\mathbf{a})$ projects onto the subspace spanned by $\psi_+(\mathbf{a})$. We conclude that the density operators corresponding to the surface of the Bloch sphere are the pure states.

The qubit density operator $\rho(\mathbf{a})$ is a pure state if and only if $|\mathbf{a}| = 1$. In this case, we have

$$\rho(\mathbf{a}) = \frac{1}{2}\left(\mathbf{1}_2 + \mathbf{a}\cdot\boldsymbol{\sigma}\right) = |\psi_+(\mathbf{a})\rangle\langle\psi_+(\mathbf{a})| \quad \text{for } |\mathbf{a}| = 1. \tag{5.116}$$

The vector $\psi_+(\mathbf{a})$ was defined in (4.19).

Let $0 < |\mathbf{a}| < 1$, and define $\hat{\mathbf{a}} = \mathbf{a}/|\mathbf{a}|$ as the unit vector in the direction of $\mathbf{a}$. The vectors $\psi_\pm(\hat{\mathbf{a}})$ defined in (4.19) are the eigenvectors of $\rho(\mathbf{a})$, and we find that the canonical form of $\rho(\mathbf{a})$ is given by

$$\rho(\mathbf{a}) = \frac{1}{2}(1+|\mathbf{a}|)|\psi_+(\hat{\mathbf{a}})\rangle\langle\psi_+(\hat{\mathbf{a}})| + \frac{1}{2}(1-|\mathbf{a}|)|\psi_-(\hat{\mathbf{a}})\rangle\langle\psi_-(\hat{\mathbf{a}})|. \tag{5.117}$$

We can assume that $\mathbf{a}$ is a linear combination of two other vectors in the Bloch sphere,

$$\mathbf{a} = \lambda\mathbf{b} + (1-\lambda)\mathbf{c}, \quad \text{with } 0 \le \lambda \le 1. \tag{5.118}$$

This means that $\mathbf{a}$ is on the straight line segment connecting $\mathbf{b}$ and $\mathbf{c}$. A linear combination of the form (5.118) is called a *convex linear combination*. Then

$$\rho(\mathbf{a}) = \lambda\rho(\mathbf{b}) + (1-\lambda)\rho(\mathbf{c}). \tag{5.119}$$

We can even choose $\mathbf{b}$ and $\mathbf{c}$ on the surface of the Bloch sphere, with $|\mathbf{b}| = |\mathbf{c}| = 1$. (Obviously, this can be done in an infinite number of ways.) Hence, any qubit density matrix $\rho(\mathbf{a})$ can be written as a convex linear combination of pure states,

$$\rho(\mathbf{a}) = \lambda|\psi_+(\mathbf{b})\rangle\langle\psi_+(\mathbf{b})| + (1-\lambda)|\psi_+(\mathbf{c})\rangle\langle\psi_+(\mathbf{c})| \tag{5.120}$$
$$\text{whenever } \mathbf{a} = \lambda\mathbf{b} + (1-\lambda)\mathbf{c} \text{ with } |\mathbf{b}| = |\mathbf{c}| = 1.$$

An example is shown in Figure 5.2.

According to Section 5.6.3, the equations (5.117) and (5.120) can be interpreted as physically different preparation procedures. In order to prepare an ensemble in the state $\rho(\mathbf{a})$, we can

(a) prepare individual copies of the qubit in the states $\psi_\pm(\hat{\mathbf{a}})$ with the probabilities $\frac{1}{2}(1 \pm |\mathbf{a}|)$, or

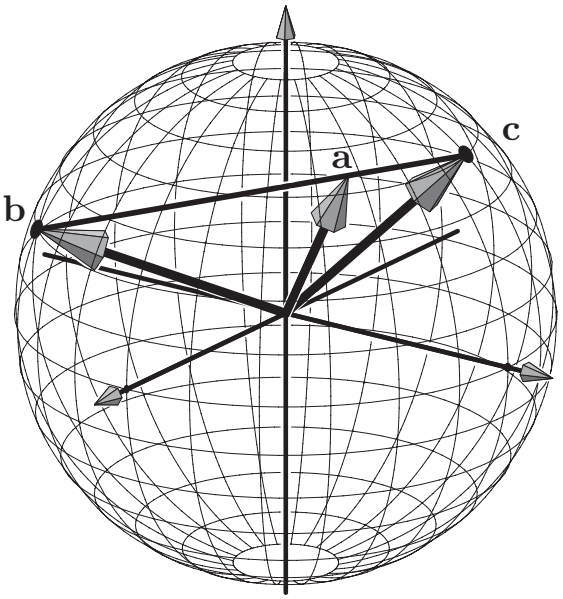


FIGURE 5.2. Density operators for qubits correspond to vectors in the Bloch sphere. The surface of the Bloch sphere corresponds to pure states; the center is the maximally mixed state. Any mixed state $\rho(\mathbf{a})$ is a (non-unique) convex linear combination of pure states. A convex linear combination of $\mathbf{b}$ and $\mathbf{c}$ is a vector $\mathbf{a}$ on the line segment joining $\mathbf{b}$ and $\mathbf{c}$.

(b) prepare individual copies of the qubit in the states $\psi_+(\mathbf{b})$ and $\psi_+(\mathbf{c})$ with the probabilities λ and $(1-\lambda)$, respectively.

The ensemble in the state $\rho(\mathbf{a})$ finally contains no information about which of the preparation procedures (a) or (b) has actually been carried through.

Ψ Many of the observations made here are true for more general systems. For any two density operators ρ_1 and ρ_2, the convex combination $\lambda\rho_1 + (1-\lambda)\rho_2$ is again a density operator. A set with these properties is called a *convex set*. Geometrically, a convex set may be visualized as a set with the property that for any two points in the set, the line segment joining the two points also belongs to the set. Hence, the set of all density operators (the set of all states of a physical system) is a convex set, for which the Bloch sphere is a good example. A point in a convex set is called an *extremal point* if it cannot be written as a convex combination of two other points. The pure states of a quantum system are just the extremal points in the convex set of states.

EXERCISE 5.26. *Given a qubit in the state $\rho(\mathbf{a})$ and a unit vector $\mathbf{n} \in \mathbb{R}^3$, show that the expectation value of the observable $\mathbf{n} \cdot \boldsymbol{\sigma}$ is given by*

$$\langle \mathbf{n} \cdot \boldsymbol{\sigma} \rangle_{\rho(\mathbf{a})} = \mathbf{a} \cdot \mathbf{n}. \tag{5.121}$$

What is the expectation value of $\mathbf{n} \cdot \boldsymbol{\sigma}$ in the state $\rho = \frac{1}{2}\mathbf{1}$?

EXERCISE 5.27. *A qubit is in a mixed state obtained by preparing $\psi_+(\mathbf{e}_x)$ with probability $1/4$ and $\psi_+(\mathbf{e}_y)$ with probability $3/4$. Find the canonical form of the density operator and determine the direction for which the probability that the qubit has spin-up is maximal. What is the probability of finding spin-up in the x-direction?*

5.8. More About Bipartite Systems

5.8.1. Normal form of the state vector

Let us return to our investigation of the states of compound systems. Using the canonical form (5.91) of the density operator ρ^A, we can also obtain a normal form for the state vector of the compound system. This normal form makes it easy to determine whether the state is entangled or not and whether ψ is uniquely determined by the states of the subsystems or not.

Let us write a state ψ of a bipartite system as a linear combination of product states as in (5.22). We choose in $\mathfrak{H}^A$ an orthonormal basis $\{\phi_j^A\}$ with respect to which ρ^A is diagonal. Let $\{\psi_j^B\}$ be an arbitrary orthonormal basis in $\mathfrak{H}^B$. Then

$$\psi = \sum_{ij} c_{ij}\, \phi_i^A \otimes \psi_j^B = \sum_i \phi_i^A \otimes \Big(\sum_j c_{ij}\, \psi_j^B \Big) = \sum_i \phi_i^A \otimes \chi_i^B, \tag{5.122}$$

where

$$\chi_i^B = \sum_j c_{ij}\, \psi_j^B. \tag{5.123}$$

Because ρ^A is diagonal in the basis $\{\phi_j^A\}$, we have, using Eq. (5.84),

$$\rho^A = \sum_j p_j\, |\phi_j^A\rangle\langle\phi_j^A| = \sum_{j,k} \sum_l \overline{c_{jl}}\, c_{kl}\, |\phi_k^A\rangle\langle\phi_j^A| \tag{5.124}$$

and hence

$$\sum_l \overline{c_{jl}}\, c_{kl} = p_j\, \delta_{jk}. \tag{5.125}$$

This expression is just the scalar product of the vectors χ_j^B and χ_k^B, because

$$\langle \chi_j^B, \chi_k^B \rangle = \sum_{lr} \overline{c_{jl}}\, c_{kr}\, \langle \psi_l^B, \psi_r^B \rangle = \sum_{lr} \overline{c_{jl}}\, c_{kr}\, \delta_{lr} = \sum_{l} \overline{c_{jl}}\, c_{kl}.$$

Hence,

$$\langle \chi_j^B, \chi_k^B \rangle = p_j\, \delta_{jk}, \tag{5.126}$$

that is, the vectors $\{\chi_j^B\}$ form an orthogonal set in $\mathfrak{H}^B$. After normalization, we obtain an orthonormal set consisting of the vectors

$$\phi_j^B = \frac{1}{\sqrt{p_j}}\, \chi_j^B. \tag{5.127}$$

Inserting this into (5.122), we obtain the normal form of ψ,

$$\psi = \sum_i \sqrt{p_i}\, \phi_i^A \otimes \phi_i^B. \tag{5.128}$$

From this we may conclude the following: ψ is separable if and only if precisely one of the summands is nonzero, otherwise it is entangled. In the separable case, we have $p_k = 1$ for precisely one index k, and $\rho^A = |\phi_k^A\rangle\langle\phi_k^A|$ is a pure state. As a consequence, we have the following result:

Pure and mixed states:

A subsystem is in a pure state if and only if the compound system is in a separable state. The subsystem is in a mixed state if and only if the state of the compound system is entangled.

If the compound system is described by (5.128), then the state of the subsystem B is given by

$$\rho^B = \sum_i p_i\, |\phi_i^B\rangle\langle\phi_i^B|. \tag{5.129}$$

Hence: For both subsystems A and B, the nonzero eigenvalues of the density operators are the same! To each eigenvector of ρ^A belonging to a nonzero eigenvalue, there is an eigenvector of ρ^B belonging to the same eigenvalue. Only the degree of degeneracy of the eigenvalue zero may be different (note that $\mathfrak{H}^A$ and $\mathfrak{H}^B$ need not have the same dimensions). Moreover, the states of A and B are either both pure or both mixed.

Normal form of states:

For a given state $\psi \in \mathfrak{H}^A \otimes \mathfrak{H}^B$ of the compound system, we can find orthonormal bases $\{\phi_j^A\}$ in $\mathfrak{H}^A$ and $\{\phi_j^B\}$ in $\mathfrak{H}^B$ such that

$$\psi = \sum_i \sqrt{p_i}\ \phi_i^A \otimes \phi_i^B, \quad p_i \geq 0, \quad \sum_i p_i = 1. \tag{5.130}$$

Equation (5.130) is called a *Gram-Schmidt representation* of ψ.

If the state of the compound system is given by (5.130), then the states of the subsystems are described by the density operators

$$\rho^A = \sum_i p_i\, |\phi_i^A\rangle\langle\phi_i^A|, \qquad \rho^B = \sum_i p_i\, |\phi_i^B\rangle\langle\phi_i^B|. \tag{5.131}$$

Hence, ρ^A and ρ^B have the same nonzero eigenvalues. ρ^A is a pure state if and only if ρ^B is a pure state.

Ψ In general, the Gram-Schmidt representation is not unique. As an example, we consider the vector ψ, given in a Gram-Schmidt representation as

$$\psi = \sum_{j=1}^n \phi_j^A \otimes \phi_j^B \tag{5.132}$$

(the corresponding density matrix ρ^A has a degenerate eigenvalue). We are going to construct another Gram-Schmidt representation of ψ with the help of an arbitrary unitary $n \times n$ matrix U. In the Hilbert space $\mathfrak{H}^A$, we consider the basis transformation defined by the transposed matrix $U^\top$ according to $\psi_j^A = \sum_l (U^\top)_{jl} \phi_l^A$. At the same time, we consider the transformation $\psi_j^B = \sum_m (U^{-1})_{jm} \phi_m^B$ in the Hilbert space $\mathfrak{H}^B$. We note that $\psi_j^A = U\phi_j^A$ and $\psi_j^B = U^\top \phi_j^B$. Hence, if the vectors ϕ_j^A and ϕ_j^B form orthonormal bases, then the unitarity of the matrix U implies that the sets $\{\psi_j^A\}$ and $\{\psi_j^B\}$ are also orthonormal bases. Now we compute

$$\begin{aligned}
\sum_{j=1}^n \psi_j^A \otimes \psi_j^B &= \sum_{j=1}^n \Big(\sum_{l=1}^n (U^\top)_{jl} \phi_l^A \Big) \otimes \Big(\sum_{m=1}^n (U^{-1})_{jm} \phi_m^B \Big) \\
&= \sum_{l,m}^n \Big(\sum_{j=1}^n U_{lj} (U^{-1})_{jm} \Big)\, \phi_l^A \otimes \phi_m^B \\
&= \sum_{l,m}^n \delta_{lm}\, \phi_l^A \otimes \phi_m^B = \sum_{l=1}^n \phi_l^A \otimes \phi_l^B = \psi.
\end{aligned} \tag{5.133}$$

Hence, we have obtained another Gram-Schmidt representation of ψ. We note that the Gram-Schmidt representation is essentially unique if all nonzero

numbers p_j are different (that is, all nonzero eigenvalues of ρ^A or ρ^B are nondegenerate). In this case, the only remaining freedom is the replacement $\phi_i^A \otimes \phi_i^B \to \mathrm{e}^{\mathrm{i}\lambda}\phi_i^A \otimes \mathrm{e}^{-\mathrm{i}\lambda}\phi_i^B$.

5.8.2. Maximally entangled states

Because the definition of ρ^A involves all the coefficients c_{ij} of the state ψ in the basis $\{\psi_i^A \otimes \psi_j^B\}$, one might ask to what extent the state ψ of the bipartite system is already determined by ρ^A and ρ^B. In physical terms, the question is the following: To what extent do measurements of the states of the subsystems A and B tell us something about the state ψ of the compound system?

Unfortunately, the connection between the states ψ of the compound system and the states ρ^A, ρ^B of the subsystems is not one-to-one. Many different states of the compound system may lead to the same states in the subsystems. For example, any Bell state (5.30) of a two-qubit system leads to the same density matrix for both subsystems:

$$\rho^A = \frac{1}{2}\,\mathbf{1}_2, \quad \rho^B = \frac{1}{2}\,\mathbf{1}_2, \quad \text{(for any Bell state)}. \tag{5.134}$$

Bell states of a two-qubit system are a special case of the following definition.

Definition 5.3. Assume that $\mathfrak{H}^A$ and $\mathfrak{H}^B$ both have dimension n. A state of the compound system is said to be *maximally entangled* if

$$\rho^A = \frac{1}{n}\,\mathbf{1}^A, \qquad \rho^B = \frac{1}{n}\,\mathbf{1}^B. \tag{5.135}$$

Let $\{\psi_1, \psi_2, \ldots, \psi_n\}$ and $\{\phi_1, \phi_2, \ldots, \phi_n\}$ be any two orthonormal bases of n-dimensional Hilbert spaces $\mathfrak{H}^A$ and $\mathfrak{H}^B$, respectively. A maximally entangled state is, for example,

$$\psi = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi_j \otimes \phi_j \qquad \text{i.e.,} \quad c_{ij} = \frac{1}{\sqrt{n}}\,\delta_{ij}. \tag{5.136}$$

Obviously, this state already has the normal form (5.130). Hence, you can easily verify Eq. (5.135), using (5.131) and the completeness of bases. For example,

$$\rho^A = \frac{1}{n} \sum_{j=1}^{n} |\psi_j\rangle\langle\psi_j| = \frac{1}{n}\,\mathbf{1}^A. \tag{5.137}$$

This is a statistical mixture where each of the pure states ψ_j contributes with the same probability (a *maximally mixed* state).

If we choose other orthonormal bases in the subspaces, we obtain another maximally entangled state of the compound system but the same density matrices ρ^A and ρ^B for the subsystems. Equation (5.137), written in terms of

another orthonormal basis $\{\psi'_j\}$, can be interpreted as a statistical mixture of the states ψ'_j, but the two representations are completely indistinguishable by measurements performed on A (as discussed in Section 5.7.2). Corresponding to the infinite number of orthonormal bases in $\mathfrak{H}^A$ and $\mathfrak{H}^B$, there is an infinite number of maximally entangled states, and all of them lead to the same state of the subsystem.

Different maximally entangled states are different pure states of the compound system. Hence, they can be distinguished by suitable ("global") measurements performed on the system as a whole. But they cannot be distinguished by local measurements, which provide only information about ρ^A and ρ^B. In general, the state ψ of a bipartite system contains more information than the states ρ^A and ρ^B of the subsystems together. This is nothing but the quantum mechanical way of saying that the whole is more than the sum of its parts.

5.8.3. Purification

We consider a quantum system A whose pure states are vectors in a Hilbert space $\mathfrak{H}^A$. We assume that the system is in a mixed state, described by a density operator ρ. We can find an orthonormal basis $\{\phi_j^A\}$ such that the density operator is given by

$$\rho = \sum_i p_j |\phi_j^A\rangle\langle\phi_j^A|. \tag{5.138}$$

Now, we take another Hilbert space $\mathfrak{H}^B$ whose dimension is at least the number m of nonzero eigenvalues of ρ (counting the degree of degeneracy). Then, we choose an (arbitrary) orthonormal system $\{\psi_j^B\}$ in $\mathfrak{H}^B$ consisting of m vectors and define the state

$$\psi = \sum_j \sqrt{p_j}\, \phi_j^A \otimes \psi_j^B. \tag{5.139}$$

By construction, our system A is now part of a compound system. The state ψ of the compound system is a pure (entangled) state with the property that the state of the subsystem A is given by $\rho^A = \rho$ (according to (5.131)). The pure state ψ of the larger system is called a *purification* of ρ. There are many purifications of ρ, because we have a complete freedom in the choice of the orthonormal system $\{\psi_j^B\}$. In fact, there is a one-to-one correspondence between orthonormal systems of m vectors in $\mathfrak{H}^B$ and the possible purifications ψ of ρ.

We see that any mixed state may be considered a pure state in a larger Hilbert space. A purification of ρ can also be characterized as a pure state ψ such that

$$\langle\psi, S \otimes \mathbf{1}\, \psi\rangle = \operatorname{Tr} S\rho \tag{5.140}$$

holds for any (bounded) subsystem observable $S \otimes \mathbf{1}$.

5.8.4. Projection postulate for mixed states

Consider a system in a mixed state ρ. We assume that the Hilbert space is finite-dimensional and consider some observable A with a discrete spectrum,

$$A = \sum_j \lambda_j P_j, \tag{5.141}$$

where λ_j are the eigenvalues, and P_j are the orthogonal projection operators onto the corresponding eigenspaces. If a measurement of A produces the eigenvalue λ_k, then, immediately after the measurement, the system is in the state

$$\rho_k = \frac{1}{\operatorname{Tr} P_k \rho}\, P_k\, \rho\, P_k. \tag{5.142}$$

This result follows from the projection postulate. Let ψ be any purification of ρ. Then, according to the projection postulate, a measurement of the observable $A \otimes \mathbf{1}$ that produces the eigenvalue λ_k projects the system into the range of the projection operator $P_k \otimes \mathbf{1}$. That is, the state of the purified bipartite system after the measurement is $\phi_k = N_k(P_k \otimes \mathbf{1})\,\psi$ (with a suitable normalization constant N_k). In order to compute the density matrix ρ_k of the original system after the measurement, we consider the expectation values of the subsystem observables,

$$\begin{aligned} \langle S \rangle &= N_k^2 \langle (P_k \otimes \mathbf{1})\,\psi, (S \otimes \mathbf{1})(P_k \otimes \mathbf{1})\,\psi \rangle \\ &= N_k^2 \langle (\psi, (P_k S P_k \otimes \mathbf{1})\,\psi \rangle = N_k^2 \operatorname{Tr} P_k S P_k \rho \end{aligned} \tag{5.143}$$

and determine the normalization constant from

$$1 = \operatorname{Tr} \mathbf{1}\rho = N_k^2 \operatorname{Tr} P_k^2 \rho = N_k^2 \operatorname{Tr} P_k \rho. \tag{5.144}$$

By definition, we must have $\langle S \rangle = \operatorname{Tr} S\rho_k$, hence

$$\operatorname{Tr} S\rho_k = \frac{1}{\operatorname{Tr} P_k \rho} \operatorname{Tr} P_k S P_k \rho = \operatorname{Tr} S\Big(\frac{1}{\operatorname{Tr} P_k \rho}\, P_k \rho P_k\Big). \tag{5.145}$$

Because this holds for all operators S, we finally obtain (5.142).

Exercise 5.28. *The state of a qubit is given by*

$$\rho = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}. \tag{5.146}$$

What is the probability of finding $+1$ in a measurement of σ_1? What is the state of the qubit after having obtained this result? Give a quick answer and verify it by a computation using (5.142).

EXERCISE 5.29. *Consider a system in an eigenspace of L^2 with angular-momentum quantum number $\ell = 1$. Denote the eigenstates of L_3 by $|m\rangle$ with $m = -1, 0, 1$. Assume that the system is in a mixed state where $|0\rangle$ has the probability $1/2$ and the two "real orbitals" $\frac{1}{\sqrt{2}}(|1\rangle \pm |-1\rangle)$ each have the probability $1/4$. Suppose you want to measure L_3, but due to experimental inaccuracies you only get the result that L_3 is non-negative. What is the probability for that result? What is the state of the system after the measurement according to the projection postulate* (5.142)*?*

5.9. Indistinguishable Particles

5.9.1. (Anti-)symmetric states of bipartite systems

We consider a bipartite system where the subsystems are of the same kind, for example, two particles in $\mathbb{R}^3$ or two qubits. More precisely, we assume that both subsystems are described by the same Hilbert space $\mathfrak{H} = \mathfrak{H}^A = \mathfrak{H}^B$. Whenever the system is in the separable state

$$\psi = \psi_1 \otimes \psi_2, \tag{5.147}$$

then the first system is in the state ψ_1 and the second system is in the state ψ_2.

Now, let us exchange the two subsystems. This means that we put the first system in the state ψ_2 and the second system in the state ψ_1. Then, the state of the combined system becomes $\psi_2 \otimes \psi_1$. In general, this state is different from (5.147),

$$\psi_1 \otimes \psi_2 \neq \psi_2 \otimes \psi_1 \quad \text{for } \psi_1 \neq \psi_2. \tag{5.148}$$

Therefore, these two states can be distinguished by suitable measurements, and this allows us to distinguish between the two subsystems. If the two systems are truly identical, then there should be absolutely no way to tell whether the first is in the state ψ_1 and the second in the state ψ_2, or vice versa. Obviously, this situation cannot be described adequately by a separable state like (5.147) with $\psi_1 \neq \psi_2$.

Identical systems:

Two quantum systems are said to be *identical* if there is no measurement that can distinguish between the two systems.

This rather vague definition just states that identical systems are *indistinguishable*. In order to investigate identical systems, we introduce the

exchange operator $\mathcal{X}$. It is defined in the Hilbert space $\mathfrak{H} \otimes \mathfrak{H}$. As usual, we define first the action of $\mathcal{X}$ on separable states,

$$\mathcal{X}\,(\psi_1 \otimes \psi_2) = \psi_2 \otimes \psi_1, \tag{5.149}$$

and extend this definition by linearity and continuity to all vectors in $\mathfrak{H} \otimes \mathfrak{H}$

$$\mathcal{X} \sum_{j,k} c_{jk}\, \phi_j \otimes \phi_k = \sum_{j,k} c_{jk}\, \phi_k \otimes \phi_j. \tag{5.150}$$

The operator $\mathcal{X}$ is unitary and self-adjoint on $\mathfrak{H} \otimes \mathfrak{H}$, and

$$\mathcal{X}^2 = \mathbf{1}. \tag{5.151}$$

Hence, the only possible eigenvalues of $\mathcal{X}$ are ± 1.

The operator $\mathcal{X}$ defines a symmetry transformation for the states of the composite system (see Section 1.1.2). We may now sharpen our definition of identical systems.

Exchange symmetry:

A bipartite system is said to be composed of identical (or indistinguishable) subsystems if all states remain unchanged under the symmetry transformation defined by the exchange operator $\mathcal{X}$.

What are the consequences of the exchange symmetry for the state vectors? The two normalized vectors ψ and $\mathcal{X}\psi$ must correspond to the same state of the composite system. Hence, these two vectors can only differ by a phase factor:

$$\mathcal{X}\,\psi = \mathrm{e}^{\mathrm{i}\lambda}\,\psi \quad \text{for some } \lambda \in [0, 2\pi) \tag{5.152}$$

(λ is called the *exchange phase*). From this we obtain

$$\psi = \mathcal{X}^2\,\psi = \mathrm{e}^{2\,\mathrm{i}\lambda}\,\psi \tag{5.153}$$

and because of (5.151) we find $\mathrm{e}^{2\,\mathrm{i}\lambda} = 1$, that is, the exchange phase can only be $\lambda = 0$ or $\lambda = \pi$. Either we have $\mathcal{X}\psi = \psi$ (the state vector is symmetric) or we have $\mathcal{X}\psi = -\psi$ (the state vector is antisymmetric).

Note that *all* state vectors must be either symmetric or antisymmetric. It is not possible that some are symmetric and others are antisymmetric, as long as we adhere to the superposition principle, because a superposition of a symmetric and an antisymmetric state vector would neither be symmetric nor antisymmetric. For example, assuming $\mathcal{X}\psi = \psi$ and $\mathcal{X}\phi = -\phi$, we find that $\mathcal{X}(\psi + \phi) = \psi - \phi$ is orthogonal to $\psi + \phi$ and thus violates the exchange symmetry. We conclude:

Symmetrization postulate:

Consider a bipartite system composed of identical subsystems. All state vectors must be either *symmetric*,

$$\mathcal{X}\,\psi = \psi, \tag{5.154}$$

or *antisymmetric*,

$$\mathcal{X}\,\psi = -\psi, \tag{5.155}$$

under exchange of the subsystems.

Systems described by symmetric state vectors are called *bosonic*, and systems with antisymmetric state vectors are called *fermionic*.



The statement in the box above is called a postulate, although it appears as a mathematical consequence of our definition of identity. However, this definition is based on a rather vague concept of subsystem exchange. What is the physical meaning of an exchange of two identical particles? Is it some physical operation or just a mathematical concept? Can we exclude that the composition of two interchange operations has to be represented by the operator $-\mathbf{1}$ (like a rotation through 2π of a spin 1/2 particle)? In this case, one would have to replace the exchange operator (5.149) by an operator $\mathcal{X}'$ with the property $(\mathcal{X}')^2 = -\mathbf{1}$. (The corresponding ray transformation is still the identity on rays, see Section 1.1.3.) This would change the mathematical definition of "indistinguishable" and the possible values of the exchange phase. Indeed, theoretical considerations like these have led to an experimental search for a violation of the symmetrization postulate (without success) and to the theoretical concept of *anyons*, a hypothetical class of particles in two dimensions (which are employed, for example, in discussions of the so-called *fractional quantum Hall effect*).

As a consequence of the symmetrization postulate, we find that the physical Hilbert space of a system composed of identical subsystems can only be a subspace of the tensor product $\mathfrak{H} \otimes \mathfrak{H}$.

Starting with a separable state vector $\psi_1 \otimes \psi_2$, it is easy to construct suitable symmetric or antisymmetric vectors in $\mathfrak{H} \otimes \mathfrak{H}$: The vector

$$\psi_{\mathrm{S}} = \frac{1}{\sqrt{2}} \Big(\psi_1 \otimes \psi_2 + \psi_2 \otimes \psi_1 \Big) \tag{5.156}$$

is symmetric whereas

$$\psi_{\mathrm{A}} = \frac{1}{\sqrt{2}} \Big(\psi_1 \otimes \psi_2 - \psi_2 \otimes \psi_1 \Big) \tag{5.157}$$

is antisymmetric. The symmetric state vectors form the eigenspace of the exchange operator $\mathcal{X}$ belonging to the eigenvalue $+1$; the antisymmetric

state vectors belong to the eigenvalue -1. We may define the projection operators onto the eigenspaces by

$$P_{\mathrm{S}} = \frac{1}{\sqrt{2}}(\mathbf{1} + \mathcal{X}) \quad \text{and} \quad P_{\mathrm{A}} = \frac{1}{\sqrt{2}}(\mathbf{1} - \mathcal{X}). \tag{5.158}$$

The *symmetrizer* P_{S} and the *antisymmetrizer* P_{A} are orthogonal projection operators. Their ranges are Hilbert spaces. The range of P_{S} is called the *symmetric subspace* or *bosonic Hilbert space*. The range of P_{A} is the *antisymmetric* or *fermionic Hilbert space*.

Symmetric and antisymmetric subspaces:

The appropriate Hilbert space of a bipartite bosonic system is the subspace of symmetric states $\mathfrak{H}_{\mathrm{S}} = P_{\mathrm{S}}(\mathfrak{H} \otimes \mathfrak{H})$. The Hilbert space of a fermionic system is $\mathfrak{H}_{\mathrm{A}} = P_{\mathrm{A}}(\mathfrak{H} \otimes \mathfrak{H})$. It consists of all antisymmetric state vectors.

From (5.158), we conclude immediately that $\mathfrak{H} \otimes \mathfrak{H} = \mathfrak{H}_{\mathrm{S}} \oplus \mathfrak{H}_{\mathrm{A}}$ (orthogonal direct sum), and hence any vector $\psi \in \mathfrak{H} \otimes \mathfrak{H}$ can be decomposed in a unique way into a symmetric part and an antisymmetric part, which are mutually orthogonal. It should be stressed, however, that according to the symmetrization postulate, the Hilbert space for a system consisting of two indistinguishable subsystems is either $\mathfrak{H}_{\mathrm{S}}$ or $\mathfrak{H}_{\mathrm{A}}$. Superpositions of symmetric and antisymmetric states are not allowed, because they violate the exchange symmetry. Similarly, the only useful observables are those that leave the symmetric and antisymmetric subspaces separately invariant.

A separable state $\psi_1 \otimes \psi_2$ cannot be antisymmetric.

All states of a fermionic system are entangled. It is not possible to describe a subsystem of a fermionic system by a pure state. Moreover, for any $\psi \in \mathfrak{H}$ we have

$$P_{\mathrm{A}}(\psi \otimes \psi) = 0 \quad \text{(Pauli's exclusion principle)}. \tag{5.159}$$

The Pauli principle states that a fermionic state cannot be obtained by antisymmetrizing a product state with two identical factors. Its usual formulation, "two parts of a fermionic system cannot be in the same state" is rather misleading, because whenever the system is in the state

$$\psi_{\mathrm{A}} = P_{\mathrm{A}}(\psi_1 \otimes \psi_2) \quad \text{with } \psi_1 \perp \psi_2 \tag{5.160}$$

then both subsystems are actually in the same (mixed) state

$$\rho^A = \rho^B = \frac{1}{2}\big(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|\big). \tag{5.161}$$

Bosonic systems are very different from fermionic systems, because

$$P_{\rm S}(\psi\otimes\psi)=\psi\otimes\psi. \tag{5.162}$$

Hence, a bipartite bosonic system can be in a separable state where both subsystems are described by the same pure state. All other states are entangled, for example, the state (5.156) with $\psi_1\neq\psi_2$,

$$\psi_{\rm S}=P_{\rm A}(\psi_1\otimes\psi_2). \tag{5.163}$$

Again, either subsystem is in the mixed state described by the density operator (5.161).

5.9.2. Example: Bosons and fermions

Consider a system of two particles in $\mathbb{R}^n$. The Hilbert space is

$$L^2(\mathbb{R}^{2n})=L^2(\mathbb{R}^n)\otimes L^2(\mathbb{R}^n). \tag{5.164}$$

For the tensor product of square integrable functions, we may use the ordinary product instead of the symbol $\otimes$. Given a basis $\{\psi_j\}$ of wave functions in $L^2(\mathbb{R}^n)$, we can form a basis consisting of the separable wave functions

$$\psi_j(\mathbf{x}^{(1)})\,\psi_k(\mathbf{x}^{(2)}) \tag{5.165}$$

in $L^2(\mathbb{R}^{2n})$. A general two-particle wave function can be expanded as

$$\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})=\sum_{j,k}c_{jk}\,\psi_j(\mathbf{x}^{(1)})\,\psi_k(\mathbf{x}^{(2)}). \tag{5.166}$$

Because the subsystems are particles, the exchange operator $\mathcal{X}$ is called the *particle interchange operator*. According to (5.150), it given by

$$\mathcal{X}\,\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})=\psi(\mathbf{x}^{(2)},\mathbf{x}^{(1)}). \tag{5.167}$$

A wave function describing a system of two identical particles is either symmetric,

$$\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})=\psi(\mathbf{x}^{(2)},\mathbf{x}^{(1)}), \tag{5.168}$$

or antisymmetric,

$$\psi(\mathbf{x}^{(1)},\mathbf{x}^{(2)})=-\psi(\mathbf{x}^{(2)},\mathbf{x}^{(1)}). \tag{5.169}$$

Examples are shown in Figure 5.3. Particles with symmetric wave functions are called *bosons*; particles with antisymmetric wave functions *fermions*.

Actually, it turns out that all particles with half-integer spin are fermions whereas all particles with integer spin are bosons. Systems of particles with spin have to be described, of course, by (anti-)symmetrized tensor products of spinor-wave functions.

The Hamiltonian operator

$$H=-\frac{1}{2\mathrm{m}}\left(\Delta^{(1)}+\Delta^{(2)}\right)+V(\mathbf{x}^{(1)})+V(\mathbf{x}^{(2)})+U(|\mathbf{x}^{(1)}-\mathbf{x}^{(2)}|). \tag{5.170}$$

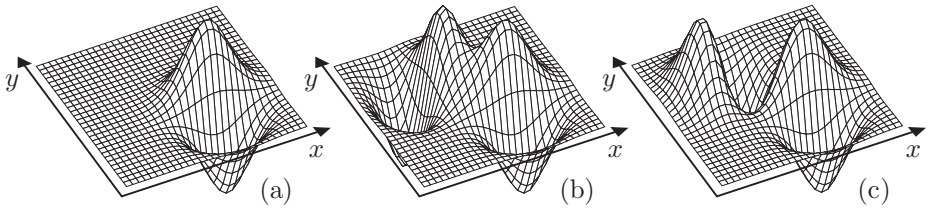
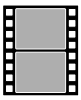


FIGURE 5.3. Some two-particle wave functions. (a) A separable wave function, (b) a symmetric, and (c) an antisymmetric wave function.

describes the energy of two particles in an external potential V. The potential U describes the interaction between the two particles. H applied to an (anti-)symmetric wave function ψ gives a function $H\psi$ that is again (anti-)symmetric. Therefore, H is a suitable Hamiltonian also for a system of two bosons or two fermions.



CD 5.6 shows symmetric and antisymmetric wave functions of two particles interacting via the harmonic oscillator force. The symmetry properties of the wave packet are preserved under the time evolution.

5.9.3. Example: Two-qubit system

The state of a system consisting of two identical qubits must be described either by a symmetric or by an antisymmetric state vector. For fermionic qubits, the only suitable vector among the states of the Bell basis is

$$\psi_{\mathrm{o}}^{-} = \frac{1}{\sqrt{2}}\Big(\psi_{+}\otimes\psi_{-} - \psi_{-}\otimes\psi_{+}\Big). \tag{5.171}$$

On the other hand, there are three orthogonal states for two bosonic qubits: ψ_{e}^{+}, ψ_{e}^{-}, and ψ_{o}^{+}. Another orthonormal basis in the subspace of a bosonic two-qubit system is $\{\psi_{+}\otimes\psi_{+}, \psi_{-}\otimes\psi_{-}, \psi_{\mathrm{o}}^{+}\}$.

We conclude that the Hilbert space of two fermionic qubits is only one-dimensional whereas two bosonic qubits have a three-dimensional Hilbert space. However, this observation is of little practical importance. In the applications, the qubit states usually only describe a part of the properties of a system. The fermionic or bosonic nature of a qubit can be accounted for by the spatial part of its wave function. Consider, for example, a fermion with spin 1/2. Its state is described by a spinor-wave function. For example, a state with spin-up is

$$\psi_{\mathrm{up}}(\mathbf{x}) = \phi(\mathbf{x})\psi_{+} = \begin{pmatrix}\phi(\mathbf{x})\\ 0\end{pmatrix} \quad \text{with } \phi\in L^2(\mathbb{R}^3). \tag{5.172}$$

A two-fermion system can have a symmetric qubit state if only its spatial part is antisymmetric. For example, from the single-particle states $\phi_1(\mathbf{x})\psi_+$ and $\phi_2(\mathbf{x})\psi_+$, we can form the antisymmetric two-particle state

$$\begin{aligned}\psi(\mathbf{x},\mathbf{y}) &= (\phi_1(\mathbf{x})\psi_+)\otimes(\phi_2(\mathbf{y})\psi_+) - (\phi_2(\mathbf{x})\psi_+)\otimes(\phi_1(\mathbf{y})\psi_+) \\ &= (\phi_1(\mathbf{x})\,\phi_2(\mathbf{y}) - \phi_2(\mathbf{x})\,\phi_1(\mathbf{y}))(\psi_+\otimes\psi_+). \end{aligned} \tag{5.173}$$

Here, the antisymmetry is contained completely in the spatial part of the two-particle wave function, while the spin-part is given by the symmetric two-qubit state $\psi_+\otimes\psi_+$. Therefore, the bosonic or fermionic nature of the subsystems is usually of no concern in a discussion of qubit systems.

It is interesting to apply the Pauli principle to a spinor-wave function like (5.173). In this two-particle state, both particles have spin up. Hence, the space-part is antisymmetric and vanishes for $\mathbf{x}=\mathbf{y}$. We conclude

Pauli's exclusion principle:
Two fermions with the same spin are never found at the same position.

5.10. Special Topic: Multiparticle Systems with Spin

For particles with spin s, the component S_3 of the spin has $2s+1$ different eigenvalues. The spinor-wave functions have $2s+1$ components, and the single-particle Hilbert space is

$$\mathfrak{H}^{(1)} = L^2(\mathbb{R}^3)^{2s+1} = \underbrace{L^2(\mathbb{R}^3)\oplus L^2(\mathbb{R}^3)\oplus\ldots\oplus L^2(\mathbb{R}^3)}_{2s+1\text{ copies}}. \tag{5.174}$$

We denote any spinor-wave function in the eigenspace of S^3 belonging to the eigenvalue m_s by $\psi(\mathbf{x};m_s)$. An arbitrary spinor-wave function is a sum $\sum_{m_s}\psi(\mathbf{x};m_s)$, because the Hilbert space is a direct sum of the eigenspaces. It is useful to introduce the abbreviation

$$\xi = (\mathbf{x}, m_s), \quad \mathbf{x}\in\mathbb{R}^n,\ m_s\in\{-s,-s+\tfrac{1}{2},\ldots,s\}, \tag{5.175}$$

and write the scalar product between two spinor-wave functions in the convenient form

$$\langle\psi,\phi\rangle = \int \overline{\psi(\xi)}\,\psi(\xi)\,d\xi \;=\; \sum_{m_s=-s}^{s}\int_{\mathbb{R}^n}\overline{\psi(\mathbf{x},s)}\,\phi(\mathbf{x},s)\,d^nx. \tag{5.176}$$

The Hilbert space for a system of N particles with spin s is the N-fold tensor product

$$\mathfrak{H}^{(N)} = \underbrace{\mathfrak{H}^{(1)} \otimes \mathfrak{H}^{(1)} \otimes \cdots \otimes \mathfrak{H}^{(1)}}_{N \text{ copies}}. \tag{5.177}$$

A wave function for a system of N particles with spin is a function of N variables $\xi^{(1)}, \ldots, \xi^{(N)}$, where $\xi^{(j)}$ describes the position coordinate and the third component of the spin of particle number j. For example, the product functions can be written as

$$\psi_1(\xi^{(1)})\, \psi_2(\xi^{(2)}) \cdots \psi_N(\xi^{(N)}). \tag{5.178}$$

Arbitrary linear combinations of the product functions yield general N-particle wave functions $\psi(\xi^{(1)}, \ldots, \xi^{(N)})$ in $\mathfrak{H}^{(N)}$.

For particles with the same physical characteristics (mass, charge, spin, or other internal quantum numbers), we postulate the physical indistinguishability in the following form: Any wave function that is obtained from $\psi(\xi^{(1)}, \ldots, \xi^{(N)})$ by an arbitrary permutation of the particles should describe the same n-particle state. A permutation of N indices is a one-to-one mapping $\boldsymbol{\pi}$ of the ordered set of numbers $\{1, 2, \ldots N\}$ onto itself. The set of all permutation of N indices is called the symmetric group S_N (it is a group with respect to the composition of the mappings). This group contains $N!$ elements (= the number of permutations of N indices). A transposition is a permutation that just exchanges two of the indices. Any permutation can be written as a composition of transpositions. The permutation is called even if it is a composition of an even number of transpositions, otherwise it is called odd. The sign of the permutation, denoted by $\operatorname{sgn} \boldsymbol{\pi}$, is $+1$ if the permutation is even and -1 if the permutation is odd.

Permutations $\boldsymbol{\pi}$ of N indices define transformations $P_{\boldsymbol{\pi}}$ in the space of N-particle wave functions,

$$P_{\boldsymbol{\pi}}\, \psi(\xi^{(1)}, \ldots, \xi^{(N)}) = \psi(\xi^{(\boldsymbol{\pi}(1))}, \ldots, \xi^{(\boldsymbol{\pi}(N))}). \tag{5.179}$$

It is easy to see that $P_{\boldsymbol{\pi}}$ is a unitary transformation in the n-particle Hilbert space.

Definition 5.4. An N-particle wave function ψ is called *symmetric* if $P_{\boldsymbol{\pi}}\, \psi = \psi$ for all permutations $\boldsymbol{\pi} \in S_N$. The wave function ψ is called *antisymmetric* if $P_{\boldsymbol{\pi}}\, \psi = \operatorname{sgn} \boldsymbol{\pi})\, \psi$.

We define the linear operators

$$P_{\mathrm{S}} = \frac{1}{N!} \sum_{\boldsymbol{\pi} \in S_N} P_{\boldsymbol{\pi}}, \qquad P_{\mathrm{A}} = \frac{1}{N!} \sum_{\boldsymbol{\pi} \in S_N} (\operatorname{sgn} \boldsymbol{\pi}) P_{\boldsymbol{\pi}}. \tag{5.180}$$

These are orthogonal projection operators in the Hilbert space $\mathfrak{H}^{(N)}$ of N-particle wave functions. For an arbitrary wave function ψ, we find that $P_{\rm S}\psi$ is symmetric whereas $P_{\rm A}\psi$ is antisymmetric. Hence, the range of $P_{\rm S}$ is called the bosonic subspace $\mathfrak{H}_{\rm S}^{(N)}$ of the N-particle Hilbert space, and the range of $P_{\rm A}$ is the fermionic subspace $\mathfrak{H}_{\rm A}^{(N)}$. The following result is obtained in the framework of relativistic quantum field theory.

Spin-statistics theorem:

Particles with integer spin are bosons, and particles with half-integer spin are fermions.

In particular, a system of N particles with spin $1/2$ is always described by an antisymmetric wave function in the fermionic Hilbert space $\mathfrak{H}_{\rm A}^{(N)}$. Such a wave function has the property

$$\psi = P_{\rm A}\,\psi. \tag{5.181}$$

We can obtain a fermionic wave function by applying the projection operator $P_{\rm A}$ to a product wave function (5.178). After multiplication with the factor $\sqrt{N!}$, we obtain the following wave function in the fermionic Hilbert space:

$$\begin{aligned}\psi_{\rm A}(\xi^{(1)},\ldots,\xi^{(N)}) &= \sqrt{N!}\,P_{\rm A}\,\psi_1(\xi^{(1)})\psi_2(\xi^{(2)})\cdots\psi_N(\xi^{(N)}) \\ &= \frac{1}{\sqrt{N!}}\begin{vmatrix}\psi_1(\xi^{(1)}) & \psi_1(\xi^{(2)}) & \cdots & \psi_1(\xi^{(N)}) \\ \psi_2(\xi^{(1)}) & \psi_2(\xi^{(2)}) & \cdots & \psi_2(\xi^{(N)}) \\ \vdots & \vdots & & \vdots \\ \psi_N(\xi^{(1)}) & \psi_N(\xi^{(2)}) & \cdots & \psi_N(\xi^{(N)})\end{vmatrix}. \end{aligned} \tag{5.182}$$

This wave function is called the *Slater determinant* of $\psi_1,\ldots,\psi_N$. It is normalized, $\|\psi_{\rm A}\| = 1$, whenever the functions $\{\psi_1,\ldots\psi_N\}$ form an orthonormal set. Starting with an orthonormal basis $\{\psi_1,\psi_2,\ldots\}$ in the one-particle Hilbert space, we obtain a basis of the N-fermion Hilbert space $\mathfrak{H}_{\rm A}^{(N)}$ by forming all possible Slater determinants out of N different basis vectors,

$$\psi_{k_1,\ldots,k_N}(\xi^{(1)},\ldots,\xi^{(N)}) = \sqrt{N!}\,P_{\rm A}\,\psi_{k_1}(\xi^{(1)})\cdots\psi_{k_N}(\xi^{(N)}). \tag{5.183}$$

Whenever two functions are equal (for example, if $k_i = k_j$), then two rows of the corresponding Slater determinant are equal, and hence the determinant vanishes. This is again Pauli's exclusion principle.

The operator of multiplication by the coordinate $x_i^{(k)}$ of the $k^{\rm th}$ particle makes no sense in the bosonic or fermionic Hilbert space, because this

operation would destroy the (anti)symmetry of the wave function. Physically meaningful operators for bosons (or fermions) are only those operators that leave the Hilbert space of bosons (or fermions) invariant, that is, the operators map the Hilbert spaces $\mathfrak{H}_{\mathrm{S}}^{(N)}$ (or $\mathfrak{H}_{\mathrm{S}}^{(N)}$) into itself. An example is the Hamiltonian for a system of N charged particles (all having mass m and charge q),

$$H = \sum_{k=1}^{N} \frac{1}{2\mathrm{m}} \left(\mathbf{p}^{(k)}\right)^2 + \sum_{j<k,j,k=1}^{N} \frac{q^2}{|\mathbf{x}^{(j)} - \mathbf{x}^{(k)}|}. \tag{5.184}$$

Here, $\mathbf{p}^{(k)}$ is the momentum operator in the kth single-particle subspace. It is proportional to the gradient with respect to the coordinates $x_i^{(k)}$. Other examples of operators that preserve the symmetry of the wave function include the total momentum and total angular-momentum operators

$$\mathbf{p} = \sum_{k=1}^{N} \mathbf{p}^{(k)}, \qquad \mathbf{J} = \sum_{k=1}^{N} \big(\mathbf{L}^{(k)} + \mathbf{S}^{(k)}\big). \tag{5.185}$$

All these operators commute with arbitrary permutations $P_{\boldsymbol{\pi}}$ (for this it is sufficient to verify that the operators commute with arbitrary transpositions), hence they commute with the projection operators P_{S} and P_{A}.

5.11. Special Topic: Addition of Angular Momenta

5.11.1. Total angular momentum

In this section, we want to combine two quantum systems with angular momentum. Let the angular-momentum operators be denoted by $\mathbf{J}^A$ and $\mathbf{J}^B$. It is an interesting and physically relevant problem to investigate the angular momentum of the combined system. Actually, we encountered this problem already when we combined the orbital angular momentum $\mathbf{J}^A = \mathbf{L}$ with the spin $\mathbf{J}^B = \mathbf{S}$ of a particle, or when we combined two qubits (spin 1/2 particles) into a single two-qubit system. Naturally, this question is very important for the quantum mechanics of atoms and molecules.

We assume that the square of the total angular momentum has a fixed constant value in both systems and denote the corresponding quantum numbers by j^A and j^B, respectively. Given these quantum numbers, we may restrict our attention to the corresponding angular-momentum eigenspaces. Hence, we assume that the Hilbert space $\mathfrak{H}^A$ is the $(2j^A + 1)$-dimensional eigenspace of $(J^A)^2$ and $\mathfrak{H}^B$ is the $(2j^B + 1)$-dimensional eigenspace of $(J^B)^2$. The Hilbert spaces of the subsystems are therefore finite-dimensional, and the theory developed in this chapter can be applied without difficulties. An orthonormal basis in $\mathfrak{H}^A$ is given by the eigenstates of J_3^A. We use Dirac's

notation and denote these eigenstates simply by $|m^A\rangle$. It is not necessary to use the quantum number j^A as a second label, because it is kept fixed throughout this section. We just have to remember that the maximal value of m^A is j^A. Similarly, the we denote the eigenstates of J_3^B by $|m^B\rangle$, where m^B goes in integer steps from $-j^B$ to $+j^B$. The states $|m^B\rangle$ form an orthonormal basis in the Hilbert space $\mathfrak{H}^B$.

We define the *total angular momentum* of the combined system by

$$\mathbf{J} = \mathbf{J}^A \otimes \mathbf{1}^B + \mathbf{1}^B \otimes \mathbf{J}^B. \tag{5.186}$$

The two subsystem observables commute with each other, because they belong to different subsystems:

$$[J_k^A \otimes \mathbf{1}^B, \mathbf{1}^B \otimes J_l^B] = 0. \tag{5.187}$$

The components J_k of the total angular momentum $\mathbf{J}$ fulfill the usual angular-momentum commutation relations

$$[J_k, J_l] = \sum_j \epsilon_{klj}\, J_j. \tag{5.188}$$

This follows immediately from the corresponding relations for the subsystem observables. $\mathbf{J}$ is a Kronecker sum as defined in Section 5.4.1. It follows from (5.56) that a rotation of the combined system can be obtained by separately rotating the subsystems:

$$\mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}} = \mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}^A} \otimes \mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}^B}. \tag{5.189}$$

Our problem is the following: What are the simultaneous eigenvalues and eigenstates of the operators J^2 and J_3?

Exercise 5.30. *Show that*

$$J^2 = (J^A)^2 \otimes \mathbf{1}^B + \mathbf{1}^A \otimes (J^B)^2 + 2\sum_{k=1}^{3} J_k^A \otimes J_k^B, \tag{5.190}$$

and that J^2 commutes neither with $J_3^A \otimes \mathbf{1}^B$ nor with $\mathbf{1}^A \otimes J_3^B$.

5.11.2. Eigenvalues of J_3

Given the orthonormal bases of the subsystems, we can form the product states

$$|m^A \otimes m^B\rangle = |m^A\rangle \otimes |m^B\rangle. \tag{5.191}$$

Here, the $\otimes$ on the left-hand side is just to remind us that these states are product states. This notation helps us to distinguish the product states from other states given by two quantum numbers. We remind the reader that the states $|m^A \otimes m^B\rangle$ form an orthonormal basis of $\mathfrak{H}^A \otimes \mathfrak{H}^B$ (see Section 5.2.2).

Moreover, it follows from (5.186) and (5.50) that the product states are eigenstates of the operator J_3,

$$J_3 \, |m^A \otimes m^B\rangle = (m^A + m^B) \, |m^A \otimes m^B\rangle. \tag{5.192}$$

Thus, the eigenvalues of J_3 are all the possible values of the sum $m^A + m^B$.

Eigenvalues of J_3:

The operator J_3 has the $2(j^A + j^B + 1)$ different eigenvalues

$$m = m^A + m^B = -j^A - j^B \,,\; -j^A - j^B + 1 \,,\; \ldots \,,\; j^A + j^B. \tag{5.193}$$

The Hilbert space $\mathfrak{H}^A \otimes \mathfrak{H}^B$ has the dimension $(2j^A+1)(2j^B+1)$, which is bigger than $2(j^A+j^B+1)$. Therefore, at least some of the eigenvalues m must be degenerate. For example, the sum $m^A+m^B = j^A+j^B-1$ can be obtained either with $(m^A, m^B) = (j^A, j^B-1)$ or with $(m^A, m^B) = (j^A-1, j^B)$. Hence, the vectors $|j^A \otimes j^B-1\rangle$ and $|j^A-1 \otimes j^B\rangle$ both belong to the same eigenvalue $j^A + j^B - 1$ of J_3. It is not too difficult to determine the degeneracy $d(m)$ for an arbitrary eigenvalue m of J_3. Assuming $j^A \geq j^B$, we find

$$d(m) = \begin{cases} 2j^B + 1, & |m| \leq j^A - j^B, \\ j^A + j^B + 1 - |m|, & j^A - j^B \leq |m| \leq j^A + j^B. \end{cases} \tag{5.194}$$

Figure 5.4 helps to determine $d(m)$: Each point represents an eigenstate of J_3, for which (J_3^A, J_3^B) have the eigenvalues (m^A, m^B). On the diagonal lines, the sum $m = m^A + m^B$ is constant, hence each diagonal line represents an eigenvalue m of J_3. The number of points on such a line is the degree of degeneracy $d(m)$ of that eigenvalue. On the dotted lines we have $|m| > j^A - j^B$. The degree of degeneracy has the maximal value $2j^B + 1$ on the dashed lines, where $|m| \leq j^A - j^B$.

EXERCISE 5.31. *Assume that $\{\psi_n\}$ is an orthonormal basis in a finite-dimensional Hilbert space. Consider two linear operators A and B and show the following: If A and B do not commute, then the ψ_n cannot all be simultaneous eigenvectors of A and B.*

EXERCISE 5.32. *Combine the results of Exercises 5.30 and 5.32 to show that the product states $|m^A \otimes m^B\rangle$ cannot all be eigenstates of J^2.*

5.11.3. The quantum numbers of the total angular momentum

Next, we are interested in the possible eigenvalues of J^2, the square of the total angular momentum of the composite system. From Section 1.5, we know that the quantum numbers j of J^2 are among the numbers $0, 1/2, 1, 3/2, \ldots$. Moreover, from Theorem 1.1 we know that whenever we find the quantum

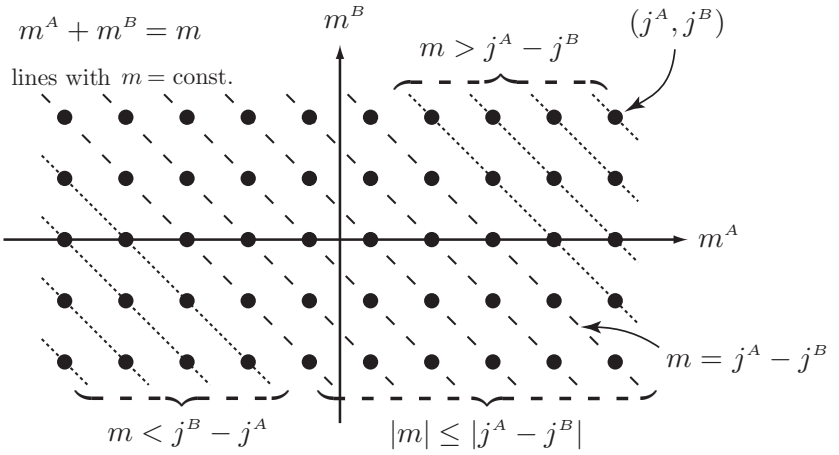


FIGURE 5.4. The points represent the product states (5.191) with all possible values of (m^A, m^B), assuming $j^A > j^B$. Each diagonal line represents an eigenvalue m of J_3.

number j, then there is a $2j+1$-dimensional eigenspace of J^2. The eigenvectors can be chosen as simultaneous eigenvectors of J_3, with eigenvalues $m = -j, -j+1, \ldots, +j$. We denote these simultaneous eigenvectors of J^2 and J_3 by $|j, m\rangle$, that is (assuming $\hbar = 1$)

$$J^2|j,m\rangle = j(j+1)|j,m\rangle, \qquad J_3|j,m\rangle = m|j,m\rangle. \tag{5.195}$$

As a first step, we consider the Hilbert space $\mathfrak{H}_1 = \mathfrak{H}^A \otimes \mathfrak{H}^B$. We know already all possible eigenvalues m of J_3; see (5.193). The largest possible eigenvalue of J_3 is $m_{\max} = j^A + j^B$. This, of course, must also be the largest possible value of the quantum number j:

$$j_{\max} = m_{\max} = j^A + j^B. \tag{5.196}$$

In the following, we write $j_{\max} = j_1$. The corresponding eigenspace of J^2 has the dimension $2j_1 + 1$. In Figure 5.5, this eigenspace corresponds to the top row of points.

As a second step, consider the orthogonal complement $\mathfrak{H}_2$ of the eigenspace with $j = j_1$ (delete the top row in Fig. 5.5). In $\mathfrak{H}_2$ we find no more state with the $m = j_1$, because there is just one state with quantum numbers $(j, m) = (j_1, j_1)$, and this state is already contained in the eigenspace with $j = j_1$.

According to (5.193), the next largest m is $j^A + j^B - 1$. Originally, there were two orthogonal states with this quantum number (according to (5.194),

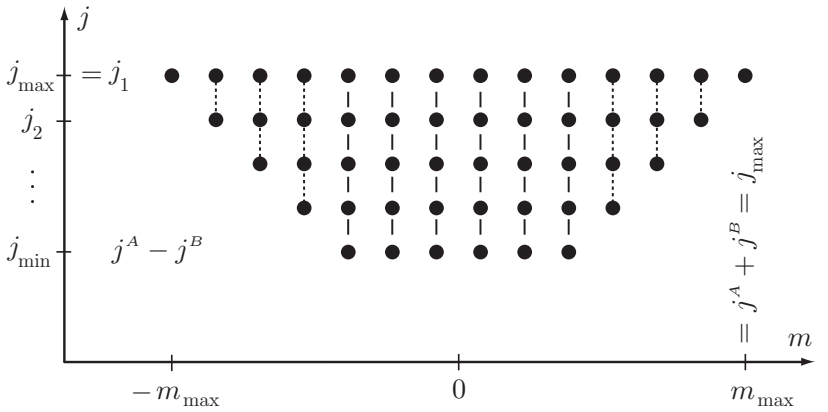


FIGURE 5.5. The points represent the simultaneous eigenstates $|j, m\rangle$ of the total angular-momentum operators J^2 and J_3. The number of points in a row is $2j + 1$, the degree of degeneracy of the eigenvalue of J^2. The number of points in a column is the degree of degeneracy $d(m)$ of the eigenvalue m of J_3 according to (5.194).

$d(j^A + j^B - 1) = 2$), but precisely one such state is already contained in the eigenspace with $j = j_1$. In $\mathfrak{H}_2$ remains a unique state with $m = j^A + j^B - 1$, and this state must therefore belong to $j = j_2 = j^A + j^B - 1$. In the corresponding eigenspace of J^2 are the states with quantum numbers $m = -j_2, -j_2 + 1, \ldots, +j_2$ (second row in Fig. 5.5).

We may iterate this procedure until we use up all dimensions in the Hilbert space $\mathfrak{H}^A \otimes \mathfrak{H}^B$. In the third step, for example, we consider the subspace $\mathfrak{H}_3$ which is orthogonal to the eigenspaces with $j = j_1$ and $j = j_2$ (delete the first two rows in Fig. 5.5). Unless $\mathfrak{H}_3 = \{0\}$ (containing only the zero-vector), the maximal value of m in $\mathfrak{H}_3$ is $j_2 - 1$, and there is precisely one state with this quantum number in $\mathfrak{H}_3$. Hence, the quantum number $j = j_3 = j_2 - 1$ occurs, and there must be a corresponding $2j_3+1$-dimensional eigenspace of J^2 (symbolized by the third row of points in Fig. 5.5).

Finally, assuming that $j^A \geq j^B$, we find in the step with number $2j^B + 1$ the angular-momentum subspace with quantum number $j = j_1 - 2j^B = j^A - j^B$. After that, no more states are left, because $d(m) = 2j^B + 1$ for all $m \leq j^A - j^B$. That is, the quantum number $j = j^A - j^B - 1$ cannot occur, and we find the following minimal value of j:

$$j_{\min} = j^A - j^B \quad (\text{if } j^A \geq j^B). \tag{5.197}$$

Whenever $j^A < j^B$, we simply exchange j^A and j^B in the argument leading

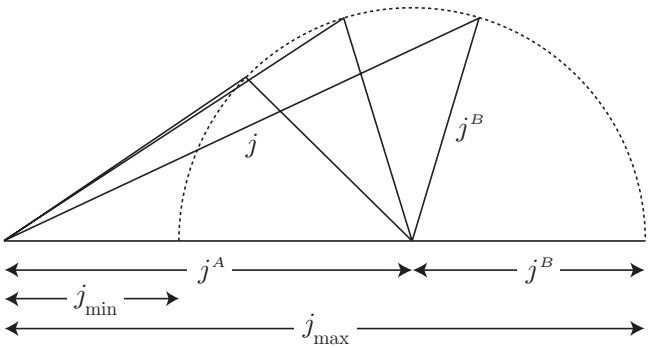


FIGURE 5.6. The triangular condition states that the values of the quantum numbers j, j^A, and j^B could be the lengths of the sides of a triangle.

to (5.197). In this case, we obtain

$$j_{\min} = j^B - j^A \quad (\text{if } j^A < j^B). \tag{5.198}$$

Let us now collect our results in the following box.

Spectrum of J^2:

Whenever the subsystems have the angular momenta j^A and j^B, then the quantum number j of the total angular momentum J^2 has the values

$$|j^A - j^B|\,,\ |j^A - j^B| + 1\,,\ \ldots\,,\ j^A + j^B. \tag{5.199}$$

Hence, the quantum numbers j^A, j^B, and j satisfy the *triangular condition*

$$|j^A - j^B| \le j \le j^A + j^B. \tag{5.200}$$

If $j^B < j^A$, then there are $2j^B + 1$ different values of j.

According to the triangular condition, the three integer or half-integer numbers j, j^A, and j^B could represent the sides of a triangle (see Fig. 5.6).

5.11.4. Clebsch-Gordan coefficients

Above, we found that an orthonormal basis of eigenvectors of J_3 is given by the product states $|m^A \otimes m^B\rangle$. Next, we want to determine the simultaneous eigenvectors $|j, m\rangle$ of J^2 and J_3. Consider the eigenspace belonging to the maximal quantum numbers $(j_{\max}, m_{\max})$. This eigenspace is one-dimensional, and from (5.193) we know that $|j^A \otimes j^B\rangle$ is contained in that eigenspace.

Hence, $|j_{\max}, m_{\max}\rangle$ must be proportional to that vector. We set

$$|j_{\max}, m_{\max}\rangle = |j^A \otimes j^B\rangle. \tag{5.201}$$

This vector is a useful starting point for the construction of the remaining eigenvectors $|j, m\rangle$.

The product states $|m^A \otimes m^B\rangle$ cannot all be eigenvectors of J^2 (see Exercise 5.30 and Exercise 5.31). In general, the simultaneous eigenvectors $|j, m\rangle$ of J^2 and of J_3 are entangled and have to be written as linear combinations of the basis states

$$|j, m\rangle = \sum_{m^A, m^B} |m^A \otimes m^B\rangle\langle m^A \otimes m^B | j, m\rangle. \tag{5.202}$$

The coefficients $\langle m^A \otimes m^B | j, m\rangle$ are called *Clebsch-Gordan coefficients*.[2]

We note that the eigenvectors $|j, m\rangle$ are only unique up to a phase factor. Given arbitrary phases θ_{jm}, the vectors $\exp(i\theta_{jm})\,|j, m\rangle$ would do equally well. The eigenvectors can be made unique by requiring that the Clebsch-Gordan coefficients all be real, and that

$$\langle j^A \otimes j - j^A \mid j, j\rangle > 0. \tag{5.203}$$

The Clebsch-Gordan coefficients are known explicitly for arbitrary values of j^A and j^B. For practical calculations, one has used tables,[3] and today these coefficients are built into computer-algebra systems.[4]

In this book, we do not derive the general formula for $\langle m^A \otimes m^B | j, m\rangle$, but we are going to consider some simple special cases below. Moreover, we note the following simple property of the Clebsch-Gordan coefficients:

Selection rule for Clebsch-Gordan coefficients:

$$\langle m^A \otimes m^B | j, m\rangle = 0 \qquad \text{if } m^A + m^B \neq m. \tag{5.204}$$

EXERCISE 5.33. *Prove the selection rule* (5.204) *for Clebsch-Gordan coefficients.*

[2] Or *Wigner coefficients* or *vector-addition coefficients* or *vector-coupling coefficients*.

[3] See, for example, Abramowitz-Stegun [**1**], Section 27.9.

[4] In *Mathematica*, the function `ClebschGordan`$[\{j^A, m^A\}, \{j^B, m^B\}, \{j, m\}]$ implements the Clebsch-Gordan coefficient $\langle m^A \otimes m^B | j, m\rangle$, with j^A and j^B being the angular momenta of the subsystems.

5.11.5. Angular momentum plus spin 1/2

In Theorem 1.1, we used ladder operators to generate a system of orthogonal eigenvectors. Consider subsystem A. The maximal eigenvalue of J_3^A is j^A. Starting from the corresponding normalized eigenvector $|j^A\rangle \in \mathfrak{H}^A$, we can construct an orthonormal basis of the $2j^A+1$-dimensional Hilbert space $\mathfrak{H}^A$ by applying the operators $J_-^A = J_1^A - \mathrm{i}J_2^A$ repeatedly,

$$|m^A - 1\rangle = \frac{1}{\sqrt{j^A(j^A+1) - m^A(m^A-1)}}\, J_-^A\, |m^A\rangle, \tag{5.205}$$

for $m^A = j^A, j^A-1, \ldots, -j^A+1$. A similar procedure starting with $|j^B\rangle \in \mathfrak{H}^B$ leads to an orthonormal basis of vectors $|m^B\rangle$ for subsystem B.

The ladder operators $J_\pm$ for the combined system can be defined in terms of the ladder operators for the subsystems by

$$J_\pm = J_\pm^A \otimes \mathbf{1}^B + \mathbf{1}^A \otimes J_\pm^B. \tag{5.206}$$

For every $|j,m\rangle$ with $|j^A - j^B| \le j \le j^A + j^B$ and $-j < m \le j$, we define

$$|j, m-1\rangle = \frac{1}{\sqrt{j(j+1) - m(m-1)}}\, J_-\, |j,m\rangle. \tag{5.207}$$

In that way, a unique orthonormal basis is obtained as soon as we have chosen the vectors $|j,j\rangle$ with the maximal value of m for a given j.

As an illustration, we consider a composite system where subsystem B has spin 1/2. We combine a system with angular momentum $j^A > 0$ (this could be an orbital angular momentum or a particle with spin) and a second system with $j^B = \frac{1}{2}$. Then, the possible values for the total angular momentum are, according to (5.199),

$$j = j^A - \tfrac{1}{2} \quad \text{and} \quad j = j^A + \tfrac{1}{2}. \tag{5.208}$$

In subsystem B, we denote the normalized eigenvector of J_3^B belonging to the eigenvalue $+\frac{1}{2}$ by $|\uparrow\rangle$ ("spin-up"), and hence

$$|\downarrow\rangle = J_-^B\, |\uparrow\rangle \tag{5.209}$$

is the normalized eigenvector belonging to the eigenvalue $-\frac{1}{2}$ ("spin-down").

Let us construct an orthonormal system in $\mathfrak{H} = \mathfrak{H}^A \otimes \mathfrak{H}^B$ by applying J_- repeatedly to the eigenvector (5.201) with maximal eigenvalues. As a first

step consider

$$|\, j^A{+}\tfrac{1}{2}\,,\, j^A{-}\tfrac{1}{2}\,\rangle = \frac{1}{\sqrt{2j^A+1}}\, J_-\,|\, j^A{+}\tfrac{1}{2}\,,\, j^A{+}\tfrac{1}{2}\,\rangle \tag{5.210}$$

$$= \frac{1}{\sqrt{2j^A+1}}\,\big(J_-^A\otimes \mathbf{1}^B + \mathbf{1}^A\otimes J_-^B\big)\,|\, j^A\otimes\uparrow\rangle \tag{5.211}$$

$$= \frac{1}{\sqrt{2j^A+1}}\,\big((J_-^A\,|j^A\rangle)\otimes|\uparrow\rangle + |j^A\rangle\otimes|\downarrow\rangle\big) \tag{5.212}$$

Using (5.205), we find $J_-^A\,|j^A\rangle = \sqrt{2j^A}\,|\, j^A{-}1\,\rangle$, hence

$$|\, j^A{+}\tfrac{1}{2}\,,\, j^A{-}\tfrac{1}{2}\,\rangle = \sqrt{\frac{2j^A}{2j^A+1}}\,|\, j^A{-}1\otimes\uparrow\rangle + \sqrt{\frac{1}{2j^A+1}}\,|\, j^A\otimes\downarrow\rangle. \tag{5.213}$$

By iterating the application of the operator J_-, we obtain all states with quantum numbers $m = j^A{-}\frac{1}{2},\, j^A{-}\frac{3}{2},\,\dots\,,\, -\big(j^A{+}\frac{1}{2}\big)$ in the subspace with $j = j^A{+}\frac{1}{2}$. By induction, you may verify the formula

$$|\, j^A{+}\tfrac{1}{2}\,,\, m\,\rangle = \sqrt{\frac{j^A+\frac{1}{2}+m}{2j^A+1}}\,|\, m{-}\tfrac{1}{2}\otimes\uparrow\rangle + \sqrt{\frac{j^A+\frac{1}{2}-m}{2j^A+1}}\,|\, m{+}\tfrac{1}{2}\otimes\downarrow\rangle \tag{5.214}$$

which actually holds for all

$$m = j^A{+}\tfrac{1}{2}\,,\, j^A{-}\tfrac{1}{2}\,,\,\dots\,,\, -\big(j^A{+}\tfrac{1}{2}\big). \tag{5.215}$$

Given $\xi = a|\psi\rangle + b|\phi\rangle$ with $|\psi\rangle$ orthogonal to $|\phi\rangle$, we find that $\chi = -b|\psi\rangle + a|\phi\rangle$ is orthogonal to ξ and $\|\xi\| = \|\chi\|$. Hence, the vector

$$|\, j^A{-}\tfrac{1}{2}\,,\, m\,\rangle = -\sqrt{\frac{j^A+\frac{1}{2}-m}{2j^A+1}}\,|\, m{-}\tfrac{1}{2}\otimes\uparrow\rangle + \sqrt{\frac{j^A+\frac{1}{2}+m}{2j^A+1}}\,|\, m{+}\tfrac{1}{2}\otimes\downarrow\rangle \tag{5.216}$$

is a normalized vector orthogonal to (5.214), whenever $|\, j^A{+}\frac{1}{2}\,,\, m\,\rangle$ consists of two non-vanishing summands, that is, for

$$m = j^A{-}\tfrac{1}{2}\,,\, j^A{-}\tfrac{3}{2}\,,\,\dots\,,\, -\big(j^A{-}\tfrac{1}{2}\big). \tag{5.217}$$

We still have to justify the notation $|\, j^A{-}\frac{1}{2}\,,\, m\,\rangle$ for the vector defined in (5.216). Without computation, you can see that this vector is an eigenvector of $J_3 = J_3^A\otimes\mathbf{1}^B + \mathbf{1}^A\otimes J_3^B$ belonging to the eigenvalue m. Moreover, you can see that for any given m according to (5.217), the vector (5.216) is orthogonal to all vectors $|\, j^A{+}\frac{1}{2}\,,\, m'\,\rangle$ with arbitrary m' in the range (5.215). Moreover, we find

$$\langle\, j^A{-}\tfrac{1}{2}\,,\, m \;|\; j^A{-}\tfrac{1}{2}\,,\, m'\,\rangle = 0 \quad \text{for } m\neq m'. \tag{5.218}$$

We conclude that these vectors span a subspace with dimension $2(j^A-\frac{1}{2})+1$ (= the number of different m's in (5.217)), and that this subspace is orthogonal to the subspace belonging to $j = j^A+\frac{1}{2}$. This subspace is therefore the eigenspace of J^2 belonging to the quantum number $j = j^A-\frac{1}{2}$. Note that the signs in (5.216) are chosen such that

$$\langle\, j^A \otimes \downarrow \mid j^A-\tfrac{1}{2}\,,\, j^A-\tfrac{1}{2}\,\rangle > 0, \tag{5.219}$$

which agrees with the convention (5.203).

From (5.214) and (5.216) we find the nonvanishing Clebsch-Gordan coefficients $\langle m^A \otimes m^B \mid j, m\rangle$ according to the following table.

TABLE 5.1. Clebsch-Gordan coefficients for $j^B = 1/2$.

	$m^A = m-\frac{1}{2}\ ,\ m^B = \frac{1}{2}$	$m^A = m+\frac{1}{2}\ ,\ m^B = -\frac{1}{2}$
$j = j^A+\frac{1}{2}$	$\sqrt{\dfrac{j+m}{2j}}$	$\sqrt{\dfrac{j-m}{2j}}$
$j = j^A-\frac{1}{2}$	$-\sqrt{\dfrac{j+1-m}{2j}}$	$\sqrt{\dfrac{j+1+m}{2j}}$

EXAMPLE 5.2. A simple example is obtained by combining two particles with spin 1/2. Here, $j^A = j^B = \frac{1}{2}$, and the possible values of the total angular momentum are $j = 1$ and $j = 0$. For $j = 1$ we obtain from (5.214) the three states with $m = -1, 0, 1$, namely

$$|1,-1\rangle = |\downarrow \otimes \downarrow\rangle, \tag{5.220}$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow \otimes \downarrow\rangle + |\downarrow \otimes \uparrow\rangle\big), \tag{5.221}$$

$$|1,1\rangle = |\uparrow \otimes \uparrow\rangle, \tag{5.222}$$

and for $j = 0$ we obtain

$$|0,0\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow \otimes \downarrow\rangle - |\downarrow \otimes \uparrow\rangle\big). \tag{5.223}$$

The three states with $j = 1$ are frequently called *triplet states*, and the single state with $j = 0$ is called *singlet state*. The vector $|1,0\rangle$ is just another way of writing the basis vector ψ_{o}^+ of the Bell basis (5.30), and the singlet state $|0,0\rangle$ is the same as ψ_{o}^-.

EXAMPLE 5.3. As a second example, we consider an orbital angular momentum $j^A = \ell$ with the eigenvectors $|m_\ell\rangle = Y_\ell^{m_\ell}$ (spherical harmonics) and combine it with the spin. In the standard representation, the spin eigenvectors are

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.224}$$

and the tensor product states are, in accordance with (5.32), given by

$$|m_\ell \otimes \uparrow\rangle = \begin{pmatrix} Y_\ell^{m_\ell} \\ 0 \end{pmatrix}, \qquad |m_\ell \otimes \downarrow\rangle = \begin{pmatrix} 0 \\ Y_\ell^{m_\ell} \end{pmatrix}. \tag{5.225}$$

The possible values of the total angular momentum are $j = \ell + \frac{1}{2}$ and $\ell - \frac{1}{2}$. For $j = \ell + \frac{1}{2}$, (5.214) becomes

$$|\ell{+}\tfrac{1}{2}\,,\, m_j\rangle = \sqrt{\frac{1}{2\ell+1}} \begin{pmatrix} \sqrt{\ell + \frac{1}{2} + m_j}\, Y_\ell^{m_j - \frac{1}{2}} \\ \sqrt{\ell + \frac{1}{2} - m_j}\, Y_\ell^{m_j + \frac{1}{2}} \end{pmatrix}. \tag{5.226}$$

With $\kappa = \ell + 1$, this is precisely the spinorial harmonic $\mathcal{Y}_{\kappa,m_j}$, as defined in (3.151). For $j = \ell - \frac{1}{2}$, (5.216) leads to

$$|\ell{-}\tfrac{1}{2}\,,\, m_j\rangle = \sqrt{\frac{1}{2\ell+1}} \begin{pmatrix} -\sqrt{\ell + \frac{1}{2} - m_j}\, Y_\ell^{m_j - \frac{1}{2}} \\ \sqrt{\ell + \frac{1}{2} + m_j}\, Y_\ell^{m_j + \frac{1}{2}} \end{pmatrix}. \tag{5.227}$$

With $\kappa = -\ell$, this agrees with the definition of $\mathcal{Y}_{\kappa,m_j}$ in (3.152).

Chapter 6

Quantum Information Theory

Chapter summary: The question of whether quantum mechanics violates the principle of locality and Einstein causality has been the subject of ongoing discussions. This is relevant for the theory of quantum communication, which makes use of the "nonlocal" correlations between the parts of a bipartite system in an entangled state. In Section 6.1, we describe the Bell states of two-qubit systems and discuss their measurement by means of exchanging locally available information.

In Section 6.2, we discuss the question of whether a manipulation of one qubit of an entangled pair can in any way influence the other qubit (which may be spatially separated from the first qubit). Closely related is the famous Einstein-Podolsky-Rosen (EPR) paradox, which is presented in Section 6.3. In Section 6.4, we describe the mathematical and statistical properties of the correlations arising from entanglement. In Section 6.5, we give an elementary derivation of Bell's inequality that is based on classical ideas about the correlation of measurement results ("local realism"). It has been confirmed experimentally that quantum mechanics violates Bell's inequality. While quantum mechanics is local in the sense required by Bell's theorem, it appears to violate the assumption of realism (that is, the assumption that observables do have values even if they are not measured).

The strange features of the correlations between the parts of an entangled system can be used for quantum information exchange. Alice and Bob can communicate rather efficiently if they share entangled pairs of qubits. The dense coding protocol describes the transmission of classical 2-bit information using a single qubit. The teleportation protocol sends one qubit with the help of two classical bits (see Section 6.6).

A quantum computer (Section 6.7) applies a sequence of unitary transformations ("reversible logic gates") to a quantum register (a system of n qubits). This transforms an initial state (the input) into a final state (the output) that is measured. In that way, any task that can be performed by a classical computer can also be performed by a quantum computer.

The ability of quantum logic gates to operate on superpositions of input states can be used to do computations with all possible classical inputs in parallel. For certain problems, one succeeds in using interference in a clever way so that the final measurement gives the desired result in a few steps, where a classical computer would need many steps. In Section 6.9, we give a detailed description of a simple algorithm, where a quantum computer is more efficient than a classical computer. An overview of important quantum algorithms concludes this chapter.

6.1. Entangled States of Two-Qubit Systems

In this chapter, we consider compound systems whose parts are simple qubits. Having dealt with the more formal aspects of composite systems in Chapter 5, it is now time to focus our interest on questions of interpretation.

6.1.1. Bases of entangled states

As explained in Section 5.2.4, the Hilbert space of the two-qubit system is $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$. We can choose the basis of product states (5.29). Whenever the compound system is in one of these four basis states, the states of the subsystems are pure states. But, we can also choose the Bell basis $\{\psi_{\mathrm{e}}^+, \psi_{\mathrm{e}}^-, \psi_{\mathrm{o}}^+, \psi_{\mathrm{o}}^-\}$ consisting of the states

$$\begin{aligned} \psi_{\mathrm{e}}^{\pm} &= \frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_+ \pm \psi_- \otimes \psi_-\big), \\ \psi_{\mathrm{o}}^{\pm} &= \frac{1}{\sqrt{2}}\big(\psi_+ \otimes \psi_- \pm \psi_- \otimes \psi_+\big). \end{aligned} \tag{6.1}$$

All Bell states are maximally entangled. Whenever the system is in a Bell state, the subsystems are in the maximally mixed state

$$\rho^A = \rho^B = \frac{1}{2}\,|\psi_+\rangle\langle\psi_+| + \frac{1}{2}\,|\psi_-\rangle\langle\psi_-| = \frac{1}{2}\,\mathbf{1}. \tag{6.2}$$

This state of a single qubit can be interpreted not only as a mixture of the states ψ_+ and ψ_- (spin-up and spin-down with respect to the z-axis), but, equivalently, as a mixture of $\psi_+(\mathbf{n})$ and $\psi_-(\mathbf{n})$ with respect to an arbitrary axis $\mathbf{n}$.

The states $\psi_{\mathrm{e,o}}^{\pm}$ are simultaneous eigenvectors of the observables

$$\sigma_3 \otimes \sigma_3 \quad \text{and} \quad \sigma_1 \otimes \sigma_1. \tag{6.3}$$

These two operators are self-adjoint in $\mathbb{C}^4$ and they commute:

$$[\sigma_3 \otimes \sigma_3, \sigma_1 \otimes \sigma_1] = 0. \tag{6.4}$$

We have

$$\begin{aligned} \sigma_3 \otimes \sigma_3\, \psi_{\mathrm{e}}^{\pm} &= \psi_{\mathrm{e}}^{\pm}, \qquad \sigma_1 \otimes \sigma_1\, \psi_{\mathrm{e}}^{\pm} = \pm\psi_{\mathrm{e}}^{\pm}, \\ \sigma_3 \otimes \sigma_3\, \psi_{\mathrm{o}}^{\pm} &= -\psi_{\mathrm{o}}^{\pm}, \quad \sigma_1 \otimes \sigma_1\, \psi_{\mathrm{o}}^{\pm} = \pm\psi_{\mathrm{o}}^{\pm}. \end{aligned} \tag{6.5}$$

So the eigenvalue of $\sigma_3 \otimes \sigma_3$ determines the index 'e' or 'o' (*parity bit*), and the eigenvalue of $\sigma_1 \otimes \sigma_1$ tells us whether the upper index is '+' or '−' (*phase bit*).

The compatibility of the observables $\sigma_3 \otimes \sigma_3$ and $\sigma_1 \otimes \sigma_1$ is remarkable, as the observables σ_3 and σ_1 do not commute in the single-qubit subspaces. It means that there exists a two-qubit experiment giving numerical values (+1

or -1) for both observables in a single run. The simultaneous measurement of the two observables can be used to prepare an arbitrary state of the Bell basis. This procedure meets great practical difficulties and will be discussed further in Section 6.1.4 below.

6.1.2. Global aspects of local measurements

It has been pointed out before that the state of a two-qubit system may contain information that is not accessible by measurements on the individual qubits. Quite generally, measurements on isolated subsystems are called *local measurements*. They have to be distinguished from *global measurements*, which are performed on the system as a whole. Only global measurements can reveal systemic properties (properties relating to the composite system as a whole).

The tensor product $S \otimes T$ of single-qubit operators S and T is a product of subsystem observables (see Section 5.4.2),

$$S \otimes T = (S \otimes \mathbf{1})\,(\mathbf{1} \otimes T). \tag{6.6}$$

An observable of this form can be measured locally, by measuring S on the first qubit and T on the second. But, the preparation of a Bell state involves the simultaneous measurement of two such tensor-product observables (6.3). This cannot be done with local measurements. In order to understand the difficulties, we need to discuss the involved procedures carefully. As you will see shortly, even the local measurement of $S \otimes T$ involves some global information exchange.

For this experiment, we need an ensemble of qubit pairs, all prepared in the same pure two-qubit state. After the preparation, we may separate the qubits without changing the state of the two-qubit system (the position of the qubits does not enter our description at all). We give one qubit of each pair to Alice in Auckland and the other to Bob in Barcelona. Due to the large spatial separation, any direct physical interaction between the qubits can be excluded. Hence, Alice and Bob can measure a property of one qubit without disturbing the other. Their measurements will be perfectly local. We ask Alice to measure the observable S on her qubits, and Bob to measure T on his qubits.



CD 5.9 shows an experiment of the type we have in mind. A source prepares a system in a certain two-qubit state, for example, a composite particle in the singlet state ψ_o^- with total angular momentum zero. Then, the particle decays into two separated qubits moving in opposite directions. Due to the conservation of the total angular momentum, the two-qubit system remains in the singlet state, although its parts get spatially separated by a large distance.

The following distinction between different variants of local measurements may prevent confusion.

(a) *Completely local measurements*: Alice and Bob ignore each other. They perform independent measurements and report only the final results of their statistical analysis.

(b) *Local measurements with information exchange*: Alice and Bob perform the measurements independently. For the final evaluation, they exchange information about the results obtained for the individual members of the ensemble.

Consider first the case (a). Alice and Bob both perform ensemble measurements on the qubits given to them. Both remain completely ignorant about the other's results. Nevertheless, Alice can easily obtain the mean value $\langle S\rangle$ of S, and Bob can determine the mean value $\langle T\rangle$. Unfortunately, it is not possible to determine the mean value of $S\otimes T$ from this information, because in general

$$\langle S\otimes T\rangle \neq \langle S\rangle\,\langle T\rangle. \tag{6.7}$$

As an example, consider $S = T = \sigma_3$ and assume that the two-qubit system was prepared in one of the Bell states (6.1). According to (6.5), the expectation value of $\sigma_3\otimes\sigma_3$ is either $+1$ or -1. But the mean value obtained by Alice is

$$\langle\sigma_3\rangle = \operatorname{Tr}\rho^A\sigma_3 = \operatorname{Tr}\sigma_3 = 0. \tag{6.8}$$

Similarly, because $\rho^A = \frac{1}{2}\mathbf{1} = \rho^B$, Bob will also obtain $\langle\sigma_3\rangle = 0$. Hence, it is not enough just to combine the expectation values obtained individually by Alice and Bob. Some additional information is missing, namely the information about the correlations of qubits belonging to the same pair.

Now, let us consider method (b). Alice and Bob try to determine the expectation value of $S\otimes T$ by performing individual measurements of S and T as in method (a). But this time, they keep the information that is necessary to identify the qubits that belong to the same pair. The qubits of each ensemble thus have to be numbered, and Alice and Bob need to record these numbers together with the measured results. Alice obtains results $s_1, s_2, \ldots$ for the observable S, and Bob gets the values $t_1, t_2, \ldots$ for T. In the final evaluation, they combine the individually measured values into the products $s_1t_1, s_2t_2, \ldots$ (the index shows us which values belong together). The mean value of these products approximates the expectation value $\langle S\otimes T\rangle$.

We summarize our observations as follows: In order to learn something about a two-qubit system from local measurements, Alice and Bob have to collaborate by recording additional information (the number of the qubit) together with the observed values and by communicating the results. We need to know which qubits belong together in order to obtain values for $S\otimes T$ from local measurements. Method (a) is sufficient to learn everything

about the subsystems. But we need method (b) to determine properties of the combined system.

6.1.3. Determining a Bell state

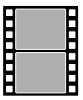
As an application of local measurements with information exchange, we describe a method to determine whether a two-qubit system is in one of the Bell states (6.1).

We consider again an ensemble of identically prepared qubit pairs. The individual qubits are sent to the "local observers" Alice and Bob who measure the observables σ_1 and σ_3 on their qubits. Because these observables are not compatible, Alice and Bob measure σ_1 on half of the qubits in the ensemble, and σ_3 on the other half. We assume that they make their choices independently and by independent tosses of a fair coin.

When the system is in a Bell state, we expect that the results of the local measurements will be $+1$ or -1 with probability $1/2$, because each subsystem is in the maximally mixed state $\rho = (1/2)\,\mathbf{1}$. Therefore, Alice and Bob will both get 0 for the expectation values of σ_1 and σ_3.

In order to learn something about the state of the two-qubit system, Alice and Bob have to proceed according to method (b), as discussed above. In the final evaluation, only those qubit-pairs can be used where Alice and Bob both measured the same observable (about half of the original ensemble). The information that allows them to identify these pairs (the ordered record of individual measurement results) makes the difference between methods (a) and (b). By combining their results, Alice and Bob can determine the expectation value of $\sigma_1 \otimes \sigma_1$ and of $\sigma_3 \otimes \sigma_3$. For a system in a Bell state, these expectation values must be either $+1$ or -1, and the state can be determined from (6.5).

> By local measurements with information exchange, Alice and Bob can determine whether or not a given ensemble of qubits is in one of the Bell states (and in which one).



CD 5.11 lets you perform a variant of the experiment just described. The source is known to prepare one of the four Bell states. Alice and Bob measure σ_1 and σ_3. You have to combine their results in order to determine which of the Bell states was prepared. Notice that without exchange of information, neither Alice nor Bob is able to learn anything about the state of the two-qubit system.

Can Alice and Bob *prepare* a Bell state with the same method? (That is, by measuring σ_1 and σ_3 separately and communicating the results.) This

question can be formulated as follows: Given an arbitrary ensemble of qubit pairs, can Alice and Bob select a subensemble for which both observables $\sigma_1 \otimes \sigma_1$ and $\sigma_3 \otimes \sigma_3$ have well-defined values? The answer is no. This is an important point, so let's have a closer look.

6.1.4. Preparing a Bell state

A Bell state is a common eigenvector of the commuting operators $\sigma_3 \otimes \sigma_3$ and $\sigma_1 \otimes \sigma_1$. We can project an arbitrary two-qubit state onto that eigenvector by first measuring the observable $\sigma_3 \otimes \sigma_3$ and then $\sigma_1 \otimes \sigma_1$ (or the other way round). The following discussion shows that this cannot be achieved by local measurements.

Two separated observers Alice and Bob can measure the value of $\sigma_3 \otimes \sigma_3$ for a single qubit pair by measuring σ_3 individually and by checking whether the results coincide (eigenvalue $+1$) or not (eigenvalue -1). After that, the qubit pair is in the corresponding eigenstate of $\sigma_3 \otimes \sigma_3$. But this procedure reveals too much! By measuring the qubits individually, we determine not only a coincidence, but also the individual values. In fact, using (6.6), Alice measures $\sigma_3 \otimes \mathbf{1}$ and Bob $\mathbf{1} \otimes \sigma_3$. As a result, the qubit pair is projected onto one of the four separable states $\psi_\pm \otimes \psi_\pm$. A subsequent measurement of $\sigma_1 \otimes \sigma_1$ would leave the state separable. We conclude that an entangled state like a Bell state cannot be prepared by measuring the qubits separately. Entanglement obviously describes information that is destroyed in a local measurement.

The eigenspace of $\sigma_3 \otimes \sigma_3$ belonging to the eigenvalue $+1$ is two-dimensional and consists of arbitrary (in general entangled) superpositions of $\psi_+ \otimes \psi_+$ and $\psi_- \otimes \psi_-$ (both spins up or both spins down). A local measurement as described above projects an entangled eigenstate onto a separable eigenstate. But, an ideal measurement of an observable should not change an eigenstate of that observable.

> A preparatory measurement of $\sigma_3 \otimes \sigma_3$ must determine whether the two spins are equal (eigenvalue $+1$) or opposite (eigenvalue -1), without determining the individual spins.

EXERCISE 6.1. *Show that* $P_{\pm,k} = \frac{1}{2}(\mathbf{1} \otimes \mathbf{1} \pm \sigma_k \otimes \sigma_k)$. *are the projections onto the eigenspaces of* $\sigma_k \otimes \sigma_k$. *Show that*

$$P_{+,3}P_{+,1} = P_{+,1}P_{+,3} = |\psi_{\mathrm{e}}^+\rangle\langle\psi_{\mathrm{e}}^+|. \tag{6.9}$$

Discuss the result and find similar formulas for the other Bell basis states.

EXERCISE 6.2. *Define the projection operators* $P^A_{\pm,k} = \frac{1}{2}(\mathbf{1} \pm \sigma_k) = P^B_{\pm,k}$ *in the Hilbert spaces of the individual qubits A and B. Show that* $P^A_{+,k} \otimes P^B_{+,k}$ *is an orthogonal projection operator. Show that*

$$P_{+,k} \neq P^A_{+,k} \otimes P^B_{+,k}. \tag{6.10}$$

Moreover, prove that $P^A_{+,1} \otimes P^B_{+,1}$ *does not commute with* $P^A_{+,3} \otimes P^B_{+,3}$ *and discuss these results.*

As explained above, a preparatory measurement of $\sigma_3 \otimes \sigma_3$ cannot be done by local measurements of the type defined in Section 6.1.2. Theoretically, a better method of measuring $\sigma_3 \otimes \sigma_3$ is to consider the total spin of the two-qubit system

$$\Sigma_3 = \mathbf{1} \otimes \sigma_3 + \sigma_3 \otimes \mathbf{1}. \tag{6.11}$$

We can measure this observable by combining the two qubits into a single composite particle that we send through a Stern-Gerlach analyzer. This is what we called a global measurement on the two-qubit system.

If the qubits are spin-1/2 particles, then the three states ψ^+_{e}, ψ^-_{e}, and ψ^+_{o} belong to the eigenspace with total angular momentum $j = 1$ whereas ψ^-_{o} is the singlet state with $j = 0$ (see Example 5.2 in Section 5.11.5). (Notice, however, that the possible values of Σ_3 are -2, 0, and 2.)

The square of Σ_3 is

$$\Sigma_3^2 = 2(\mathbf{1} \otimes \mathbf{1} + \sigma_3 \otimes \sigma_3). \tag{6.12}$$

We see that knowing the value of Σ_3^2 is equivalent to knowing the value of $\sigma_3 \otimes \sigma_3$. The Bell states $\psi^\pm_{\mathrm{o}}$ belong to the zero eigenvalue of Σ_3^2, the states $\psi^\pm_{\mathrm{e}}$ to the nonzero eigenvalue. A Stern-Gerlach analyzer for Σ_3^2 is shown in Figure 6.1. This device is basically a Stern-Gerlach apparatus for measuring Σ_3, where the partial beams belonging to the nonzero eigenvalues are recombined like in the interferometer described in Section 4.6.1. Without this recombination, the apparatus would eventually change the eigenstates of Σ_3^2 belonging to nonzero eigenvalues.

The operators Σ_3^2 and Σ_1^2 commute because of (6.4). We can now project onto one of the Bell states by measuring the simultaneous values of Σ_3^2 (the parity bit) and of Σ_1^2 (the phase bit). We show the necessary arrangement of Stern-Gerlach devices in Figure 6.2. A setup for measuring the Bell states of a two-qubit system is called a *Bell-state analyzer*.

A Bell state can only be prepared by a global measurement, for example, by a simultaneous measurement of Σ_3^2 and Σ_1^2, where Σ_j are the components of the total spin of the two-qubit system.

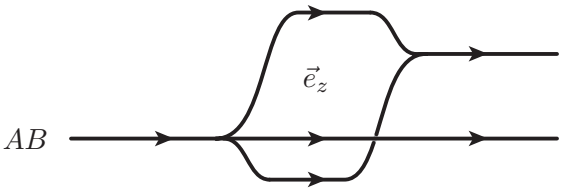


FIGURE 6.1. A measuring device for the square of the component S_3 of the total spin of a 2-qubit system. An inhomogeneous magnetic field splits a beam of combined qubits into three components (spin 1). The partial beams corresponding to nonzero spin-values are refocused in the manner of a Stern-Gerlach interferometer.

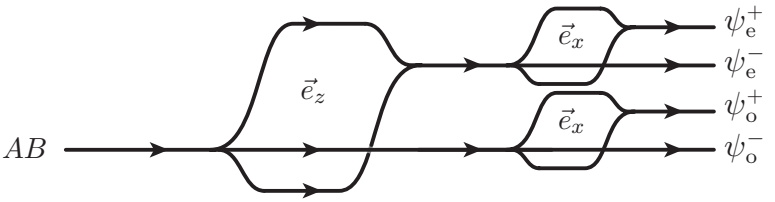


FIGURE 6.2. A Bell-state analyzer can (in principle) be built by combining measurement devices for Σ_3^2 and Σ_1^2.

The preparation and measurement of the entangled basis states is a major problem in the practical realization of two-qubit systems. The method described above requires that a (temporary) bound state of the two-particle system has to be formed without disturbing the spins of the particles (that is, without disturbing the spin-state of the system). This might be difficult in practice, but it illustrates that Bell basis measurements can be done at least in principle.[1]

6.2. Local and Nonlocal

6.2.1. Nonlocal interaction?

A measurement performed by Alice on a qubit of an entangled pair changes the state of Bob's qubit, without Bob being able to notice this change. Consider a two-qubit system, say, in the state ψ_{e}^{+}. Assume that Alice measures the spin in z-direction and finds "spin-up." This measurement corresponds

[1]Methods to produce Bell states in quantum optics are described in [**2**].

to the projection operator

$$P_+ \otimes \mathbf{1} \quad \text{with } P_+ = |\psi_+\rangle\langle\psi_+|. \tag{6.13}$$

The effect of this projection operator on ψ_{e}^+ is easily computed:

$$(P_+ \otimes \mathbf{1})\psi_{\mathrm{e}}^+ = \frac{1}{\sqrt{2}}\big((P_+\psi_+) \otimes \psi_+ + (P_+\psi_-) \otimes \psi_-\big) = \frac{1}{\sqrt{2}}\psi_+ \otimes \psi_+. \tag{6.14}$$

If Alice measures spin-up, this projects the state of the two-qubit system into the separable state $\psi_+ \otimes \psi_+$. Hence, Bob's qubit, which was in the mixed state $\rho^B = \mathbf{1}$, is in the state ψ_+ immediately after the measurement.

> A measurement performed by Alice may change the state of Bob's qubit, provided the two-qubit system is in an entangled state.

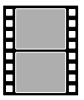
We want to show that this is in no way paradoxical and cannot be interpreted in terms of an instantaneous transfer of information.

First of all, we note that Bob cannot determine the state of his qubit in a single measurement, hence there is no way for him to tell if anything has changed as a consequence of Alice's measurement. Therefore, we must consider the situation in the context of an ensemble measurement. Assume that Alice and Bob share an ensemble of qubits in the state ψ_{e}^+. The projection operator P_+ is actually a Stern-Gerlach filter that allows Alice to select a subensemble of qubits with spin-up. But Bob does not know for which of his qubits the other has passed the spin-up test by Alice. Bob still owns a whole ensemble of qubits, and this ensemble just realizes the state $\rho = (1/2)\mathbf{1}$. Nothing has happened on Bob's side. There is no statistical test that can help Bob to decide whether Alice has performed any measurement or not.

The measurements done by Alice indeed select among Bob's qubits a subensemble characterized by the state ψ_+. Now, assume that Alice and Bob follow method (b) described in Section 6.1.2. That is, Alice finally tells Bob which of her qubits gave spin-up. Then he too is able to select the corresponding qubits from his ensemble. In that way, he obtains a subensemble that is characterized by the state ψ_+. The information acquired by Alice allows Bob to perform the projection P_+ on his qubits manually. It is remarkable that purely classical information is sufficient to select a subensemble in a pure state from qubits in a mixed state ρ. The difference between a pure and a mixed state is classical information.

The projection operator P_+ applied to Alice's ensemble has no effect on Bob's ensemble, unless we use the information gained by Alice to filter the corresponding qubits on Bob's side. As long as this information is not added to Bob's ensemble, the projection operator $P_+ \otimes \mathbf{1}$ is not really applied to the two-qubit system. In short: If Alice proceeds according to method (a)

(Section 6.1.2), she just performs the projection P_+ on her qubits and *nothing* happens to Bob's qubits. In order to measure the two-qubit observable $P_+ \otimes \mathbf{1}$, Alice and Bob have to proceed according to method (b) (that is, by exchanging information via classical communication) thereby applying the projection also to the ensemble on Bob's side. In practice, the transmission of information has to obey the constraints imposed by the theory of relativity.



CD 5.12 illustrates the discussion in this section. As soon as Alice measures the spin of her qubit, she can predict Bob's result. But this fact changes nothing on Bob's side. He still finds his qubits in a maximally mixed state. As soon as Alice sends information about her results to Bob, he is able to select a subensemble of qubits in a pure state.

I think we can learn from this an important lesson: A quantum mechanical state is the collection of physically available information about a system. The state of Bob's ensemble can only be changed if we put some information into it. When Alice performs her measurement, the information is not available at Bob's location, hence the state of Bob's qubit ensemble is not changed automatically by Alice's actions. The information gained by Alice first has to be transferred to Bob by means of some physical process (classical communication). Local measurements in a bipartite system do not establish a "nonlocal interaction" or a "superluminal communication," as it is often claimed.

6.2.2. Local manipulations

By means of purely local unitary transformations, Alice can change the state of the two-qubit system. She can, for example, apply the unitary transformation σ_1 to her qubit (or rather to all qubits of her ensemble). This changes the state of the compound system according to

$$\begin{aligned}(\sigma_1 \otimes \mathbf{1})\psi_{\mathrm{e}}^{\pm} &= \frac{1}{\sqrt{2}}\big((\sigma_1\psi_+)\otimes\psi_+ + (\sigma_1\psi_-)\otimes\psi_-\big)\\ &= \frac{1}{\sqrt{2}}(\psi_-\otimes\psi_+ \pm \psi_+\otimes\psi_-) = \pm\psi_{\mathrm{o}}^{\pm}, \qquad (6.15)\\ (\sigma_1 \otimes \mathbf{1})\psi_{\mathrm{o}}^{\pm} &= \pm\psi_{\mathrm{e}}^{\pm}. \qquad (6.16)\end{aligned}$$

The second formula follows immediately from the first using $\sigma_1^2 = 1$.

EXERCISE 6.3. *Show that if Alice performs the unitary transformation σ_3 on her qubit, she can change the phase bit of a maximally entangled state*

of the two-qubit system, that is,

$$(\sigma_3 \otimes \mathbf{1})\psi_{\mathrm{e}}^{\pm} = \psi_{\mathrm{e}}^{\mp}, \qquad (\sigma_3 \otimes \mathbf{1})\psi_{\mathrm{o}}^{\pm} = \psi_{\mathrm{o}}^{\mp}. \tag{6.17}$$

No local unitary transformation of the form $U \otimes \mathbf{1}$ can in any way influence subsystem B. The state of Bob's qubit is not changed by Alice's manipulations. You can see this most clearly if you write the state ψ of the compound system in the Gram-Schmidt form (5.130)

$$\psi = \sum_i \sqrt{p_i}\ \phi_i^A \otimes \phi_i^B \tag{6.18}$$

and apply $U \otimes \mathbf{1}$ to this expression. This gives

$$(U \otimes \mathbf{1})\,\psi = \sum_i \sqrt{p_i}\ (U\phi_i^A) \otimes \phi_i^B. \tag{6.19}$$

The resulting vector is again in the Gram-Schmidt form, because $\{U\phi_i^A\}$ is just another orthonormal basis in $\mathfrak{H}^A$. Hence, $U \otimes \mathbf{1}$ has no influence on the subsystem B, which, according to (5.131), remains in the state

$$\rho^B = \sum_i p_i\, |\phi_i^B\rangle\langle\phi_i^B|. \tag{6.20}$$

The two states ψ and $(U\otimes\mathbf{1})\psi$ of the compound system are just two different purifications of ρ^B.

A local unitary transformation U is a product of subsystem transformations,

$$U = U^A \otimes U^B, \quad \text{with } U^A,\, U^B \text{ unitary.} \tag{6.21}$$

Such a transformation cannot change the entanglement of the system. If ψ is separable, then $U\psi$ is also separable. If ψ is entangled, so is $U\psi$.

6.3. The Einstein-Podolsky-Rosen Paradox

Entangled two-qubit states may be used to illustrate a paradox described by Einstein, Podolsky, and Rosen (EPR) in 1935. The main question is the following. Given two incompatible observables S and T, quantum mechanics does not predict the values of T in an experiment where only S is measured. Is this because of our insufficient knowledge of reality, or is it a property of reality itself? Is quantum mechanics incomplete, or could it be that the observable simply does not have a definite value?

Einstein, Podolsky, and Rosen strongly favored the point of view that the statistical interpretation of a wave function describes the state of knowledge of an observer and not so much an actual property of the system. Although they did not question the correctness of quantum mechanics, it was their

goal to attack the completeness of quantum mechanics as claimed by the Copenhagen interpretation. They reasoned as follows.

First, they attempted to give a definition of physical reality:

Physical reality:

If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.

For example, if the state is described by an eigenvector of an observable S belonging to the eigenvalue λ, then in a measurement of the observable S, the value λ will be found with probability 1 (that is, with certainty). Hence, the value of S may be regarded as an element of reality.

The EPR paper also contains a definition of what the authors consider to be a complete theory.

Completeness of a theory:

A theory is complete if every element of the physical reality has a counterpart in the physical theory.

If two observables, for example the components σ_3 and σ_1 of the spin, do not commute, then knowledge of σ_3 makes σ_1 completely undetermined. The principal question is whether this uncertainty is due to an incomplete knowledge of the observer, or whether this is a "defect" of reality. The Copenhagen interpretation considers quantum mechanics a complete theory and hence takes the point of view that the values of σ_1 and σ_3 are not simultaneously elements of reality.

Suppose we could show that, by the definition of physical reality given above, we can find a situation where the values of σ_1 and σ_3 are simultaneous elements of the physical reality. Then, it is shown that quantum theory is incomplete, because nothing in the theory corresponds to these values.

EPR considered an entangled pair as an example of this situation. Suppose Alice and Bob share an ensemble of entangled qubit pairs. We assume that the state of the compound system is ψ_{o}^{-}. Among the entangled states, Bell's singlet state ψ_{o}^{-} is distinguished by the property

$$(\sigma_k \otimes \mathbf{1})\,\psi_{\mathrm{o}}^{-} = -(\mathbf{1} \otimes \sigma_k)\,\psi_{\mathrm{o}}^{-}, \quad k = 1, 2, 3. \tag{6.22}$$

You can verify this property by an explicit calculation. We may formulate this result as follows.

Anticorrelation of spin in Bell's singlet state:

Consider a two-qubit system in the state $\psi_{\rm o}^-$. Then, the measurement of an arbitrary spin component σ_k of qubit A gives always the opposite of the result obtained by measuring σ_k on qubit B. We say the values of σ_k for qubit A and B are *strictly anticorrelated*.

It has to be stressed that this is indeed a prediction for the coincidence of measurement results for individual qubit pairs. This can be seen most clearly by noting that $\psi_{\rm o}^-$ is a simultaneous eigenstate of the observables $\sigma_k \otimes \sigma_k$ (see Section 6.1.2),

$$(\sigma_k \otimes \sigma_k)\, \psi_{\rm o}^- = -\psi_{\rm o}^-, \quad k = 1, 2, 3. \tag{6.23}$$

Therefore, a measurement of $\sigma_k \otimes \sigma_k$ in the state $\psi_{\rm o}^-$ gives -1 with certainty. Measuring σ_k on the individual qubits of a pair gives opposite values for every single qubit pair.



CD 5.9 shows a two-qubit system in the state $\psi_{\rm o}^-$. We simulate a sequence of elementary measurements of the observable $\boldsymbol{\sigma} \cdot \mathbf{n} \otimes \boldsymbol{\sigma} \cdot \mathbf{n}$ with respect to three different directions $\mathbf{n}$. The results for the subsystems are strictly anticorrelated. CD 5.10 shows a classical two-particle system whose three observable properties show the same anti-correlations between the subsystems.

For a two-qubit system in the state $\psi_{\rm o}^-$, the state of either subsystem is the maximally mixed state $\rho = \frac{1}{2}\,\mathbf{1}$. Hence, according to quantum mechanics, the value of σ_3 is completely undetermined for both qubits of the entangled pair. Now, let us assume that Alice measures σ_3. As usual, we assume that Alice and Bob are spatially separated, so that Bob's qubit remains undisturbed by anything Alice does with her qubit. Whenever she finds spin-up in an elementary experiment, then she knows that Bob would obtain spin-down for the other qubit of the pair. Hence, without disturbing Bob's qubit in any way, Alice can predict the value of σ_3 for Bob's qubit. Thus, the value of σ_3 for Bob's qubit must exist beforehand, because the measurement done by Alice cannot influence Bob's qubit. The value of σ_3 for Bob's qubit is thus an element of reality whether Bob performs a measurement of σ_3 or not. In the quantum mechanical formalism, this statement is expressed by the fact that the observables $\sigma_3 \otimes \sigma_3$ and $\sigma_3 \otimes \mathbf{1}$ are compatible. Moreover, a measurement of any of these observables in the state $\psi_{\rm o}^-$ projects onto the same state of the two-qubit system, either onto $\psi_+ \otimes \psi_-$ or onto $\psi_- \otimes \psi_+$.

But Alice could also perform a measurement of, say, σ_1 instead of σ_3. Because of the spatial separation, the decision to measure σ_1 should not change the qubit on Bob's side. Hence the value of σ_3 for Bob's qubit should

remain an element of reality, no matter, what Alice decides to do. Moreover, if Alice measures σ_1, then the value of σ_1 for Bob's qubit is also an element of reality. EPR concluded that therefore the values of both observables σ_1 and σ_3 simultaneously belong to the same reality. Because nothing in the quantum mechanical theory lets us predict these values, quantum mechanics must be incomplete—at least in the opinion of EPR.

One might object that Alice still cannot measure the simultaneous values of σ_1 and σ_3. That is true, but if Alice and Bob work together, Alice could measure the value of σ_1 on her qubit and Bob could measure the value of σ_3 on the second qubit of the entangled pair. Because the result of Alice's measurement lets us predict the value of σ_1 for Bob's qubit, there is no need for Bob to determine σ_1, and he can determine σ_3 instead. This possibility of speaking about the simultaneous values of σ_1 and σ_3 (which are therefore an element of the EPR-reality) without the possibility of predicting them lets the theory appear incomplete. The following box describes what quantum mechanics has to say about this situation.

The incompleteness of quantum mechanics:

The two-qubit observable $\sigma_1 \otimes \sigma_3$ is not compatible with $\sigma_1 \otimes \sigma_1$ or with $\mathbf{1} \otimes \sigma_1$. Hence, in an experiment where Alice measures σ_1 and Bob measures σ_3, no value is associated with those other observables, in particular, the value of σ_1 for Bob's qubit remains undetermined.

A statement describing what "Bob would have obtained if he had performed another measurement instead" is called *counterfactual*. It refers to an experimental setup that has never been realized. Without recourse to counterfactual arguments, it is not possible to speak about the simultaneous values of noncompatible observables (like different components of the spin, or like position and momentum). Quantum mechanics is indeed incomplete in the sense of EPR, but it is incomplete only in view of counterfactual statements. Quantum mechanics does give a "complete" description of experimental situations that are actually realized. In a measurement of $\sigma_1 \otimes \sigma_1$, the value of σ_1 for Bob's qubit is indeed an element of the quantum reality, in a measurement of $\sigma_1 \otimes \sigma_3$, it is not.

According to quantum mechanics, an observable (like a component of the spin of Bob's qubit) does not have a definite value if actually no attempt is made to measure this observable. Recall that the state of the system is the collection of physically available information about this system (in an operational sense). In a situation where a measurement of an observable cannot be performed (because some other, noncompatible observable is measured instead), the information about the value of the observable does

not exist, and hence the corresponding property does not exist in the state of the system. All possibilities for that property are left open. We have learned numerous situations where the presence of two possibilities leads to interference and hence to observable effects (Schrödinger-cat state, double-slit experiment). Assuming that actually one of the possibilities is realized destroys the very reason for that interference. We want to stress that this "denial of counterfactuals" is built deeply into the formalism of quantum mechanics: a priori, an observable is not represented by a value, but by an operator in the Hilbert space of the system, and hence by all its possible values simultaneously. The idea that the possible values are actually described by some hidden parameters will be put to a test in Section 6.5 below.

6.4. Correlations Arising from Entangled States

Entanglement contains information about the system that can be expressed as correlations between results of measurements on the subsystems. Einstein, Podolsky, and Rosen used these correlations to construct a paradox. The objection by EPR is serious enough that it is worth investigating the correlations in entangled qubit pairs further.

6.4.1. Joint probabilities in two-qubit systems

A measurement of an observable $\hat{S}$ is a random experiment. With respect to that experiment, we define a random variable S that describes the measured values of the observable $\hat{S}$ (see Section 4.2.1). In this section and in the following, we use, wherever necessary, the hat ($\hat{}$) to distinguish an observable from the associated random variable.

Given a two-qubit system, we choose two unit vectors $\mathbf{n}$ and $\mathbf{m}$ in $\mathbb{R}^3$ and define the subsystem observables

$$\hat{S} = \boldsymbol{\sigma} \cdot \mathbf{n} \otimes \mathbf{1}^B, \quad \hat{T} = \mathbf{1}^A \otimes \boldsymbol{\sigma} \cdot \mathbf{m}. \tag{6.24}$$

Because of (5.61), these observables are compatible. Therefore, they can be measured simultaneously, and the results are described by two ± 1-valued random variables S and T.

We denote the *joint probability* that the measurement of $\hat{S}$ gives $S = \epsilon$ and the measurement of $\hat{T}$ gives $T = \delta$ by

$$\mathrm{prob}(S{=}\epsilon, T{=}\delta), \quad \epsilon, \delta = \pm 1. \tag{6.25}$$

Note that the joint probability can only be defined for compatible observables.

Given the state ψ of the two-qubit system, we can compute the joint probability (6.25) as follows. Let $P_\epsilon \otimes \mathbf{1}$ be the projection operator onto the eigenspace of $\hat{S}$ belonging to the eigenvalue ϵ. Similarly, define $\mathbf{1} \otimes Q_\delta$ as the

projection operator onto the eigenspace of $\hat{T}$ belonging to the eigenvalue δ. From (6.24) we conclude

$$P_{\pm 1} = \frac{1}{2}\big(\mathbf{1} \pm \boldsymbol{\sigma} \cdot \mathbf{n}\big), \quad Q_{\pm 1} = \frac{1}{2}\big(\mathbf{1} \pm \boldsymbol{\sigma} \cdot \mathbf{m}\big). \tag{6.26}$$

We have $P_1 + P_{-1} = \mathbf{1}$, and $Q_1 + Q_{-1} = \mathbf{1}$. The operator

$$(P_\epsilon \otimes \mathbf{1})(\mathbf{1} \otimes Q_\delta) = P_\epsilon \otimes Q_\delta. \tag{6.27}$$

is also an orthogonal projection operator, because it is the product of two commuting projection operators. It projects onto the simultaneous eigenspace of $\hat{S}$ and $\hat{T}$ belonging to the eigenvalues (ϵ, δ), and hence

$$\text{prob}(S{=}\epsilon, T{=}\delta) = \langle \psi \,,\, P_\epsilon \otimes Q_\delta \, \psi \rangle. \tag{6.28}$$

Two random variables S and T are *independent* if their joint probability equals a product of independent factors:

$$\text{prob}(S{=}\epsilon, T{=}\delta) = \text{prob}(S{=}\epsilon)\, \text{prob}(T{=}\delta). \tag{6.29}$$

In that case, the outcome of a measurement of $\hat{S}$ and the result of a simultaneous measurement of $\hat{T}$ are completely independent of each other. This is indeed the case whenever the state ψ of the two-qubit system is separable. From $\psi = \psi^A \otimes \psi^B$ it follows that

$$\begin{aligned} \text{prob}(S{=}\epsilon, T{=}\delta) &= \langle \psi^A \otimes \psi^B \,,\, (P_\epsilon \otimes Q_\delta)\, \psi^A \otimes \psi^B \rangle \\ &= \langle \psi^A, P_\epsilon \, \psi^A \rangle \, \langle \psi^B \,, Q_\delta \, \psi^B \rangle \\ &= \text{prob}(S{=}\epsilon)\ \text{prob}(T{=}\delta). \end{aligned} \tag{6.30}$$

For an entangled state of the two-qubit system, we expect correlations between the observables of the subsystems. A correlation is an association between two observables that lets us predict, to some extent, the value of one observable by a measurement of the other. The best example is the state ψ_{o}^-, as discussed in Section 6.3 (see Exercise 6.4).

We are going to describe the correlations between two commuting observables $\hat{S}$ and $\hat{T}$ by the probability that a measurement of $\hat{S}$ and $\hat{T}$ gives the same result. This is called the probability for *coincidence* and is denoted by $\text{prob}(S{=}T)$. For observables with ± 1 as the only eigenvalues, we have

$$\text{prob}(S{=}T) = \frac{1}{2}\,\big(1 + \langle ST \rangle\big). \tag{6.31}$$

The expectation value $\langle ST \rangle$ is the *correlation coefficient* of the two ± 1-valued random variables S and T. Note that the expectation value of a product makes sense only if the operators commute; otherwise $\langle ST \rangle = \langle \psi, \hat{S}\hat{T}\psi \rangle$ need not be a real number.

EXERCISE 6.4. *Consider a two-qubit system in the state* $\psi_{\rm o}^-$. *Choose* $\hat{S} = \boldsymbol{\sigma}\cdot\mathbf{n}\otimes\mathbf{1}$ *and* $\hat{T} = \mathbf{1}\otimes\boldsymbol{\sigma}\cdot\mathbf{n}$. *Show that*

$$\text{prob}(S{=}1, T{=}{-}1) = 1 \;\neq\; \text{prob}(S{=}1)\ \text{prob}(T{=}{-}1) = \frac{1}{4}. \tag{6.32}$$

EXERCISE 6.5. *Let* S *and* T *be random variables with* ± 1 *as the only possible values. Show that the correlation coefficient* $\langle ST\rangle$ *can be written as*

$$\begin{aligned}\langle ST\rangle = {}& \text{prob}(S{=}1, T{=}1) + \text{prob}(S{=}{-}1, T{=}{-}1)\\ & - \text{prob}(S{=}1, T{=}{-}1) - \text{prob}(S{=}{-}1, T{=}1).\end{aligned} \tag{6.33}$$

EXERCISE 6.6. *Show that for any* ± 1*-valued random variables* S *and* T *the following is true:*

$$\text{prob}(S{=}T) = \text{prob}(S{=}1, T{=}1) + \text{prob}(S{=}{-}1, T{=}{-}1) \tag{6.34}$$

$$= \frac{1}{2}\big(1 + \langle ST\rangle\big). \tag{6.35}$$

EXERCISE 6.7. *Use* (6.33) *to show that if* S *and* T *are independent, then* S *and* T *are uncorrelated, that is,*

$$\langle ST\rangle = \langle S\rangle\langle T\rangle. \tag{6.36}$$

EXERCISE 6.8. *Assume that the two-qubit system is in the state* $\psi_{\rm o}^-$. *Compute the correlation coefficient* $\langle ST\rangle$ *for* $\hat{S} = \sigma_3\otimes\mathbf{1}$ *and* $\hat{T} = \mathbf{1}\otimes\sigma_1$.

EXERCISE 6.9. *Let* $\hat{S}$ *be a qubit observable with the eigenvalues* ± 1 *and corresponding eigenvectors* $\psi_\pm$. *Define the orthogonal projection operators* $P_\pm = |\psi_\pm\rangle\langle\psi_\pm|$. *Show that*

$$P_+ + P_- = \mathbf{1}, \quad P_+ - P_- = \hat{S}, \quad P_\pm = \frac{1}{2}\,(1 \pm \hat{S}). \tag{6.37}$$

EXERCISE 6.10. *Consider a two-qubit system. Let* $\hat{S}$ *be an observable in the Hilbert space of qubit* A *with eigenprojections* $P_\pm$ *as described in the previous exercise. Assume that* $\hat{T}$ *is a similar observable for qubit* B *and denote the corresponding projection operators by* $Q_\pm$. *Show that*

$$P_+\otimes Q_+ + P_-\otimes Q_- = \frac{1}{2}\,(\mathbf{1} + \hat{S}\otimes\hat{T}), \tag{6.38}$$

$$P_+\otimes Q_- + P_-\otimes Q_+ = \frac{1}{2}\,(\mathbf{1} - \hat{S}\otimes\hat{T}). \tag{6.39}$$

Are these orthogonal projection operators?

EXERCISE 6.11. *Assume that the two-qubit system is in the state* $\psi_{\rm o}^-$. *Compute the correlation coefficient* $\langle ST\rangle$ *for* $\hat{S} = \sigma_3\otimes\mathbf{1}$ *and* $\hat{T} = \mathbf{1}\otimes\sigma_1$.

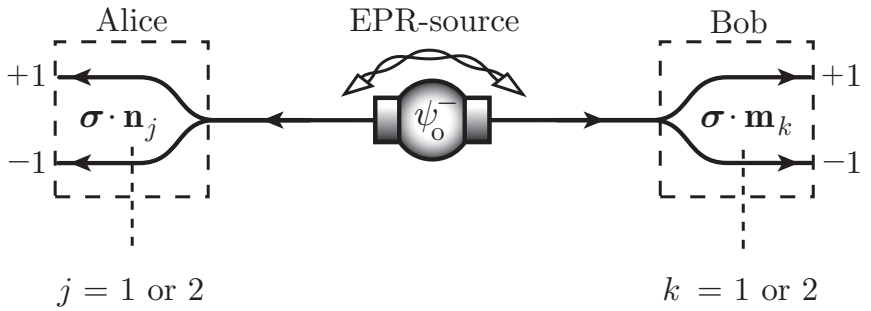


FIGURE 6.3. Setup for an experiment testing EPR correlations.

6.4.2. A protocol for verifying joint probabilities

In order to verify the predictions about prob($S{=}\epsilon, T{=}\delta$), we have to perform an ensemble measurement on a large number of identical qubit pairs. Schematically, a possible setup for this experiment is shown in Figure 6.3. A source prepares pairs of qubits in a certain two-qubit state. As usual, one qubit is sent to Alice, the other to Bob. In order to exclude any mutual influence of measurements on the subsystems, we may assume that Alice and Bob are spatially separated by a large distance and well isolated from each other. Thus, we make sure that the only connection between the subsystems is the preparation as a two-qubit state, that is, their common origin.

Quantum mechanics does not distinguish between ensemble measurements performed in parallel (same time, different places) or sequentially (same place, one after another). In this setup we perform sequential trials, corresponding to the way that this experiment is actually carried through. This provides an automatic numbering of the qubits, which is useful because we want to compare the results for the qubits belonging to the same pair (method (b) in Section 6.1.2).

In view of the discussion in th following section, we now introduce four observables by choosing four directions $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{m}_1$, $\mathbf{m}_2$ in $\mathbb{R}^3$ and setting

$$\hat{S}_j = \boldsymbol{\sigma}\cdot\mathbf{n}_j \otimes \mathbf{1}^B, \quad \hat{T}_k = \mathbf{1}^A \otimes \boldsymbol{\sigma}\cdot\mathbf{m}_k, \quad j,k = 1,2. \tag{6.40}$$

Notice that each observable $\hat{S}_j$ is compatible with each observable $\hat{T}_k$, because these operators act on different parts of the system. Hence, one of the observables $\hat{S}_j$ can be measured simultaneously with one of the observables $\hat{T}_k$. It is, however, impossible to obtain values for $\hat{S}_1$ *and* $\hat{S}_2$ in a single run of the experiment whenever $\mathbf{n}_1 \neq \mathbf{n}_2$. In quantum mechanics, the random variables S_1 and S_2 are never defined in the same experiment.

In the next section, we are going to derive the inequality (6.53) involving the four coincidence probabilities prob($S_1{=}T_1$), prob($S_1{=}T_2$), prob($S_2{=}T_1$),

and prob(S_2=T_2). In an experiment, however, where the values of $\hat{S}_1$ and $\hat{T}_1$ are measured, the random variables S_2 and T_2 are not defined. In order to measure these probabilities, we therefore need four different ensemble measurements. Usually, one proposes the following protocol to determine all joint probabilities

$$\text{prob}(S_j{=}\epsilon, T_k{=}\delta), \quad j,k = 1,2, \quad \epsilon,\delta = \pm 1. \tag{6.41}$$

This protocol is known as the *EPR protocol*.

In each trial, the source emits two qubits prepared in a certain two-qubit state. One qubit is sent to Alice, the other to Bob. When the qubits are on their way, Alice chooses a random value for the index j (1 or 2 with probability 1/2), and when the qubit arrives, she measures the corresponding observable S_j. Similarly and independently, Bob makes a random choice for k and measures T_k on his qubit. The large distance and the delayed choice prevents any unwanted communication between Alice and Bob and guarantees that they will make their decisions independently. This procedure is repeated many times with identically prepared qubit pairs. According to these conditions, any particular pair of observables $(\hat{S}_j, \hat{T}_k)$ is measured in about a quarter of all trials.



The experimental setup described in this section can be seen in CD 5.13. In this simulation, the source generates a pair of qubits in the state ψ_o^-. Alice and Bob both measure the component of the spin with respect to two randomly chosen directions. The experiment serves to falsify the CHSH inequality discussed in Section 6.5 below.

When all measurements have been performed, Alice can estimate the probabilities prob(S_j=ϵ) for both observables $\hat{S}_j$ ($j = 1,2$) and for $\epsilon = \pm 1$, Likewise, without consulting Alice, Bob can determine the four probabilities prob(T_k=δ) (with $k = 1,2$ and $\delta = \pm 1$). But, in order to compute the joint probabilities for S_j and T_k, Alice and Bob must get together to compare their results for the individual qubit pairs. The joint probabilities are estimated by counting:

$$\text{prob}(S_j{=}\epsilon, T_k{=}\delta) \approx \frac{N(S_j{=}\epsilon, T_k{=}\delta)}{N_{jk}}. \tag{6.42}$$

Here, N_{jk} is the total number of those trials where Alice had measured $\hat{S}_j$ and Bob had measured $\hat{T}_k$. $N(S_j{=}\epsilon, T_k{=}\delta)$ is the number with the desired result S_j=ϵ, and T_k=δ. Note that the quantum mechanical probability only refers to the subset of trials where the observables $\hat{S}_j$ and $\hat{T}_k$ are actually measured (in the other trials, the random variables S_j and T_k are not defined). Hence, the number N_{jk} in the denominator is not the total number

of trials. From the point of view of quantum mechanics, the set of trials according to the EPR protocol decomposes into four disjoint subsets. On each subset, different pairs of random variables are defined because different sets of observables are measured.

A probability that refers only to a subset of trials where some condition is satisfied, is usually called a *conditional probability*. Within the random experiment according to the EPR protocol, the quantum mechanical probability $\mathrm{prob}(S_j{=}\epsilon, T_k{=}\delta)$ could thus be interpreted as the conditional probability for the event $S_j = \epsilon$ and $T_k = \delta$, provided that the observables $\hat{S}_j$ and $\hat{T}_k$ are measured. Note, however, that the same event is not even defined without this condition.

EXERCISE 6.12. *The joint probabilities* $\mathrm{prob}(S_j{=}\epsilon, T_k{=}\delta)$ *have certain properties that we can expect to be true a priori. Using* (6.28)*, prove and explain the property*

$$\sum_{\epsilon,\delta=\pm 1} \mathrm{prob}(S_j{=}\epsilon, T_k{=}\delta) = 1, \quad \textit{for all } j,k = 1,2. \tag{6.43}$$

EXERCISE 6.13. *Show that according to quantum mechanics, the joint probabilities satisfy the locality condition*

$$\mathrm{prob}(\hat{S}_j{=}\epsilon) = \sum_{\delta=\pm 1} \mathrm{prob}(\hat{S}_j{=}\epsilon, \hat{T}_1{=}\delta) = \sum_{\delta=\pm 1} \mathrm{prob}(\hat{S}_j{=}\epsilon, \hat{T}_2{=}\delta). \tag{6.44}$$

This result means the following: The probability that Alice obtains the result $\hat{S}_j = \epsilon$ *does not depend on whether Bob measures the observable* $\hat{T}_1$ *or* $\hat{T}_2$*.*

6.5. Bell Inequalities and Local Hidden Variables

In 1964 J.S. Bell derived a famous inequality for the correlations in a two-qubit system that can be tested by the experimental setup described in Section 6.4.2. The proof of Bell's theorem needs, in particular, an assumption of realism and an assumption of locality, both well established from the point of view of classical physics. Moreover, Bell showed that quantum mechanics predicts a clear violation of Bell's inequality. Experiments carried out in 1982 by A. Aspect and coworkers (and by many others since) confirm the predictions of quantum mechanics.

It is often claimed that quantum mechanics is a nonlocal theory, because the violation of Bell's inequality can explained by a nonlocal influence between spatially separated qubits. But, the assumption of locality needed for Bell's theorem is perfectly satisfied by quantum mechanics.

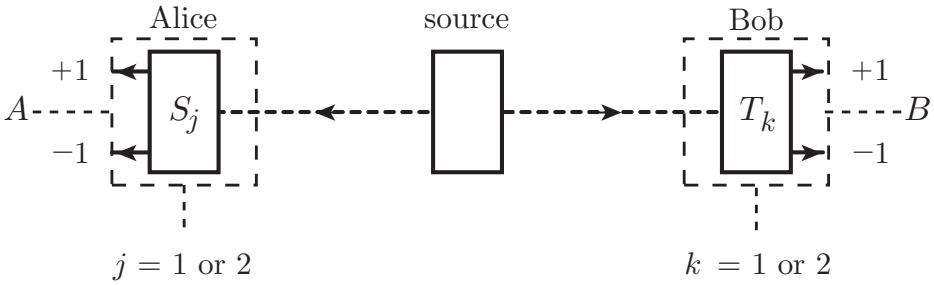


FIGURE 6.4. The general EPR setup. A source provides some information Alice and Bob independently use to produce the outputs A and B (± 1-valued random variables). On either side, the output is generated by sending the information through a machine (measurement device) that can operate in two possible modes. The mode of operation is determined by the two random variables j and k.

On the other hand, realism (as required in the formulation of Bell's inequality) assumes that the value of an observable $\hat{S}_1$ exists in every elementary experiment, even if another observable $\hat{S}_2$ is measured that is not compatible with $\hat{S}_1$. This assumptions means that we associate random variable S_1 and S_2 with both observables. Only the value of the actually measured observable is known, the other value remains unknown. This is precisely the sort of counterfactual realism discussed in Section 6.3. This type of realism has no counterpart in the formalism of quantum mechanics (where observables are represented by operators and not by values).

Hence, quantum mechanics can violate Bell's inequality without violating any principle of locality. Quantum mechanics is local but not realistic. The experimentally verified violation of Bell's inequality rather indicates that any realistic theory behind quantum mechanics (that is, a hidden variable theory) would have to be nonlocal.

6.5.1. General setup for measuring correlations

We want to describe an experiment for measuring the subsystem correlations in very general terms. The protocol for this experiment is the same as the EPR protocol described in Section 6.4.2. The experimental setup according to Figure 6.4 is a special case of the setup discussed here.[2]

[2]Our discussion in part follows the paper by Gill, R.D., Weihs, G., Zeilinger, A., and Zukowski M., Europhys. Lett. **61**, 282-283 (2003), quant-ph/0204169. Note, however, that their definition of "locality" is expressed in terms of hidden variables and differs slightly from ours, which is expressed in terms of statistical independence (see below).

A general EPR experiment consists in a sequence of trials with the following experimental setup. For each trial, a source generates some message and sends it to Alice and to Bob. The message could be anything: classical bits and bytes or physical particles or electromagnetic waves. Alice and Bob have measurement devices A and B that convert the incoming message either into the output $+1$ or into -1. The two devices A and B are regarded as black boxes. Depending on the physical realization, these could be Stern-Gerlach apparatus, computers, or coin-tossing machines. (It does not matter whether these devices work in a deterministic or random way).

The devices A and B can operate in two different modes, depending (say) on the position of a switch. In case of Stern-Gerlach devices, the two modes correspond to two possible directions of the magnetic field. For each trial, the mode of operation is chosen randomly. It is assumed that Alice and Bob make these choices independently, for example, by tossing a coin, in such a way that each of the possible combinations $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ occurs with the same probability $1/4$.

The whole setup is designed to make sure that on both sides only the locally available information is used to generate the output. Alice is not allowed to use any information generated on Bob's side, and vice versa. The only thing that connects Alice and Bob is the common origin of the message they receive.

6.5.2. Realism and hidden variables

Mathematically, we may describe the results obtained by Alice and Bob as ± 1-valued random variables A and B. There are two more random variables, namely the indices j and k describing the settings chosen by Alice and Bob for their devices. The indices j and k have the values 1 and 2 with equal probability $1/2$.

Let us now make our first assumption. It is related to the idea of hidden variables.

Assumption of realism:

The possible outputs A and B can be described by four ± 1-valued random variables S_1, S_2, T_1, and T_2. These variables describe the output for the corresponding settings:

$$A = S_j, \quad B = T_k. \tag{6.45}$$

According to the EPR setup described in the previous section, only two of the four quantities are actually observed in each trial. But random variables have (by definition) well-defined values for each trial, whether they

are observed or not. Hence, the assumption of realism makes it meaningful to formulate counterfactual statements like "Alice would have observed the value of S_2 if she had chosen the setting $j = 2$ instead of $j = 1$ for her device."

The assumption of realism is very natural from the point of view of classical physics: Here, the source emits classical information (bits and bytes) or classical physical systems. In principle, Alice can always duplicate this message and send the two copies of the input to two identical clones of her device. She may set one device to mode "1", the other to mode "2", and thus generate the outputs for both settings simultaneously in a single trial. By tossing a coin, Alice now decides which output she keeps and which she ignores. In each trial, the values of S_1 and S_2 can both be assumed to exist, although according to the EPR protocol, only one of these values is actually recorded.

As explained at the end of Section 6.3, the assumption of realism is very problematic (to say the least) in quantum mechanics, because it attempts to justify counterfactual statements. Whenever the observables corresponding to S_1 and S_2 are not compatible, quantum mechanics says nothing about the value of S_2 in a single trial where S_1 is measured. All but the outputs that are actually observed must be regarded as *hidden variables*.

The procedure of duplicating the message is not possible in quantum mechanics, not even in a thought experiment. Alice cannot duplicate an unknown qubit in order to determine incompatible properties simultaneously. Actually, the impossibility to duplicate a quantum system in an unknown state is a law of nature. It is called the no-cloning theorem and will be proved in Section 6.8.4.

6.5.3. Statistical independence

The random variables j and k determine the mode of the devices in each trial. It is assumed that the mechanisms generating the numbers j and k are not in any way related to the mechanisms generating the output of the devices. This is what we expect if Alice and Bob make their choices by independent tosses of fair coins.

Assumption of independence:

The set of random variables $\{S_1, S_2, T_1, T_2\}$ is statistically independent from the set $\{j, k\}$.

Here, it is useful to form the "vector-valued" random variables

$$X = (S_1, S_2, T_1, T_2), \quad Y = (j, k). \tag{6.46}$$

Note that X has 2^4 different values x, corresponding to the 2^4 possible quadruples of values of S_1, S_2, T_1, and T_2. The four possible values y of Y describe the four possible settings of the two devices: $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$.

The assumption of independence means that X and Y are independent random variables. The statistical independence allows one to compare two different probabilities. One is the probability $\text{prob}(X{=}x)$ for the event that X has some given value x. The other is the *conditional probability* $\text{prob}(X{=}x \mid Y{=}y)$ for the event $X{=}x$ given that Y has the value y.

Given an ensemble with a total of N trials, the statistical probability $\text{prob}(X{=}x)$ is for large N approximated by the fraction

$$\text{prob}(X{=}x) \approx \frac{N(X{=}x)}{N}, \tag{6.47}$$

where $N(X{=}x)$ is the number of trials where the indicated event occurred.

The conditional probability is approximated by the relative frequency of the event in the subensemble of trials where $Y = y$:

$$\text{prob}(X{=}x \mid Y{=}y) \approx \frac{N(X{=}x, Y{=}y)}{N(Y{=}y)}. \tag{6.48}$$

The statistical independence of X and Y is the statement that

$$\text{prob}(X{=}x) = \text{prob}(X{=}x \mid Y{=}y) \quad \text{all } x \text{ and } y. \tag{6.49}$$

If two vector-valued random variables X and Y are independent, then any component of X is independent of any component of Y. For example, the random variable j is independent of S_1 and S_2. This means that how Alice chooses the value of j (coin tossing) is not related to how the potential outcomes S_1 and S_2 are generated. In particular, the decision, which output is to be recorded, must not be influenced by the actual values of S_1 or S_2.

Another consequence of the assumption of independence is locality. The random variable k is statistically independent of S_1 and of S_2. Thus, the mechanism generating the possible outputs for Alice does not influence and is not influenced by the choice of the setting k of Bob's device. This is what we expect if the labs of Alice and Bob are spatially separated and well isolated from each other.

Assumption of locality:

The *values* of the two potential outcomes on one side and the *setting* chosen on the other side are independent random variables.

Exercise 6.13 describes the locality condition satisfied by quantum mechanics. It means that the values of an observable measured by Alice do not

depend on which measurement is performed by Bob. Indeed, as we learned from Eq. (6.44), the probability that Alice gets a certain result $S_j = \epsilon$ is not changed by anything that Bob does. Quantum mechanics, however, knows nothing about the values of observables that are not measured. In short:

> Quantum mechanics supports the assumption of locality, but not the assumption of realism.

The assumption of independence actually allows us to determine the probabilities $\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta)$ within the EPR protocol. This protocol only allows us to determine the conditional probability $\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta \mid j{=}a, k{=}b)$. The assumption of independence implies that

$$\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta) = \mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta \mid j{=}a, k{=}b) \tag{6.50}$$

for all choices of a, $b = 1$ or 2 and ϵ, $\delta = +1$ or -1.

In order to determine $\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta)$ experimentally, we cannot use the formula

$$\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta) \approx \frac{N(S_a{=}\epsilon, T_b{=}\delta)}{N} \tag{6.51}$$

because we do not know the total number $N(S_a{=}\epsilon, T_b{=}\delta)$ of events with $S_a{=}\epsilon$ and $T_b{=}\delta$ (Alice and Bob ignore half of the values in each run).

However, from the measurement records of Alice and Bob, we obtain the fraction of the events with $S_a{=}\epsilon$ and $T_b{=}\delta$ among the trials with settings $j{=}a$ and $k{=}b$. Hence, we can approximate the conditional probability as follows:

$$\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta \mid j{=}a, k{=}b) \approx \frac{N(S_a{=}\epsilon, T_b{=}\delta, j{=}a, k{=}b)}{N(j{=}a, k{=}b)}. \tag{6.52}$$

By (6.50), the conditional probability $\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta \mid j{=}a, k{=}b)$ is equal to the unconditional probability $\mathrm{prob}(S_a{=}\epsilon, T_b{=}\delta)$. But unlike (6.51), the right-hand side of (6.52) contains only numbers that can be determined within the EPR protocol.

6.5.4. Bell's theorem

Bell was the first to derive inequalities for the joint probabilities that can discriminate between quantum mechanics and local realistic theories. Here, we present a variant of Bell's inequality that is directly applicable to our experimental setup. It is due to Clauser, Horne, Shimony, and Holt (CHSH).

CHSH inequality:

Let S_1, S_2, T_1, T_2 be random variables that have ± 1 as the only possible values. Then the probabilities $\mathrm{prob}(S_j = T_k)$ satisfy the inequality

$$\mathrm{prob}(S_1{=}T_2) \leq \mathrm{prob}(S_1{=}T_1) + \mathrm{prob}(S_2{=}T_1) + \mathrm{prob}(S_2{=}T_2). \quad (6.53)$$

Note that these are actually four inequalities, because we can choose an arbitrary pair of observables S_j, T_k on the left side of the inequality.

PROOF. We have the following implication:

$$(S_1{\neq}T_1) \text{ and } (S_2{\neq}T_1) \text{ and } (S_2{\neq}T_2) \quad \Rightarrow \quad (S_1{\neq}T_2). \quad (6.54)$$

You may check this by evaluating the left side of the implication for both possible values of S_1. From elementary logic, we know that the statement $A \Rightarrow B$ is equivalent with $\neg B \Rightarrow \neg A$. Hence,

$$(S_1{=}T_2) \quad \Rightarrow \quad (S_1{=}T_1) \text{ or } (S_2{=}T_1) \text{ or } (S_2{=}T_2). \quad (6.55)$$

Denote the set of events where $S_1{=}T_2$ by $\{S_1{=}T_2\}$, and so forth. The implication above shows that this set is a subset of the union of three other sets:

$$\{S_1{=}T_2\} \subset \{S_1{=}T_1\} \cup \{S_2{=}T_1\} \cup \{S_2{=}T_2\}. \quad (6.56)$$

From this, the CHSH inequality follows as an elementary property of probabilities. □

Whereas the proof of the CHSH inequality is rather simple, the question of whether this inequality can be applied in a description of the EPR experiment is quite non-trivial and has caused ongoing discussions. In the derivation of the CHSH inequality, we considered the simultaneous occurrence of the events $S_1 = T_2$ and $S_2 = T_2$ in the same trial. This is justified by the assumption of realism, which is not supported by quantum mechanics. But, perhaps quantum mechanics is incomplete, as claimed by Einstein, Podolski, and Rosen. Perhaps, $\hat{S}_2$ actually has a value (the hidden variable S_2) in an experiment where $\hat{S}_1$ is determined. According to the definition of EPR, the value S_2 is an element of reality, even if Alice decides to measure $\hat{S}_1$, because S_2 *could* be determined by Bob by exploiting the antisymmetry. (In fact, Bob measures $\hat{T}_1$ or $\hat{T}_2$, hence S_2 remains undetermined.) We see that we cannot speak about both random variables S_1 and S_2 without recourse to counterfactual arguments.

Bell's theorem gives us a method to distinguish between quantum theory and *local realism*. Here, the expression "local realism" refers to any theory that satisfies the assumptions of realism and independence (which implies, in particular, an assumption of locality).

Bell's theorem:

Consider the setup and protocol described in Section 6.5.1. Assume that the possible outputs A and B are determined by a local realistic theory, as described in the assumptions of realism and independence in Sections 6.5.2 and 6.5.3. Define the conditional probabilities

$$P_{\text{same}}(a,b) = \text{prob}(A{=}B \mid j{=}a, k{=}b) \tag{6.57}$$

for Alice and Bob to obtain the same result, given the settings a and b of their devices (a, $b = 1$ or 2). These probabilities satisfy the following inequality:

$$P_{\text{same}}(1,2) \leq P_{\text{same}}(1,1) + P_{\text{same}}(2,1) + P_{\text{same}}(2,2). \tag{6.58}$$

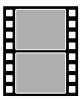
PROOF. By the assumption of realism, the conditional probability of the event $A{=}B$ given that $j{=}a, k{=}b$ is described in terms of random variables:

$$P_{\text{same}}(a,b) = \text{prob}(S_a{=}T_b \mid j{=}a, k{=}b). \tag{6.59}$$

By the assumption of independence (which again involves realism), the conditional probability is equal to the unconditional probability

$$\text{prob}(S_a{=}T_b \mid j{=}a, k{=}b) = \text{prob}(S_a{=}T_b). \tag{6.60}$$

With $P_{\text{same}}(a,b) = \text{prob}(S_a{=}T_b)$, the result (6.58) follows immediately from the CHSH inequality (6.53). □



In CD 5.14, you can test the CHSH inequality (6.58). Here, the anticorrelation between the two particles is simulated by hidden variables. Hence, the outcome of the measurements is determined by information that is generated during the state preparation and carried by the particles to the measurement devices. This simulation presents an example of the optimal case, where equality holds in (6.58). Hence, in a finite number of trials, the relative frequencies might violate the CHSH inequality due to statistical effects. In contrast, CD 5.13 shows that quantum mechanics violates the CHSH inequality by a wide margin. Thus, a violation due to statistical errors in a finite number of experiments is highly improbable. We use qubits in the state ψ_o^- and choose the observables S_j and T_k as described in the next section.

6.5.5. Violation of the Bell inequality

Using the experimental setup of Section 6.4.2, we want to test the validity of Bell's theorem in quantum mechanics. First, we consider the proof of Bell's theorem and note that the formula (6.59) remains meaningful in quantum

mechanics. The quantity $P_{\text{same}}(a,b)$ describes the conditional probability that Alice and Bob obtain the same result for a fixed choice (a,b) of settings. Under this condition, Alice measures $\hat{S}_a$ and Bob measures $\hat{T}_b$, and hence the random variables S_a and T_b are well defined. Actually, the conditional probability $P_{\text{same}}(a,b)$ is equal to the *quantum mechanical* probability that the two compatible observables $\hat{S}_a$ and $\hat{T}_b$ produce the same value in a joint measurement,

$$P_{\text{same}}(a,b) = \text{prob}(S_a = T_b). \tag{6.61}$$

But, within the EPR protocol, it is not allowed to interpret the quantum mechanical probability $\text{prob}(S_a = T_b)$ as an unconditional probability (see the discussion at the end of Section 6.4.2). The reason is that in the other trials, other pairs of observables than $\hat{S}_a$ and $\hat{T}_b$ are measured, and at least one of the random variables S_a or T_b does not exist (as far as quantum mechanics is concerned). At this point, the proof of Bell's theorem breaks down. It does not apply to quantum mechanics.

This consideration does not prove that the assertion (6.58) is false. However, given the state of the two-qubit system, quantum mechanics allows us to compute the probabilities $\text{prob}(S_a{=}T_b)$. Here, we assume that the two-qubit system is in the Bell state ψ_{o}^-. Moreover, we assume that the observables $\hat{S}_a$ and $\hat{T}_b$ are given by Eq. (6.40) as the components of the spin in the direction of unit vectors $\mathbf{n}_a$ and $\mathbf{m}_b$, respectively.

According to (6.35), we have to compute

$$\text{prob}(S_j{=}T_k) = \frac{1}{2}(1 + \langle S_j T_k \rangle) \tag{6.62}$$

for all choices of j and k. Therefore, we first consider the expectation value of

$$(\boldsymbol{\sigma}\cdot\mathbf{n}\otimes\mathbf{1})\,(\mathbf{1}\otimes\boldsymbol{\sigma}\cdot\mathbf{m}) = \boldsymbol{\sigma}\cdot\mathbf{n}\otimes\boldsymbol{\sigma}\cdot\mathbf{m} \tag{6.63}$$

in the state ψ_{o}^- (for arbitrary unit vectors $\mathbf{n}$ and $\mathbf{m}$). Because of Eq. (6.22), we find that

$$(\mathbf{1}\otimes\boldsymbol{\sigma}\cdot\mathbf{m})\,\psi_{\text{o}}^- = -(\boldsymbol{\sigma}\cdot\mathbf{m}\otimes\mathbf{1})\,\psi_{\text{o}}^-. \tag{6.64}$$

Now we obtain

$$\begin{aligned}
&\langle \psi_{\text{o}}^-, (\boldsymbol{\sigma}\cdot\mathbf{n}\otimes\mathbf{1})\,(\mathbf{1}\otimes\boldsymbol{\sigma}\cdot\mathbf{m})\,\psi_{\text{o}}^- \rangle \\
&\qquad = -\langle \psi_{\text{o}}^-, (\boldsymbol{\sigma}\cdot\mathbf{n}\otimes\mathbf{1})\,(\boldsymbol{\sigma}\cdot\mathbf{m}\otimes\mathbf{1})\,\psi_{\text{o}}^- \rangle \\
&\qquad = -\langle \psi_{\text{o}}^-, (\boldsymbol{\sigma}\cdot\mathbf{n}\,\boldsymbol{\sigma}\cdot\mathbf{m}\otimes\mathbf{1})\,\psi_{\text{o}}^- \rangle \\
&\qquad = -\text{Tr}\,(\rho^A\,\boldsymbol{\sigma}\cdot\mathbf{n}\,\boldsymbol{\sigma}\cdot\mathbf{m}).
\end{aligned} \tag{6.65}$$

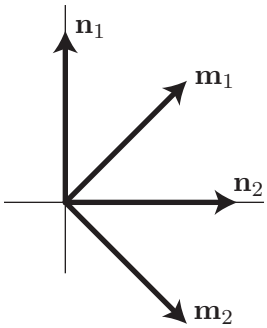


FIGURE 6.5. Choice of unit vectors illustrating the violation of Bell's inequality.

Here, we used (5.85). Whenever the two-qubit system is in a Bell state, the state of the first qubit is $\rho^A = (1/2)\,\mathbf{1}$. Hence,

$$\begin{aligned} -\mathrm{Tr}\,(\rho^A\,\boldsymbol{\sigma}\cdot\mathbf{n}\,\boldsymbol{\sigma}\cdot\mathbf{m}) &= -\frac{1}{2}\sum_{i,j} n_i m_j\,\mathrm{Tr}\,(\sigma_i\sigma_j) \\ &= -\sum_{i,j} n_i m_j \delta_{ij} = -\sum_i n_i m_i \\ &= -\mathbf{n}\cdot\mathbf{m} = -\cos\alpha \end{aligned} \tag{6.66}$$

where α is the angle between the unit vectors $\mathbf{n}$ and $\mathbf{m}$.

Hence, we obtain for the expectation values of the observables $\hat{S}_j\hat{T}_k$ in the state ψ_{o}^- the result

$$\langle S_j T_k\rangle = \langle \psi_{\mathrm{o}}^-, \hat{S}_j\hat{T}_k\psi_{\mathrm{o}}^-\rangle = -\cos\alpha_{jk}, \tag{6.67}$$

where α_{jk} is the angle between $\mathbf{n}_j$ and $\mathbf{m}_k$. Therefore,

$$\begin{aligned} P_{\mathrm{same}}(j,k) = \mathrm{prob}(S_j{=}T_k) &= \frac{1}{2}(1+\langle S_jT_k\rangle) = \frac{1}{2}(1-\cos\alpha_{jk}) \\ &= \sin^2\frac{\alpha_{jk}}{2}, \quad j,k=1,2. \end{aligned} \tag{6.68}$$

Now, we assume that the vectors $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{m}_1$, and $\mathbf{m}_2$ are arranged as in Figure 6.5, so that

$$\alpha_{11} = \alpha_{21} = \alpha_{22} = \frac{\pi}{4}, \quad \alpha_{12} = \frac{3\pi}{4}. \tag{6.69}$$

Hence,

$$P_{\mathrm{same}}(1,1) = P_{\mathrm{same}}(2,1) = P_{\mathrm{same}}(2,2) = \frac{1}{4}(2-\sqrt{2}), \tag{6.70}$$

$$P_{\mathrm{same}}(1,2) = \frac{1}{4}(2+\sqrt{2}). \tag{6.71}$$

These probabilities satisfy the inequality

$$P_{\text{same}}(1,2) - P_{\text{same}}(1,1) - P_{\text{same}}(2,1) - P_{\text{same}}(2,2) = \frac{1}{4}(2+\sqrt{2}) - \frac{3}{4}(2-\sqrt{2}) = \sqrt{2} - 1 > 0. \tag{6.72}$$

This is a clear violation of the CHSH inequality (6.53) by a wide margin. This violation has been confirmed in many experiments.

Quantum mechanics is incompatible with a local realistic hidden variable model.



In CD 5.15 and CD 5.16, we explore another variant of Bell's theorem. We consider three observables A, B, and C with values ± 1. Bell's inequality states that

$$\text{prob}(A{=}B) + \text{prob}(B{=}C) + \text{prob}(C{=}A) \geq 1. \tag{6.73}$$

CD 5.15 shows the violation of this inequality by two qubits in the state ψ_{o}^{-} for appropriately chosen directions of the spin. CD 5.16 shows an example of a classical two-particle system, where each particle has three well-defined properties, all being strictly anticorrelated. Such a system must satisfy (6.73).

The violation of Bell's theorem—meanwhile confirmed experimentally beyond reasonable doubt—has led to an ongoing discussion. It has been speculated that the origin of this contradiction between quantum mechanics and Bell's theorem lies in a nonlocal character of quantum mechanics. But, we have seen that the locality requirement satisfied by quantum mechanics (see Exercise 6.13) is essentially the same as the one used in the proof of Bell's theorem. The reason why Bell's theorem cannot be applied to quantum mechanics is that quantum mechanics violates the assumption of realism, as pointed out in Section 6.5.3. However, locality (sometimes called Einstein causality) does play a role: One can mimic a violation of Bell's theorem with classical systems if one allows a certain amount of communication between Alice and Bob (or between the subsystems). Therefore, it is assumed that any hidden variable theory for quantum mechanics has to be nonlocal.

6.6. Entanglement-Assisted Communication

Having investigated the nature of entanglement, it is now time to use it. The correlations between entangled subsystems can be exploited to enhance the possibilities for communication between Alice and Bob. Here, we present two methods that can be applied whenever Alice and Bob share an entangled pair of qubits: The dense coding protocol transmits two classical bits

of information by sending just one qubit, and the teleportation protocol transmits one qubit by sending two classical bits.

6.6.1. Dense coding

For the purpose of transmitting information, qubits are at least as good as classical bits. By the following procedure, Alice and Bob can exchange one bit of information with the help of one qubit: Alice and Bob agree that they both measure the spin with respect to the z-axis. Alice prepares a qubit in one of the two σ_3-eigenstates, then she sends the qubit to Bob. Bob performs a measurement of σ_3 in order to find out which of the two eigenstates had been prepared (a one-bit information).

Not more than one bit of information can be sent with the help of a single qubit, because any qubit-observable has at most two different eigenvalues. However, Alice and Bob can make use of entanglement to achieve a better result using the *dense coding protocol*[3] described below:

> Whenever Alice and Bob share a maximally entangled qubit pair, then Alice can transmit a two-bit message by sending her qubit to Bob.

The method is called the dense coding protocol, because only one qubit conveys two bits of classical information.

We use the experimental setup described in Figure 6.6. The protocol uses the fact that Alice can change an entangled two-particle state by a unitary subsystem transformation, as discussed in Section 6.2.1. Assume that the two-qubit pair has been prepared in one of the states of the Bell basis, say, $\psi_{\rm e}^+$. Alice gets qubit A and Bob takes qubit B.

In order to prepare the information for Bob, Alice applies a unitary transformation to qubit A. The following unitary operations change $\psi_{\rm e}^+$ into any of the four states of the Bell basis:

$$\begin{aligned} U_1 &= \mathbf{1}, \qquad (U_1 \otimes \mathbf{1})\, \psi_{\rm e}^+ = \psi_{\rm e}^+, \\ U_2 &= \sigma_1, \qquad (U_2 \otimes \mathbf{1})\, \psi_{\rm e}^+ = \psi_{\rm o}^+, \\ U_3 &= \sigma_3, \qquad (U_3 \otimes \mathbf{1})\, \psi_{\rm e}^+ = \psi_{\rm e}^-, \\ U_4 &= \sigma_3\sigma_1, \quad (U_4 \otimes \mathbf{1})\, \psi_{\rm e}^+ = \psi_{\rm o}^-, \end{aligned} \tag{6.74}$$

This follows immediately from Eqs. (6.15) and (6.17). Having done one of these unitary operations, Alice sends her modified qubit to Bob. Bob now puts the two qubits together and performs a Bell-state measurement (as described in Section 6.1.4). The result enables him to conclude which of the four transformations Alice did (a two-bit information).

[3] Bennett, C.H., and Wiesner, S.J., Phys.Rev.Lett. **69**, 2881-2884 (1992).

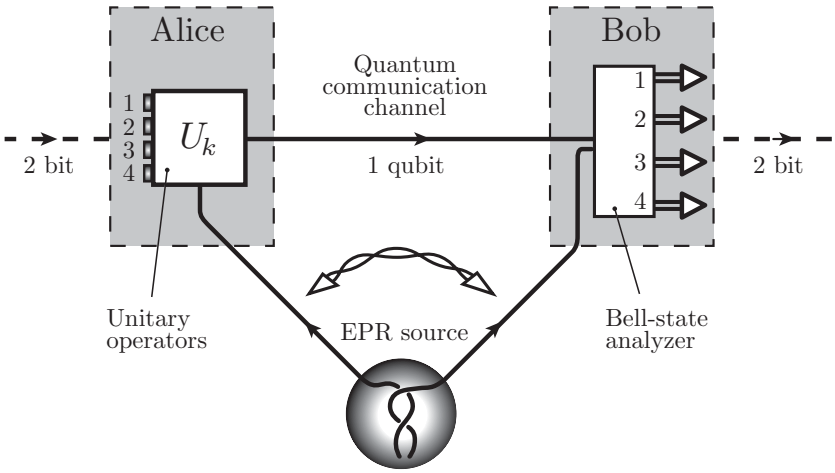


FIGURE 6.6. The dense coding protocol. Two bits of classical information are sent by transmitting a single qubit. Sender and receiver already share a two-qubit system in a maximally entangled state.

As noted above, a single qubit does not carry that much information. The four single-qubit states $U_k\psi$, $k = 1, 2, 3, 4$ (with arbitrary $\psi \in \mathbb{C}^2$) cannot be linearly independent, and hence cannot be distinguished by a single measurement. In order to send a two-bit information using only one qubit, it is necessary that Alice and Bob already share an entangled pair. The entanglement contains the additional information necessary to distinguish the four transformed states of qubit A by a single measurement in an unambiguous way. In a sense, however, the heuristic equation 1 bit = 1 qubit still remains true, because Bob has to use two qubits in order to interpret a two-bit message.

6.6.2. Quantum state teleportation

Alice and Bob share a pair of maximally entangled qubits A and B. Now, Alice receives a second qubit C. The following quantum teleportation protocol enables Alice to transmit the quantum information contained in system C to Bob. Alice just needs to send a classical two-bit message to Bob, the additional information needed by Bob to reconstruct the quantum state of C is contained in the entanglement of the pair AB. During the whole process, the state of qubit C is unknown to Alice and Bob.

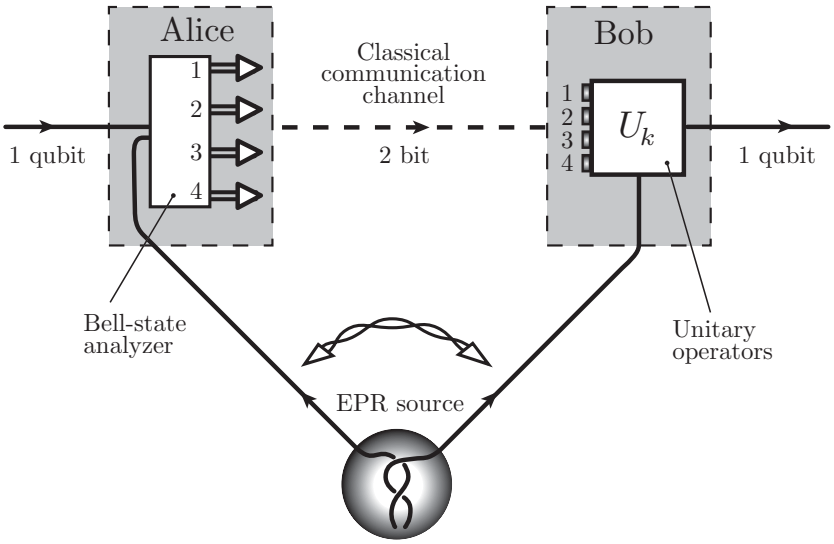


FIGURE 6.7. The teleportation protocol. A qubit is sent by transmitting two bits of classical information. Sender and receiver already share a maximally entangled qubit pair.

Whenever Alice and Bob share a maximally entangled qubit pair, then Alice can transmit the state of a third qubit (that is, the quantum information unit) by sending two classical bits to Bob.

Figure 6.7 shows the experimental setup for the teleportation protocol.[4] Without loss of generality, we may assume that the entangled pair is in the state ψ_{e}^{+}. If Alice receives another qubit, we have in fact a three-qubit problem. Let the state of qubit C be $\psi = a_+\psi_+ + a_-\psi_-$, with arbitrary complex coefficients $a_\pm$ satisfying $|a_+|^2 + |a_-|^2 = 1$. The state of the three-qubit system is the tensor product

$$\begin{aligned}\psi \otimes \psi_e^+ &= \frac{1}{\sqrt{2}}\,(a_+\psi_+ + a_-\psi_-) \otimes (\psi_+ \otimes \psi_+ + \psi_- \otimes \psi_-) \\ &= \frac{1}{\sqrt{2}}\,a_+\,(\psi_+ \otimes \psi_+ \otimes \psi_+ + \psi_+ \otimes \psi_- \otimes \psi_-) + \\ &\quad \frac{1}{\sqrt{2}}a_-\,(\psi_- \otimes \psi_+ \otimes \psi_+ + \psi_- \otimes \psi_- \otimes \psi_-) \qquad (6.75)\end{aligned}$$

[4]Bennett, C.H., Brassard, G., Crepeau, C., Jozsa, R., Peres, A., and Wootters, W.K., Phys.Rev.Lett. **70**, 1895–1899 (1993).

in the Hilbert space $\mathfrak{H}^C \otimes \mathfrak{H}^A \otimes \mathfrak{H}^B$. In the triple tensor products, the first factor always refers to qubit C and the second and third factors to qubits A and B. Alice now puts the qubits C and A together and performs a Bell-state measurement on the system CA (as described in Section 6.1.4). A little calculation will show us that she will measure any of the four basis states with equal probability: In the Hilbert space of the system CA, we write the product states $\psi_+ \otimes \psi_+$, $\psi_+ \otimes \psi_-$, and so forth, as linear combinations of Bell states, for example,

$$\frac{1}{\sqrt{2}}\,\psi_+ \otimes \psi_+ = \frac{1}{2}(\psi_{\mathrm{e}}^+ + \psi_{\mathrm{e}}^-). \tag{6.76}$$

Here, $\psi_{\mathrm{e}}^\pm$ and $\psi_{\mathrm{o}}^\pm$ refer to the Bell states of system CA. Then, we insert these linear combinations for the first two factors in each summand of (6.75). We obtain

$$\begin{aligned}(6.75) = {} & \frac{1}{2}\,a_+ \Big((\psi_{\mathrm{e}}^+ + \psi_{\mathrm{e}}^-) \otimes \psi_+ + (\psi_{\mathrm{o}}^+ + \psi_{\mathrm{o}}^-) \otimes \psi_- \Big) + \\ & \frac{1}{2}\,a_- \Big((\psi_{\mathrm{o}}^+ - \psi_{\mathrm{o}}^-) \otimes \psi_+ + (\psi_{\mathrm{e}}^+ - \psi_{\mathrm{e}}^-) \otimes \psi_- \Big) \\ = {} & \frac{1}{2} \Big(\psi_{\mathrm{e}}^+ \otimes (a_+\psi_+ + a_-\psi_-) + \psi_{\mathrm{e}}^- \otimes (a_+\psi_+ - a_-\psi_-) + \\ & \psi_{\mathrm{o}}^+ \otimes (a_-\psi_+ + a_+\psi_-) + \psi_{\mathrm{o}}^- \otimes (-a_-\psi_+ + a_+\psi_-) \Big). \end{aligned} \tag{6.77}$$

This is a superposition of four mutually orthogonal tensor product states. If Alice determines the Bell state of the pair CA (as described in Section 6.1.4), she will find one of the four basis states with equal probability 1/4. This measurement changes the entangled state (6.77) of the compound system $(CA)B$ into a product state (see Section 6.2.1). That means that after the measurement, the qubit B is known to be in one of the following states:

(1)	$a_+\psi_+ + a_-\psi_- = \psi$	if the measurement gave ψ_{e}^+,
(2)	$a_+\psi_+ - a_-\psi_- = \sigma_3\psi$	if the measurement gave ψ_{e}^-,
(3)	$a_-\psi_+ + a_+\psi_- = \sigma_1\psi$	if the measurement gave ψ_{o}^+,
(4)	$-a_-\psi_+ + a_+\psi_- = \sigma_1\sigma_3\psi$	if the measurement gave ψ_{o}^-.

Now Alice tells Bob (by sending a classical two-bit message, that is, one of the numbers $1, 2, 3, 4$) which of the results she obtained. Finally, Bob performs the corresponding unitary transformation, $U_1 = \mathbf{1}$, $U_2 = \sigma_3$, $U_3 = \sigma_1$, or $U_4 = \sigma_3\sigma_1$, to convert the state of his qubit to ψ. Then he has qubit B in the state ψ, that is, qubit B is now an exact copy of qubit C in its original state.

6.7. Quantum Computers

The theory of quantum computers is based on ideas of Richard P. Feynman[5] and David Deutsch.[6] It has been argued that a quantum computer might outperform a classical computer by making use of superpositions and interference of n-qubit states. Indeed, more recently one has described a few computational problems where a quantum computer is more efficient than any classical computer. Some of these examples are discussed at the end of this chapter.

In the theory of (classical) computers, a finite array of bits is called a *register*. An n-bit register can have 2^n different states. The states are labeled by an n-tuple of 0s and 1s. For example, 01101 is a state of a 5-bit register. In abstract terms, a (classical) *computation* is simply a mapping of an n-bit register to an m-bit register. In a real computer, information is actually encoded in physical systems. Hence, a computation should be viewed as a physical process.

As pointed out in Section 4.1.1, the ultimate goal of miniaturization will be to encode information in qubits. In the theory of quantum computers, an n-qubit system is called a *quantum register*. A computation will be a physical process that transforms the state of a quantum register. In quantum mechanics, a physical process is described by a unitary time evolution in the state-space of the system. Hence, we are going to describe a quantum computation as a unitary transformation of the state of an n-qubit register.

The Hilbert space of quantum register consisting of n qubits is

$$\mathfrak{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2. \tag{6.78}$$

This Hilbert space has the dimension $N = 2^n$, hence it is isomorphic to $\mathbb{C}^N$. The basis states in a one-qubit subspace are, in the context of quantum computing, denoted by

$$\psi_+ = |0\rangle, \quad \psi_- = |1\rangle, \tag{6.79}$$

in order to reflect the usual notation of bits ('0' or '1') in classical information theory. For n-qubit states we use an abbreviated notation, for example,

$$|01101\rangle = |0\rangle \otimes |1\rangle \otimes |1\rangle \otimes |0\rangle \otimes |1\rangle, \quad \text{etc.} \tag{6.80}$$

We see that the 2^n product-basis states of a quantum register correspond precisely to the 2^n possible states of a classical n-bit register. In this notation, the four states of the product basis in the two-qubit Hilbert space are given by

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}. \tag{6.81}$$

[5]Feynman, R.P., Int.J.Theor.Phys. **21**, 467–488 (1982).

[6]Deutsch, D., Proc.Roy.Soc.London **A400**, 97–117 (1985).

and the Bell basis states are

$$\psi_{\mathrm{e}}^{+} = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big), \quad \text{etc.} \tag{6.82}$$

Another frequent notation uses the decimal number that corresponds to the binary string representing a product-basis state. For example, we may interpret the binary string 01101 as the binary representation of the decimal number 13 and denote the state (6.80) by $|13\rangle$.

Here are the basic assumptions (made by theorists) on the realizability of quantum computing.

Basic assumptions of quantum information theory:

For a system consisting of n qubits, it is, at least in principle, possible to perform the following actions:

(A) An arbitrary unitary transformation,
(B) A complete measurement with respect to some orthogonal basis.

If $\{\psi_1, \psi_2, \ldots, \psi_N\}$ is an orthonormal basis in the Hilbert space $\mathbb{C}^N$ of n qubits (with $N = 2^n$), then a *complete measurement* with respect to this basis is defined as a measurement of the simultaneous values of the commuting observables $|\psi_1\rangle\langle\psi_1|, \ldots, |\psi_N\rangle\langle\psi_N|$.

The basic assumptions above imply that it is possible to perform a complete measurement with respect to any orthonormal basis, because any two orthonormal bases are related by a unitary transformation. Assume, for example, that in the two-qubit system it is possible to perform a complete measurement with respect to the product basis. Then, you can also perform a complete measurement with respect to the Bell basis if the following unitary transformation is done first:

$$U_{\mathrm{BP}} \ : \quad \begin{cases} \psi_{\mathrm{e}}^{+} \to |00\rangle \\ \psi_{\mathrm{o}}^{+} \to |01\rangle \\ \psi_{\mathrm{e}}^{-} \to |10\rangle \\ \psi_{\mathrm{o}}^{-} \to |11\rangle \end{cases} \tag{6.83}$$

In quite general terms, a quantum computer is some device that takes an n-qubit system in some state ψ_{in} as input and performs a sequence of unitary transformations resulting in a final state ψ_{out}. This part of the computation is reversible, because all unitary transformations are invertible. Without restriction, we may assume that the initial state is one of the states of the product basis, because a suitable unitary transformation can change it to any other n-qubit state. In that way, comparison with classical computing is easier, because we can interpret the initial state as a "classical input".

Any state of the product basis is described by a classical bit string, that is, by an n-tuple of 0s and 1s. As long as a quantum computer works with product-basis states, it operates like a classical computer by manipulating strings of bits. The power of a quantum computer comes from the fact that it can perform a computation with a superposition of "classical states" and produce a result that depends on the interference of all these states. The ability of a quantum computer to work with superpositions is called *quantum parallelism*.

The final step of quantum computation is irreversible. It consists in a complete measurement of the final state ψ_{out} with respect to the product basis. This measurement gives one of the basis states, that is, a classical bit string, which may be considered the result of the computation.

Quantum computation:

A typical quantum computation proceeds as follows:

(1) One prepares an n-qubit register in one of the states of the product basis.
(2) The quantum computer performs a sequence of unitary transformations ("computational steps").
(3) The final state of the n-qubit register is measured with respect to the product basis. This gives one of the basis states (that is, a string of n bits) as the result of the computation.

The final state of the quantum computation remains inaccessible to the observer, because it cannot be determined with a single measurement (or a finite number of measurements). The final measurement thus gives a particular result only with a certain probability. We must require that the probability to obtain the right answer is strictly larger than 1/2, so that the result can be amplified by repeating the computation several times.

6.8. Logic Gates

A sequence of unitary transformations of a quantum register constitutes a *quantum algorithm*. We want to determine which unitary transformations may be considered as elementary computational steps, sufficiently versatile for the construction of useful algorithms. In classical computing, elementary computational steps involving a few bits are frequently called *logic gates*. Some well-known logic gates are AND, OR, and NOT. (Actually, these gates are sufficient to perform any computation.)

In the following, we are going to define a few logic gates for qubits. To this purpose, we distinguish between local and global transformations. A

unitary transformation of a multipartite system is called *local* if it is of the form

$$U = U_1 \otimes U_2 \otimes \cdots \otimes U_n, \tag{6.84}$$

otherwise it is called *nonlocal* or *global*. A local unitary transformation can be performed by transforming each of the qubits individually. In that way, the entanglement of the system cannot be changed. On the other hand, a global unitary transformation can turn a product state into an entangled state and vice versa.

6.8.1. Single-qubit gates: The Hadamard transformation

The classical NOT-gate changes one bit of input to its opposite such that 0 becomes 1 and 1 becomes 0. The *quantum* NOT *gate* is the unitary operator

$$U_{\text{NOT}} = \sigma_1, \tag{6.85}$$

which maps $|1\rangle \to |0\rangle$, and $|0\rangle \to |1\rangle$. Another quantum gate is

$$U_{\sqrt{\text{NOT}}} = \frac{1+\mathrm{i}}{2}\,(1 - \mathrm{i}\sigma_1), \tag{6.86}$$

which has the property

$$(U_{\sqrt{\text{NOT}}})^2 = U_{\text{NOT}}. \tag{6.87}$$

This gate has no counterpart in classical information theory, where the only possible one-bit transformations are the identity (doing nothing) or the NOT-gate.

Another single-qubit gate without a classical analog is the Hadamard transformation defined by

$$U_{\text{H}} : \quad \begin{cases} |0\rangle \to \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) \\ |1\rangle \to \frac{1}{\sqrt{2}}\big(|0\rangle - |1\rangle\big) \end{cases}. \tag{6.88}$$

In the basis $\{|0\rangle, |1\rangle\}$ of the one-bit Hilbert space $\mathbb{C}^2$, the Hadamard transformation corresponds to the matrix defined in (4.38), that is,

$$U_{\text{H}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\,(\sigma_1 + \sigma_3). \tag{6.89}$$

This transformation is particularly useful for quantum computing, because if we apply U_{H} to each of the n qubits in the initial state $|00\ldots0\rangle$, we obtain

$$\begin{aligned}|\psi\rangle &= U_{\mathrm{H}} \otimes U_{\mathrm{H}} \otimes \cdots \otimes U_{\mathrm{H}} \, |00\ldots0\rangle \\ &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \cdots \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ &= \frac{1}{2^{n/2}} \left(|00\ldots0\rangle + |00\ldots1\rangle + \ldots + |11\ldots1\rangle\right) \\ &= \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle. \end{aligned} \tag{6.90}$$

Hence, this operation transforms the initial state of the qubit register into a superposition of all separable input states (which is still separable). All the 2^n states of the product basis appear with equal probability $1/2^n$ in this superposition. Applying n one-bit transformations to a classical register just gives one of the possible states of the register. Applying n one-qubit transformations to a quantum register as in (6.90) gives a superposition of all 2^n basis states. This gives us the advantage that further computational steps (that is, unitary transformations of the quantum register) can now be performed with all possible classical states at once, whereas a classical computation can only proceed with one of the states of the n-bit register (quantum parallelism).

But we still have to deal with the following problem. Whereas a single measurement of a classical register reveals its state, a single measurement of a quantum register just produces a random result containing (almost) no information about its state prior to the measurement. A successful quantum algorithm uses interference among the states to amplify the desired result while suppressing others (see the examples in Section 6.9 below).

Exercise 6.14. *Show that the action of the Hadamard transformation on an arbitrary n-qubit state $|x\rangle$ (where $x = x_0x_1\ldots x_{n-1}$, $x_i \in \{0,1\}$) is given by*

$$|x\rangle \longrightarrow \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} (-1)^{x.y} \, |y\rangle, \tag{6.91}$$

where $x.y$ denotes a bit-by-bit scalar product, that is,

$$x.y = x_0y_0 + x_1y_1 + \ldots x_{n-1}y_{n-1}. \tag{6.92}$$

Hence, the matrix of the n-qubit Hadamard transformation with respect to the basis $\{|0\rangle, |1\rangle, \ldots, |2^n-1\rangle\}$ is the $2^n \times 2^n$ matrix

$$H_{xy} = \frac{1}{2^{n/2}} \, (-1)^{x.y}. \tag{6.93}$$

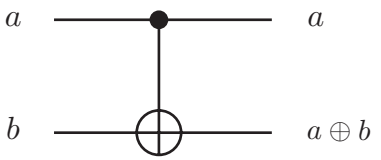


FIGURE 6.8. The controlled-not gate CNOT in classical information theory. $\oplus$ is addition modulo 2.

6.8.2. The controlled-NOT gate

In order to perform a global unitary transformation, we have to bring two or more qubits into contact, for example, in a scattering process. An elementary global unitary transformation of a two-qubit system is the controlled-NOT, or CNOT transformation U_{CNOT}. It is defined as the linear operator that acts on the vectors of the product basis as follows:

$$U_{\text{CNOT}} : \quad \begin{cases} |00\rangle \to |00\rangle \\ |01\rangle \to |01\rangle \\ |10\rangle \to |11\rangle \\ |11\rangle \to |10\rangle \end{cases} \tag{6.94}$$

In the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, this operator corresponds to the matrix

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \sigma_1 \end{pmatrix}. \tag{6.95}$$

If U_{CNOT} is applied twice, it just reproduces the input, that is,

$$U_{\text{CNOT}}^2 = \mathbf{1}_4. \tag{6.96}$$

The classical analog is called a CNOT-gate. It switches the second bit whenever the first bit is 1, otherwise it does nothing. A graphical representation of the CNOT-gate is depicted in Figure 6.8. It is customary to use the same wiring diagram for the quantum-mechanical CNOT-gate (see Fig. 6.9).

Here, the symbol $\oplus$ denotes the integer addition modulo 2, that is,

$$0 \oplus 0 = 0, \quad 0 \oplus 1 = 1, \quad 1 \oplus 0 = 1, \quad 1 \oplus 1 = 0. \tag{6.97}$$

The operation $a, b \to a \oplus b$ is the classical XOR-gate (exclusive or). The XOR-gate is not reversible and thus cannot directly be implemented as a unitary operator in a two-qubit Hilbert space. Therefore, the classical CNOT-gate keeps a record of the first bit together with the result of the XOR-operation.

An immediate generalization of the quantum CNOT-gate is the controlled unitary transformation. Its action on the states of the product basis is

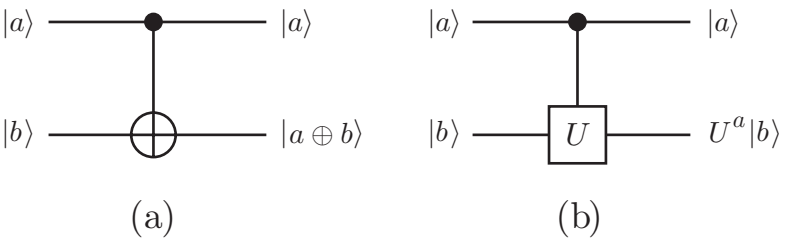


FIGURE 6.9. (a) The quantum controlled-not gate CNOT. (b) The controlled unitary transformation. The gates are unitary operators in the two-qubit Hilbert space. The labels indicate the action of these operators on the product states $|ab\rangle = |a\rangle \otimes |b\rangle$, where $a, b = 0, 1$.

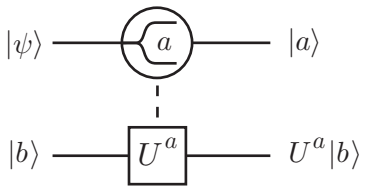


FIGURE 6.10. This version of a classically controlled unitary gate has to be distinguished from the controlled unitary transformation in Figure 6.9. The dashed line symbolizes the classical bit $a = 0$ or 1 that is the result of the measurement of the state $|\psi\rangle$ of qubit A.

defined by

$$|ab\rangle \to |a\rangle \otimes U^a\,|b\rangle \quad \text{for } a, b \in \{0, 1\}, \tag{6.98}$$

and the corresponding wiring diagram is given in Figure 6.9. The CNOT-gate is obtained by setting $U = \sigma_1$.

EXERCISE 6.15. *Find the matrix representation of the controlled unitary transformation* (6.98) *in the product basis.*

The application of the CNOT-gate has to be distinguished carefully from the following procedure (see Fig. 6.10): Perform a measurement of qubit A and, according to the result, apply the transformation $U^0 = \mathbf{1}$ or $U^1 = U$ to qubit B. This procedure can be carried through with the help of local operations in combination with classical communication. It does not define a unitary transformation of the two-qubit system. Nevertheless, it gives the same output as U_{CNOT} in the case that the input is one of the states of the

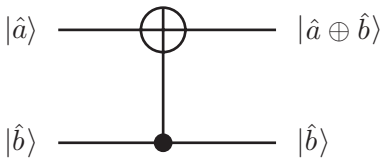


FIGURE 6.11. The action of the controlled-not gate CNOT in the basis formed by the states $|\hat{a}\rangle = U_{\rm H}\,|a\rangle$, with $a, b = 0, 1$.

product basis. But, for superposition states, the output depends on the statistical outcome of the measurement.

The unitary transformation $U_{\rm CNOT}$ is nonlocal (global), and cannot be done without bringing the qubits together. The nonlocality follows from the fact that the transformation $U_{\rm CNOT}$ changes the entanglement. For example, $U_{\rm CNOT}$ turns the separable state $U_{\rm H}|0\rangle \otimes |0\rangle$ into the maximally entangled Bell-basis state $\psi_{\rm e}^{+}$.

More generally,

$$U_{\rm PB} = U_{\rm CNOT}\,(U_{\rm H} \otimes \mathbf{1}) \tag{6.99}$$

is the unitary transformation that maps the product basis states into the Bell-basis states. Hence, the operator $U_{\rm PB}$ is the inverse of the operator $U_{\rm BP}$ defined in (6.83). Using $U_{\rm H}^2 = \mathbf{1}_2$ and $U_{\rm CNOT}^2 = \mathbf{1}_4$, we obtain

$$U_{\rm BP} = U_{\rm PB}^{-1} = (U_{\rm H} \otimes \mathbf{1})\,U_{\rm CNOT}. \tag{6.100}$$

Hence, a Bell-basis analyzer could be built by first applying the $U_{\rm CNOT}$ operator to a state ψ of the two-qubit system, followed by a Hadamard transformation of the first qubit. This maps each of the Bell states onto a corresponding product state. Finally, a measurement in the product basis (which can be done separately on the two qubits) shows the Bell state of ψ. Similarly, an EPR-source can be built by applying (6.99) to a product state (e.g., $|00\rangle$).

The commonly used wiring diagram for $U_{\rm CNOT}$, Figure 6.9, suggests that the first qubit is never changed by $U_{\rm CNOT}$. This is not true. Consider, as an example, the application of $U_{\rm CNOT}$ on products of $|\hat{0}\rangle = U_{\rm H}\,|0\rangle$ and $|\hat{1}\rangle = U_{\rm H}\,|1\rangle$. We obtain

$$U_{\rm CNOT}\ :\quad \begin{cases} |\hat{0}\hat{0}\rangle \to |\hat{0}\hat{0}\rangle \\ |\hat{0}\hat{1}\rangle \to |\hat{1}\hat{1}\rangle \\ |\hat{1}\hat{0}\rangle \to |\hat{1}\hat{0}\rangle \\ |\hat{1}\hat{1}\rangle \to |\hat{0}\hat{1}\rangle \end{cases} \tag{6.101}$$

and now the first qubit is changed depending on the state of the second qubit. We see that the wiring diagram depends on the chosen basis. With respect to the basis $\{|\hat{0}\hat{0}\rangle, |\hat{0}\hat{1}\rangle, |\hat{1}\hat{0}\rangle, |\hat{1}\hat{1}\rangle\}$, the wiring diagram for $U_{\rm CNOT}$ actually should be turned upside down (see Fig. 6.11).

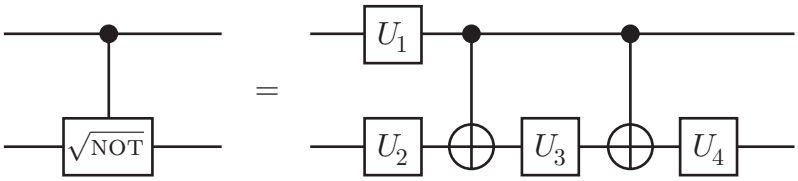


FIGURE 6.12. Representation of the controlled-$\sqrt{\text{NOT}}$-gate by elementary quantum gates.

6.8.3. Sequences of quantum logic gates

An arbitrary unitary transformation in the n-qubit Hilbert space can be composed of single-qubit transformation and the two-qubit transformation CNOT.

As an example for the universality of the CNOT-gate, consider Figure 6.12, where the controlled-$\sqrt{\text{NOT}}$-gate is decomposed into a sequence of single-qubit operations and CNOT-gates. The unitary operators in the wiring diagram are given by

$$U_1 = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\pi/4} \end{pmatrix} \qquad U_2 = U\big(\frac{\pi}{2}\,\mathbf{e}_z\big)\,U\big(-\frac{\pi}{4}\,\mathbf{e}_y\big)$$
$$U_3 = U\big(\frac{\pi}{4}\,\mathbf{e}_y\big) \qquad U_4 = U\big(-\frac{\pi}{2}\,\mathbf{e}_z\big)$$

in terms of the rotation matrices $U(\vec{\alpha})$ defined in (4.28).

An important logical gate is the TOFFOLI-gate shown in Figure 6.13(a). The gate operates on three bits, and the definition in classical information theory is

$$(a, b, c) \longrightarrow (a, b, ab \oplus c). \tag{6.102}$$

The last bit (the target bit) is negated if and only if the first two bits (the control bits) are both 1. Note again that the unitary action of the quantum mechanical gate has to be distinguished from the classically controlled application of a unitary operator to the third qubit, as described in Figure 6.13(b). Here, the result of a measurement on the first qubits determines which unitary operator is chosen. This operation occurs, for example, in the teleportation protocol (see Fig. 6.14).

The TOFFOLI-gate, which is sometimes also called the controlled-controlled-NOT-gate, is important because it realizes several elementary logical

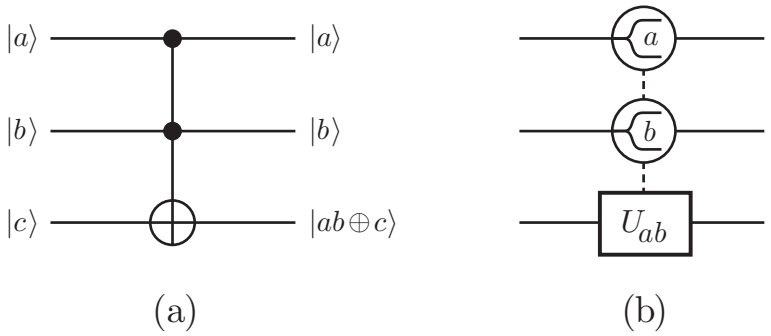


FIGURE 6.13. (a) TOFFOLI-gate. (b) Classically controlled unitary transformation: The transformation U_{ab} is applied to the third qubit if a measurement of the first two qubits in the product basis gives $|ab\rangle$.

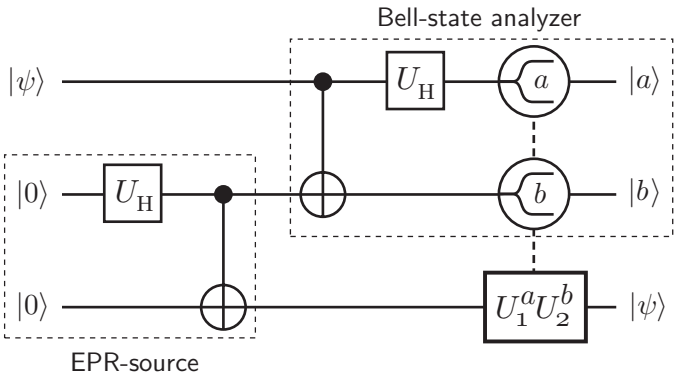


FIGURE 6.14. The teleportation protocol as a wiring diagram

operations in a reversible way. For example,

for $c = 0$:	$(a, b, 0) \to (a, b, ab)$	AND-gate,
for $b = c = 1$:	$(1, 1, c) \to (1, 1, 1 \oplus c)$	NOT-gate,
for $a = 1$:	$(1, b, c) \to (1, b, b \oplus c)$	XOR-gate,
for $b = 1$, $c = 0$:	$(a, 1, 0) \to (a, 1, a)$	FANOUT-gate.

We see that all basic logical operations can be implemented with TOFFOLI-gates. Therefore, a combination of TOFFOLI-gates can be used to perform an arbitrary computation with logically reversible operations. The cost for the reversibility is the inevitable accumulation of "junk" bits that have to be saved at each gate.

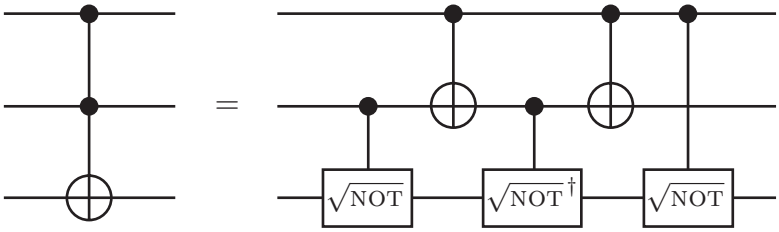


FIGURE 6.15. Representation of the TOFFOLI-gate by elementary quantum gates.

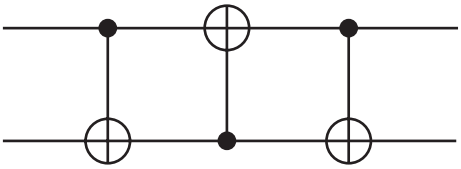


FIGURE 6.16. The SWAP-gate.

Figure 6.15 (together with Fig. 6.12) illustrates that the TOFFOLI-gate can be built from two-qubit CNOT-gates and single-qubit transformations.

EXERCISE 6.16. *Describe the action of the gate depicted in Figure 6.16.*

6.8.4. An impossible gate

The CNOT-gate has the property

$$U_{\text{CNOT}}\left(|a\rangle \otimes |0\rangle\right) = |a\rangle \otimes |a\rangle \quad \text{for } a = 0, 1. \tag{6.103}$$

This copying process, however, does not work for superpositions such as $|\psi\rangle = a|0\rangle + b|1\rangle$. For example, the state $U_{\text{H}}|0\rangle$ becomes the entangled state

$$U_{\text{CNOT}}\left(U_{\text{H}}|0\rangle \otimes |0\rangle\right) = \psi_{\text{e}}^{+} \neq U_{\text{H}}|0\rangle \otimes U_{\text{H}}|0\rangle. \tag{6.104}$$

It is impossible to construct a unitary transformation that does a better job. This is the no-cloning theorem in its simplest form.

No-cloning theorem:

A unitary transformation that maps $|\psi\rangle \otimes |0\rangle$ to $|\psi\rangle \otimes |\psi\rangle$ for all ψ does not exist.

For a proof of the no-cloning theorem we assume that there exists a unitary operator U such that

$$U\,|\psi\rangle \otimes |0\rangle = |\psi\rangle \otimes |\psi\rangle, \qquad U\,|\phi\rangle \otimes |0\rangle = |\phi\rangle \otimes |\phi\rangle \tag{6.105}$$

holds for all ψ and ϕ. Because of the assumed unitarity, the scalar product must be preserved. From the definition of the scalar product for separable states we obtain

$$\langle\psi,\phi\rangle\langle 0,0\rangle = \langle\psi,\phi\rangle = \langle\psi,\phi\rangle\langle\psi,\phi\rangle \tag{6.106}$$

which would be wrong whenever $\langle\psi,\phi\rangle$ is neither 1 nor 0. Hence, such an operator U does not exist.

6.9. Quantum Algorithms

6.9.1. Function evaluation

The TOFFOLI and CNOT gates implement all basic logic operations, hence, at least in principle, we are able to compute any Boolean function

$$f : \mathbb{Z}_2^n \to \mathbb{Z}_2^m, \quad \text{where } \mathbb{Z}_2 = \{0,1\}. \tag{6.107}$$

($\mathbb{Z}_2^n$ is the set of n-bit integers.) By definition, a Boolean function maps an n-bit register onto an m-bit register. Let us consider here as a simple example a function f that associates with each n-bit register a value in $\{0,1\}$. Certainly, this function is not invertible. In order to implement the function evaluation in a unitary way, we repeat the trick of keeping a record of the input. The mapping

$$U_f : |xa\rangle = |x\rangle \otimes |a\rangle \longrightarrow |x\rangle \otimes |a \oplus f(x)\rangle \tag{6.108}$$

is defined for $x \in \mathbb{Z}_2^n$ and $a \in \{0,1\}$ and hence for all separable basis states of a $n+1$-qubit register. By assuming that U_f is a linear transformation, we extend its domain of definition from the set of basis vectors to all vectors of the $n+1$-qubit Hilbert space. Obviously U_f is unitary, because it maps the set of basis states onto itself. Hence, it can be implemented by a sequence of elementary logical operations as described previously. From now on, we consider the function evaluation as a black box, which maps any $n+1$-qubit state to another state of the $n+1$-qubit register. This black box is usually called an *oracle*. If we want to know the value of the function f for a particular binary string $x \in \mathbb{Z}_2^n$, then we prepare the input state $|x0\rangle$ and send it through the black box. This gives the output $|x\rangle \otimes |f(x)\rangle$, and we simply measure the state of the last qubit in order to determine the value $f(x) = 0$ or 1. In that respect, the quantum algorithm for the computation of f is not at all better than a classical algorithm.

The power of quantum computing comes into play if we apply the black box to a superposition state. Applying Hadamard transformations to the first n-qubits in the product state $|00\ldots0\rangle$, we obtain a superposition of all

possible n-qubit states. Still, we can send this superposition into the black box:

$$U_f \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x0\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} U_f|x0\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |xf(x)\rangle. \tag{6.109}$$

Now all possible outputs $f(x)$ for $x \in \mathbb{Z}_2^n$ contribute to the final superposition. In a single run, the black box has obtained all 2^n results simultaneously. This is called quantum parallelism. It does not appear to be of much help, however, because a measurement of the final state in the product basis would only reveal one of the results, chosen randomly. After the measurement, the output register is projected into the corresponding product state, which contains no further information about f.

A superposition state is, however, capable of holding global information about the function f. The following examples illustrate that it is possible to take advantage of quantum parallelism in order to extract this global information in an efficient way.

6.9.2. Simple quantum algorithms

We define the set D of all functions

$$f : \mathbb{Z}_2^n \to \mathbb{Z}_2 = \{0, 1\} \tag{6.110}$$

that are either constant or balanced. A function is balanced if we have $f(x) = 0$ for exactly 2^{n-1} values of $x \in \mathbb{Z}_2^n$ (that is, for exactly half of all possible inputs) and $f(x) = 1$ for the other 2^{n-1} input values. f is constant if for all x, the value $f(x)$ is the same (either 0 or 1).

EXERCISE 6.17. *Let $a = a_1 a_2 \dots a_n \in \mathbb{Z}_2^n$ be an n-bit string $\neq 0$ (at least one of the bits a_j must be nonzero). Show that the function*

$$g_a(x) = a.x \tag{6.111}$$

is balanced. Here, the scalar product $a.x$ is defined modulo 2,

$$a.x = a_1x_1 + a_2x_2 + \ldots + a_nx_n \mod 2 \tag{6.112}$$

$$= a_1x_1 \oplus a_2x_2 \oplus \ldots \oplus a_nx_n. \tag{6.113}$$

EXERCISE 6.18. *Consider a balanced function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and show that*

$$\sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} = 0. \tag{6.114}$$

As a consequence of the previous exercise, prove the formula

$$\sum_{x \in \mathbb{Z}_2^n} (-1)^{a.x} (-1)^{b.x} = \begin{cases} 2^n & \text{for } a = b, \\ 0 & \text{for } a \neq b. \end{cases} \tag{6.115}$$

Here is the Deutsch-Jozsa problem: It is promised that a given function f belongs to D. How can you tell whether f is constant or balanced? Classically, it will be necessary to compute the values $f(x)$ for several inputs x. If you are lucky, you can indeed know the answer after evaluating f for two different inputs x (in case you get two different results). In the worst case, however, you will need $2^{n-1}+1$ function evaluations in order to be sure. It is clear that no classical computer can do better than that. But what about the quantum computer?

The quantum computer would not be better than a classical computer if you use the black box (unitary operator) just to compute the values of f for different inputs x. But, by exploiting superposition and interference in a clever way, it is indeed possible to get the answer with a single call to the black box.

We prepare the input state

$$\underbrace{|0\rangle \otimes \cdots \otimes |0\rangle}_{n \text{ times}} \otimes |1\rangle = |\underbrace{0 \ldots 0}_{n \text{ times}}\, 1\rangle \tag{6.116}$$

and send each qubit through a Hadamard gate in order to obtain a superposition of all input states. By (6.90) we obtain

$$\underbrace{U_{\mathrm{H}} \otimes \cdots \otimes U_{\mathrm{H}}}_{n+1 \text{ times}} |0 \ldots 01\rangle = \frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes U_{\mathrm{H}}|1\rangle. \tag{6.117}$$

Next, this superposition of all possible inputs is sent through the black box U_f. What will happen? We know that according to (6.108)

$$|x0\rangle \xrightarrow{U_f} |x\, f(x)\rangle, \qquad |x1\rangle \xrightarrow{U_f} |x\, (1 \oplus f(x))\rangle$$

and hence

$$\begin{aligned} |x\rangle \otimes U_{\mathrm{H}}|1\rangle &= |x\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\ &\xrightarrow{U_f} \frac{1}{\sqrt{2}} \begin{cases} |x0\rangle - |x1\rangle & \text{if } f(x) = 0 \\ |x1\rangle - |x0\rangle & \text{if } f(x) = 1 \end{cases} \\ &= (-1)^{f(x)} |x\rangle \otimes U_{\mathrm{H}}|1\rangle. \end{aligned}$$

By the linearity of U_f, we can compute the action on the superposition (6.117),

$$\frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes U_{\mathrm{H}}|1\rangle \xrightarrow{U_f} \frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} |x\rangle \otimes U_{\mathrm{H}}|1\rangle. \tag{6.118}$$

Finally, we perform a Hadamard transformation with the first n qubits (the last qubit is no longer of interest). Using (6.91) we obtain

$$\begin{aligned}
&\underbrace{U_{\mathrm{H}} \otimes \cdots \otimes U_{\mathrm{H}}}_{n \text{ times}} \frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} |x\rangle \\
&\quad = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \sum_{y \in \mathbb{Z}_2^n} (-1)^{x.y} |y\rangle \\
&\quad = \sum_{y \in \mathbb{Z}_2^n} c_y \, |y\rangle, \quad \text{with} \quad c_y = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)+x.y}.
\end{aligned} \tag{6.119}$$

(Here, $x.y$ is the scalar product modulo 2 as defined in Exercise 6.17.) The probability that the final measurement of the first n qubits gives $|y\rangle = |0, \ldots 0\rangle$ is

$$|c_{0\ldots0}|^2 = \Big| \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \Big|^2 = \begin{cases} 1 & \text{if } f \text{ is constant,} \\ 0 & \text{if } f \text{ is balanced.} \end{cases} \tag{6.120}$$

Hence, a single measurement of the first n qubits tells you whether f is balanced or not.

Deutsch-Jozsa problem:

Assume that $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ is either constant or balanced. A quantum computer can determine this property with a single evaluation of f. A classical computer has to evaluate f at least twice and at most $2^{n-1} + 1$ times.

A variant of the Deutsch-Jozsa problem is the *Bernstein-Vazirani problem*. We consider the function $f(x) = a.x$ with $a \neq 0$ (see Exercises 6.17 and 6.18). For this function, the coefficient c_y in (6.119) becomes, according to (6.115)

$$c_y = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{a.x} (-1)^{x.y} = \begin{cases} 1 & \text{for } a = y, \\ 0 & \text{for } a \neq y. \end{cases} \tag{6.121}$$

Hence, the final state (6.119) of the first n qubits in the Deutsch-Jozsa algorithm simply becomes $|a\rangle$. A single measurement of the $|x\rangle$ register thus gives the result a. Classically, we would need to evaluate the function f for the n inputs

$$x^{(j)} = 0 \ldots 010 \ldots 0 \quad (\text{with '1' on the } j^{\text{th}} \text{ place}) \tag{6.122}$$

This gives all the binary digits $a_j = f(x^{(j)}) = a.x^{(j)}$ of a. There is no faster way to compute a with a classical computer, because any evaluation of $\mathbb{Z}_2$-valued function f reveals only one bit of information, and hence we need at least n bits of information to determine the n-bit string a.

Bernstein-Vazirani problem:

Consider the function $f(x) = a.x$ from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$ with unknown $a \in \mathbb{Z}_2^n$. A quantum computer allows computation of a with a single evaluation of f. A classical computer has to evaluate f precisely n times.

Note that when comparing classical with quantum computing, the time needed to compute the function f is not taken into account.

6.9.3. Further applications of quantum computers

The quantum algorithms described in the previous section show that for some problems, quantum computers are more powerful than classical computers. Here, we list some further examples that are treated in detail in the specialized literature about this subject[7]:

6.9.3.1. *Grover's algorithm.* In 1996, Lov Grover considered the problem of searching an unsorted database of N items (looking for a needle in a haystack). As an example, imagine the task of finding in a phone directory the entry belonging to a particular phone number.

In an abstract language, we can formulate this problem as follows: We assume that the database contains $N = 2^n$ items labeled by $x \in \mathbb{Z}_2^n$. It is our task to retrieve one of these items, say, $x = a$. Clearly, there has to be a test that tells us whether we have picked the right one. This test may be represented by a function $f_a : \mathbb{Z}_2^n \to \mathbb{Z}_2$, given by

$$f_a(x) = \begin{cases} 1 & \text{for } x = a, \\ 0 & \text{for } x \neq a. \end{cases} \tag{6.123}$$

Again, the function should be imagined as a black box (oracle). A database query is the evaluation of $f_a(x)$ for some input x. Grover's problem is to determine the unknown element $a \in \mathbb{Z}_2^n$ with a reasonably high probability with as few queries as possible.

The best classical algorithm is to examine the items one by one until one hits the right one by chance. Hence, one has to examine an average of $N/2$ elements to find a with a probability of 50%. Grover's method exploits quantum parallelism, that is, f_a is evaluated for the superposition of all possible input states. A sequence of state rotations ("Grover iterations") then

[7] See, for example, [**6**], [**5**], [**2**], and [**12**].

enlarges the probability amplitude of the correct result by constructive interference while supressing the wrong results. The number of database queries required by Grover's algorithm is therefore only of order $\sqrt{N}$. Moreover, it has been shown that Grover's algorithm is optimal. We say that Grover's algorithm provides a quadratic speedup with respect to any classical algorithm.

6.9.3.2. *Simon's algorithm.* We consider a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ for which there exists $a \in \mathbb{Z}_2^n$, $a \neq 0$, such that

$$f(x \oplus a) = f(x), \quad \text{for all } x \in \mathbb{Z}_2^n. \tag{6.124}$$

Note that f is actually 2-to-1, because $(x \oplus a) \oplus a = x$. It is our task to determine the period a with as few function evaluations as possible. Simon's algorithm, which is quite similar to the algorithms presented in Section 6.9.2, provides an exponential speedup over any classical algorithm. The number of function evaluations required by Simon's algorithm is of order n; the number required by any classical algorithm is of order 2^n. This is an example where the classical problem belongs to another *complexity class* as the corresponding problem in quantum computing.

6.9.3.3. *Period finding and the quantum Fourier transform.* We write the input x in decimal representation as an integer in the interval $[0, 2^n - 1]$. We consider a periodic function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2^m$. Such a function is called periodic if there is an integer r (the period) in the interval $[1, 2^n - 1]$, such that

$$f(x + r) = f(x), \quad \text{whenever } x \text{ and } x + r \text{ are in } [0, 2^n - 1]. \tag{6.125}$$

In order to determine an unknown period by classical methods, one has to evaluate the function a number of times that increases exponentially with n. Peter Shor described a quantum algorithm that finds the period with a polynomial number of function evaluations. His method is based on an efficient quantum implementation of the Fourier transform. The quantum Fourier transform is the unitary transformation whose action on the basis vectors $|x\rangle$ is defined by

$$U_{\mathrm{QFT}}|x\rangle = \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} \mathrm{e}^{2\pi \mathrm{i} x.y/N} \, |y\rangle, \quad \text{with } N = 2^n. \tag{6.126}$$

The quantum Fourier transformation provides an exponential speedup compared to the classical FFT algorithm (FFT = fast Fourier transformation).

6.9.3.4. *Shor's algorithm.* In 1994, Peter Shor published a quantum algorithm for finding the prime factors of an N-digit number. Making extensive use of superposition and entanglement, the algorithm solves the problem in a time of order N^3 (roughly). It is believed that any classical factoring algorithm would need a time depending exponentially on N (but this is not proved). Shor's algorithm is based on a quantum algorithm for finding the period of a function using the quantum Fourier transformation. The other parts of Shor's factorization method can be performed on classical computers. Shor's algorithm is of considerable interest, because the computational hardness of the factorization of large integers is a key feature of the popular RSA method for encrypting messages.

6.9.3.5. *Error correction.* Quantum algorithms are particularly vulnerable to perturbations. Coherent superpositions of quantum states are extremely difficult to preserve. As soon as one of the qubits interacts with the surrounding, the qubit has to be considered a part of a much larger system, and its state is turned from a superposition into a statistical mixture. This destroys at least a part of quantum parallelism, and the specific advantage of a quantum computer is lost. Therefore, the possibility of quantum error correction, that is, the restoration of the exact quantum state of a perturbed qubit, is of particular importance. Peter Shor showed in 1995 that quantum error correction can be implemented in a way similar to error correction in classical algorithms. The algorithm uses redundant qubits to detect errors and to restore the perturbed qubit's original state.

It is certain that quantum effects will sooner or later play a major role in the realization of computers. Whether quantum algorithms will ever be used in the solution of real-world problems remains to be seen. In order to have some practical advantage over classical algorithms, the computation has to be done with many qubits, and with a clock rate similar to classical computers. At the time of this writing, we are very far from achieving this goal, but at least it appears that no physical law or principle speaks against the realization of quantum computers.

Chapter 7

Relativistic Systems in One Dimension

Chapter summary: The fundamental equation of relativistic quantum physics is the Dirac equation. It combines the special theory of relativity with quantum mechanics. In this chapter, we restrict our attention to free particles moving in one space dimension, leaving the more general case to the next chapter. In one dimension, relativistic particles are described by wave functions with two components, which are comparatively easy to visualize and, nevertheless, show many features of higher-dimensional solutions that have puzzled a whole generation of physicists.

Dirac originally derived his equation by linearizing the classical relativistic energy-momentum relation using matrix-valued coefficients (Section 7.2). A direct generalization of the nonrelativistic interpretation of wave functions in terms of position probability densities leads to a natural Hilbert space for the Dirac equation and to a tentative one-particle interpretation (Section 7.3). In Section 7.4, we find plane-wave solutions of the Dirac equation and form (approximately) localized wave packets as "continuous superpositions" of plane waves. It turns out that the Dirac operator (the energy observable of a free particle) has a spectrum consisting of positive and negative energies. Correspondingly, the Hilbert space can be decomposed as a direct sum of subspaces with positive and negative energy (Section 7.5).

Surprisingly, the kinematics of wave packets according to the Dirac equation turns out to be a rich and interesting subject (Section 7.6). The presence of a limiting velocity (the velocity of light) leads to characteristic distortions of wave packets in position space. Moreover, a wave packet with negative energy moves in a direction opposite to its average momentum. A very peculiar effect called Zitterbewegung—an oscillating motion of the average position—occurs under certain conditions for superpositions of positive and negative energies. This and other interference effects are discussed in Section 7.7. In Section 7.8, we describe wave packets in energy space and in velocity space. These are representations where the energy or the (classical) velocity are diagonal. The energy and velocity distributions are important for an analysis and understanding of a wave packet's time evolution.

The free-particle Dirac equation is relativistically invariant. The group of Poincaré transformations has a unitary representation in the Hilbert space. We discuss the Lorentz contraction of wave packets and find interesting interference effects for a velocity transformation of superpositions involving positive and negative energies (Section 7.9).

7.1. Introduction

The replacement of the momentum by the operator $-\mathrm{i}\hbar\nabla$ in the nonrelativistic expression for the kinetic energy leads to the free-particle Hamiltonian of quantum mechanics. We want to follow an analogous approach in the relativistic case. According to the special theory of relativity, the relation between the energy and the momentum for a free particle is

$$E = \sqrt{c^2\, p^2 + \mathrm{m}^2\, c^4}. \tag{7.1}$$

For particles at rest ($p^2 = 0$), this becomes Einstein's famous equation $E = \mathrm{m}c^2$. We can obtain an operator in position space for the relativistic kinetic energy by applying the usual substitution rule $\mathbf{p} \to -\mathrm{i}\hbar\nabla$ to this expression:

$$H_{\mathrm{sqrt}} = \sqrt{-c^2\hbar^2\Delta + \mathrm{m}^2 c^4}. \tag{7.2}$$

The equation

$$\mathrm{i}\hbar\,\frac{\partial}{\partial t}\,\psi(\mathbf{x}, t) = H_{\mathrm{sqrt}}\,\psi(\mathbf{x}, t) \tag{7.3}$$

is often called the *square-root Klein-Gordon equation*. Several arguments have been put forward against this equation:

(1) There is no easy way to modify this equation to incorporate electromagnetic fields in a way that is compatible with the special theory of relativity.
(2) The operator $H_{\mathrm{sqrt}} - \gamma/r$, where γ is the Coulomb coupling constant defined in (2.6), gives wrong (inaccurate) eigenvalues for the hydrogen atom.
(3) The square root of a differential operator is a nonlocal operator. Hence, the time derivative of ψ at a point $\mathbf{x}$ is, according to (7.3), related to the values of $\psi(\mathbf{y}, t)$ at all points $\mathbf{y} \in \mathbb{R}^3$.
(4) The time evolution generated by this operator is acausal. A wave packet that is initially strictly localized in a finite region of space instantaneously spreads over the entire $\mathbb{R}^3$.
(5) The solutions of the square-root Klein-Gordon equation are scalar wave functions. Real electrons have spin, and in position space they should be described by a matrix-wave equation.

In 1926, Paul A.M. Dirac followed another approach to derive a relativistic wave equation. We are going to explain Dirac's idea in the simple case of particles moving in one space dimension. The generalization to two and three dimensions will turn out to be rather straightforward. Many effects can be explained for one-dimensional wave packets, which are so much easier to visualize.

7.2. The Free Dirac Equation

7.2.1. The relativistic energy-momentum relation

In our treatment of particles with spin, we used spinor wave functions with two components. Correspondingly, the Hamiltonian describing the energy of a particle with spin is a matrix operator. Thus, it is not so far-fetched to look for a matrix Hamiltonian also in the relativistic case. Dirac had the ingenious idea that the additional degrees of freedom offered by matrix operators would allow him to "linearize" the square-root in (7.1). He wrote the energy of a particle with momentum p in one dimension as

$$E = \alpha\, cp + \beta\, \mathrm{m}c^2, \tag{7.4}$$

with two unknown coefficients α and β. The corresponding expression in quantum mechanics was then obtained with the usual substitution rule $p \longrightarrow -\mathrm{i}\hbar d/dx$.

Let us compute the square of the expression (7.4). To this purpose, we assume that the coefficients α and β are square matrices with constant coefficients. We have to be careful not to set $\alpha\beta$ equal to $\beta\alpha$, because two matrices do not necessarily commute. But we can assume that a constant-coefficient matrix commutes with the momentum p (or with the momentum operator $p = -\mathrm{i}\hbar\, d/dx$). Hence, we obtain

$$E^2 = \left(\alpha\, cp + \beta\, \mathrm{m}c^2\right)^2 = \alpha^2 c^2 p^2 + \alpha\beta p\mathrm{m}c^3 + \beta\alpha p\mathrm{m}c^3 + \beta^2 \mathrm{m}^2 c^2. \tag{7.5}$$

Next. we compare this result with the classical expression for the square of the energy, namely

$$E^2 = c^2 p^2 + \mathrm{m}^2 c^4. \tag{7.6}$$

We require that the two expressions for E^2 are equal. This tells us something about α and β: For example, we see that it is impossible to meet this requirement if the matrices α and β commute. The right sides in (7.5) and (7.6) are only equal if the following relations hold:

$$\alpha\beta + \beta\alpha = \mathbf{0}, \qquad \alpha^2 = \beta^2 = \mathbf{1}. \tag{7.7}$$

Here, $\mathbf{0}$ and $\mathbf{1}$ denote the zero and unit matrices, respectively. We also note that the matrices α and β should be Hermitian. Otherwise, the expression corresponding to (7.4) in quantum mechanics has no chance to become a self-adjoint operator.

It is not so difficult to find Hermitian matrices satisfying the relations (7.7). In fact, we know matrices with these properties already. Take, for example, the Pauli matrices $\alpha = \sigma_1$ and $\beta = \sigma_3$, and the relations are satisfied automatically. Thus, we have reached the following conclusion.

Dirac's energy momentum relation:

The expression

$$E = \sigma_1 cp + \sigma_3 \mathrm{m}c^2 = \begin{pmatrix} \mathrm{m}c^2 & cp \\ cp & -\mathrm{m}c^2 \end{pmatrix} \tag{7.8}$$

relates the energy and the momentum of a particle moving in one dimension in such a way that

$$E^2 = (c^2p^2 + \mathrm{m}^2c^4)\mathbf{1}. \tag{7.9}$$

You can see immediately that the choice $\alpha = \sigma_1$ and $\beta = \sigma_3$ is not unique, because any other pair of Pauli matrices would serve our purpose equally as well. For later reference, we call the choice made in (7.8) the *standard representation*.

EXERCISE 7.1. *Show that the dimensions of any two matrices α and β satisfying* (7.7) *must be an even number. Hint: Show that their eigenvalues must be ± 1 and compute the trace in* (7.7).

7.2.2. The Dirac operator

The Hamiltonian operator obtained with the linearized energy (7.4) is the *free-particle Dirac operator* in one space dimension,

$$H_0 = \alpha c\Big(-\mathrm{i}\hbar\,\frac{d}{dx}\Big) + \beta \mathrm{m}c^2. \tag{7.10}$$

The free-particle Dirac operator is a matrix-differential operator representing the energy observable of a free particle. In the standard representation, we have $\alpha = \sigma_1$, $\beta = \sigma_3$, and

$$H_0 = \begin{pmatrix} \mathrm{m}c^2 & -\mathrm{i}\hbar c\,\frac{d}{dx} \\ -\mathrm{i}\hbar c\,\frac{d}{dx} & -\mathrm{m}c^2 \end{pmatrix}. \tag{7.11}$$

The Dirac operator acts as a linear operator on vector-valued functions of the type

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix} \in \mathbb{C}^2. \tag{7.12}$$

At each point x, the wave function ψ is now a vector with two components. Each component is a complex number. The operator H_0 is applied to ψ via the rules of matrix multiplication,

$$H_0\,\psi(x) = \begin{pmatrix} \mathrm{m}c^2\,\psi_1(x) - \mathrm{i}\hbar c\,\psi_2'(x) \\ -\mathrm{i}\hbar c\,\psi_1'(x) - \mathrm{m}c^2\,\psi_2(x) \end{pmatrix}, \tag{7.13}$$

where the prime denotes the differentiation with respect to x.

The evolution equation with the Dirac operator as the generator of the time evolution is called the *Dirac equation*.

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\,\psi(x,t) = H_0\,\psi(x,t). \tag{7.14}$$

As usual, we are going to use units with $\hbar = 1$ from now on.

EXERCISE 7.2. *Show that the variable substitution* $(t, x) \to (\hbar t, \hbar x)$ *removes the constant* $\hbar$ *from the free-particle Dirac equation* (7.14).

EXERCISE 7.3. *Show that any two Pauli matrices* (σ_j, σ_k) *with* $j \neq k$ *can be obtained from* (σ_1, σ_3) *via a unitary transformation* U *(depending on* j *and* k*). As a consequence,*

$$\sigma_j cp + \sigma_k \mathrm{m}c^2 = UH_0U^\dagger, \tag{7.15}$$

where H_0 *is the free-particle Dirac operator* (7.11) *in the standard representation. Hint: Use the theory described in Section 4.4. Rotate the* x_1*- and* x_3*-axes into the* x_j*- and* x_k*-axes, respectively.*

7.3. Dirac Spinors and State Space

7.3.1. A Hilbert space for the Dirac equation

In order to apply the methods and techniques of quantum theory, we need to define a Hilbert space for the Dirac equation. A suitable state space for the one-dimensional Dirac equation must consist of vector-valued functions

$$\psi(x) = \begin{pmatrix}\psi_1(x)\\ \psi_2(x)\end{pmatrix} \tag{7.16}$$

with two components in order to match the dimension of the Dirac matrices. We assume that the components are square-integrable functions,

$$\int_{-\infty}^{\infty} |\psi_j(x)|^2\,dx < \infty \quad (j = 1, 2). \tag{7.17}$$

The set of all ψ with two square-integrable component functions forms the Hilbert space

$$L^2(\mathbb{R})^2 = L^2(\mathbb{R}) \oplus L^2(\mathbb{R}). \tag{7.18}$$

We encountered a Hilbert space of two-component functions already in Section 3.5.1, where we introduced wave functions describing particles with spin 1/2. In the present context, the elements ψ of the Hilbert space $L^2(\mathbb{R})^2$ are frequently called *Dirac spinors*.[1]

[1] In case of one-dimensional systems, the name "spinors" is badly chosen, because it will turn out that the two components have nothing to do with the spin.

The scalar product between two Dirac spinors is

$$\langle \psi, \phi \rangle = \int_{\mathbb{R}} \big\langle \psi(x)\,,\, \phi(x) \big\rangle_2 \, dx. \tag{7.19}$$

Here, the integrand is the $\mathbb{C}^2$-scalar product of the $\mathbb{C}^2$-vectors $\psi(x)$ and $\phi(x)$. It is defined as in (3.48) by

$$\big\langle \psi(x)\,,\, \phi(x) \big\rangle_2 = \overline{\psi_1(x)}\, \phi_1(x) + \overline{\psi_2(x)}\, \phi_2(x). \tag{7.20}$$

Accordingly, the norm $\|\psi\|$ of a Dirac spinor ψ is given by

$$\|\psi\|^2 = \langle \psi, \psi \rangle = \int_{\mathbb{R}} |\psi(x)|_2^2 \, dx, \tag{7.21}$$

and $|\cdot|_2$ is the norm of a vector in $\mathbb{C}^2$ (see (3.49)), that is,

$$|\psi(x)|_2 = \sqrt{|\psi_1(x)|^2 + |\psi_2(x)|^2}. \tag{7.22}$$

Figure 7.1 shows our favorite method of visualizing Dirac spinors.[2] The example shows a spinor where both components are Gaussian functions. The graph actually combines plots of three curves: phase-colored plots of $\psi_1(x)$ and $\psi_2(x)$, where the second component is plotted upside-down and with less saturation, and a filled plot of $|\psi(x)|_2$ is in the background. Any method that visualizes the components individually depends on the chosen representation of Dirac matrices. We use the standard representation $(\alpha, \beta) = (\sigma_1, \sigma_3)$ for all visualizations on the CD-ROM. Any change of representation $\psi \to U\psi$ (where U is a unitary matrix) would change how the Dirac spinor is split into components.



CD 6.1.1 discusses various methods of visualizing Dirac spinors. We prefer to combine separate plots of the upper and the lower components into one single graph as in Figure 7.1. It has to be stressed, however, that any such method depends on the particular representation of Dirac matrices.

7.3.2. The standard interpretation

Choosing a Hilbert space of square-integrable functions was perfectly natural for the Schrödinger equation: The requirement of square-integrability was inevitable because $|\psi(x)|^2$ was interpreted as a position probability density. Hence, a wavefunction describing a particle that is somewhere in space with certainty must satisfy the normalization condition

$$\|\psi\|^2 = \int_{\mathbb{R}} |\psi(x)|^2 \, dx = 1. \tag{7.23}$$

[2]Unfortunately, Figure 7.1 is reproduced here as a gray-scale image. Many color images of Dirac spinors can be found on the CD-ROM.

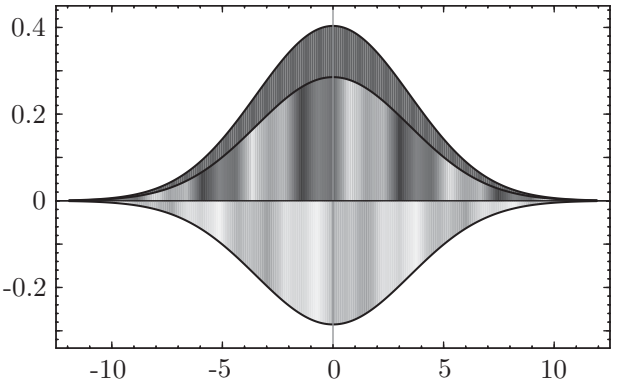


FIGURE 7.1. A Gaussian Dirac spinor in the standard representation. Both components are Gaussian functions with some average momentum. The average momentum determines the distance between the equally colored stripes (wavelength) and the order of colors as in nonrelativistic quantum mechanics (see CD 6.9.3 for a color version of this image).

But is square-integrability also reasonable for the solutions of the Dirac equation?

It makes perfect sense to require square-integrability for Dirac spinors if we can interpret $|\psi(x)|_2^2$ as the position probability density of a relativistic particle. This is certainly the most direct generalization of the nonrelativistic interpretation, therefore it is called the *standard interpretation*.

Standard interpretation:

A relativistic particle in one space dimension is described by a square-integrable Dirac spinor ψ, and the expression

$$\int_B |\psi(x)|_2^2 \, dx \tag{7.24}$$

is the probability of finding the particle in the region $B \subset \mathbb{R}$. Accordingly, the function $x \to |\psi(x)|_2^2$ will be called the *standard position probability density*.

The standard interpretation is a consistent "one-particle interpretation," because the time evolution generated by the Dirac equation is unitary (see Section 7.4.3 below). If $\psi(t)$ describes a one-particle state at the initial time $t = 0$, then there is one and only one particle for all times: A normalized initial wave packet remains normalized, $\|\psi(t)\| = 1$ for all t.

The standard interpretation has to be regarded as a tentative interpretation rule that is not universally accepted. The strange behavior of the solutions of the Dirac equation (to be discussed below) has led physicists to look for alternatives. For example, one can define a suitable unitary transformation U and interpret $|(U\psi)(x)|_2^2$ as the position probability density. Presently, it is not possible to reach a final verdict, because none of the many possible interpretations is without flaws. In fact, any one-particle formalism has to be regarded as a compromise between simplicity and accuracy, because at high energies additional particles and antiparticles are created. Hence, only a formalism with a variable number of particles (quantum-field theory) can be self-consistent. Concerning the interpretation of the one-particle Dirac equation, I'm going to vote for a pragmatic point of view in Section 8.4.3 below.

7.3.3. The relativistic momentum space

By a direct generalization of the corresponding nonrelativistic ideas, we also obtain a tentative interpretation rule for the momentum observable. We denote by $\hat{\psi}$ the Fourier transform of a Dirac spinor ψ. For vector-valued wave functions, the Fourier transformation is applied to each component separately,

$$\hat{\psi}_j(k) = \big(\mathcal{F}\psi_j\big)(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}kx}\, \psi_j(x)\, dx, \quad j = 1, 2. \tag{7.25}$$

The result $\hat{\psi}$ is again a Dirac spinor, and the Fourier transformation can be defined as a one-to-one mapping of $L^2(\mathbb{R})^2$ into itself, with the property $\|\psi\| = \|\hat{\psi}\|$ (Fourier-Plancherel relation; see Book One, Chapter 2). The Fourier transform as a unitary mapping in the Hilbert space $L^2(\mathbb{R})^2$ is again denoted by the letter $\mathcal{F}$.

The inverse Fourier transformation $\mathcal{F}^{-1} : \hat{\psi} \to \psi$ describes each spinor-component ψ_j as a continuous superposition of plane waves $\exp(\mathrm{i}kx)$,

$$\psi_j(x) = \big(\mathcal{F}^{-1}\hat{\psi}_j\big)(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}kx}\, \hat{\psi}_j(k)\, dk, \quad j = 1, 2. \tag{7.26}$$

By analogy with nonrelativistic quantum mechanics, we may interpret $\hat{\psi}$ as the representation of ψ in momentum space. Hence, we define $|\hat{\psi}(k)|_2^2$ as the *standard momentum probability density*.

Standard interpretation (continued):

For a particle described by a square-integrable Dirac spinor ψ, the expression

$$\int_G |\hat{\psi}(p)|_2^2 \, dp \tag{7.27}$$

is the probability of finding the momentum of the particle in $G \subset \mathbb{R}$.

In favor of this interpretation, we state that it is consistent with choosing $p = -\mathrm{i}d/dx$ as the momentum operator. It would be hard to accept anything else, because this interpretation of the momentum was built into the theory from the very beginning; see Equation (7.4). The momentum operator acts component-wise, that is, the differentiation is applied to each component of the spinor. As in all other physical theories, the momentum operator generates translations in position space (see Book One, Chapter 6),

$$\big(\mathrm{e}^{-\mathrm{i}pa}\psi\big)(x) = \psi(x-a), \tag{7.28}$$

where $\mathrm{e}^{-\mathrm{i}pa}$ again acts component-wise on Dirac spinors.

We would also like to remind the reader that the Fourier transform $\mathcal{F}$ turns a differential operator into a multiplication operator. As in Book One we find

$$\mathcal{F}\big(-\mathrm{i}\frac{d}{dx}\psi\big)(k) = k(\mathcal{F}\psi)(k) = k\hat{\psi}(k). \tag{7.29}$$

We can even evaluate the action of the free-particle Dirac operator H_0 in momentum space. It becomes multiplication with the k-dependent matrix

$$\mathbf{h}_0(k) = c\alpha k + \beta \mathrm{m}c^2 = \begin{pmatrix} \mathrm{m}c^2 & ck \\ ck & -\mathrm{m}c^2 \end{pmatrix} \tag{7.30}$$

(the matrix form refers to the standard representation). That is,

$$(\mathcal{F}H_0\psi)(k) = \mathbf{h}_0(k)\,(\mathcal{F}\psi)(k) \quad \text{or} \quad \mathcal{F}H_0\mathcal{F}^{-1}\hat{\psi}(k) = \mathbf{h}_0(k)\,\hat{\psi}(k) \tag{7.31}$$

which we write briefly as

$$\mathcal{F}H_0\mathcal{F}^{-1} = \mathbf{h}_0(k). \tag{7.32}$$

7.3.4. A solution of the Dirac equation

The behavior of solutions of the time-dependent Dirac equation is rather strange. A numerical solution of this equation that starts from some reasonable initial condition (a Gaussian wave packet) reveals a quite unexpected behavior that is typical also for solutions of the Dirac equation in higher dimensions. An example of this behavior is shown in Figure 7.2.

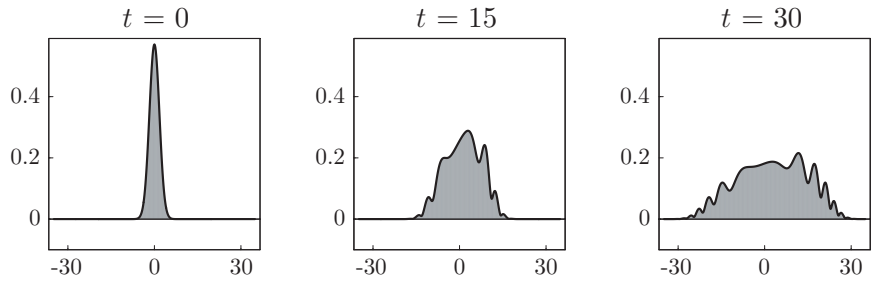


FIGURE 7.2. Snapshots from the time evolution of a Gaussian wave packet according to the one-dimensional Dirac equation (position probability density).



CD 6.1.2 shows a numerically obtained solution of the Dirac equation, similar to the one in Figure 7.2. Despite the "nice" initial condition, the solution shows strange oscillations and self-interference effects.

Figure 7.2 shows three snapshots from the time evolution $\psi(t)$ of the initial function

$$\psi_0(x) = N \exp(-x^2/2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{7.33}$$

This is a spinor with Gaussian initial functions in both components. More precisely, the pictures show the absolute value $|\psi(x,t)|_2$ of the spinor at time t. According to the standard interpretation, the square of this quantity is the position probability density. We see that the shape of the wave packet at later times shows strange distortions. The wiggles are similar to those caused by interference phenomena.

7.4. Plane Waves and Wave Packets

In this section, we determine *plane wave solutions* of the Dirac equation. They belong to a fixed value of the momentum, but their position is completely undetermined. Localized wave packets can be formed in complete analogy to our treatment of the Schrödinger equation (see Book One, Chapter 3). This procedure involves the Fourier transformation.

7.4.1. Diagonalization in momentum space

The Fourier transformation $\mathcal{F}$ converts the free-particle Dirac operator H_0 (a matrix-differential operator) into the matrix-multiplication operator $\mathbf{h}_0(k)$ given by

$$\mathbf{h}_0(k) = \mathcal{F} H_0 \mathcal{F}^{-1}(k) = c\alpha k + \beta \mathrm{m}c^2. \tag{7.34}$$

For each $k \in \mathbb{R}$, the matrix $\mathbf{h}_0(k)$ is a Hermitian two-by-two matrix. It can be diagonalized with the help of a unitary matrix, but the diagonalization procedure depends on the particular representation of the Dirac matrices α and β. Using the methods of linear algebra, we find a unitary matrix $\mathbf{u}(k)$ that diagonalizes $\mathbf{h}_0(k)$ in the standard representation. The choice of $\mathbf{u}(k)$ is not unique, because we can always multiply it by a phase factor. Here, we consider the matrix

$$\mathbf{u}(k) = d_+(k)\,\mathbf{1}_2 + d_-(k)\,\beta\alpha \tag{7.35}$$

where

$$d_+(k) = \frac{1}{\sqrt{2}}\Big(1+\frac{\mathrm{m}c^2}{\lambda(k)}\Big)^{\frac{1}{2}}, \quad d_-(k) = \frac{\mathrm{sgn}(k)}{\sqrt{2}}\Big(1-\frac{\mathrm{m}c^2}{\lambda(k)}\Big)^{\frac{1}{2}}, \tag{7.36}$$

and

$$\lambda(k) = \sqrt{c^2k^2 + \mathrm{m}^2c^4} \qquad \text{(positive square-root)}. \tag{7.37}$$

It may be checked that

$$\mathbf{u}(k)\,\mathbf{h}_0(k)\,\mathbf{u}(k)^{-1} = \lambda(k)\,\beta. \tag{7.38}$$

This equation is independent of the representation. In the standard representation, $\lambda(k)\,\beta$ is diagonal, and $\mathbf{u}(k)$ is indeed the matrix that diagonalizes $\mathbf{h}_0(k)$. It may be checked that $\mathbf{u}(k)$ is unitary,

$$\mathbf{u}(k)^\dagger = \mathbf{u}(k)^{-1} = d_+(k)\,\mathbf{1}_2 - d_-(k)\,\beta\alpha. \tag{7.39}$$

From (7.38) we conclude that the eigenvalues of the matrix $\mathbf{h}_0(k)$ are $+\lambda(k)$ and $-\lambda(k)$, because the Dirac matrix β has the eigenvalues $+1$ and -1. The set of all possible eigenvalues of $\mathbf{h}_0(k)$ for each k defines the *spectrum* $\sigma(H_0)$ of the free-particle Dirac operator:

$$\sigma(H_0) = \{\pm\lambda(k) \mid k \in \mathbb{R}\} = (-\infty, -\mathrm{m}c^2] \cup [\mathrm{m}c^2, \infty). \tag{7.40}$$

You will certainly recognize the relativistic energy-momentum relation in the expression (7.37). Both signs of the square root of (7.9) play a role in (7.40) and hence in Dirac's theory. The occurrence of the negative eigenvalue has remarkable consequences that we are going to describe in the next sections.

Exercise 7.4. *Verify the following properties of the quantities $d_\pm$, which are defined in* (7.36)*:*

$$d_+^2 + d_-^2 = 1, \quad d_+^2 - d_-^2 = \frac{\mathrm{m}c^2}{\lambda(k)}, \quad 2d_+d_- = \frac{ck}{\lambda(k)}. \tag{7.41}$$

Exercise 7.5. *For the velocity v of a free particle, one has the relation*

$$\frac{v^2}{c^2} = \frac{c^2k^2}{\lambda(k)^2}. \tag{7.42}$$

Show that

$$d_+^2 - d_-^2 = \sqrt{1 - \frac{v^2}{c^2}}, \qquad 2d_+d_- = \frac{ck(E)}{E} = \operatorname{sgn}(k)\,\frac{|v|}{c}. \tag{7.43}$$

EXERCISE 7.6. *Investigate the limits of $d_\pm(k)$ and of the matrices $\mathbf{h}_0(k)$ and $\mathbf{u}(k)$, as k goes to 0.*

EXERCISE 7.7. *Show that the inverse of the matrix $\mathbf{u}(\mathbf{k})$ is given by* (7.39)*.*

EXERCISE 7.8. *Verify the relation* (7.38)*.*

7.4.2. Eigenvectors and plane-wave solutions

The eigenvectors of the matrix $\mathbf{h}_0(k)$ are the columns of the matrix $\mathbf{u}(k)^\dagger$. In the standard representation, we have

$$\mathbf{u}(k)^\dagger = d_+(k)\mathbf{1}_2 - \mathrm{i}\, d_-(k)\,\sigma_2. \tag{7.44}$$

Hence, the eigenvector of $\mathbf{h}_0(k)$ belonging to the positive eigenvalue $+\lambda(k)$ is

$$u_{\mathrm{pos}}(k) = \mathbf{u}^\dagger(k)\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}d_+(k)\\d_-(k)\end{pmatrix}, \tag{7.45}$$

and the eigenvector belonging to $-\lambda(k)$ is

$$u_{\mathrm{neg}}(k) = \mathbf{u}^\dagger(k)\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-d_-(k)\\d_+(k)\end{pmatrix}. \tag{7.46}$$

Now, it is fairly easy to construct plane-wave solutions. From the relation $H_0\,\mathrm{e}^{\mathrm{i}kx} = \mathbf{h}_0(k)\,\mathrm{e}^{\mathrm{i}kx}$, we conclude immediately that the stationary plane waves

$$u_{\mathrm{pos}}(k;x) = \frac{1}{\sqrt{2\pi}}\,u_{\mathrm{pos}}(k)\,\mathrm{e}^{\mathrm{i}kx}, \quad u_{\mathrm{neg}}(k;x) = \frac{1}{\sqrt{2\pi}}\,u_{\mathrm{neg}}(k)\,\mathrm{e}^{\mathrm{i}kx} \tag{7.47}$$

are eigenfunctions[3] of H_0,

$$H_0\,u_{\mathrm{pos}}(k;x) = \lambda(k)\,u_{\mathrm{pos}}(k;x), \quad H_0\,u_{\mathrm{neg}}(k;x) = -\lambda(k)\,u_{\mathrm{neg}}(k;x). \tag{7.48}$$

In order to obtain a solution of the time-dependent Dirac equation, we only have to multiply the stationary plane wave with the corresponding time factor $\exp(\pm\mathrm{i}\lambda(k)t)$ and arrive at the following result.

[3]The plane waves are not eigenvectors in the sense of Book One, Definition 5.1, because they are not square-integrable. Hence, they do not belong to the domain of H_0.

Plane-wave solutions of the Dirac equation:

The free-particle Dirac equation in the standard representation

$$\mathrm{i}\frac{\partial}{\partial t}\,\psi(x,t) = -\mathrm{i}c\sigma_1\frac{\partial}{\partial x}\psi(x,t) + \sigma_3 \mathrm{m}c^2\psi(x,t) \tag{7.49}$$

has for each $k \in \mathbb{R}$ the following two plane-wave solutions

$$u_{\mathrm{pos}}(k;x,t) = \frac{1}{\sqrt{2\pi}}\,u_{\mathrm{pos}}(k)\,\mathrm{e}^{\mathrm{i}kx-\mathrm{i}\lambda(k)t}, \tag{7.50}$$

$$u_{\mathrm{neg}}(k;x,t) = \frac{1}{\sqrt{2\pi}}\,u_{\mathrm{neg}}(k)\,\mathrm{e}^{\mathrm{i}kx+\mathrm{i}\lambda(k)t}, \tag{7.51}$$

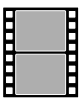
with $u_{\mathrm{pos}}(k)$ and $u_{\mathrm{neg}}(k)$ given by (7.45) and (7.46).

We call $u_{\mathrm{pos}}(k;x,t)$ a *plane wave with positive energy* and $u_{\mathrm{neg}}(k;x,t)$ a *plane wave with negative energy*. For the moment, these are just names. It is not yet clear what the occurrence of negative energies means physically. We are going to investigate the behavior of these "solutions with negative energy" more closely before we assign a meaning to them in Section 8.4.1.

Let us first have a closer look at the two types of plane waves. Because of

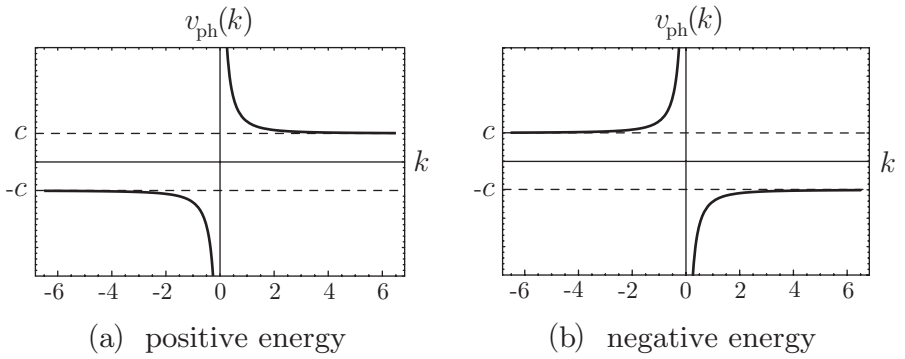
$$|d_+(k)| > |d_-(k)|, \tag{7.52}$$

the upper component of u_{pos} is always larger in absolute value than the lower component.[4] On the other hand, the lower component of u_{neg} is, for all k, larger than the upper component. From the behavior of $d_-(k)$ in the limit $k \to 0$, we conclude that the smaller component of the plane waves vanishes in the limit $k \to 0$.



CD 6.2 investigates the stationary plane waves in the standard representation, as defined in (7.47). Observe the relation between wavelength and momentum, which is the same as nonrelativistic quantum mechanics. The relation between the size of the upper and the lower component is characteristic for the sign of the energy. Moreover, caused by the additional minus in (7.46), positive and negative energies are distinguished by different phase relations between upper and lower components. These properties will still be recognizable in the visualizations of wave packets.

[4] Of course, this observation is only relevant with respect to the chosen standard representation. Changing the representation of Dirac matrices (that is, multiplying all solutions with a constant unitary matrix) gives an equivalent description with other relations between upper and lower components. But, nevertheless, the relative size of the components in the standard representation is useful for identifying the type of solution in a visualization.



(a) positive energy (b) negative energy

FIGURE 7.3. Phase velocity of plane waves as a function of the momentum k. The phase velocity is always larger than the velocity of light. For negative energy, the sign of the velocity is opposite to the sign of the momentum.

The phase velocity of a plane wave $\exp(\mathrm{i}kx - \mathrm{i}\lambda(k)t)$ is given by $\lambda(k)/k$. In our case we have

$$v_{\rm ph}(k) = \pm\frac{\sqrt{c^2k^2 + \mathrm{m}^2c^4}}{k} \longrightarrow \begin{cases} \pm c, & \text{as } k \to \infty, \\ \pm\infty, & \text{as } k \to 0 \text{ (from above).} \end{cases} \tag{7.53}$$

Here, the sign depends on the sign of the energy. For a plane wave with negative energy, the sign of the phase velocity is opposite to the sign of the momentum. Figure 7.3 shows a plot of the phase velocity as a function of k. We see that the phase velocity approaches a finite limit $\pm c$, the velocity of light, as the momentum k goes to infinity. It appears strange that the phase is always faster than light, $|v_{\rm ph}(k)| > c$ (for all k). But the phase velocity is physically not observable and carries no information from one point to another, hence there is no conflict with the theory of relativity.

It might also seem strange that the phase velocity increases when the momentum k decreases and that in the limit of zero momentum, the phase velocity even tends to infinity. But this effect already occurs in nonrelativistic quantum mechanics. Consider the Schrödinger equation with a constant potential $V(x) = V_0 = \lim_{k\to 0} E(k)$, where $E(k) = k^2/2\mathrm{m} + V_0$ is the energy as a function of k. The corresponding plane wave $\exp(\mathrm{i}kx - \mathrm{i}E(k)t)$ has the phase velocity $(k^2/2\mathrm{m} + V_0)/k$, which diverges like V_0/k, as $k \to 0$. For the Dirac equation, a plane wave with energy $\pm\lambda(k)$ feels an "effective constant potential" $\pm\lim_{k\to 0}\lambda(k) = \pm\mathrm{m}c^2$.

As in the case of the Schrödinger equation, the phase velocity of relativistic plane waves can be changed by adding a constant to the Dirac operator.

For example, we may subtract the rest energy and consider the Dirac equation with the Hamiltonian $H_0 - \mathrm{m}c^2$. Then, for sufficiently small k, the phase velocity of a positive energy solution behaves similarly to the phase velocity according to the free-particle Schrödinger equation. If we are only interested in the plane waves with negative energy, we may *add* the constant $\mathrm{m}c^2$ to the Dirac operator.



CD 6.3 shows the time evolution of plane waves with various momenta and positive and negative energies. The time-dependent Dirac equation is linear, and hence the superposition of plane waves with different momenta is again a solution. CD 6.3.3–CD 6.3.5 present a few typical examples of superpositions.

EXERCISE 7.9. *For each $k \in \mathbb{R}$, show that the two vectors $u_{\mathrm{pos}}(k)$ and $u_{\mathrm{neg}}(k)$ form an orthonormal basis of $\mathbb{C}^2$, that is, show that the equations*

$$|u_{\mathrm{pos}}(k)|_2 = 1, \quad |u_{\mathrm{neg}}(k)|_2 = 1, \quad \langle u_{\mathrm{pos}}(k), u_{\mathrm{neg}}(k)\rangle_2 = 0 \tag{7.54}$$

hold for all $k \in \mathbb{R}$ (consider also the case $k = 0$).

EXERCISE 7.10. *Verify by a direct calculation that* (7.50) *and* (7.51) *are solutions of* (7.49).

EXERCISE 7.11. *Find the plane-wave solutions of the Dirac equation with a constant potential V_0. For this, replace the Dirac operator H_0 by $H_0 + V_0$. All energies get shifted by V_0. Compute the phase velocities of the two types of plane-wave solutions and discuss their behavior as a function of k. What do you get for $V_0 = \pm\mathrm{m}c^2$?*

7.4.3. Building wave packets

The time-dependent free-particle Dirac equation can now be solved in exactly the same way as the Schrödinger equation in Book One. We perform a Fourier transformation $\mathcal{F}$ of the initial spinor

$$\psi(x, t{=}0) = \phi(x), \tag{7.55}$$

which is assumed to be square-integrable. The result is the square-integrable spinor in momentum space

$$\hat{\phi}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}kx}\, \phi(x)\, dx. \tag{7.56}$$

Next, we write this vector as a linear combination of the basis vectors $u_{\mathrm{pos}}(k)$ and $u_{\mathrm{neg}}(k)$,

$$\hat{\phi}(k) = \hat{\phi}^+(k)\, u_{\mathrm{pos}}(k) + \hat{\phi}^-(k)\, u_{\mathrm{neg}}(k), \tag{7.57}$$

where the expansion coefficients are given by $\mathbb{C}^2$-scalar products

$$\hat{\phi}^+(k) = \langle u_{\mathrm{pos}}(k), \hat{\phi}(k)\rangle_2, \quad \hat{\phi}^-(k) = \langle u_{\mathrm{neg}}(k), \hat{\phi}(k)\rangle_2. \tag{7.58}$$

The next step is to multiply the parts of $\hat{\phi}(k)$ by the appropriate time factors $\exp(\mp \mathrm{i}\lambda(k)t)$,

$$\hat{\psi}(k,t) = \hat{\phi}^+(k)\, u_{\mathrm{pos}}(k)\, \mathrm{e}^{-\mathrm{i}\lambda(k)t} + \hat{\phi}^-(k)\, u_{\mathrm{neg}}(k)\, \mathrm{e}^{\mathrm{i}\lambda(k)t}. \tag{7.59}$$

Finally, we perform an inverse Fourier transform

$$\begin{aligned}\psi(x,t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}kx}\, \hat{\psi}(k,t)\, dk \\ &= \int_{\mathbb{R}} \Big(\hat{\phi}^+(k)\, u_{\mathrm{pos}}(k;x,t) + \hat{\phi}^-(k)\, u_{\mathrm{neg}}(k;x,t) \Big)\, dk. \end{aligned} \tag{7.60}$$

and obtain the solution of the time-dependent free-particle Dirac equation (in position space) with the initial condition $\psi(t{=}0) = \phi$.

Ψ Equation (7.60) defines a solution of the Dirac equation for all initial conditions $\phi \in L^2(\mathbb{R})^2$, provided that the integral is defined as a limit $n \to \infty$ of integrals over finite intervals from $-n$ to n. The limit $n \to \infty$ is also needed for a proper definition of the Fourier integral (7.56) as a unitary transformation with the domain $L^2(\mathbb{R})^2$ (as explained in Book One, Section 2.5.4). The solution thus obtained for initial conditions not in the domain of the Dirac operator H_0 is a so-called *mild solution* of the Dirac equation (see Book One, Section 6.1). An initial function in the domain of H_0 gives a *strict solution* (a solution that is in the domain of H_0 for all times). The domain of H_0 can be described as the set of those square-integrable Dirac spinors ϕ, for which $\lambda(k)\, \hat{\phi}(k)$ is still a square-integrable function of k.

Let us collect the main result of this section in the following box.

Solution of the Dirac equation:

Let $\phi \in L^2(\mathbb{R})^2$. The time evolution generated by the free-particle Dirac operator H_0 is unitary. It is given by

$$\begin{aligned}\psi(x,t) &= \mathrm{e}^{-\mathrm{i}H_0 t}\, \phi(x) \\ &= \int_{-\infty}^{\infty} \Big(\hat{\phi}^+(k)\, u_{\mathrm{pos}}(k;x,t) + \hat{\phi}^-(k)\, u_{\mathrm{neg}}(k;x,t) \Big)\, dk, \end{aligned} \tag{7.61}$$

where

$$\hat{\phi}^{\pm}(k) = \big\langle u_{\substack{\mathrm{pos} \\ \mathrm{neg}}}(k), \hat{\phi}(k) \big\rangle_2, \tag{7.62}$$

and $\hat{\phi}$ is the Fourier transform of the initial vector ϕ.

EXERCISE 7.12. *Show that $\hat{\phi}^+(k)$ and $\hat{\phi}^-(k)$ are the components of the Dirac spinor $\mathbf{u}(k)\,\hat{\phi}(k)$. Thus*

$$\mathbf{u}(k)\,\hat{\phi}(k) = \begin{pmatrix} \hat{\phi}^+(k) \\ \hat{\phi}^-(k) \end{pmatrix}. \tag{7.63}$$

EXERCISE 7.13. *Using the Fourier-Plancherel relation, prove that*

$$\|\phi\|^2 = \int_{-\infty}^{\infty} \Big(|\hat{\phi}^+(k)|^2 + |\hat{\phi}^-(k)|^2 \Big)\, dk. \tag{7.64}$$

EXERCISE 7.14. *Use* (7.64) *to prove that the norm of the solution* (7.61) *is independent of t,*

$$\|\psi(\cdot, t)\|^2 = \|\phi\|^2. \tag{7.65}$$

EXERCISE 7.15. *A generalization of the previous exercise is the following. Assume that ϕ and ψ are two Dirac spinors. Define $\hat{\phi}^\pm(k)$ and $\hat{\psi}^\pm(k)$ as before and prove that*

$$\langle \phi, \psi \rangle = \int_{-\infty}^{\infty} \Big(\overline{\hat{\phi}^+(k)}\, \hat{\psi}^+(k) + \overline{\hat{\phi}^-(k)}\, \hat{\psi}^-(k) \Big)\, dk. \tag{7.66}$$

EXERCISE 7.16. *For an initial spinor in the domain of H_0, show that $\hat{\psi}(k,t)$ given by* (7.59) *is a solution of the Dirac equation in momentum space,*

$$\mathrm{i}\, \frac{d}{dt}\, \hat{\psi}(k,t) = \mathbf{h}_0(k)\, \hat{\psi}(k,t). \tag{7.67}$$

7.5. Subspaces with Positive and Negative Energies

From Equation (7.61) at time $t = 0$ we see that any square-integrable spinor ϕ can be written as a continuous superposition of the stationary plane waves $u_{\mathrm{pos}}(k;x)$ and $u_{\mathrm{neg}}(k;x)$,

$$\begin{aligned} \phi(x) &= \int_{\mathbb{R}} \hat{\phi}^+(k)\, u_{\mathrm{pos}}(k;x)\, dk + \int_{\mathbb{R}} \hat{\phi}^-(k)\, u_{\mathrm{neg}}(k;x)\, dk \\ &= \phi_{\mathrm{pos}}(x) + \phi_{\mathrm{neg}}(x). \end{aligned} \tag{7.68}$$

The first summand is a wave packet made exclusively out of plane waves with positive energy. Any such wave packet will be called a *wave packet with positive energy*. It has the general form

$$\psi(x) = \int_{-\infty}^{\infty} f(k)\, u_{\mathrm{pos}}(k;x)\, dk, \tag{7.69}$$

where $f(k)$ is a suitable function that describes the distribution of momenta in the wave packet. Obviously, the wave packet ψ has the property that

$$\hat{\psi}^+(k) = \big\langle u_{\rm pos}(k), \hat{\psi}(k) \big\rangle_2 = f(k), \quad \hat{\psi}^-(k) = \big\langle u_{\rm neg}(k), \hat{\psi}(k) \big\rangle_2 = 0. \quad (7.70)$$

We can apply the Dirac operator H_0 to this wave packet. Using (7.48) we find

$$H_0\, \psi(x) = \int_{-\infty}^{\infty} f(k)\, H_0\, u_{\rm pos}(k;x)\, dk = \int_{-\infty}^{\infty} \lambda(k)\, f(k)\, u_{\rm pos}(k;x)\, dk. \quad (7.71)$$

The action of the Dirac operator H_0 on ψ can thus be described as the multiplication of $f(k)$ by $\lambda(k)$. We may now easily calculate the expectation value of the Dirac operator H_0. In view of (7.66) we find, using $\lambda(k) > {\rm m}c^2$, that

$$\langle \psi, H_0\, \psi \rangle = \int_{-\infty}^{\infty} \overline{f(k)}\, \lambda(k)\, f(k)\, dk = \int_{-\infty}^{\infty} \lambda(k)\, |f(k)|^2\, dk > {\rm m}c^2, \quad (7.72)$$

and this is sufficient reason to call ψ a wave packet with positive energy.

Similarly, a *wave packet with negative energy* is a superposition of plane waves with negative energy, like

$$\phi(x) = \int_{-\infty}^{\infty} g(k)\, u_{\rm neg}(k;x)\, dk. \quad (7.73)$$

For this wave packet we find

$$\hat{\phi}^+(k) = 0 \quad \text{and} \quad \hat{\phi}^-(k) = g(k). \quad (7.74)$$

The action of the Dirac operator H_0 on ϕ amounts to a multiplication of $g(k)$ with $-\lambda(k)$. Hence, the expectation value of H_0 in the state ϕ is negative,

$$\langle \phi, H_0\, \phi \rangle < -{\rm m}c^2. \quad (7.75)$$

Using (7.66), we find immediately that the scalar product between a wave packet ψ with positive energy and a wave packet ϕ with negative energy is

$$\langle \psi, \phi \rangle = \int_{-\infty}^{\infty} \Big(\overline{f(k)} \cdot 0 + 0 \cdot g(k) \Big)\, dk = 0. \quad (7.76)$$

A wave packet with positive energy is always orthogonal to a wave packet with negative energy.

Thus, every wave packet ϕ in the Hilbert space $L^2(\mathbb{R})^2$ can be split in a unique way into a sum of two orthogonal parts $\phi_{\rm pos}$ and $\phi_{\rm neg}$, where $\phi_{\rm pos}$ is a wave packet with positive energy, and $\phi_{\rm neg}$ is a wave packet with negative energy. Figure 7.4 shows this decomposition for the wave packet

$$\psi(x) = \Big(\frac{1}{3\pi} \Big)^{1/4} \begin{pmatrix} \exp(-x^2/6) \\ 0 \end{pmatrix}. \quad (7.77)$$

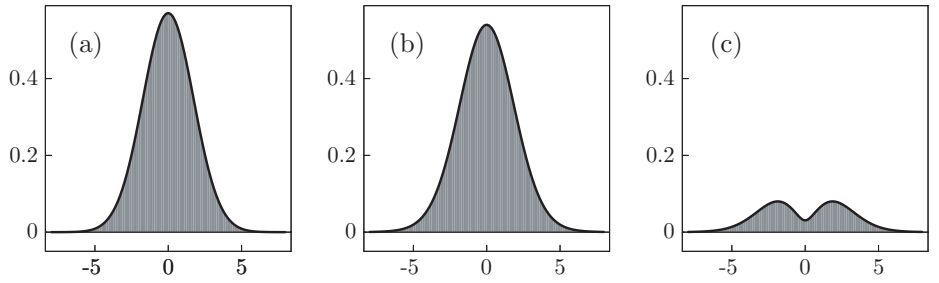


FIGURE 7.4. The Dirac spinor of Equation (7.77) and its parts with positive and negative energy. (a) $|\psi(x)|_2$. (b) $|\psi_{\rm pos}(x)|_2$. (c) $|\psi_{\rm neg}(x)|_2$.

The set of all wave packets with positive energy is itself a Hilbert space, a subspace of $L^2(\mathbb{R})^2$. This subspace will be called the positive-energy subspace $\mathfrak{H}_{\rm pos}(H_0)$. Similarly, the wave packets with negative energy form the negative-energy subspace $\mathfrak{H}_{\rm neg}(H_0)$.

Positive/negative-energy subspaces:

The Hilbert space $L^2(\mathbb{R})^2$ is the direct sum of two orthogonal Hilbert subspaces,

$$L^2(\mathbb{R})^2 = \mathfrak{H}_{\rm pos}(H_0) \oplus \mathfrak{H}_{\rm neg}(H_0). \tag{7.78}$$

This means that every square-integrable spinor ψ has a representation

$$\psi = \psi_{\rm pos} + \psi_{\rm neg}, \qquad \langle \psi_{\rm pos}, \psi_{\rm neg} \rangle = 0 \tag{7.79}$$

as a sum of two uniquely determined, mutually orthogonal parts with

$$\langle H_0 \rangle_{\psi_{\rm pos}} > {\rm m}c^2 > 0, \qquad \langle H_0 \rangle_{\psi_{\rm neg}} < -{\rm m}c^2 < 0. \tag{7.80}$$

In momentum space, the decomposition of a wave packet into positive- and negative-energy parts is described by (7.57). The functions

$$\hat\psi_{\rm pos}(k) = \hat\psi^+(k)\, u_{\rm pos}(k), \quad \text{and} \quad \hat\psi_{\rm neg}(k) = \hat\psi^-(k)\, u_{\rm neg}(k) \tag{7.81}$$

are the Fourier transforms of $\psi_{\rm pos}(x)$ and $\psi_{\rm neg}(x)$. It follows from the orthogonality of positive- and negative-energy parts and from the unitarity of the Fourier transformation that

$$\|\psi\|^2 = \|\psi_{\rm pos}\|^2 + \|\psi_{\rm neg}\|^2 = \|\hat\psi_{\rm pos}\|^2 + \|\hat\psi_{\rm neg}\|^2. \tag{7.82}$$

The Dirac operator in momentum space is just multiplication by $\pm\lambda(k)$ on the parts of positive and negative energy,

$$\mathcal{F}\,(H_0\psi)(k) = \lambda(k)\,\hat\psi_{\rm pos}(k) - \lambda(k)\hat\psi_{\rm neg}(k). \tag{7.83}$$

Similar results hold for functions of H_0. In particular, the time evolution $\exp(-\mathrm{i}H_0 t)$ is described by

$$\mathcal{F}\,(\mathrm{e}^{-\mathrm{i}H_0 t}\,\psi)(k) = \mathrm{e}^{-\mathrm{i}\lambda(k)t}\,\hat{\psi}_{\mathrm{pos}}(k) + \mathrm{e}^{\mathrm{i}\lambda(k)t}\,\hat{\psi}_{\mathrm{neg}}(k). \tag{7.84}$$

We conclude from this that a wave packet with $\hat{\psi}_{\mathrm{neg}} = 0$ stays in the positive-energy subspace for all times.

> For the solutions of the free-particle Dirac equation, the sign of the energy is a conserved quantity. A wave packet that initially has positive energies, has positive energies for all times.

We note that this result remains true for the Dirac equation with time-independent external fields, because it follows directly from the conservation of the energy.

EXERCISE 7.17. *Consider the operator*

$$P_{\mathrm{pos}} = \frac{1}{2}\Big(\mathbf{1} + \frac{H_0}{\sqrt{H_0^2}}\Big). \tag{7.85}$$

Describe its action in momentum space

$$P_{\mathrm{pos}}(k) = \mathcal{F}\,P_{\mathrm{pos}}\,\mathcal{F}^{-1} \tag{7.86}$$

and compute $\mathbf{u}(k)\,P_{\mathrm{pos}}(k)\,\mathbf{u}(k)^{-1}$*, where* $\mathbf{u}(k)$ *is the matrix defined in* (7.35)*.*

EXERCISE 7.18. *Show that in the standard representation*

$$P_{\mathrm{pos}}(k) = u_{\mathrm{pos}}(k)\cdot u_{\mathrm{pos}}(k)^{\dagger} = \begin{pmatrix} d_+(k) \\ d_-(k) \end{pmatrix} \cdot \big(d_+(k)\,,\, d_-(k)\big) \tag{7.87}$$

and derive a similar relation involving u_{neg}*.*

EXERCISE 7.19. *Show that* P_{pos} *is an orthogonal projection operator, that is,*

$$P_{\mathrm{pos}}^2 = P_{\mathrm{pos}} = P_{\mathrm{pos}}^{\dagger}. \tag{7.88}$$

Compute its action on a wave packet with positive (negative) energy and show that

$$P_{\mathrm{pos}}\,L^2(\mathbb{R})^2 = \mathfrak{H}_{\mathrm{pos}}. \tag{7.89}$$

EXERCISE 7.20. *Find a similar operator* P_{neg} *that projects onto the subspace* $\mathfrak{H}_{\mathrm{neg}}$ *and show that*

$$P_{\mathrm{pos}}\,P_{\mathrm{neg}} = P_{\mathrm{neg}}\,P_{\mathrm{pos}} = 0, \quad P_{\mathrm{pos}} + P_{\mathrm{neg}} = \mathbf{1}. \tag{7.90}$$

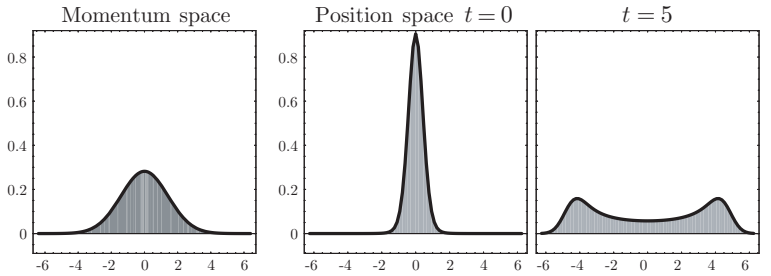


FIGURE 7.5. (a) Dirac spinor in momentum space. It has positive energy and a wide Gaussian momentum distribution stretching into the relativistic regime. (b) The position probability density is initially similar to a Gaussian, but at later times develops "shock fronts" that travel approximately with the velocity of light (units with $\mathrm{m} = c = \hbar = 1$).

EXERCISE 7.21. *Show that the free-particle Dirac operator acts like the square-root Klein-Gordon operator on a wave packet with positive energy,*

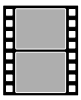
$$H_0 \psi_{\text{pos}} = (c^2 p^2 + \mathrm{m}^2 c^4)^{1/2} \, \psi_{\text{pos}}. \tag{7.91}$$

Describe its action on a wave packet with negative energy.

7.6. Kinematics of Wave Packets

7.6.1. The limiting velocity

The Dirac equation has solutions that behave in a quite familiar way. Any wave packet that consists only of positive energies and that has a momentum distribution corresponding to nonrelativistic velocities behaves approximately in the same way as a solution of the Schrödinger equation.



CD 6.4.1 shows a wave packet with positive energy. It has a momentum distribution in a neighborhood of $k = 0$, and all contributing momenta correspond to nonrelativistic velocities $|v| < c/2$. As a consequence, the wave packet evolves similar to a solution of the Schrödinger equation familiar from Book One. On the other hand, CD 6.4.2 and CD 6.4.3 show the distortions arising from the contributions of relativistic velocities.

For the nonrelativistic Schrödinger equation of a free particle, Gaussian wave packets have a particular importance. (A Gaussian is a function of the

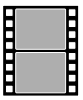
type $\exp(ax^2 + bx + c)$ where a has a negative real part). If the initial function is a Gaussian, then the wave packet at later times is again a Gaussian function. In momentum space, the free nonrelativistic time evolution is just multiplication by the (Gaussian) phase factor $\exp(-\mathrm{i}k^2t/2)$. For the Dirac equation, Gaussian functions are no longer convenient. For example, the Gaussian Dirac spinors

$$\begin{pmatrix} \exp(-x^2) \\ \exp(-x^2) \end{pmatrix}, \quad \begin{pmatrix} \exp(-x^2) \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \exp(-x^2) \end{pmatrix} \tag{7.92}$$

all have nonzero parts with positive and negative energy, which are not Gaussian functions. In momentum space, the time evolution multiplies the positive-energy part with the (non-Gaussian) phase factor $\exp(-\mathrm{i}\lambda(k)t)$ and the negative-energy part with $\exp(+\mathrm{i}\lambda(k)t)$. Consequently, the wave packet at later times is not a Gaussian spinor any longer.

The non-Gaussian distortion is most clearly visible if the wave packet has a very wide momentum distribution. An example is provided by Figure 7.5. It shows a wave packet with positive energy and a Gaussian momentum distribution. The momentum distribution is symmetric around $k = 0$ and extends far into the region $|k| > 1$. For the visualization, we use scaled units with $\mathrm{m} = c = 1$, hence $v = k/\sqrt{k^2 + 1}$ is measured in units of the speed of light, and the momenta $|k| > 1$ correspond to relativistic velocities $|v| > 0.7$.

We note that a linear relation between velocity and momentum would be necessary in order to preserve the Gaussian shape in position space (as is the case in nonrelativistic quantum mechanics). But according to the special theory of relativity, no part of a wave packet can move faster than light. Instead, the components with high momenta all move with similar velocities close to $\pm c$. Therefore, these fast parts cannot spread as usual and form wave fronts in position space, as shown in Figure 7.5(b).



CD 6.5 shows a few examples of propagating wave packets with positive average momentum. The wave packet in CD 6.5.1 behaves essentially like a solution of the Schrödinger equation. The time evolution in momentum space is shown in CD 6.5.2, illustrating the fact that the momentum distribution is constant in time. The wave packet in CD 6.5.3 moves with relativistic speed and shows virtually no dispersion. CD 6.5.4 shows a wave packet with a wide momentum distribution ranging from slow to relativistic momenta (similar to Figure 7.6). As a result, the Gaussian-like initial shape gets heavily distorted in position space.

An example of a propagating wave packet is shown in Figure 7.6. This illustration also shows the position and velocity distributions in the wave

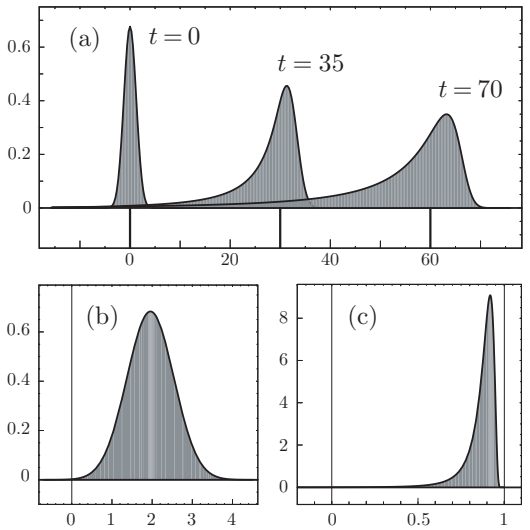


FIGURE 7.6. Non-Gaussian distortion of a moving wave packet with a wide momentum distribution and positive energy. (a) Snapshots of the time evolution at $t = 0$, $t = 35$, and $t = 70$. (b) Momentum distribution. (c) Velocity distribution ($c = 1$).

packet. Though the momentum distribution has a Gaussian shape, the velocity distribution is distorted by the presence of a limit for the propagation speed of any signal. (Here, we use scaled units with $c = 1$). We see that the position distribution asymptotically approaches a scaled version of the velocity distribution. (This kinematical fact was observed earlier in the context of nonrelativistic quantum mechanics, see Book One, Section 3.6.)

The illustrations belonging to this section all show wave packets with positive energy. The initial wave packet is a superposition of the form (7.69). It is a characteristic feature of these wave packets that positive momenta correspond to positive velocities. The velocity distribution $\rho(v)$ of a wave packet with positive energies is obtained from the momentum distribution $|\hat{\psi}(k)|^2$ by a variable substitution:

$$\psi \in \mathfrak{H}_{\text{pos}}(H_0): \quad \rho(v) = \bigl|\hat{\psi}\bigl(k(v)\bigr)\bigr|^2, \tag{7.93}$$

where

$$k(v) = \frac{\mathrm{m}v}{\sqrt{1 - v^2/c^2}} \tag{7.94}$$

is the momentum as a function of the velocity.

EXERCISE 7.22. *Find a Dirac spinor in momentum space with positive energy and a Gaussian momentum distribution, that is,*

$$|\hat{\psi}(k)|_2 = \big(|\hat{\psi}_1(k)|^2 + |\hat{\psi}_2(k)|^2\big)^{1/2} = \exp(-k^2/2). \tag{7.95}$$

Does this spinor have a Gaussian position distribution?

7.6.2. Negative-energy wave packets



CD 6.6 shows the time evolution of wave packets with negative energy. Observe that in the standard representation, you can recognize the sign of the energy from the relative size of the upper and lower components. For a negative-energy wave packet, the lower component is larger than the upper component. CD 6.6.1 shows a wave packet with average momentum zero. CD 6.6.2 shows a wave packet with a positive average momentum, similar to the one in Figure 7.7. It is interesting that this wave packet moves toward the left, that is, in a direction opposite to its momentum. CD 6.6.3 shows the corresponding time evolution in momentum space.

A wave packet with negative energy moves in a direction opposite to its average momentum. Hence, the velocity distribution $\rho(v)$ of a wave packet with negative energy is obtained from its momentum distribution $|\hat{\psi}(k)|^2$ by the following substitution:

$$\psi \in \mathfrak{H}_{\text{neg}}(H_0): \quad \rho(v) = \big|\hat{\psi}\big(-k(v)\big)\big|^2, \tag{7.96}$$

where $k(v)$ is the momentum as a function of the velocity as defined in (7.94).

Figure 7.7 shows some snapshots from the motion of a wave packet with negative energy. This wave packet moves toward the left, although it is composed of positive momenta only.

The modulus of the momentum in Figure 7.7 is larger than in the example of the previous section, Figure 7.6. Hence, the velocity distribution (see Fig. 7.7(c)) is a narrow peak around an average velocity close to $-c$. But, a narrow velocity distribution means less spreading in position space. Indeed, Figure 7.7(a) shows less dispersion compared to the example in Figure 7.6. At very high energies, the momentum is approximately proportional to the energy, $E \approx ck$. In position space, these high-energy wave packets move essentially without spreading. The limiting case $E = ck$ is exactly realized for electromagnetic waves, which move with the speed of light and have no dispersion at all.

In general, a wave packet has nonvanishing parts with positive and negative energies. Figure 7.8 shows, as an example, the motion of the initial spinor

$$\psi(x) = N \exp(-x^2/2 + \mathrm{i}\, k_0\, x) \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{7.97}$$

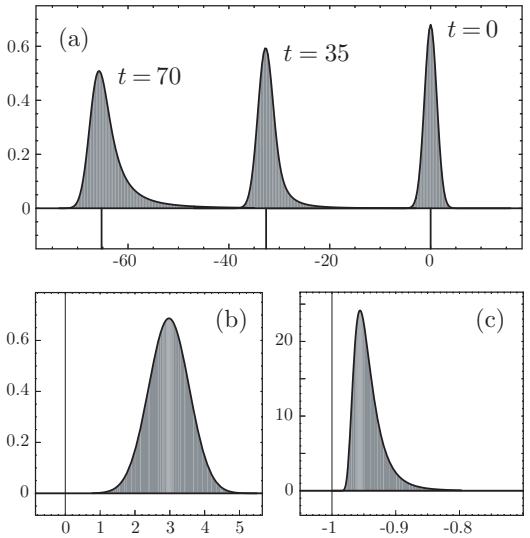
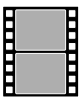


FIGURE 7.7. Non-Gaussian distortion of a moving wave packet with negative energy. The wave packet has positive momentum but negative velocity. (a) Snapshots of the time evolution. (b) Momentum distribution. (c) Velocity distribution ($c = 1$).

This wave packet has a Gaussian momentum distribution around a positive mean value k_0 and a Gaussian position distribution. It is a superposition of parts with positive and negative energies. The part with positive energies moves to the right (because k_0 is positive). The part with negative energy moves in a direction opposite to its momentum. As a consequence, the wave packet will split into two parts. Indeed, we can see this behavior in Figure 7.8.



Several examples of wave packets that split into parts with positive and negative energies moving in opposite directions are shown in CD 6.9 and in CD 6.11. In all these examples, the momentum distribution is centered around some positive momentum, hence the part moving to the left has negative energy.

7.7. Zitterbewegung

7.7.1. The standard position operator

The standard interpretation described in Section 7.3.2 leads naturally to the assumption that the operator x (multiplication by the variable x) represents

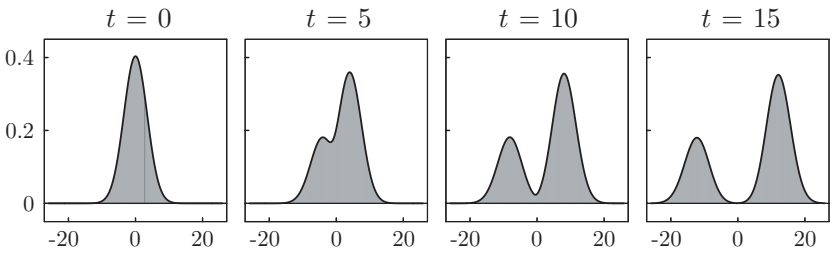


FIGURE 7.8. Strange behavior of a wave packet with positive momentum. The initial wave packet (7.97) has only an upper component and a momentum distribution around a positive mean value. It splits into a part with positive energy moving to the right and a (smaller) part with negative energy moving to the left.

the position observable. This position operator, however, is not generally accepted among physicists, and there have been many suggestions for alternative position operators. In order to distinguish x from other possible choices, we make the following definition.[5]

Definition:

The linear operator x of component-wise multiplication by the position variable x,

$$(x\psi)(x) = \begin{pmatrix} x\,\psi_1(x) \\ x\,\psi_2(x) \end{pmatrix} \tag{7.98}$$

is called the *standard position operator*.

The (dense) domain of x consists of those spinors in $L^2(\mathbb{R})^2$, for which $x\psi$ is again a square-integrable Dirac spinor.

It is interesting to examine the expectation value of the standard position for the wave packet shown in Figure 7.2. The expectation value of x in the state $\psi(t)$ is (according to our interpretation) the average position of the particle at time t. It is given by

$$\overline{x}(t) = \langle x \rangle_{\psi(t)} = \int_{-\infty}^{\infty} x\, |\psi(x,t)|_2^2\, dx. \tag{7.99}$$

Figure 7.9 shows this function for the initial spinor ψ defined in (7.33). It shows an asymptotically damped rapid oscillation around some mean value.

[5]As we prefer not to introduce too many symbols, we denote, as usual, a multiplication operator by the same symbol as the expression by which it multiplies.

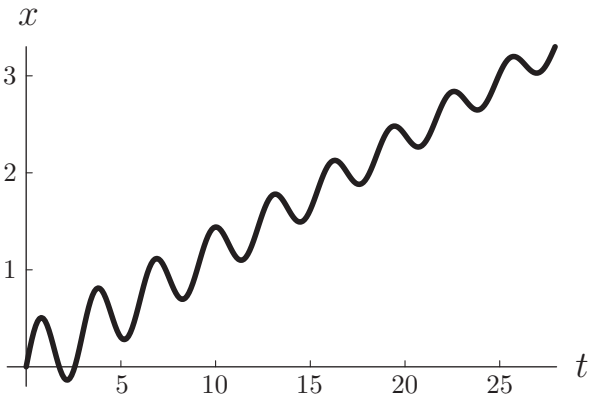


FIGURE 7.9. The expectation value of the position of the wave packet in Figure 7.2 oscillates around a mean-value.

Despite the very symmetric initial condition, the mean value has a slow positive drift-velocity. The oscillatory motion of Dirac spinors was discovered in the year 1930 by Schrödinger, who called it "Zitterbewegung" (a German word which means "quivering motion"). This is a rather strange phenomenon, because it says that a free particle does not move with a constant velocity—a clear violation of Newton's second law.

The Zitterbewegung is strange enough to deserve a closer investigation. Is it a physically measurable phenomenon or is it just an artifact of the formalism and its interpretation?



Examples of wave packets showing Zitterbewegung are CD 6.7.2, all of CD 6.9, CD 6.10, CD 6.11, and CD 6.12.3 and CD 6.12.4. The strange behavior seen in CD 6.12.1 and CD 6.12.2, however, cannot be attributed to Zitterbewegung.

7.7.2. Velocity of the standard position

The time-dependence of wave packets moving according to the Dirac equation usually cannot be determined explicitly. But we can describe the temporal behavior of the standard position operator in order to gain an impression of quantum-relativistic kinematics.

We remind the reader of the fact that the expectation value of x in the state $\psi(t)$ is given by

$$\langle x\rangle_{\psi(t)} = \langle x(t)\rangle_{\psi}, \quad \text{where} \quad x(t) = \mathrm{e}^{\mathrm{i}H_0\,t}\, x\, \mathrm{e}^{-\mathrm{i}H_0\,t}, \tag{7.100}$$

and H_0 is the free-particle Dirac operator according to (7.10). The time-dependent position operator $x(t)$ tells us how the average position of the

spinor changes with time. In the following, we are going to derive an explicit expression for $x(t)$. This will, in particular, describe the Zitterbewegung (see Fig. 7.9).

The time evolution of an observable can be computed from the Heisenberg equation (see Book One, Section 7.4.1). For the standard position operator $x(t)$, we obtain

$$\begin{aligned}\frac{d}{dt}\,x(t) &= \mathrm{e}^{\mathrm{i}H_0 t}\,\mathrm{i}[H_0, x]\,\mathrm{e}^{-\mathrm{i}H_0 t} = \mathrm{e}^{\mathrm{i}H_0 t}\,\mathrm{i}[c\alpha p, x]\,\mathrm{e}^{-\mathrm{i}H_0 t} \\ &= \mathrm{i}c\,\mathrm{e}^{\mathrm{i}H_0 t}\,\alpha\,[-\mathrm{i}d/dx\,, x]\,\mathrm{e}^{-\mathrm{i}H_0 t} = c\,\mathrm{e}^{\mathrm{i}H_0 t}\,\alpha\,\mathrm{e}^{-\mathrm{i}H_0 t} = c\,\alpha(t). \qquad (7.101)\end{aligned}$$

We call $c\alpha$ the *standard velocity operator*, because it is the time-derivative of the standard position operator. In nonrelativistic quantum mechanics, the velocity operator is equal to the momentum operator divided by the mass. Hence, the velocity is a constant of motion under the nonrelativistic free time evolution—in agreement with Newton's second law, which characterizes the free motion by a constant velocity. In relativistic quantum mechanics, however, Eq. (7.101) leads to the following apparently paradoxical conclusions:

(1) The standard velocity operator is not related to the momentum of the particle, instead, it is c times the Dirac matrix α. In the standard representation, the observable corresponding to the Pauli matrix σ_1 is not the spin, but rather a velocity.

(2) The operator $\alpha(t)$ is unitarily equivalent to the Dirac matrix α, and hence it has the eigenvalues ± 1. A measurement of the standard velocity at any time t therefore can only give the results $\pm c$. It appears as if Dirac particles can only move with the velocity of light.

(3) The standard velocity operator $c\alpha(t)$ depends on time in a nontrivial way. A little calculation shows

$$\frac{d}{dt}\,\alpha(t) = \mathrm{e}^{\mathrm{i}H_0 t}\,\mathrm{i}[H_0, \alpha]\,e^{-\mathrm{i}H_0 t} = 2\mathrm{i}H_0\alpha(t) - 2\mathrm{i}cp. \qquad (7.102)$$

7.7.3. The classical velocity operator

The results about the velocity obtained in the previous section are not what we would have expected from classical relativistic kinematics. In particular, the statements (2) and (3) at the end of the previous section appear paradoxical, and one is tempted to look for alternative interpretations of the Dirac equation that are connected with "more reasonable" velocity operators. Comparison with classical relativistic mechanics leads to the definition of the classical velocity operator v_{cl}, which describes the average position of Dirac wave packets (see (7.118) below).

In classical mechanics, the connection between the kinetic energy E, the momentum p, and the velocity v can be described by the formula

$$p = \frac{m\,v}{\sqrt{1 - v^2/c^2}} \quad \text{or} \quad v = \frac{c^2\,p}{\sqrt{c^2p^2 + \mathrm{m}^2c^4}} = \frac{c^2\,p}{E}. \tag{7.103}$$

In Dirac's theory, the classical expression for E is replaced by the Dirac operator H_0. Translating the classical expression for the velocity into quantum mechanics thus leads to the operator

$$v_{\mathrm{cl}} = c^2\,p\,H_0^{-1}, \tag{7.104}$$

which we call the *classical velocity operator*. Its definition involves the inverse of the free-particle Dirac operator. In momentum space, where H_0 is the operator of multiplication with the matrix $\mathbf{h}_0(k)$, the inverse operator H_0^{-1} is just the multiplication with the k-dependent matrix $\mathbf{h}_0(k)^{-1}$. Hence, v_{cl} is a matrix-multiplication operator in momentum space,

$$v_{\mathrm{cl}}(k) = c^2\,k\,\mathbf{h}_0(k)^{-1}. \tag{7.105}$$

We note that the operators H_0^{-1} and v_{cl} commute with the momentum operator p and with H_0. In particular, v_{cl} is a constant of motion for the free time evolution.

Exercise 7.23. *Prove the equation*

$$H_0^{-1} = \frac{H_0}{c^2p^2 + \mathrm{m}^2c^4}. \tag{7.106}$$

Exercise 7.24. *In momentum space, the classical velocity is given by the matrix $c^2k\,\mathbf{h}_0(k)^{-1}$. Find the eigenvalues $w_j(k)$ of this matrix as functions of k. Show that the range $\{w_j(k) \mid k \in \mathbb{R},\, j = 1, 2\}$ of possible values of the classical velocity (that is, the spectrum of the operator v_{cl}) is the continuous range of numbers from $-c$ to $+c$.*

7.7.4. Time evolution of the standard velocity

Let us define the operator that describes the difference between the Dirac velocity $c\alpha$ and the classical velocity:

$$F = c\alpha - v_{\mathrm{cl}}. \tag{7.107}$$

Comparison with (7.102) shows that this operator is related to the time-derivative of the standard velocity operator,

$$\frac{d}{dt}\,c\alpha(t) = 2\mathrm{i}H_0F. \tag{7.108}$$

We may call this operator the *standard acceleration*, because it is the second derivative of the standard position operator.

The operator F anticommutes with the Dirac operator H_0, that is, the relation

$$F\,H_0 = -H_0\,F \tag{7.109}$$

holds on the domain of H_0. As a consequence, we find (see Exercise 7.26)

$$F\,\mathrm{e}^{-\mathrm{i}H_0 t} = \mathrm{e}^{\mathrm{i}H_0 t}\,F. \tag{7.110}$$

Now we conclude immediately, that

$$F(t) = \mathrm{e}^{\mathrm{i}H_0 t}\,F\,\mathrm{e}^{-\mathrm{i}H_0 t} = \mathrm{e}^{2\mathrm{i}H_0 t}\,F = F\,\mathrm{e}^{-2\mathrm{i}H_0 t}. \tag{7.111}$$

Hence, we obtain

$$\frac{d}{dt}\,F(t) = 2\mathrm{i}H_0 F, \tag{7.112}$$

which is precisely the standard acceleration (7.108). We can even determine the indefinite integral of $F(t)$,

$$\int^t F(s)\,ds = \frac{1}{2\mathrm{i}}\,H_0^{-1}\,F(t) \tag{7.113}$$

One can verify this formula immediately by a differentiation. By integrating Eq. (7.112) (or directly from the definition (7.107)) we obtain

$$c\alpha(t) = v_{\mathrm{cl}} + F(t). \tag{7.114}$$

We can even integrate the standard velocity operator $c\alpha(t)$ with respect to t. The result, which will be given in the next section, describes the time-dependence of the standard position operator $x(t)$.

EXERCISE 7.25. *Verify that F anticommutes with H_0. Prove that*

$$FH_0^{-1} = -H_0^{-1}F. \tag{7.115}$$

EXERCISE 7.26. *Define the bounded operators*

$$A_1(t) = F\,\mathrm{e}^{-\mathrm{i}H_0 t} \quad \textit{and} \quad A_2(t) = \mathrm{e}^{\mathrm{i}H_0 t}\,F. \tag{7.116}$$

Show that both operators obey the same differential equation $-\mathrm{i}dA/dt = H_0 A$ with the same initial condition. Hence, $A_1 = A_2$ (on the domain of H_0).

7.7.5. Time evolution of the standard position

When integrating (7.114), we make use of the fact that v_{cl} is a constant of motion (it commutes with H_0) and that $x(t{=}0) = x$. We obtain, using (7.113) and (7.110),

$$\begin{aligned} x(t) &= x + \int_0^t c\alpha(s)\,ds = x + v_{\mathrm{cl}}\,t + \int_0^t F(s)\,ds \\ &= x + v_{\mathrm{cl}}\,t + \frac{1}{2\mathrm{i}}\,H_0^{-1}\left(\mathrm{e}^{2\mathrm{i}H_0\ t} - \mathbf{1}\right)F. \end{aligned} \tag{7.117}$$

Time evolution of x:

Under the time evolution generated by the free-particle Dirac operator, the standard position x behaves as follows:

$$x(t) = x - (2\mathrm{i}\,H_0)^{-1}\,F + v_{\mathrm{cl}}\,t + (2\mathrm{i}\,H_0)^{-1}\,\mathrm{e}^{2\mathrm{i}H_0 t}\,F. \tag{7.118}$$

The domain of the standard position operator does not depend on time.

With this result, we can easily determine the expectation value of $x(t)$ in the state ψ. According to (7.100), this is equal to the expectation value of x in the state $\psi(t)$ that evolves from $\psi(0) = \psi$. Using the properties of F we find

$$\langle \psi\,,\,(2\mathrm{i}\,H_0)^{-1}\,\mathrm{e}^{2\mathrm{i}H_0 t}\,F\,\psi\rangle = \langle \mathrm{e}^{-\mathrm{i}H_0 t}\psi\,,\,(2\mathrm{i}\,H_0)^{-1}\,F\,\mathrm{e}^{-\mathrm{i}H_0 t}\psi\rangle \tag{7.119}$$

and hence

$$\langle x(t)\rangle_\psi = \langle x - (2\mathrm{i}\,H_0)^{-1}\,F\,\rangle_\psi + t\,\langle v_{\mathrm{cl}}\rangle_\psi + \langle (2\mathrm{i}\,H_0)^{-1}\,F\,\rangle_{\psi(t)}. \tag{7.120}$$

Obviously, the second term, which is proportional to t, is responsible for the slow drift of the wave packet in Figure 7.2. The last term describes the Zitterbewegung,

$$z_\psi(t) = \langle (2\mathrm{i}\,H_0)^{-1}\,F\,\rangle_{\psi(t)}. \tag{7.121}$$



All sections on the CD-ROM which show wave packets with Zitterbewegung also contain a plot of the average position $\langle x(t)\rangle_\psi$ as a function of time (see, for example, CD 6.7.2). Moreover, in the visualizations of wave packets in position space, the average position is always indicated as a vertical gray line.

7.7.6. Superpositions and Zitterbewegung

Equation (7.109) shows that the operator F maps the subspace with positive energy into the subspace with negative energy, and vice versa

$$\psi \in \mathfrak{H}_{\mathrm{pos}} \Rightarrow F\psi \in \mathfrak{H}_{\mathrm{neg}}, \qquad \psi \in \mathfrak{H}_{\mathrm{neg}} \Rightarrow F\psi \in \mathfrak{H}_{\mathrm{pos}}, \tag{7.122}$$

The operator $(2\mathrm{i}\,H_0)^{-1}\,F$ whose expectation value describes the Zitterbewegung according to (7.121) has the same property. Whenever ψ has positive energies, then

$$\phi = (2\mathrm{i}\,H_0)^{-1}\,F\,\psi \tag{7.123}$$

has negative energies, and vice versa. We conclude immediately that whenever ψ is a positive-energy wave packet, then

$$z_\psi(t) = \langle \psi(t)\,,\,(2\mathrm{i}\,H_0)^{-1}\,F\,\psi(t)\rangle = \langle \psi(t)\,,\,\phi(t)\rangle = 0, \tag{7.124}$$

because a state ψ with positive energy is always orthogonal to a state ϕ with negative energy. The same is true for negative-energy states. Hence,

$$z_\psi(t) = 0 \quad \text{for all } \psi \in \mathfrak{H}_{\text{pos}} \text{ and all } \psi \in \mathfrak{H}_{\text{neg}}. \tag{7.125}$$

We note that any solution of the free-particle Dirac equation with $z_\psi(t) = 0$ for all t behaves in accordance with classical relativistic kinematics:

$$\langle x(t)\rangle_\psi = \langle x\rangle_\psi + t\,\langle v_{\text{cl}}\rangle_\psi \quad \text{whenever } z_\psi(t) = 0 \text{ for all t.} \tag{7.126}$$

All solutions belonging to a definite sign of the energy show this "reasonable" behavior.

In general, however, a square-integrable Dirac spinor in $L^2(\mathbb{R})^2$ is a superposition of positive- and negative-energy parts, $\psi = \psi_{\text{pos}} + \psi_{\text{neg}}$, and we expect that the term describing Zitterbewegung is nonzero

$$\begin{aligned} z_\psi(t) = &\left\langle \psi_{\text{neg}}(t)\,,\, (2\mathrm{i}\,H_0)^{-1}\,F\,\psi_{\text{pos}}(t)\right\rangle \\ &+ \left\langle \psi_{\text{pos}}(t)\,,\, (2\mathrm{i}\,H_0)^{-1}\,F\,\psi_{\text{neg}}(t)\right\rangle. \end{aligned} \tag{7.127}$$

The operator $(2\mathrm{i}\,H_0)^{-1}\,F$ is a nonlocal operator in position space but is a simple matrix-multiplication operator in momentum space

$$\begin{aligned} \mathcal{F}\,(2\mathrm{i}\,H_0)^{-1}\,F\,\mathcal{F}^{-1} &= (2\mathrm{i})^{-1}\,\mathbf{h}_0(k)^{-1}\left(c\alpha - c^2\,k\,\mathbf{h}_0(k)^{-1}\right) \\ &= \frac{1}{2\mathrm{i}}\,\frac{\mathrm{m}c}{k^2+\mathrm{m}^2c^2}\,\beta\,\alpha = \frac{1}{2}\,\frac{\mathrm{m}c}{k^2+\mathrm{m}^2c^2}\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}. \end{aligned} \tag{7.128}$$

The last expression is only valid in the standard representation. Zitterbewegung vanishes, whenever the spinors $\hat{\psi}_{\text{pos}}(k)$ and $\hat{\psi}_{\text{neg}}(k)$ do not overlap (have disjoint support in momentum space). An example is shown in Figure 7.10.

In case of the wave packet

$$\psi_1(x) = N\,\exp(-x^2/2)\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{7.129}$$

the parts with positive and negative energy are located in the same region of momentum space, and therefore the spinor shows Zitterbewegung.

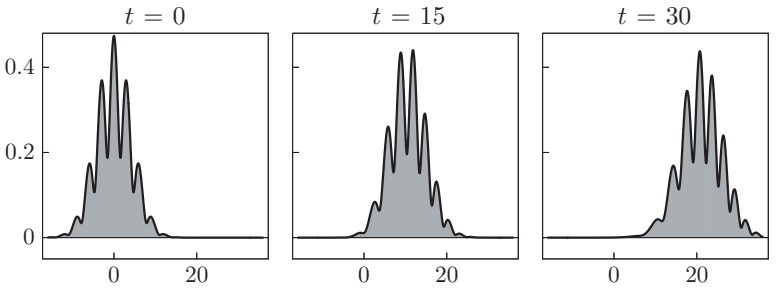


FIGURE 7.10. Time evolution of a Dirac wave packet showing $|\psi(x,t)|_2^2$ (the standard position probability density). The parts with positive and negative energy have the same velocity distribution but a disjoint momentum distribution. Although there is no Zitterbewegung, there is an ongoing interference between the parts with positive and negative energy.

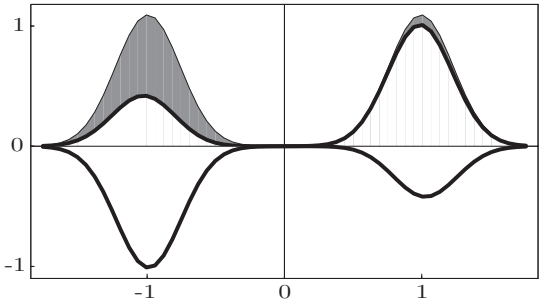


FIGURE 7.11. Image of a wave packet in momentum space, showing the upper component, the lower component (plotted upside down), and $|\psi|_2$ (the gray curve in the background). The part with positive momentum (on the right side) has positive energy. The part with negative momentum (on the left side) has negative energy. Hence, both parts move in the same direction and keep interfering in position space, as shown in Figure 7.10.



An initial function like (7.129) is explored in CD 6.7. The first topic CD 6.7.1 allows you to explore the various parts with positive and negative energy. It is interesting that although this wave packet has a completely symmetric momentum distribution with average momentum $k = 0$, the average velocity is positive. The expectation value of the position oscillates around a mean value that slowly drifts to the right. CD 6.7.2 shows the time evolution in position space, and CD 6.7.3 shows the time evolution in momentum space. While the momentum probability distribution remains constant, the upper and lower components of the Fourier transform exhibit the Zitterbewegung in momentum space.

On the other hand, the spinor

$$\psi_2(x) = P_{\text{pos}}\,\psi(x)\mathrm{e}^{\mathrm{i}k_0 x} + P_{\text{neg}}\,\psi(x)\mathrm{e}^{-\mathrm{i}k_0 x} \tag{7.130}$$

has a positive energy part that is shifted to the mean value k_0 in momentum space and a negative energy part that is shifted to $-k_0$ in momentum space. When k_0 is sufficiently large, the overlap of the two parts is approximately zero, and there is no noticeable Zitterbewegung (see Fig. 7.11). Note, however, that the two parts with opposite momenta correspond to parts with the same velocity in position space. Hence, the parts with positive and negative energy move together in position space and keep interfering with each other. The interference pattern is clearly visible in Figure 7.10.



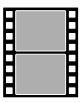
CD 6.12.1 shows the time evolution of an initial function similar to (7.130). Here, a small remaining overlap in momentum space causes a Zitterbewegung with a very small amplitude. In CD 6.12.2 we see a wave packet, where the parts with positive and negative energy are widely separated in momentum space, as in Figure 7.11. Although there is an ongoing interference of positive and negative energies in position space (as in Fig. 7.10), the average position moves essentially without Zitterbewegung.

In case of the wave packet

$$\psi_3(x) = N\,\exp(-x^2/2 + \mathrm{i}k_0 x)\,\begin{pmatrix}1\\0\end{pmatrix}, \tag{7.131}$$

the parts with positive and negative energy do overlap in momentum space. In position space, however, these parts move away from each other in opposite directions and ψ_{neg} and ψ_{pos} become (approximately) orthogonal to each other (see Fig. 7.8). At the same time, but to a lesser extent, the function $(2\mathrm{i}\,H_0)^{-1}\,F\,\psi_{\text{pos}}$ becomes more and more orthogonal to ψ_{neg} so that the Zitterbewegung gradually vanishes. Indeed, one can show quite generally that

$$z_\psi(t) \to 0 \quad \text{for any } \psi \in L^2(\mathbb{R})^2. \tag{7.132}$$



CD 6.9 shows examples of the type defined in (7.131) with a rapidly vanishing Zitterbewegung. The oscillation of the average position vanishes as soon as the parts with positive and negative energy become sufficiently separated in position space. The wave functions in CD 6.8 are symmetric with respect to reflections at $x = 0$. Hence, the expectation value of the position remains at the origin for all times and shows no Zitterbewegung. However, the time evolution of the wave packet develops the interference patterns characteristic of a superposition of positive and negative energy wave packets.

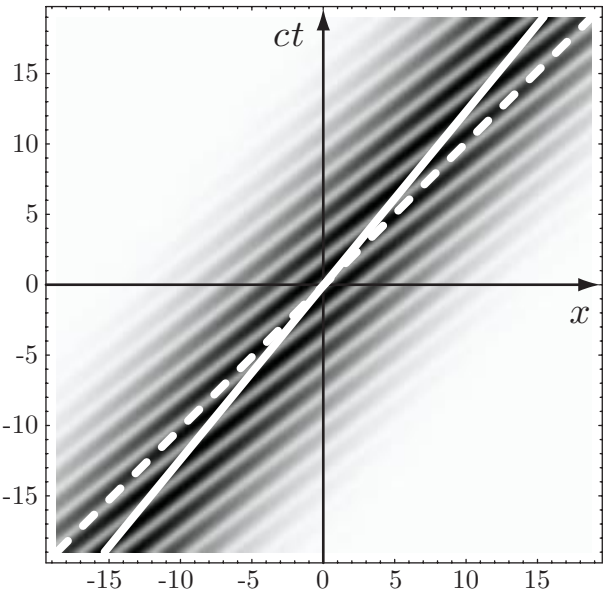


FIGURE 7.12. Space-time diagram of the motion of the wave packet in Figure 7.10. The local maxima of the position density move faster than light.

7.7.7. Superluminal motion?

The motion of a classical particle in one space dimension is described by a function $t \mapsto x(t)$. The *space-time trajectory* or *world line* is the set of all points $(x(t), ct)$ that are occupied by the particle in space-time. (Here, the time-coordinate is multiplied by c to give it the dimension of a length.) A world line is usually plotted as a curve in a *space-time diagram* with x as the horizontal coordinate and ct as the vertical coordinate. For example, the ct-axis is the world line of a particle that is at rest at $x = 0$. The angle between the world line and the vertical ct-axis is a measure for the velocity of the particle. All world lines of massive particles must be steeper than the world line of a photon (which has an angle of 45°).

Figure 7.12 is a space-time diagram of a wave packet. It shows the dependence of the standard position probability density on space and time coordinates. This image shows the wave packet of Figure 7.10, whose position density has several local maxima. We see that these ripples move at superluminal speed; the world line of a photon (the dashed white line) is steeper than the dark stripes of the position probability density. The solid white line is the world line of the average position $\langle x(t) \rangle_\psi$. The classical

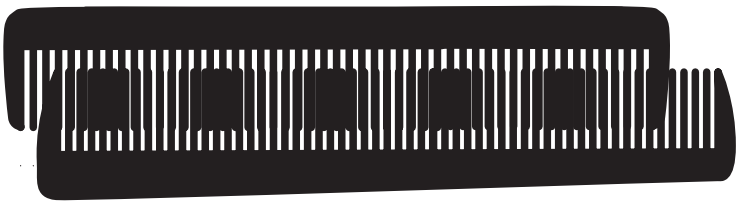


FIGURE 7.13. A Moiré pattern produced by two hair combs.

velocity of the particle is slower than the velocity of light. As explained in the previous section, the wave packet shows practically no Zitterbewegung.



CD 6.13.1 shows an animation of the wave packet discussed here similar to the one of Figure 7.10. Here, we compare the velocity of the interference ripples of this wave packet with the velocity of light. CD 6.13.2 shows that the velocity of the ripples is equal to the average phase velocity of parts with positive and negative energy. According to Section 7.4.2, the phase velocity is faster than the velocity of light. CD 6.13.3 is a classical example of an interference pattern (a Moiré pattern, see below) that moves faster than light. CD 6.13.4 shows that no information can propagate with the interference ripples. The information about a sudden perturbation of one of the ripples can at most spread with the velocity of light.

The superluminal motion of the ripples of a wave packet does not contradict the theory of relativity. These ripples are the result of the interference of slower moving wave packets with positive and negative energies. An interference pattern does not carry information from one point to another, and hence it can move with arbitrary speed.

A classical example of this effect is the *Moiré pattern* produced by overlaying two slightly different periodic patterns. For example, you can observe a Moiré pattern by looking through the teeth of two pocket combs, one placed in front of the other, as shown in Figure 7.13. Slide one comb slowly and watch how the Moiré pattern moves quickly.

Figure 7.14 shows the world lines of two patterns. Both patterns are one-dimensional regular grids of black dots. With respect to the given coordinate system, one of the patterns is at rest (the world lines of the black dots are vertical). The other pattern slides slowly to the right. Its world lines are slightly tilted.

In the given coordinate system, the patterns have a slightly different periodicity length. The superposition of the two pattern therefore leads to a Moiré pattern, a periodic succession of dark and bright areas in space. The

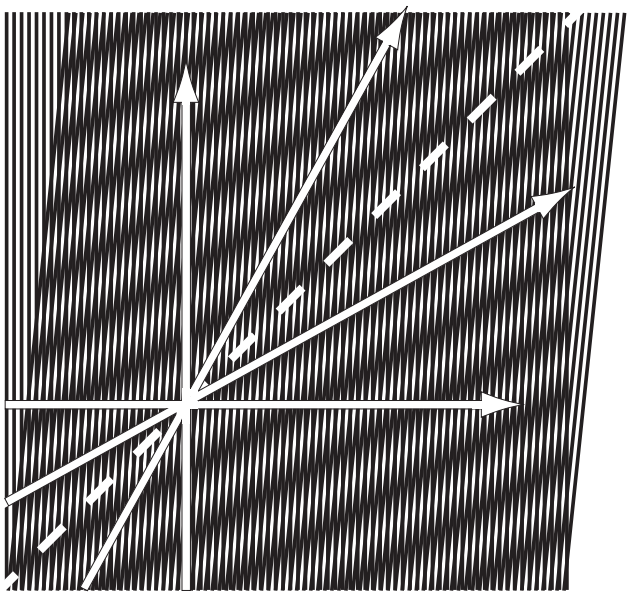


FIGURE 7.14. Space-time diagram of two one-dimensional patterns producing a Moiré pattern moving at superluminal velocity. In the moving coordinate system represented by the tilted coordinate axis, the Moiré pattern has no spatial structure.)

slowly drifting motion of one of generating patterns causes the Moiré pattern to move very quickly. Actually, the dark areas move much faster than light.

According to the special theory of relativity, a moving coordinate frame is given by tilted coordinate axis, as indicated in Figure 7.14. We can find a moving coordinate frame where both patterns have the same periodicity length, and hence the Moiré pattern completely looses its spatial structure. But due to the relative motion of the generating patterns, the Moiré pattern still depends on time. An observer in the moving frame has uniformly dark and bright times in periodic succession.

7.8. Special Topic: Energy Representation and Velocity Space

For a careful analysis of the behavior of spinor wave packets, it is useful to visualize the probability distributions of the energy and the classical velocity. This section provides the necessary theoretical and mathematical background.

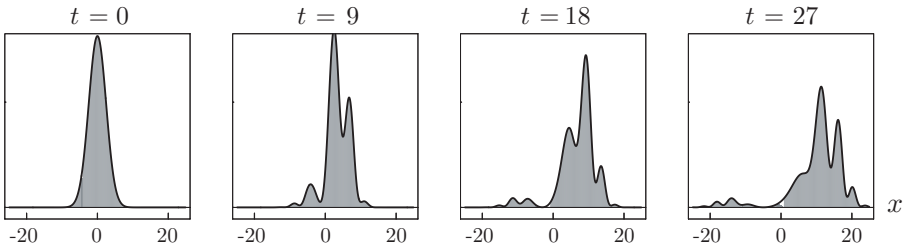
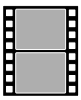


FIGURE 7.15. Time evolution of a wave packet with Gaussian initial shape according to the free-particle Dirac equation. The images show the standard position probability distribution at the times $t = 0, 9, 18, 27$.

As an example, consider Figure 7.15. It shows a Gaussian initial spinor that develops an interesting interference pattern during its time evolution. The energy distribution in this wave packet is plotted in Figure 7.16. It shows that the wave packet has equally sized parts with positive and negative energy. The interference of these parts is responsible for the ripples of the position probability density in position space. The range of allowed energies has a gap from $-\mathrm{m}c^2$ to $+\mathrm{m}c^2$, but in the dimensionless units used in Figure 7.16 we have $\mathrm{m} = c = 1$, and the gap extends from -1 to $+1$.

The velocity distribution of this wave packet is plotted in Figure 7.17 (left image). More precisely, this is the probability density of the values of the classical velocity operator v_{cl}. These values must be within the interval $(-c, c)$. (In dimensionless units, the values are strictly between -1 and $+1$.) The shape of the velocity distribution describes the averaged shape of the position distribution at late times. This shows the comparison with the position probability density at a late time (see the right part of Fig. 7.17). (A similar observation has been made in Book One, Sections 3.6 and 3.7.)

Both the energy and velocity distributions do not depend on time, because H_0 and v_{cl} are conserved quantities.



CD 6.10.1 shows an animation of the wave packet in Figure 7.15. The right-moving part and the left-moving part are both superpositions of positive and negative energies.

7.8.1. Variable substitutions

Some relevant physical observables for a free particle in one space dimension are the energy E, the momentum k, and the velocity v. In Dirac's quantum mechanics, all these quantities may be positive or negative. According to

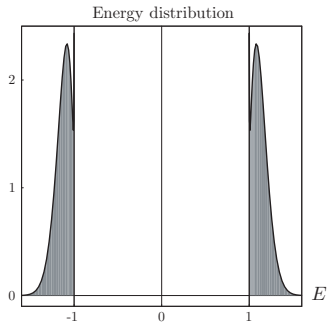


FIGURE 7.16. The stationary energy distribution of the wave packet in Figure 7.15 shows parts with positive and negative energies.

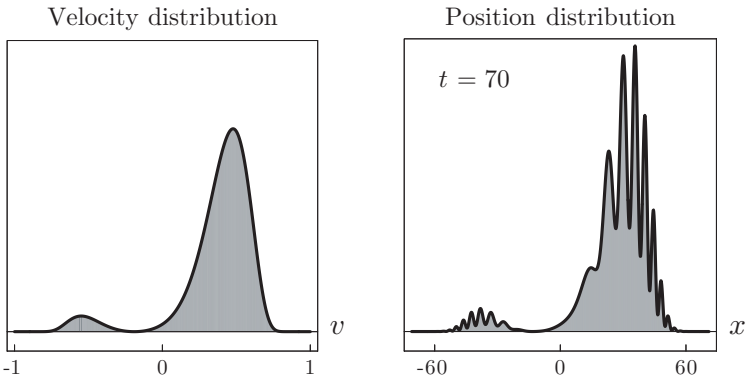


FIGURE 7.17. Comparison of the velocity and position distributions at a late time for the wave packet in Fig. 7.15.

the special theory of relativity, these observables are connected by

$$v = \frac{c^2 k}{E}, \tag{7.133}$$

and for a particle with mass m, we have

$$E^2 = c^2 k^2 + \mathrm{m}^2 c^4. \tag{7.134}$$

These two relations allow us to express any of these quantities as a function of any other and the sign of the third. For example, you may check that

$$v = \frac{c^2 k}{\sqrt{c^2 k^2 + \mathrm{m}^2 c^4}} \operatorname{sgn} E = \frac{c\sqrt{E^2 - \mathrm{m}^2 c^4}}{E} \operatorname{sgn} k. \tag{7.135}$$

EXERCISE 7.27. *Find the formulas that express k as a function of v and* $\operatorname{sgn} E$ *or as a function of E and* $\operatorname{sgn} v$. *Similarly, prove that*

$$E = \sqrt{c^2k^2 + \mathrm{m}^2c^4}\,\operatorname{sgn} k\,\operatorname{sgn} v = \frac{\mathrm{m}c^2}{\sqrt{1 - v^2/c^2}}\,\operatorname{sgn} k\,\operatorname{sgn} v. \tag{7.136}$$

7.8.2. Energy representation

For scattering theory, it is sometimes useful to use the energy as a variable. We can obtain this energy representation essentially by a variable substitution. When we replace the variable k by the variable E, one has to take into account that k is not only a function of E, but also depends on the sign of v, that is, on the direction of motion. We define

$$\kappa(E) = \frac{E}{c}\left(1 - \frac{\mathrm{m}^2c^4}{E^2}\right)^{\frac{1}{2}}, \quad \text{for all } E \in \sigma(H_0), \tag{7.137}$$

so that $k = \kappa(E)\operatorname{sgn} v$. Here, $\sigma(H_0)$ denotes the energy spectrum of the Dirac operator H_0, as defined in (7.40). Let us now define a plane wave with a positive velocity (the corresponding wave packets will move to the right). Using $u_{\mathrm{pos,neg}}$ from (7.47), we define the stationary solutions

$$\underset{\rightarrow}{\omega}(E;x) = n(E)\begin{cases} u_{\mathrm{pos}}(\kappa(E),x) & \text{for } E > \mathrm{m}c^2, \\ u_{\mathrm{neg}}(\kappa(E),x) & \text{for } E < -\mathrm{m}c^2, \end{cases} \tag{7.138}$$

and the time-dependent plane wave with positive velocity

$$\underset{\rightarrow}{\omega}(E;x,t) = \underset{\rightarrow}{\omega}(E;x)\,\mathrm{e}^{-\mathrm{i}\,Et}. \tag{7.139}$$

For $E < 0$, the momentum $\kappa(E)$ in the argument of the plane wave u_{neg} is negative. This corresponds to a positive velocity, as discussed before. The factor $n(E)$ is a normalization constant. We choose

$$n(E) = \sqrt{\frac{E}{c\kappa(E)}}\,. \tag{7.140}$$

Similarly, we choose the plane-wave solution with negative velocity as

$$\underset{\leftarrow}{\omega}(E;x) = n(E)\begin{cases} u_{\mathrm{pos}}(-\kappa(E),x) & \text{for } E > \mathrm{m}c^2, \\ u_{\mathrm{neg}}(-\kappa(E),x) & \text{for } E < -\mathrm{m}c^2, \end{cases} \tag{7.141}$$

$$\underset{\leftarrow}{\omega}(E;x,t) = \underset{\leftarrow}{\omega}(E;x)\,\mathrm{e}^{-\mathrm{i}\,Et}. \tag{7.142}$$

The arrow, of course, indicates the direction of motion.

It is useful to define

$$b_{\pm}(E) = \sqrt{\frac{1}{2}\Big(1 \pm \frac{\mathrm{m}c^2}{E}\Big)}, \quad \text{for } E \in \sigma(H_0). \tag{7.143}$$

These functions can be expressed in terms of the functions $d_\pm(k)$ defined in (7.36):

$$b_\pm(E) = \begin{cases} d_\pm(\kappa(E)) & \text{for } E > \mathrm{m}c^2, \\ \mp d_\mp(\kappa(E)) & \text{for } E < -\mathrm{m}c^2. \end{cases} \tag{7.144}$$

Now, the definitions of $\underset{\leftarrow}{\omega}$ and $\underset{\rightarrow}{\omega}$ above can be rewritten as

$$\underset{\rightarrow}{\omega}(E;x,t) = \frac{1}{\sqrt{2\pi}}\, \underset{\rightarrow}{\omega}(E)\, \mathrm{e}^{\mathrm{i}\,\kappa(E)\,x - \mathrm{i}Et}, \tag{7.145}$$

$$\underset{\leftarrow}{\omega}(E;x,t) = \frac{1}{\sqrt{2\pi}}\, \underset{\leftarrow}{\omega}(E)\, \mathrm{e}^{-\mathrm{i}\,\kappa(E)\,x - \mathrm{i}Et}, \tag{7.146}$$

where

$$\underset{\rightarrow}{\omega}(E) = n(E) \begin{pmatrix} b_+(E) \\ b_-(E) \end{pmatrix}, \tag{7.147}$$

$$\underset{\leftarrow}{\omega}(E) = n(E) \operatorname{sgn} E \begin{pmatrix} b_+(E) \\ -b_-(E) \end{pmatrix}. \tag{7.148}$$

We note the formulas

$$n(E)\, b_+(E) = \frac{1}{\sqrt{2}} \Big(\frac{E + \mathrm{m}c^2}{E - \mathrm{m}c^2} \Big)^{1/4}, \quad n(E)\, b_-(E) = \frac{1}{2\, n(E)\, b_+(E)}, \tag{7.149}$$

and

$$\underset{\leftarrow}{\omega}(E) = \operatorname{sgn} E\, \sigma_3\, \underset{\rightarrow}{\omega}(E), \quad \underset{\leftarrow}{\omega}(E;x,t) = \operatorname{sgn} E\, \sigma_3\, \underset{\rightarrow}{\omega}(E,-x,t). \tag{7.150}$$

We can form wave packets as usual. For example, take any square-integrable function $\underset{\rightarrow}{g}(E)$ and build the wave packet

$$\underset{\rightarrow}{\psi}(x,t) = \int_{\sigma(H_0)} \underset{\rightarrow}{g}(E)\, \underset{\rightarrow}{\omega}(E;x,t)\, dE. \tag{7.151}$$

Then, this wave packet moves to the right. Similarly, the wave packet

$$\underset{\leftarrow}{\psi}(x,t) = \int_{\sigma(H_0)} \underset{\leftarrow}{g}(E)\, \underset{\leftarrow}{\omega}(E;x,t)\, dE \tag{7.152}$$

moves to the left. The normalization constant $n(E)$ was chosen in order to have the relation

$$\big\| \underset{\rightarrow}{\psi} \big\|^2 = \int_{\sigma(H_0)} \big| \underset{\rightarrow}{g}(E) \big|^2\, dE, \tag{7.153}$$

(independent of t) and similarly for $\| \underset{\leftarrow}{\psi} \|$. Moreover, $\underset{\rightarrow}{\psi}$ and $\underset{\leftarrow}{\psi}$ are orthogonal. Any wave packet ψ can thus be decomposed into a right-moving and a left-moving part,

$$\psi = \underset{\rightarrow}{\psi} + \underset{\leftarrow}{\psi}, \qquad \langle \underset{\rightarrow}{\psi}, \underset{\leftarrow}{\psi} \rangle = 0, \tag{7.154}$$

and the corresponding functions $\underset{\rightarrow}{g}(E)$ and $\underset{\leftarrow}{g}(E)$ are given by the $\mathbb{C}^2$-scalar products

$$\underset{\rightarrow}{g}(E) = \big\langle \underset{\rightarrow}{\omega}(E), \hat{\psi}(\kappa(E)) \big\rangle_2 , \tag{7.155}$$

$$\underset{\leftarrow}{g}(E) = \big\langle \underset{\leftarrow}{\omega}(E), \hat{\psi}(-\kappa(E)) \big\rangle_2 . \tag{7.156}$$

The vector-valued function g defined for $E \in \sigma(H_0)$ by

$$g(E) = \begin{pmatrix} \underset{\rightarrow}{g}(E) \\ \underset{\leftarrow}{g}(E) \end{pmatrix}, \tag{7.157}$$

is called the energy representation of ψ. When ψ is square-integrable, so is g, and

$$\|\psi\|^2 = \big\| \underset{\rightarrow}{\psi} \big\|^2 + \big\| \underset{\leftarrow}{\psi} \big\|^2 = \int_{\sigma(H_0)} \Big(\big| \underset{\rightarrow}{g}(E) \big|^2 + \big| \underset{\leftarrow}{g}(E) \big|^2 \Big)\, dE = \|g\|^2. \tag{7.158}$$

Hence, the mapping $\psi \in L^2(\mathbb{R}, dx)^2$ to $g \in L^2(\sigma(H_0), dE)^2$ is unitary. In the energy representation, the Dirac operator acts as a multiplication by the variable E. This means that the function $E\, g(E)$ is the energy representation of $H_0\psi(x)$, assuming that $g(E)$ is the energy representation of $\psi(x)$.

7.8.3. Velocity representation

For the sake of completeness, we list some formulas for the representation, where the classical velocity v is used as a variable. We define

$$k(v) = \frac{\mathrm{m}\, v}{\sqrt{1 - v^2/c^2}}, \quad -c < v < c. \tag{7.159}$$

We start with a Dirac spinor $\hat{\psi}$ in momentum space and write

$$\hat{\psi}(k) = \hat{\psi}_{\mathrm{pos}}(k) + \hat{\psi}_{\mathrm{neg}}(k), \quad \hat{\psi}_{\substack{\mathrm{pos} \\ \mathrm{neg}}}(k) = \frac{1}{2} \Big(\mathbf{1} \pm \frac{\mathbf{h}_0(k)}{\lambda(k)} \Big)\, \hat{\psi}(k). \tag{7.160}$$

Then we define

$$f_{\mathrm{pos}}(v) = \frac{\sqrt{\mathrm{m}}}{\big(1 - v^2/c^2\big)^{3/4}}\, \hat{\psi}_{\mathrm{pos}}\big(k(v)\big), \tag{7.161}$$

$$f_{\mathrm{neg}}(v) = \frac{\sqrt{\mathrm{m}}}{\big(1 - v^2/c^2\big)^{3/4}}\, \sigma_3\, \hat{\psi}_{\mathrm{neg}}\big(-k(v)\big). \tag{7.162}$$

For each v in the interval $(-c, c)$, these two vectors are orthogonal in $\mathbb{C}^2$,

$$\big\langle f_{\mathrm{pos}}(v), f_{\mathrm{neg}}(v) \big\rangle_2 = 0. \tag{7.163}$$

We define

$$f(v) = f_{\mathrm{pos}}(v) + f_{\mathrm{neg}}(v) \tag{7.164}$$

as the velocity representation of ψ. The factors in (7.161) and (7.162) have been chosen such that

$$\int_{-c}^{c} |f(v)|_2^2 \, dv = \int_{-\infty}^{\infty} |\hat{\psi}(k)|_2^2 \, dk. \tag{7.165}$$

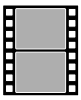
Hence, the transformation $\hat{\psi} \to f$ is a unitary transformation from the momentum space $L^2(\mathbb{R}, dk)^2$ into the "velocity space" $L^2((-c,c), dv)^2$. The action of the classical velocity operator v_{cl} in momentum space is given by

$$v_{\text{cl}}(k)\, \hat{\psi}(k) = \frac{c^2 k}{\lambda(k)}\, \hat{\psi}_{\text{pos}}(k) - \frac{c^2 k}{\lambda(k)}\, \hat{\psi}_{\text{neg}}(k). \tag{7.166}$$

Making the transition to the velocity representation, we find

$$v_{\text{cl}}(k)\, \hat{\psi}(k) \to \frac{c^2 k(v)}{\lambda\big(k(v)\big)}\, f_{\text{pos}}(v) - \frac{c^2(-k(v))}{\lambda\big(k(v)\big)}\, f_{\text{neg}}(k) = v\, f(v). \tag{7.167}$$

Hence, in velocity space the classical velocity operator acts as a multiplication by the variable v.



All sections CD 6.7–CD 6.12 contain visualizations of the energy representation $g(E)$ and the velocity representation $f(v)$ of the wave packet under consideration. Moreover, you can explore the initial wave packet $\psi(x)$ and its various parts in both position and momentum space: The parts with positive/negative energy $\psi_{\text{pos}}(x)$, $\psi_{\text{neg}}(x)$, the parts with positive/negative velocity $\underset{\rightarrow}{\psi}(x)$, $\underset{\leftarrow}{\psi}(x)$, and their Fourier transforms.

7.9. Relativistic Invariance

7.9.1. Lorentz transformations

An *event* is something that happens at a definite time at a definite place, that is, at a point in space-time. Examples of events are the emission of a photon or the detection of a particle. The coordinates of an event are described with respect to a suitable *inertial frame*.[6] For simplicity, we consider a one-dimensional situation. We use space-time coordinates $(ct, x) \in \mathbb{R}^2$, where x is the space coordinate and ct is the time coordinate of the event. The factor c denotes the velocity of light. It gives the first coordinate ct the dimension of a length. The coordinate transformations $I \to I'$ between all possible inertial frames are called Poincaré transformations. The *principle of relativity* states that all inertial frames are equivalent for the description

[6] An inertial frame is usually defined as a space-time coordinate system in which the law of inertia holds.

of nature. Hence, the Poincaré transformations should be represented by symmetry transformations within any physical theory.

Let us start in an inertial frame I, where the space-time coordinates of an event are denoted by (ct, x). Consider a second inertial frame I', where the same event has the space-time coordinates (ct', x'). We assume that I' moves with a velocity $-v$ with respect to the first frame I, and that at time $t = 0$ the origin of I' is at $x = 0$. Then, any book on special relativity tells you that the transformation

$$ct' = \gamma(v)\left(ct + \frac{v}{c}\,x\right), \qquad x' = \gamma(v)\left(x + \frac{v}{c}\,ct\right) \tag{7.168}$$

gives the coordinates (ct', x') in terms of the coordinates (ct, x). Here

$$\gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}. \tag{7.169}$$

The Poincaré transformation (7.168) is called a one-dimensional *Lorentz-boost* or *velocity transformation*. In order to be consistent with our presentation of rotations, we rather want to consider the "active" point of view, where (ct', x') are the coordinates of a particle in I that has been boosted with velocity $+v$.

In view of $-1 < v/c < 1$, it is useful to introduce the parameter ω such that

$$\frac{v}{c} = \tanh\omega, \qquad \gamma(v) = \cosh\omega. \tag{7.170}$$

Then the Lorentz boost can be written as

$$ct \to ct' = ct\,\cosh\omega + x\,\sinh\omega, \tag{7.171}$$

$$x \to x' = ct\,\sinh\omega + x\,\cosh\omega. \tag{7.172}$$

We denote the two-by-two matrix of a Lorentz boost by

$$\Lambda(\omega) = \begin{pmatrix} \cosh\omega & \sinh\omega \\ \sinh\omega & \cosh\omega \end{pmatrix}. \tag{7.173}$$

The parametrization by ω has the following advantage. If we compose two Lorentz boosts, the corresponding parameters ω are simply added,

$$\Lambda(\omega_2)\,\Lambda(\omega_1) = \Lambda(\omega_1 + \omega_2). \tag{7.174}$$

Hence, the result of this composition is again a Lorentz boost. The resulting boost has the velocity

$$v_3 = \frac{v_1 + v_2}{1 + \dfrac{v_1 v_2}{c^2}}. \tag{7.175}$$

This relativistic law for adding velocities follows immediately from the identity

$$\tanh(\omega_1 + \omega_2) = \frac{\tanh\omega_1 + \tanh\omega_2}{1 + \tanh\omega_1\,\tanh\omega_2}. \tag{7.176}$$

In the vector space $\mathbb{R}^2$, we may define the Minkowski scalar product

$$a \cdot b = a^0\, b^0 - a^1\, b^1, \quad a = \begin{pmatrix} a^0 \\ a^1 \end{pmatrix} \in \mathbb{R}^2, \quad b = \begin{pmatrix} b^0 \\ b^1 \end{pmatrix} \in \mathbb{R}^2. \tag{7.177}$$

You may verify that

$$\Lambda(\omega)a \cdot \Lambda(\omega)b = a \cdot b. \tag{7.178}$$

One may define a Lorentz transformation as a linear map of $\mathbb{R}^2$ into itself that leaves this scalar product invariant. The set of all Lorentz transformations contains

$\Lambda(\omega) : \mathbb{R}^2 \to \mathbb{R}^2,\ (\omega \in \mathbb{R})$	all Lorentz boosts,
$\Lambda_P : (ct, x) \to (ct, -x)$	the space-reflection (parity transform),
$\Lambda_T : (ct, x) \to (-ct, x)$	the time reversal,
$\Lambda_{PT} : (ct, x) \to (-ct, -x)$	the space-time inversion,

together with all possible compositions (matrix products) of these mappings. This set is called the *Lorentz group* in two dimensions.

7.9.2. Transformation of the Dirac equation

Lorentz transformations are symmetry transformations, and they should be implemented as unitary transformations in the Hilbert space of the Dirac equation. Because a velocity transformation mixes the space and time coordinates, we have to consider time-dependent spinors, that is, solutions $\psi(x, t)$ of the Dirac equation. We want to implement the Lorentz transformation as a unitary mapping $\psi \to \varphi$ of Dirac spinors in such a way that if ψ is a solution of the Dirac equation, then so is the Lorentz-transformed function ϕ.

We start by applying the *inverse* Lorentz transformation to the arguments x and t of a solution $\psi(x, t)$ of the Dirac equation. This is analogous to our definition of a rotated wave function in Section 1.2.2. In Chapter 1, we showed that for any solution $\psi(\mathbf{x}, t)$ of a spherically symmetric Schrödinger equation, the rotated function $\psi(\mathbf{x}', t)$, with $\mathbf{x}' = \mathbf{R}(\boldsymbol{\alpha})^{-1}\mathbf{x}$, gives another solution.

Let us substitute x and t in $\psi(x, t)$ by x' and t', where

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \Lambda^{-1} \begin{pmatrix} ct \\ x \end{pmatrix}, \tag{7.179}$$

and where Λ is any Lorentz transformation. In order to investigate the effect of this substitution, it is convenient to write the free-particle Dirac equation in a form that involves the space and time coordinates in a more symmetric

way. The free-particle Dirac equation (7.14) reads

$$\Big(\mathrm{i}\,\frac{\partial}{\partial t} + \mathrm{i}\,c\alpha\,\frac{\partial}{\partial x} - \beta \mathrm{m}c^2\Big)\,\psi(x,t) = 0. \tag{7.180}$$

We multiply this equation by β/c from the left and obtain, using $\beta^2 = \mathbf{1}$,

$$\Big(\mathrm{i}\,\beta\,\frac{\partial}{\partial ct} + \mathrm{i}\,\beta\alpha\,\frac{\partial}{\partial x} - \mathrm{m}c\Big)\,\psi(x,t) = 0. \tag{7.181}$$

Let us introduce the abbreviations

$$\partial = \begin{pmatrix} (1/c)\,\partial/\partial t \\ -\partial/\partial x \end{pmatrix}, \qquad \gamma = \begin{pmatrix} \beta \\ \beta\alpha \end{pmatrix}. \tag{7.182}$$

Define the product $\gamma \cdot \partial$ as in (7.177). Then we can write (7.181) in the form

$$\big(\mathrm{i}\,\gamma \cdot \partial - \mathrm{m}c\big)\,\psi(x,t) = 0. \tag{7.183}$$

Next, we want to investigate the effect of the variable substitution (7.179) on the Dirac equation. Replacing (x,t) with (x',t') gives

$$\big(\mathrm{i}\,\gamma \cdot \partial' - \mathrm{m}c\big)\,\psi(x',t') = 0, \tag{7.184}$$

where

$$\partial' = \begin{pmatrix} (1/c)\,\partial/\partial t' \\ -\partial/\partial x' \end{pmatrix}. \tag{7.185}$$

We want to find out whether the function $\psi(x',t')$ is related to a solution of the original Dirac equation. To that purpose, we must express ∂' in terms of ∂. In order to be specific, we consider a proper Lorentz transformation and use

$$\Lambda^{-1} = \begin{pmatrix} \cosh\omega & -\sinh\omega \\ -\sinh\omega & \cosh\omega \end{pmatrix}. \tag{7.186}$$

Let us compute $\partial\,\psi(x',t')$. By applying the chain rule, we obtain for the first component of $\partial\psi(x',t')$

$$\begin{aligned} \frac{1}{c}\,\frac{\partial}{\partial t}\,\psi(x',t') &= \frac{1}{c}\,\frac{\partial}{\partial t'}\,\psi(x',t')\,\frac{\partial t'}{\partial t} + \frac{\partial}{\partial x'}\,\psi(x',t')\,\frac{1}{c}\,\frac{\partial x'}{\partial t} \\ &= \Big(\cosh\omega\,\frac{1}{c}\,\frac{\partial}{\partial t'} - \sinh\omega\,\frac{\partial}{\partial x'}\Big)\,\psi(x',t'). \end{aligned} \tag{7.187}$$

Similarly, for the second component $-\partial/\partial x$ of ∂, we obtain

$$-\frac{\partial}{\partial x}\,\psi(x',t') = \Big(\sinh\omega\,\frac{1}{c}\,\frac{\partial}{\partial t'} - \cosh\omega\,\frac{\partial}{\partial x'}\Big)\,\psi(x',t'). \tag{7.188}$$

We can combine these results into

$$\partial\,\psi(x',t') = \big(\Lambda\,\partial'\big)\,\psi(x',t'). \tag{7.189}$$

Multiplying this equation from the left by Λ^{-1} gives

$$\big(\Lambda^{-1}\partial\big)\,\psi(x',t') = \partial'\,\psi(x',t'). \tag{7.190}$$

Hence, (7.184) becomes

$$\big(\mathrm{i}\,\gamma\cdot(\Lambda^{-1}\partial) - \mathrm{m}c\big)\,\psi(x',t') = 0. \tag{7.191}$$

Hence, the function $\psi(x',t')$ is not a solution of the Dirac equation (7.183).

In order to define the Lorentz transformation of a spinor, the variable substitution alone is not sufficient.

7.9.3. Representation of Lorentz boosts

Before we proceed, we want to investigate the properties of the matrix

$$\gamma\cdot u = \beta\,u^0 - \beta\alpha\,u^1. \tag{7.192}$$

Note that $\gamma\cdot\partial$ appears in the Dirac equation (7.183). We are going to need the following result.

Lorentz boosts:
Let $\omega\in\mathbb{R}$, $u = \begin{pmatrix} u^0 \\ u^1 \end{pmatrix} \in \mathbb{R}^2$, and $\gamma = \begin{pmatrix} \beta \\ \beta\alpha \end{pmatrix}$.
Then

$$\exp\Big(\frac{\omega}{2}\alpha\Big)\;\gamma\cdot u\;\exp\Big(-\frac{\omega}{2}\alpha\Big) = \gamma\cdot v \tag{7.193}$$

with $v = \Lambda(\omega)\,u$.

Proof. In order to prove this useful formula, we first note that the exponential function of the matrix $\omega\alpha/2$ is given by the usual power series

$$\exp\Big(\frac{\omega}{2}\alpha\Big) = \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\frac{\omega}{2}\alpha\Big)^n. \tag{7.194}$$

Using the property $\alpha^2 = \mathbf{1}_2$, we find $\alpha^n = \mathbf{1}_2$ if $n = 2k$ is even, and $\alpha^n = \alpha$ if $n = 2k+1$ is odd. Hence, the power series becomes

$$\mathbf{1}_2\sum_{k=0}^{\infty}\frac{\omega^{2k}}{(2k)!} + \alpha\sum_{k=0}^{\infty}\frac{\omega^{2k+1}}{(2k+1)!}. \tag{7.195}$$

Comparing this with the power series of cosh and sinh, we find the result

$$\exp\Big(\frac{\omega}{2}\alpha\Big) = \mathbf{1}_2\,\cosh\frac{\omega}{2} + \alpha\,\sinh\frac{\omega}{2}. \tag{7.196}$$

Now, the matrix α anticommutes with β and as a consequence we obtain

$$\exp\Big(\frac{\omega}{2}\alpha\Big)\,\beta = \beta\,\exp\Big(-\frac{\omega}{2}\alpha\Big). \tag{7.197}$$

Likewise, α anticommutes with the matrix $\gamma \cdot u$. But then

$$\begin{aligned}
&\exp\Big(\frac{\omega}{2}\alpha\Big)\,\gamma\cdot u\,\exp\Big(-\frac{\omega}{2}\alpha\Big)\\
&\qquad=\exp\Big(\frac{\omega}{2}\Big)\,\exp\Big(\frac{\omega}{2}\Big)\,\gamma\cdot u=\exp(\omega\alpha)\,\gamma\cdot u\\
&\qquad=(\cosh\omega+\alpha\,\sinh\omega)\,(\beta\,u^0-\beta\alpha\,u^1)\\
&\qquad=(\beta\,(u^0\,\cosh\omega+u^1\,\sinh\omega)-\beta\alpha\,(u^0\,\sinh\omega+u^1\,\cosh\omega)\\
&\qquad=\gamma\cdot\big(\Lambda(\omega)\,u\big).
\end{aligned}\tag{7.198}$$

This proves (7.193). □

EXERCISE 7.28. *Show that in the standard representation $\alpha=\sigma_1$, we have* $\exp(\omega\alpha)=\Lambda(\omega)$.

7.9.4. Invariance of the free-particle Dirac equation

With these results, we can prove the invariance of the Dirac equation under Lorentz boosts. From Eq. (7.193) we conclude that

$$\gamma\cdot(\Lambda(\omega)^{-1}\partial)=\exp\Big(-\frac{\omega}{2}\alpha\Big)\,\gamma\cdot\partial\,\exp\Big(\frac{\omega}{2}\alpha\Big)\tag{7.199}$$

and hence Equation (7.191) can be written as

$$\Big(\mathrm{i}\,\exp\Big(-\frac{\omega}{2}\alpha\Big)\,\gamma\cdot\partial\,\exp\Big(\frac{\omega}{2}\alpha\Big)-\mathrm{m}c\Big)\,\psi(x',t')=0.\tag{7.200}$$

Multiplying this equation by $\exp(\omega\alpha/2)$ from the left leads to

$$\big(\mathrm{i}\,\gamma\cdot\partial-\mathrm{m}c\big)\,\exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi(x',t')=0.\tag{7.201}$$

This is again the Dirac equation (7.183). Hence, the function

$$\begin{aligned}
&\varphi(x,t)=\exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi(x',t')\\
&=\exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi\big(x\,\cosh\omega-ct\,\sinh\omega\,,\,t\,\cosh\omega-(x/c)\,\sinh\omega\big)
\end{aligned}\tag{7.202}$$

is a solution of the Dirac equation whenever $\psi(x,t)$ is.

Lorentz invariance of the free-particle Dirac equation:

Consider a solution $\psi(x,t)$ of the free-particle Dirac equation. A Lorentz boost applied to ψ gives

$$\varphi(x,t) = \exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi(x',t') \tag{7.203}$$

with

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \Lambda(\omega)^{-1}\,\begin{pmatrix} ct \\ x \end{pmatrix}. \tag{7.204}$$

The function $\varphi(x,t)$ is a solution of the free-particle Dirac equation.

7.9.5. Unitary implementation of Lorentz boosts

Let us see how the Lorentz transformations are implemented in the Hilbert space $L^2(\mathbb{R})^2$ of the Dirac equation. Consider a Dirac spinor (at time $t=0$)

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}. \tag{7.205}$$

An active Lorentz transformation converts this spinor into

$$\varphi(x) = \exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi\big(x\,\cosh\omega\,,\,-(x/c)\,\sinh\omega\big) \tag{7.206}$$

Here, $\psi(x,t) = \exp(-\mathrm{i}H_0 t)\psi(x)$ is the solution of the Dirac equation with initial condition $\psi(x,0)=\psi(x)$. Now, we are going to investigate the properties of the mapping $U_L(\omega):\psi\to\varphi$.

The infinitesimal generator of a Lorentz boost:

A Lorentz boost of a Dirac spinor is a unitary transformation in the Hilbert space $L^2(\mathbb{R})^2$,

$$U_L(\omega)\,\psi(x) = \exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi\big(x\,\cosh\omega\,,\,-(x/c)\,\sinh\omega\big). \tag{7.207}$$

The infinitesimal generator is given by

$$N = \frac{1}{2}(H_0\,x + x\,H_0), \tag{7.208}$$

(which is self-adjoint on a suitable dense domain). We have

$$U_L(\omega) = \exp\Big(\mathrm{i}\,\omega\frac{N}{c}\Big). \tag{7.209}$$

PROOF. The generator N is obtained as usual by differentiation of $U_L(\omega)$ with respect to ω at $\omega = 0$

$$\frac{N}{c}\,\psi = -\mathrm{i}\,\frac{d}{d\omega}\,\exp\Big(\mathrm{i}\,\omega\,\frac{N}{c}\Big)\,\psi\,\Big|_{\omega=0}. \tag{7.210}$$

Hence, we compute

$$\begin{aligned} -\mathrm{i}\,\frac{d}{d\omega}\exp\Big(\frac{\omega}{2}\alpha\Big)\,\psi\big(x\,\cosh\omega, -(x/c)\,\sinh\omega\big) \\ = -\,\mathrm{i}\exp\Big(\frac{\omega}{2}\alpha\Big)\,\Big(\frac{\alpha}{2} + x\,(\sinh\omega)\,\partial_1 - (x/c)\,(\cosh\omega)\,\partial_2\Big) \\ \psi\big(x\,\cosh\omega, -(x/c)\,\sinh\omega\big). \end{aligned} \tag{7.211}$$

In the last step, we have used the chain rule. Here, ∂_1 is the derivative of the Dirac spinor with respect to the first argument, and ∂_2 denotes the derivative with respect to the second argument. Any solution ψ of the Dirac equation satisfies

$$\mathrm{i}\,\partial_2\,\psi = H_0\,\psi, \tag{7.212}$$

where H_0 is the free-particle Dirac operator. Substituting this in (7.211) and evaluating (7.211) at $\omega = 0$ gives

$$\frac{N}{c} = -\mathrm{i}\,\frac{\alpha}{2} + x\,\frac{H_0}{c} = \frac{1}{2c}(H_0\,x + x\,H_0). \tag{7.213}$$

This proves (7.208) above. Though the symmetry of N is rather obvious, the proof of self-adjointness on a suitable domain is a bit more technical and is omitted here. The self-adjointness of N proves that $U_L(\omega)$ is a strongly continuous one-parameter unitary group. □

We note that similar considerations lead to the conclusion that the parity transformation Λ_P can be implemented as the unitary transformation

$$U_P\,\psi(ct, x) = \beta\,\psi(ct, -x). \tag{7.214}$$

It is, however, necessary to realize the time reversal as well as the space-time inversion via antiunitary operators. All these operators thus define symmetry transformations of the Dirac theory. Moreover, the free-particle Dirac equation is invariant with respect to all Lorentz transformations (including the discrete transformations Λ_P, Λ_T, and Λ_{PT}.

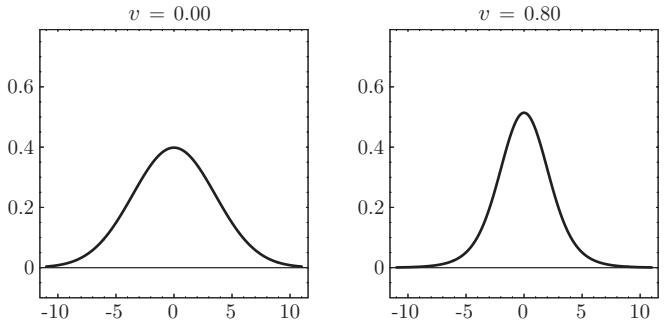


FIGURE 7.18. Behavior of a positive-energy wave packet under a velocity transformation. The right image shows the wave packet after a Lorentz transformation (boost) with the indicated velocity v (measured in units of c). The relative motion of observer and wave packet causes a Lorentz contraction.



We note that a spinor-wave packet whose components are perfectly symmetric with respect to reflections at the origin need not be invariant under a parity transformation. An example is the wave packet shown in CD 6.7. Here, the initial wave packet satisfies $\psi(x) = \psi(-x)$, but for the parity-invariance, we would need $\psi(x) = \beta\,\psi(-x)$. Indeed, the time evolution of this wave packet reveals its asymmetry (the average velocity of this wave packet is positive). On the other hand, the wave packets in CD 6.8 are all invariant with respect to a parity transformation.

It is interesting to see the Lorentz boost of a wave packet. Consider a wave packet with positive energy that is at rest in some inertial frame. The average velocity of this wave packet is zero, and the Fourier transform is localized around the origin in momentum space. In Figure 7.18, we show the wave packet with positive energy and its Lorentz transformation. After the transformation, the wave packet has the average velocity $v = 0.8c$ in the positive x-direction. The effect of the Lorentz contraction is clearly visible.



The interactive images in CD 6.14 show the effect of (active) Lorentz boosts on a wave packet with positive energy and on a wave packet with negative energy. For a wave packet with negative energy, an active boost with a positive velocity produces a wave packet with negative momentum.

Next, consider a wave packet with positive and negative energies, like the spinor that is a Gaussian function in both components. A Lorentz transformation shifts the positive-energy part toward positive momenta whereas the

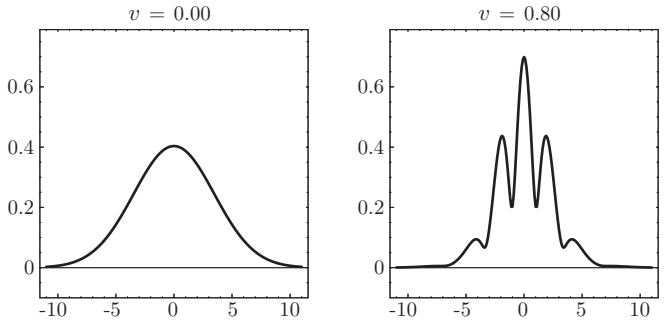
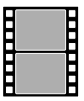


FIGURE 7.19. A superposition of positive and negative energy wave packets before and after a Lorentz boost. Apart from the Lorentz contraction, the relative motion of observer and wave packet produces interference effects. The interference is caused by the separation of the negative and positive-energy parts in momentum space.

negative-energy part is shifted in the opposite direction, because for negative energies the momentum is minus the velocity. The interference of these two parts causes the ripples in position space that are shown in Figure 7.19. Note: This image can also be interpreted as a snapshot of the wave packet as it is seen from a moving inertial frame (with velocity $-v$). The interferences are caused only by the relative motion of the wave packet and the observer.

Figure 7.20 shows a space-time diagram of the standard position probability density, that is, a density plot of the function $|\psi(x,t)|_2^2$. At $t = 0$ (along the x-axis), the wave packet has a Gaussian shape as shown in Figure 7.19 (left image). During the time evolution (observed with respect to the (x,t)-coordinate system) the wave packet develops interference ripples (see also Fig. 7.2). A moving observer describes the position probability density with respect to the (x',t')-coordinate system. As is well-known from the special theory of relativity, the events happening at $t' = 0$ in the moving coordinate system (on the x'-axis) are not simultaneous with respect to the (x,t)-coordinates. On the inclined x'-axis, the position probability density has the interference ripples visible in the right image of Figure 7.19.



CD 6.15 explores the effect of Lorentz boosts applied to wave packets that are superpositions of positive and negative energies. In momentum space, a velocity transformation shifts the parts with positive and negative energies in opposite directions (CD 6.15.3). This causes an interference in position space.

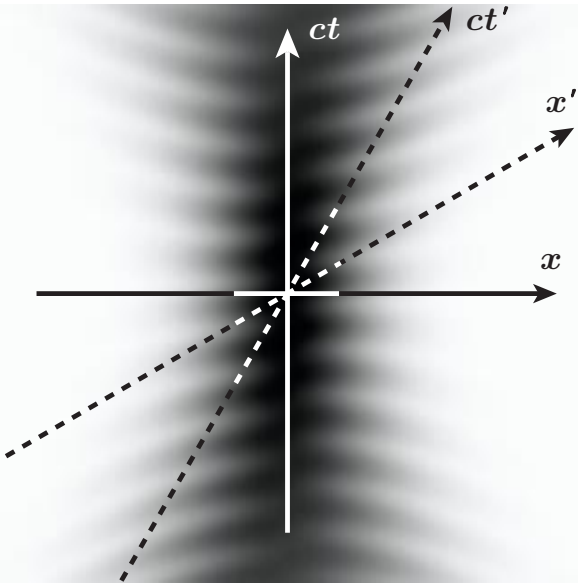


FIGURE 7.20. Space-time diagram of the position density of the wave packet in Figure 7.19. At $t = 0$ the density has a Gaussian shape. If viewed from a moving frame of reference at $t' = 0$ (that is, along the x'-axis) the density has ripples.

Chapter 8

The Dirac Equation

Chapter summary: In this chapter, we obtain the Dirac equation with external fields in three dimensions (Section 8.1). The number of components of the Dirac spinors is doubled compared to the one-dimensional situation. Roughly speaking, there are two spin components for each sign of the energy.

In the presence of external fields, the Dirac equation cannot be invariant under Lorentz transformations. But the Dirac equation is covariant in the sense that a Lorentz-transformed solution of the Dirac equation is a solution of the Dirac equation with an appropriately transformed potential energy (Section 8.2). It is possible to classify the potential functions according to their behavior under Poincaré transformations as scalar, electromagnetic, and tensor fields (Section 8.3).

Wave packets with negative energy behave quite differently from wave packets with positive energy if put into an external electromagnetic field. By introducing the operation of charge transformation, one can see that a wave packet with negative energy actually describes the behavior of a particle with positive energy but with opposite charge (Section 8.4). We can thus interpret the solutions with negative energy as antiparticles. But this interpretation only works in situations where the splitting of the Hilbert space in electronic and positronic states is meaningful and unambiguous. A counter-example showing the limits of Dirac theory is the Klein paradox, where particles starting as electrons may end up as positrons.

In Section 8.5, we investigate the connection between Dirac's theory and the nonrelativistic Pauli equation for particles with spin. The eigenvalues and eigenfunctions of the Dirac equation tend to their nonrelativistic counterparts as $c \to \infty$. We derive some formulas that let us compute relativistic perturbations of nonrelativistic energies up to first order in $1/c^2$.

The role of spherical symmetry in relativistic quantum mechanics is as important as in nonrelativistic quantum mechanics. In Section 8.6, we describe the angular-momentum subspaces of Dirac's theory. The radial Dirac equation becomes a system of two ordinary differential equations.

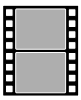
The hydrogen atom is perhaps the most important testing ground for any quantum mechanical theory. Fortunately, the Dirac equation for a hydrogen-like system can be solved analytically, and the results are in almost perfect agreement with the measurements. This success was one of the main reasons for the quick acceptance of the Dirac equation. In Section 8.8, we solve the radial Dirac equation for the hydrogen atom by factorization methods ("supersymmetry"). The higher symmetry

of the Coulomb problem in the relativistic case is related to the conservation of the Biedenharn-Johnson-Lippmann operator.

8.1. The Dirac Equation

8.1.1. Free Dirac operator in two and three dimensions

Many results obtained for the one-dimensional Dirac equation are completely analogous in higher space dimensions.



CD 6.16 and CD 6.17 show solutions of the free-particle Dirac equation in two dimensions. These solutions show Zitterbewegung and interference effects similar to one-dimensional solutions. The visualization of a $\mathbb{C}^2$-valued function is, of course, much more difficult in two dimensions. Here, we decided to visualize the upper and the lower component in separate plots.

The Dirac operator in higher dimensions is a matrix-differential operator that depends linearly on the momentum $c\,\mathbf{p}$ and the rest energy $\mathrm{m}c^2$, that is,

$$H_0 = c\,\boldsymbol{\alpha}\cdot\mathbf{p} + \beta\,\mathrm{m}c^2. \tag{8.1}$$

Again, $\mathbf{p} = -\mathrm{i}\,\nabla$ is the usual momentum operator and

$$\boldsymbol{\alpha}\cdot\mathbf{p} = \sum_{k=1}^{n} \alpha_k\, p_k, \quad p_k = -\mathrm{i}\,\frac{\partial}{\partial x_k}, \tag{8.2}$$

where n is the space dimension. Note that throughout this chapter, we use units with $\hbar = 1$. The $n+1$ quantities $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and β (the *Dirac matrices*) are matrices with constant coefficients, hence they commute with the momentum operator $\mathbf{p}$,

$$[\alpha_j, p_k] = [\beta, p_k] = 0, \quad j,k = 1,\ldots,n. \tag{8.3}$$

All these operators act on spinor-wave functions with several components (depending on the dimension of the Dirac matrices).

We want the Dirac matrices to be Hermitian, otherwise the Dirac operator would not be self-adjoint. And we want the square of H_0 to satisfy the classical relativistic relation

$$H_0^2 = c^2p^2 + \mathrm{m}^2c^4 = -c^2\Delta + \mathrm{m}^2c^4. \tag{8.4}$$

This imposes the following conditions on the Dirac matrices

$$\begin{aligned} \alpha_j\alpha_k + \alpha_k\alpha_j &= 2\delta_{jk}\,\mathbf{1} \quad \text{with } j,k = 1,\ldots,n. \\ \alpha_j\beta + \beta\alpha_j &= \mathbf{0} \quad \text{with } j = 1,\ldots,n. \\ \beta^2 &= \mathbf{1}. \end{aligned} \tag{8.5}$$

Now, we have to distinguish between two and three space dimensions. For $n = 2$, we need three anticommuting Hermitian matrices. A good example of such a set is already known to us, namely, the Pauli matrices

$$\alpha_1 = \sigma_1, \quad \alpha_2 = \sigma_2, \quad \beta = \sigma_3. \tag{8.6}$$

In three dimensions we need four anticommuting matrices. But it is not possible to find a fourth Hermitian 2×2 matrix that anticommutes with the three Pauli matrices. For $n = 3$, we thus have to consider higher dimensional matrices. A suitable set of four anticommuting matrices is given by the 4×4 matrices

$$\alpha_j = \begin{pmatrix} \mathbf{0}_2 & \sigma_j \\ \sigma_j & \mathbf{0}_2 \end{pmatrix} \quad j = 1, 2, 3, \qquad \beta = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & -\mathbf{1}_2 \end{pmatrix}. \tag{8.7}$$

Other realizations of the anticommutation relations (8.5) can be obtained from the set (8.7) by transformations $U\beta U^\dagger$, $U\alpha_j U^\dagger$ (where U is an arbitrary unitary matrix). It is usually best not to rely on a particular representation of Dirac matrices. All that one really needs are the anticommutation relations. However, for the purpose of visualization, we need to choose a particular set of Dirac matrices, and we shall refer to the choice (8.7) or (8.6) as the standard representation.

It is not possible to obtain solutions of the two- or three-dimensional Dirac equation as products of solutions of the one-dimensional Dirac equation. The free-particle Dirac operator in higher dimensions is not just a direct sum of lower-dimensional Dirac operators (as it is the case with the free-particle Schrödinger operator).

The Dirac equation in one and two space dimensions is solved by two-component spinors. The Dirac equation in three space dimensions is an equation involving four-component spinors. The Hilbert space of the Dirac equation in three space dimensions is therefore

$$\mathfrak{H} = L^2(\mathbb{R}^3)^4, \tag{8.8}$$

the space of square-integrable spinors with four components. In the standard representation, the free-particle Dirac operator is the matrix-differential operator

$$H_0 = \begin{pmatrix} \mathrm{m}c^2 & 0 & cp_3 & c(p_1 - \mathrm{i}p_2) \\ 0 & -\mathrm{m}c^2 & c(p_1 + \mathrm{i}p_2) & -cp_3 \\ cp_3 & c(p_1 - \mathrm{i}p_2) & \mathrm{m}c^2 & 0 \\ c(p_1 + \mathrm{i}p_2) & -cp_3 & 0 & -\mathrm{m}c^2 \end{pmatrix}. \tag{8.9}$$

8.1.2. Properties of the free-particle Dirac operator

The free-particle Dirac operator H_0 is best analyzed in momentum space where it is just a matrix-multiplication operator. With the help of the unitary Fourier transform $\mathcal{F}$ (applied to each component of a Dirac spinor), we convert the Dirac operator into

$$\mathcal{F}\, H_0\, \mathcal{F}^{-1} = \mathbf{h}_0(\mathbf{k}) = c\,\boldsymbol{\alpha}\cdot\mathbf{k} + \beta \mathrm{m}c^2 = \begin{pmatrix} \mathrm{m}c^2 \mathbf{1}_2 & c\,\boldsymbol{\sigma}\cdot\mathbf{k} \\ c\,\boldsymbol{\sigma}\cdot\mathbf{k} & -\mathrm{m}c^2\,\mathbf{1}_2 \end{pmatrix}. \tag{8.10}$$

Hence, the matrix-differential operator H_0 and the matrix-multiplication operator $\mathbf{h}_0(\mathbf{k})$ are unitarily equivalent.

In the standard representation, the matrix $\mathbf{h}_0(\mathbf{k})$ can be diagonalized with the unitary matrix

$$\mathbf{u}(\mathbf{k}) = a_+(k)\,\mathbf{1}_4 + \frac{a_-(k)}{k}\,\beta\,\boldsymbol{\alpha}\cdot\mathbf{k}, \tag{8.11}$$

where $k = |\mathbf{k}|$ and

$$a_\pm(k) = \frac{1}{\sqrt{2}}\Big(1 \pm \frac{\mathrm{m}c^2}{E(k)}\Big)^{1/2}, \tag{8.12}$$

with

$$E(k) = \sqrt{c^2\,k^2 + \mathrm{m}^2\,c^4} \qquad \text{(positive square-root)}. \tag{8.13}$$

It may be checked by an explicit calculation using the commutation relations (8.5) that

$$\mathbf{u}(\mathbf{k})\,\mathbf{h}_0(\mathbf{k})\,\mathbf{u}(\mathbf{k})^{-1} = E(k)\,\beta. \tag{8.14}$$

Hence, for each $\mathbf{k} \in \mathbb{R}^3$, the matrix $\mathbf{u}(\mathbf{k})$ is just the unitary matrix that diagonalizes the Hermitian 4×4 matrix $\mathbf{h}_0(\mathbf{k})$ in the standard representation (where β is represented by a diagonal matrix).

The Dirac matrix β has the eigenvalues $+1$ and -1, hence the eigenvalues of the matrix $\mathbf{h}_0(\mathbf{k})$ are $+E(k)$ and $-E(k)$. The set of all possible eigenvalues for all k is the spectrum of the multiplication operator $\mathbf{h}_0(\mathbf{k})$ and hence also of the free-particle Dirac operator H_0, which is unitarily equivalent with $\mathbf{h}_0(\mathbf{k})$:

$$\sigma(H_0) = \{\pm E(k) \mid k \geq 0\} = (-\infty, -\mathrm{m}c^2] \cup [\mathrm{m}c^2, \infty). \tag{8.15}$$

Let $u_{\mathrm{pos}}(k)$ be an eigenvector of $\mathbf{h}_0(k)$ belonging to the eigenvalue $E(k)$, and let $u_{\mathrm{neg}}(k)$ be an eigenvector belonging to the eigenvalue $-E(k)$. Then

$$u_{\substack{\mathrm{pos}\\ \mathrm{neg}}}(\mathbf{k};\mathbf{x}) = u_{\substack{\mathrm{pos}\\ \mathrm{neg}}}(k)\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \tag{8.16}$$

are solutions of the stationary free-particle Dirac equation,

$$H_0\,u_{\mathrm{pos}}(\mathbf{k};\mathbf{x}) = E(k)\,u_{\mathrm{pos}}(\mathbf{k};\mathbf{x}), \quad H_0\,u_{\mathrm{neg}}(\mathbf{k};\mathbf{x}) = -E(k)\,u_{\mathrm{neg}}(\mathbf{k};\mathbf{x}). \tag{8.17}$$

The functions u_{pos} and u_{neg} are analogs of stationary plane waves. The spectrum $\sigma(H_0)$ is the set of all energies for which the stationary free-particle Dirac equation has plane-wave-like solutions (out of which square-integrable wave packets can be formed). According to (8.15), $\sigma(H_0)$ is the continuum of all real numbers except the numbers in the *spectral gap*, the open interval $(-\mathrm{m}c^2, \mathrm{m}c^2)$. The Dirac equation in three dimensions hence also has solutions with negative energy.

The Hilbert space of the Dirac equation in three space dimensions is

$$\mathfrak{H} = L^2(\mathbb{R}^3)^4, \tag{8.18}$$

the space of square-integrable spinors with four components. This Hilbert space can be decomposed into the subspace of positive and negative energies,

$$L^2(\mathbb{R}^3)^4 = \mathfrak{H}_{\text{pos}} \oplus \mathfrak{H}_{\text{neg}}. \tag{8.19}$$

The projection operator that projects onto the subspace of positive energies is given by

$$P_{\text{pos}} = \frac{1}{2}\Big(1 + \frac{H_0}{\sqrt{H_0^2}}\Big), \tag{8.20}$$

Similarly,

$$P_{\text{neg}} = \frac{1}{2}\Big(1 - \frac{H_0}{\sqrt{H_0^2}}\Big), \tag{8.21}$$

is the projection operator onto negative energies according to the free-particle Dirac equation. We have

$$P_{\text{pos}}\, L^2(\mathbb{R}^3)^4 = \mathfrak{H}_{\text{pos}}, \quad P_{\text{neg}}\, L^2(\mathbb{R}^3)^4 = \mathfrak{H}_{\text{neg}}. \tag{8.22}$$

8.1.3. Electromagnetic fields

In order to introduce an electromagnetic field, the Dirac equation is modified as follows (see also Book One, Section 4,6): For a charged particle in a magnetic field, we replace the canonical momentum $\mathbf{p}$ by

$$\mathbf{p} - \frac{e}{c}\,\mathbf{A}(\mathbf{x}), \tag{8.23}$$

where e is the charge of the particle, and $\mathbf{A}(\mathbf{x})$ is the magnetic vector potential. In classical mechanics and nonrelativistic quantum mechanics, (8.23) is just mass times velocity.

In the presence of an electric field, we add the electrostatic potential energy $e\,\phi_{\text{el}}(\mathbf{x})$ to the energy. The operator describing the electrostatic potential energy is defined as the operator that multiplies each component of a spinor with the function $e\,\phi_{\text{el}}(\mathbf{x})$. Hence, we take

$$\mathrm{V}_{\text{el}}(\mathbf{x}) = e\,\phi_{\text{el}}(\mathbf{x})\,\mathbf{1}_4. \tag{8.24}$$

The Dirac operator in an external electromagnetic field becomes

$$H = c\,\boldsymbol{\alpha}\cdot\Big(\mathbf{p} - \frac{e}{c}\,\mathbf{A}(\mathbf{x})\Big) + \beta\,\mathrm{m}c^2 + \mathrm{V}_{\mathrm{el}}(\mathbf{x}) = H_0 + \mathrm{V}_{\mathrm{elm}}(\mathbf{x}). \tag{8.25}$$

Here, we have introduced the electromagnetic potential matrix

$$\mathrm{V}_{\mathrm{elm}}(\mathbf{x}) = e\,\phi_{\mathrm{el}}(\mathbf{x})\,\mathbf{1}_4 - e\,\boldsymbol{\alpha}\cdot\mathbf{A}(\mathbf{x}). \tag{8.26}$$

In the standard representation, this matrix is

$$\mathrm{V}_{\mathrm{elm}} = e\begin{pmatrix} \phi_{\mathrm{el}} & 0 & -A_3 & -A_1 + \mathrm{i}A_2 \\ 0 & \phi_{\mathrm{el}} & -A_1 - \mathrm{i}A_2 & -e\,A_3 \\ -A_3 & -A_1 + \mathrm{i}A_2 & \phi_{\mathrm{el}} & 0 \\ -A_1 - \mathrm{i}A_2 & -A_3 & 0 & \phi_{\mathrm{el}} \end{pmatrix}. \tag{8.27}$$

This 4×4 matrix multiplication operator has to be added to the matrix differential operator H_0 in (8.9).

There are two strong reasons why this choice of the Dirac operator is a good one:

(1) The Dirac equation with electromagnetic fields is covariant with respect to Poincaré transformations, and hence is consistent with the special theory of relativity and in particular with electromagnetism (see Section 8.2).
(2) The Dirac equation with the Hamiltonian H given by (8.25) predicts energy levels for the hydrogen atom which are in excellent agreement with the experimental observations (see Section 8.8). The same is true for other well-known external fields, for example, the constant magnetic field.

8.2. Relativistic Covariance

The special theory of relativity describes the symmetry transformations linking physically equivalent inertial systems. These transformations are called Poincaré transformations. In this chapter, we define the Poincaré transformations, show how they can be implemented as symmetry transformations in Dirac's relativistic quantum mechanics, and investigate how external fields have to be transformed.

8.2.1. Poincaré transformations

The *Minkowski space* is the four-dimensional space-time consisting of space-time points $x = (x_0, x_1, x_2, x_3) = (ct, \mathbf{x})$. It is equipped with the *metric*[1]

$$\langle x, y\rangle = x_0\,y_0 - \mathbf{x}\cdot\mathbf{y}. \tag{8.28}$$

[1]Here, the word "metric" refers to a non-definite scalar product.

A *Lorentz transformation*, described by a 4×4 matrix Λ, is a linear mapping that preserves the metric in Minkowski space,

$$\langle \Lambda\, x, \Lambda\, y\rangle = \langle x, y\rangle. \tag{8.29}$$

The set of all Lorentz transformations forms a group, the *Lorentz group*.

A special set of Lorentz transformations are the *boosts* or *velocity transformations*. We characterize a boost by the velocity of a particle after the boost,

$$\mathbf{v} = c\,\mathbf{n}\,\tanh\omega. \tag{8.30}$$

The matrix describing a boost is

$$\Lambda(\mathbf{v}) = \begin{pmatrix} \cosh\omega & \mathbf{n}^\top \sinh\omega \\ \mathbf{n}\sinh\omega & \mathbf{1}_3 - \mathbf{n}\mathbf{n}^\top(1-\cosh\omega)\end{pmatrix}. \tag{8.31}$$

Here, we interpret the unit vector $\mathbf{n}$ describing the direction of the boost as a column vector, its transpose $\mathbf{n}^\top$ is a row vector, and $\mathbf{n}\mathbf{n}^\top$ is the 3×3 matrix $(\mathbf{n}\mathbf{n}^\top)_{ik} = n_i n_k$. The inverse matrix is

$$\Lambda^{-1}(\mathbf{v}) = \Lambda(-\mathbf{v}). \tag{8.32}$$

The group of *rotations* is a subgroup of the Lorentz group. A rotation matrix is given by

$$\Lambda\big(\mathbf{R}(\varphi\mathbf{n})\big) = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{R}(\varphi\mathbf{n})\end{pmatrix} \tag{8.33}$$

where $\varphi\mathbf{n}$ is the rotation vector and $\mathbf{R}(\varphi\mathbf{n})$ is an orthogonal 3×3 rotation matrix as defined in (1.11).

Other examples of Lorentz transformations are the *parity transformation* $\mathbf{x}\to-\mathbf{x}$ and the *time-reversal operation* $t\to-t$. These *discrete Lorentz transformations* require a separate treatment, and the Lorentz group without the discrete transformations is called the *proper orthochronous Lorentz group*. Any proper orthochronous Lorentz transformation can be written as a composition of a boost and a rotation. The set of boosts alone do not form a subgroup of the Lorentz group, because the composition of two boosts along different directions is not a pure boost (this fact is related to the Thomas precession mentioned in Section 3.9.3).

A *Poincaré transformation* $\mathcal{P} = (\Lambda, a)$ consists of a Lorentz transformation Λ and a translation by a in space-time:

$$\mathcal{P}(x) = \Lambda\, x + a. \tag{8.34}$$

The set of all Poincaré transformations forms a 10-dimensional Lie group. This means that 10 continuously varying parameters are necessary to describe the group elements (3 for rotations, 3 for boosts, and 4 for the space-time translations).

8.2.2. Poincaré covariance of the Dirac equation

We want to describe the action of a Poincaré transformation on a solution of the Dirac equation

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(\mathbf{x},t) = \big(c\,\boldsymbol{\alpha}\cdot\nabla + \beta\,\mathrm{m}c^2 + \mathrm{V}(\mathbf{x},t)\big)\psi(\mathbf{x},t) \tag{8.35}$$

where $V(\mathbf{x},t)$ is (for each t) some Hermitian matrix multiplication operator. In the following, we only consider solutions ψ that are differentiable functions of the space-time coordinates. For notational convenience, we write

$$\psi(\mathbf{x},t) = \Psi(x) \quad \text{with } x = (ct,\mathbf{x}). \tag{8.36}$$

Multiplying the Dirac equation (8.35) from the left by the Dirac matrix β and dividing by c gives the equation

$$\mathrm{i}\,\langle\gamma,\partial\rangle\,\Psi(x) = \mathrm{m}c\,\psi(x) + V_{\mathrm{cov}}(x)\,\Psi(x), \tag{8.37}$$

called the *Dirac equation in covariant form*. Here, $\gamma = (\gamma_0,\boldsymbol{\gamma})$ is a four-dimensional vector whose components are matrices (the Dirac gamma matrices),

$$\gamma = (\gamma_0,\gamma_1,\gamma_2,\gamma_3) = (\gamma_0,\boldsymbol{\gamma}), \tag{8.38}$$

with

$$\gamma_0 = \beta, \qquad \boldsymbol{\gamma} = \beta\boldsymbol{\alpha}. \tag{8.39}$$

Hence, $\langle\gamma,x\rangle = c\beta t - \beta\boldsymbol{\alpha}\cdot\mathbf{x}$ is a 4×4 matrix for every x in Minkowski space. The symbol ∂ combines the derivatives with respect to time- and space-coordinates into a four-vector,

$$\partial = \big(\,\frac{1}{c}\,\frac{\partial}{\partial t},\,-\nabla\,\big). \tag{8.40}$$

The covariant potential matrix V_{cov} is defined as

$$V_{\mathrm{cov}}(x) = \frac{1}{c}\,\beta\,V(\mathbf{x},t), \quad \text{where} \quad x = (ct,\mathbf{x}). \tag{8.41}$$

From $\beta^2 = \mathbf{1}_4$ we see immediately that

$$V(\mathbf{x},t) = c\,\beta\,V_{\mathrm{cov}}(x). \tag{8.42}$$

The Dirac equation in covariant form (8.37) is useful for the investigation of relativistic invariance, because mc is a scalar (which by definition of a scalar is invariant under Lorentz transformations) and the term $\langle\gamma,\partial\rangle$ is written in the form of a Minkowski scalar product. Hence, if γ and ∂ were ordinary vectors in Minkowski space, the invariance of this term would be already guaranteed by (8.29).

For a given Poincaré transformation $\mathcal{P} = (\Lambda,a)$ and a given solution $\Psi(x)$ of the Dirac equation, we define a transformed solution in analogy to (7.203),

$$\Psi'(x) = \mathbf{T}(\Lambda)\,\Psi\big(\mathcal{P}^{-1}(x)\big) = \mathbf{T}(\Lambda)\,\Psi\big(\Lambda^{-1}(x-a)\big). \tag{8.43}$$

Here, $\mathbf{T}(\Lambda)$ is a representation of the Lorentz transformation Λ in terms of four-dimensional matrices in the space of four-component spinors. The representation matrices $\mathbf{T}(\Lambda)$ for Lorentz transformations will be chosen in such a way that the following relation holds

$$\mathbf{T}(\Lambda)\,\langle \gamma, x\rangle\,\mathbf{T}(\Lambda)^{-1} = \langle \gamma, \Lambda\, x\rangle = \langle \Lambda^{-1}\gamma, x\rangle. \tag{8.44}$$

Below, we present a choice of representation matrices $\mathbf{T}(\Lambda)$ satisfying this requirement. The relation (8.44) is equivalent to the matrix-relation

$$\Lambda\gamma = \mathbf{T}(\Lambda)^{-1}\,\gamma\,\mathbf{T}(\Lambda). \tag{8.45}$$

On the left-hand side, Λ acts as a 4×4 matrix on the four-vector γ, thereby producing another four-vector whose components are matrices. Likewise, the right-hand side of this equation is a four-vector with matrix-valued components, and the kth component is an ordinary matrix product of the three 4×4 matrices $\mathbf{T}(\Lambda)^{-1}$, γ_k, and $\mathbf{T}(\Lambda)$.

The Dirac equation is compatible with Poincaré transformations in the following sense:

Covariance of the Dirac equation:

Let $\Psi(x)$ be a solution of the Dirac equation

$$\mathrm{i}\,\big\langle \gamma, \partial\big\rangle\,\Psi(x) = \big(\mathrm{m}c + V_{\mathrm{cov}}(x)\big)\,\Psi(x). \tag{8.46}$$

Then, the transformed spinor

$$\Psi'(x) = \mathbf{T}(\Lambda)\,\Psi\big(\Lambda^{-1}(x-a)\big) \tag{8.47}$$

is a solution of the Dirac equation with the transformed potential

$$V'_{\mathrm{cov}}(x) = \mathbf{T}(\Lambda)\,V_{\mathrm{cov}}(\Lambda^{-1}(x-a))\,\mathbf{T}(\Lambda)^{-1}. \tag{8.48}$$

It follows that the free-particle Dirac equation is invariant under all Poincaré transformations.

Proof. In order to perform a Poincaré transformation of (8.46), we replace x everywhere by $x' = \Lambda^{-1}(x-a)$ and multiply both sides of (8.46) by $\mathbf{T}(\Lambda)$. This gives

$$\begin{aligned}\mathrm{i}\,\mathbf{T}(\Lambda)\,\big\langle \gamma, \partial'\big\rangle\,\Psi(x') &= \mathrm{m}c\,\mathbf{T}(\Lambda)\,\Psi(x') + \mathbf{T}(\Lambda)\,V_{\mathrm{cov}}(x')\,\Psi(x'),\\ &= \mathrm{m}c\,\mathbf{T}(\Lambda)\,\Psi(x') + V'_{\mathrm{cov}}(x)\,\mathbf{T}(\Lambda)\,\Psi(x'),\end{aligned} \tag{8.49}$$

where ∂' denotes the derivative with respect to x', and V' is the Poincaré-transformed potential according to (8.48). The chain rule of differentiation implies that

$$\partial\,\Psi(x') = \Lambda^{-1}\,\partial'\,\Psi(x') \tag{8.50}$$

with Λ^{-1} coming from the derivative of x' with respect to x. Hence,

$$\begin{aligned}\mathbf{T}(\Lambda)\left\langle \gamma, \partial' \right\rangle \psi(x') &= \mathbf{T}(\Lambda)\left\langle \gamma, \Lambda^{-1}\partial \right\rangle \psi(x') \\ &= \mathbf{T}(\Lambda)\left\langle \gamma, \Lambda^{-1}\partial \right\rangle \mathbf{T}(\Lambda)^{-1}\mathbf{T}(\Lambda)\,\psi(x'). \end{aligned} \tag{8.51}$$

Finally, we apply the algebraic property (8.44) to obtain

$$\mathbf{T}(\Lambda)\left\langle \gamma, \Lambda^{-1}\partial \right\rangle \mathbf{T}(\Lambda)^{-1} = \left\langle \gamma, \Lambda\Lambda^{-1}\partial \right\rangle = \langle \gamma, \partial \rangle. \tag{8.52}$$

Combining everything gives the equation

$$\mathrm{i}\,\langle \gamma, \partial \rangle\, \mathbf{T}(\Lambda)\Psi(x') = \big(\mathrm{m}c\,\mathbf{T}(\Lambda) + V'_{\mathrm{cov}}(x)\big)\,\mathbf{T}(\Lambda)\,\Psi(x') \tag{8.53}$$

which shows that $\Psi'(x) = \mathbf{T}(\Lambda)\Psi(x')$ is indeed a solution of the Dirac equation with the Poincaré-transformed potential. This is what we wanted to prove. □

The behavior of the covariant potential V_{cov} under a Poincaré transformation described in (8.48) implies that the potential matrix $V(x) = c\,\gamma_0\,V_{\mathrm{cov}}(x)$ is transformed into

$$V'(x) = \gamma_0\mathbf{T}(\Lambda)\,\gamma_0 V(\Lambda^{-1}(x-a))\mathbf{T}(\Lambda)^{-1}. \tag{8.54}$$

Special examples of Lorentz transformations are described next.

8.2.3. Velocity transformations

For Lorentz boosts $\Lambda(\mathbf{v})$, we define the representation matrices

$$\mathbf{T}\big(\Lambda(\mathbf{v})\big) = \mathrm{e}^{(\omega/2)\,\boldsymbol{\alpha}\cdot\mathbf{n}}, \quad \text{where } \mathbf{v} = c\,\mathbf{n}\,\tanh\,\omega. \tag{8.55}$$

The matrix-exponential $\exp(\omega\,\boldsymbol{\alpha}\cdot\mathbf{n}/2)$ is computed as usual via the power series of the exponential function and gives

$$\mathrm{e}^{(\omega/2)\,\boldsymbol{\alpha}\cdot\mathbf{n}} = \mathbf{1}_4\,\cosh\frac{\omega}{2} + \boldsymbol{\alpha}\cdot\mathbf{n}\,\sinh\frac{\omega}{2}. \tag{8.56}$$

These matrices are not unitary[2] in $\mathbb{C}^4$. Nevertheless, the boosts define unitary operators in $L^2(\mathbb{R}^3)^4$.

Consider the transformation $\Psi(x) \to \Psi'(x)$ defined in (8.47). Here, we assume that Ψ is a solution of the free-particle Dirac equation. Then, by a computation analogous to the one-dimensional case (Section 7.9.5), we obtain the generators of the Lorentz boosts as

$$\mathbf{N} = \frac{1}{2}(H_0\,\mathbf{x} + \mathbf{x}\,H_0). \tag{8.57}$$

[2]It can be shown that finite-dimensional representations of non-compact Lie groups cannot be unitary. The Lorentz transformations constitute the non-compact part of the Lorentz group, because the parameter ω varies in the unbounded (non-compact) set $\mathbb{R}$. See [**9**] for more details concerning the representation theory of the Lorentz and Poincaré groups.

The three components (N_1, N_2, N_3) of $\mathbf{N}$ define self-adjoint operators, showing that the transformations $\Psi(x) \to \Psi'(x)$ define a unitary representation of Lorentz boosts in $L^2(\mathbb{R}^3)^4$:

$$\Psi'(x) = \exp\Big(\mathrm{i}\,\frac{\omega}{c}\,\mathbf{n}\cdot\mathbf{N}\Big)\,\Psi(x) = \mathrm{e}^{(\omega/2)\,\boldsymbol{\alpha}\cdot\mathbf{n}}\,\Psi\big(\Lambda^{-1}(\mathbf{v})\,x\big). \tag{8.58}$$

Here, it is necessary to consider solutions of the free-particle Dirac equation, because the structure of the Lorentz group requires certain commutation relations for its generators (which are elements of the Lie-algebra). The generators of boosts, in particular, satisfy the commutation relations

$$\frac{1}{c^2}\,[N_j, N_k] = -\mathrm{i}\sum_{l=1}^{3}\epsilon_{jkl}\,J_l, \tag{8.59}$$

where J_l is the generator of a rotation (that is, an angular-momentum operator). It is clear that, in general, (8.59) cannot hold if the free-particle Dirac operator H_0 is replaced by the Dirac operator H in an external field. Hence, in the presence of an external field, (8.47) does not define a representation of the Poincaré group.

8.2.4. Rotations

As in nonrelativistic quantum mechanics, the generators of the rotations are the angular-momentum operators. But here, the angular momentum is a sum of an orbital angular momentum and the spin,

$$\mathbf{J} = \mathbf{L} + \mathbf{S}, \quad \mathbf{L} = \mathbf{x}\times\mathbf{p}, \quad \mathbf{S} = \frac{1}{4}\boldsymbol{\alpha}\times\boldsymbol{\alpha}. \tag{8.60}$$

In the standard representation, the spin $\mathbf{S}$ can be written in terms of the Pauli matrices as

$$\mathbf{S} = \frac{1}{2}\begin{pmatrix}\boldsymbol{\sigma} & \mathbf{0}_2\\ \mathbf{0}_2 & \boldsymbol{\sigma}\end{pmatrix}. \tag{8.61}$$

This is quite analogous to the definition of the spin in Chapter 3, see (3.39). We note that (each component of) the spin $\mathbf{S}$ commutes with the orbital angular momentum and hence

$$\exp(-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}) = \exp(-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S})\,\exp(-\mathrm{i}\varphi\,\mathbf{n}\cdot\mathbf{L}). \tag{8.62}$$

Here, the unitary operator generated by $\mathbf{L}$ acts only on the argument of the wave function (as in nonrelativistic quantum mechanics) whereas the unitary matrix $\exp(-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S})$ only affects the spinor components. Hence, we find that a rotated spinor wave function is given by

$$\begin{aligned}\big(\mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}}\,\psi\big)(\mathbf{x}) &= \mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}}\,\psi\big(\mathbf{R}(\varphi\mathbf{n})^{-1}\mathbf{x}\big)\\ &= \mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}}\,\Psi\big(\Lambda(\mathbf{R}(\varphi\mathbf{n}))^{-1}x\big),\end{aligned} \tag{8.63}$$

where $\mathbf{R}(\varphi\mathbf{n})$ is the orthogonal 3×3 rotation matrix defined in (1.11). Using the power series representation of the exponential function and the property $(\mathbf{n}\cdot\mathbf{S})^2 = 1/4$, we easily obtain

$$\mathbf{T}\big(\Lambda(\mathbf{R}(\varphi\mathbf{n}))\big) = \mathrm{e}^{-\mathrm{i}\varphi\,\mathbf{n}\cdot\mathbf{S}} = \mathbf{1}_4\,\cos\frac{\varphi}{2} - 2\mathrm{i}\,\mathbf{n}\cdot\mathbf{S}\,\sin\frac{\varphi}{2}. \tag{8.64}$$

These matrices are unitary in $\mathbb{C}^4$. Note that the angle $\varphi/2$ appears in the matrix (8.64), as it did in (4.28). It means that a rotation through 2π is represented by the matrix -1.

We finally note that more general Lorentz transformations are now obtained easily because any proper orthochronous Lorentz transformation can be written in a unique way as the product of a boost and a rotation:

$$\Lambda = \Lambda(\mathbf{v})\,\Lambda\big(\mathbf{R}(\varphi\mathbf{n})\big). \tag{8.65}$$

8.2.5. Unitary representation

The set of solutions of the free-particle Dirac equation implements a unitary representation of the orthochronous Poincaré group. For example, the time evolution according to the free-particle Dirac equation is a special Poincaré transformation, namely a translation in the time-direction of Minkowski space. It is a unitary transformation generated by the free-particle Dirac operator H_0:

$$\big(\mathrm{e}^{\mathrm{i}H_0 t}\Psi\big)(x_0,\mathbf{x}) = \Psi(x_0 - ct,\mathbf{x}), \quad \text{translation by } a = (ct,0,0,0). \tag{8.66}$$

Here, it is necessary to consider the free time evolution, because the Lie-algebra relations of the Poincaré group require that the generators of space and time translations commute, $[H_0, p_k] = 0$. This would not be true if the generator H_0 of time translations were replaced with H. Hence, in general, the transformations given by (8.47) do not define a representation of the Poincaré group.

Unitary representation of Poincaré group:

In the Hilbert space $L^2(\mathbb{R}^3)^4$ of the Dirac equation we can define a unitary representation of the orthochronous Poincaré group (all Poincaré transformations except the time reversal). Let $\psi(\mathbf{x}) = \Psi(0,\mathbf{x})$, where

$$\Psi(ct,\mathbf{x}) = \Psi(x) = \exp(-\mathrm{i}H_0 t)\,\psi(\mathbf{x}) = \psi(\mathbf{x},t). \tag{8.67}$$

Then, the representation $U(\Lambda,a)$ of a Poincaré transformation (Λ,a) is defined by

$$U(\Lambda,a)\psi(\mathbf{x}) = \mathbf{T}(\Lambda)\,\Psi\big(\Lambda^{-1}(x-a)\big) \quad \text{with } x = (0,\mathbf{x}). \tag{8.68}$$

The parity transformation P can be represented by

$$(P\psi)(\mathbf{x}) = \beta\psi(-\mathbf{x}). \tag{8.69}$$

According to a result of E.P. Wigner, the time reversal T has to be represented by an antiunitary transformation. There are several (nonequivalent) ways to implement the discrete transformations $\{P, T, PT\}$ in the Hilbert space of the Dirac equation.

8.3. Classification of External Fields

8.3.1. Poincaré transformations of external fields

With the help of $\gamma_0 = \beta$ and $\gamma_k = \beta\alpha_k$, $k = 1, 2, 3$, we define the following sixteen matrices $\Gamma_1, \ldots, \Gamma_{16}$,

$$\begin{aligned}
&\Gamma_1 : && \mathbf{1}_4,\\
&\Gamma_2, \ldots, \Gamma_5 : && \gamma_0,\ \mathrm{i}\gamma_1,\ \mathrm{i}\gamma_2,\ \mathrm{i}\gamma_3,\\
&\Gamma_6, \ldots, \Gamma_{11} : && \gamma_0\gamma_1,\ \gamma_0\gamma_2,\ \gamma_0\gamma_3,\ \mathrm{i}\gamma_2\gamma_3,\ \mathrm{i}\gamma_3\gamma_1,\ \mathrm{i}\gamma_1\gamma_2,\\
&\Gamma_{12}, \ldots, \Gamma_{15} : && \gamma_1\gamma_2\gamma_3,\ \mathrm{i}\gamma_0\gamma_2\gamma_3,\ \mathrm{i}\gamma_0\gamma_3\gamma_1,\ \mathrm{i}\gamma_0\gamma_1\gamma_2,\\
&\Gamma_{16} : && \gamma_5 = \mathrm{i}\gamma_0\gamma_1\gamma_2\gamma_3.
\end{aligned} \tag{8.70}$$

These matrices form a basis in the (real) vector space of all Hermitian 4×4 matrices. Any Hermitian matrix Γ can be written as a linear combination $\Gamma = \sum c_j \Gamma_j$, with real-valued coefficients $c_i \in \mathbb{R}$. Similarly, any potential matrix $V(x)$ (a Hermitian 4×4 matrix) can be written as a linear combination of the Gamma-matrices Γ_j with real-valued coefficient functions:

$$V(x) = \sum_{j=1}^{16} v_j(x)\, \Gamma_j \quad (v_j \text{ real-valued}). \tag{8.71}$$

The behavior under Poincaré transformations is best determined for the covariant potential

$$V_{\mathrm{cov}}(x) = \frac{1}{c} \sum_j v_j(x)\, \gamma_0\, \Gamma_j. \tag{8.72}$$

According to (8.48), a Poincaré transformation converts this expression into

$$V'_{\mathrm{cov}}(x) = \frac{1}{c} \sum_{j=1}^{16} v_j(\Lambda^{-1}(x-a))\, \mathbf{T}(\Lambda)\, \gamma_0 \Gamma_j\, \mathbf{T}(\Lambda)^{-1}. \tag{8.73}$$

We can expand the transformed external field as

$$V'_{\mathrm{cov}}(x) = \frac{1}{c} \sum_j v'_j(x)\, \gamma_0\, \Gamma_j. \tag{8.74}$$

If the Dirac equation should be covariant under Poincaré transformations, it has to be assumed that $V'_{\rm cov}(x)$ describes the external field at the space-time point x after a Poincaré transformation. The transformation law (8.45) for the Dirac-γ-matrices now implies a certain transformation law for the coefficient functions v_j. This will be illustrated with several examples below.

8.3.2. Electromagnetic vector potential

In Section 8.1.3, it has been claimed that the potential matrix

$$V(x) = e\,\phi_{\rm el}(x)\,\mathbf{1}_4 - e\,\boldsymbol{\alpha}\cdot\mathbf{A}(x) \tag{8.75}$$

describes a charged particle in an electromagnetic field $(\phi_{\rm el}, \mathbf{A})$. If we expand this matrix in the basis of Γ-matrices, we see that the electromagnetic potentials are the coefficients of the matrices Γ_1, Γ_6, Γ_7, and Γ_8. The corresponding covariant potential can be written as

$$V_{\rm cov}(x) = \frac{e}{c}\,\langle\gamma, A(x)\rangle, \tag{8.76}$$

where $A(x)$ is an abbreviation for $(\phi_{\rm el}(x), \mathbf{A}(x))$. After a Poincaré transformation, this becomes

$$\begin{aligned} V'_{\rm cov}(x) &= \frac{e}{c}\,\mathbf{T}(\Lambda)\,\langle\gamma, A(\Lambda^{-1}(x-a))\rangle\,\mathbf{T}(\Lambda)^{-1} \\ &= \frac{e}{c}\,\langle\gamma, \Lambda A(\Lambda^{-1}(x-a))\rangle. \end{aligned} \tag{8.77}$$

Here, we used the relation (8.44) again. We see that the requirement of relativistic covariance is that the fields $A(x) = (\phi_{\rm el}(x), \mathbf{A}(x))$ are transformed according to

$$A'(x) = \Lambda\,A\big(\Lambda^{-1}(x-a)\big). \tag{8.78}$$

Indeed, this is precisely the behavior of electromagnetic potentials as expected from Maxwell's equations. We note that the electric and magnetic field strengths are

$$\mathbf{E}(x) = -\nabla\phi_{\rm el}(x) - \frac{1}{c}\,\frac{\partial\mathbf{A}(x)}{\partial t}, \qquad \mathbf{B}(x) = \nabla\times\mathbf{A}(x). \tag{8.79}$$

If we perform a Poincaré transformation of the electromagnetic potentials according to (8.78), then the corresponding new field strengths $\mathbf{E}'$ and $\mathbf{B}'$ will again satisfy Maxwell's equations.

Electromagnetic potentials and covariance of Dirac's equation:
Let ψ be a solution of the Dirac equation with electromagnetic potentials A. The Poincaré-transformed solution ψ' is a solution of the Dirac equation with new electromagnetic potentials A' that are obtained by a Poincaré transformation from the original potentials. The transformation of the potentials is in accordance with the transformation law required by Maxwell's equations.

This compatibility of the transformation behavior gives us the right to call the coefficient functions of Γ_1, Γ_6, Γ_7, and Γ_8 in (8.71) "electromagnetic potentials."

In a similar way, we can give a meaning to the other coefficient functions.

8.3.3. Scalar potential

Consider a real-valued function ϕ_{sc} of $x = (ct, \mathbf{x})$ and define the potential matrix

$$V(x) = \beta\phi_{\text{sc}}(x). \tag{8.80}$$

The corresponding covariant potential is

$$V_{\text{cov}}(x) = \tfrac{1}{c}\,\phi_{\text{sc}}(x)\,\mathbf{1}_4. \tag{8.81}$$

A Poincaré transformation $V_{\text{cov}} \to V'_{\text{cov}}$ according to (8.48) requires the replacement

$$\phi_{\text{sc}}(x) \to \phi_{\text{sc}}\big(\Lambda^{-1}(x-a)\big), \tag{8.82}$$

and hence ϕ_{sc} must behave like a scalar field under a Poincaré transformation. The Dirac operator with a scalar potential becomes

$$H = c\,\boldsymbol{\alpha}\cdot\mathbf{p} + \beta\big(\mathrm{mc}^2 + \phi_{\text{sc}}(x)\big). \tag{8.83}$$

This shows that a scalar potential ϕ_{sc} acts like an x-dependent rest mass.

8.3.4. Anomalous magnetic moment

We have seen that γ combines with a four-vector field A into a covariant potential matrix. Similarly, products of γ matrices combine with tensor fields. For example,

$$V_{\text{cov}}(x) = \frac{\mu_{\text{a}}}{2c}\sum_{\mu,\nu=0}^{3} F^{\mu\nu}(x)\,\gamma_\mu\,\gamma_\nu \tag{8.84}$$

has the right behavior for the Dirac equation if the coefficient functions $F^{\mu\nu}(x)$ behave like a contravariant tensor field of second rank. That means

that we can take $F^{\mu\nu}$ to be the electromagnetic field tensor. This leads to the following potential

$$V(x) = \mu_{\mathrm{a}} \left(\mathrm{i}\beta \boldsymbol{\alpha} \cdot \mathbf{E}(x) - 2\beta\, \mathbf{S} \cdot \mathbf{B}(x) \right). \tag{8.85}$$

In (8.71), the coefficient functions of the matrices $\Gamma_3, \Gamma_4, \Gamma_5$ and $\Gamma_{13}, \Gamma_{14}, \Gamma_{15}$ can thus be interpreted as the electric and magnetic field strengths.

Physically, the potential (8.85) describes a particle with an additional anomalous magnetic moment. The magnitude of the anomalous moment is given by the coupling constant μ_{a} in units of the Bohr-magneton.

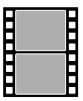
The remaining types of potential matrices (pseudoscalar-, pseudovector-, and pseudotensor-fields) occur less frequently in atomic physics and will not be discussed here.

8.4. Positive and Negative Energies

8.4.1. Negative-energy wave packets are positrons

In the presence of an external field, solutions with positive energy behave differently from solutions with negative energy. We consider a potential $V(\mathbf{x})$ that is very small outside a bounded region, like the electrostatic barrier shown in Figure 8.1. We put a well-localized wave packet in a region where the potential can be neglected. The energy of this wave packet is essentially determined by the free-particle Dirac operator. Hence, a wave packet in the subspace $\mathfrak{H}_{\mathrm{pos}}$ (according to the free-particle Dirac equation) is a good approximation of a wave packet with positive total energy. Similarly, a wave packet $\psi \in \mathfrak{H}_{\mathrm{neg}}$ that is (approximately) localized far away from the region with the potential, has a negative total energy.

Figure 8.1 compares a positive-energy wave packet with a negative-energy wave packet when they hit an electrostatic barrier. The upper row in Figure 8.1 shows that the positive-energy wave packet behaves as expected: The wave packet gets reflected at the positive (repulsive) potential barrier, because its energies are lower than the height of the barrier. The lower row in Figure 8.1 shows a negative-energy wave packet with the same initial shape and initial velocity. It hits the same potential barrier, but instead of being reflected it gets accelerated and passes quickly through the barrier.



CD 6.18 is an animated version of Figure 8.1, showing phase-colored plots of the spinor-wave packets. This helps to distinguish between positive and negative energy solutions from the beginning. The positive-energy wave packet is reflected at a positive potential barrier (as expected). The negative-energy wave packet behaves as if it would describe a particle with the opposite charge.

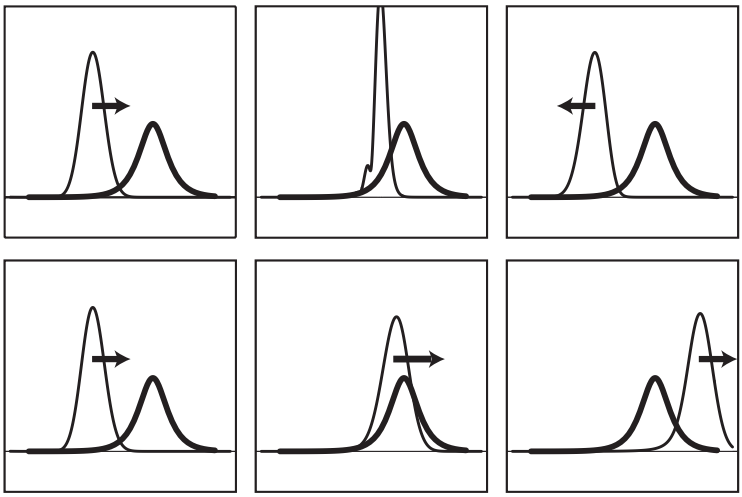


FIGURE 8.1. Wave packet hitting a "speed bump" (electrostatic barrier, thick line). Upper row: Three snapshots from the time evolution of a wave packet with positive energy. Lower row: Same for a wave packet with negative energy. Both wave packets have the same initial velocity distribution.

A negative-energy wave packet behaves like a particle with the same mass but with opposite charge. In short: If the positive-energy solutions describe an electron, then the solutions with negative energy behave like a positron. Indeed, we can give a formal proof of this observation.

We denote the Dirac operator for a particle with mass m and charge e in an external field $(\phi_{\mathrm{el}}, \mathbf{A})$ by

$$H(e, \mathrm{m}) = c\,\boldsymbol{\alpha} \cdot \big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x}, t)\big) + \beta\,\mathrm{m}c^2 + e\phi_{\mathrm{el}}(\mathbf{x}, t). \tag{8.86}$$

Consider the antiunitary transformation

$$C\psi = U_C\,\overline{\psi}. \tag{8.87}$$

Here, the bar denotes a complex conjugation, and U_C is a unitary 4×4 matrix satisfying

$$\beta U_C = -U_C\overline{\beta}, \quad \alpha_k U_C = U_C\overline{\alpha_k}, \quad k = 1, 2, 3. \tag{8.88}$$

In the standard representation, we may choose

$$U_C = \mathrm{i}\beta\alpha_2. \tag{8.89}$$

The operator C is an antiunitary transformation in the Hilbert space of the Dirac equation. Hence, it is a symmetry transformation by the definition in

Chapter 1. In particular, all transition probabilities are left invariant, that is,

$$|\langle C\psi, C\phi\rangle|^2 = |\langle \psi, \phi\rangle|^2. \tag{8.90}$$

A little calculation shows the following result: Whenever $\psi(t)$ is a solution of the Dirac equation with Hamiltonian $H(e, \mathrm{m})$, then $C\psi(t)$ is a solution of the Dirac equation with Hamiltonian $H(-e, \mathrm{m})$. We have

$$C\, H(e, \mathrm{m})\, C^{-1} = -H(-e, \mathrm{m}). \tag{8.91}$$

Therefore, the symmetry transformation C is called *charge conjugation*. C maps the negative-energy subspace of $H(e, \mathrm{m})$ onto the positive-energy subspace of the Dirac operator $H(-e, \mathrm{m})$ for a particle with the same mass but opposite charge. Hence, if ψ describes an electron, then $C\psi$ describes a positron.

In analogy with (7.22), we denote by $|\cdot|_4$ the $\mathbb{C}^4$-norm of a four spinor. The unitarity of the matrix U_C immediately implies

$$\big|C\psi(\mathbf{x})\big|_4 = \big|\, U_C\, \overline{\psi(\mathbf{x})}\,\big|_4 = \big|\,\overline{\psi(\mathbf{x})}\,\big|_4 = \big|\psi(\mathbf{x})\big|_4. \tag{8.92}$$

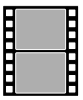
The position distribution of a spinor-wave packet cannot be distinguished from the position distribution of the charge-transformed wave packet. We conclude:

> The motion of a negative-energy spinor-wave packet ψ is indistinguishable from the motion of the positive-energy spinor-wave packet $C\psi$.

In view of this discussion, we may modify our standard one-particle interpretation in the following sense:

The Dirac equation describes two kinds of particles: States in the positive-energy subspace describe electrons, and states in the negative-energy subspace actually describe antiparticles with positive energy.

This statement is only meaningful for situations where the sign of the energy clearly distinguishes between electronic and positronic behavior. This is not always the case, as shown in the following section.



CD 6.19 shows the behavior of various wave packets in a linear electrostatic potential (constant electric field). We expect electronic behavior, whenever the energy of the wave packet is higher than $V(x) + \mathrm{m}c^2$ in the region where the wave packet is (approximately) localized. We expect positronic behavior for wave packets whose energy is less than $V(x) - \mathrm{m}c^2$. A general wave packet quickly splits into two parts, one part moving in the direction of the field's gradient (electronic wave packet) and one moving in the opposite direction (positronic behavior).

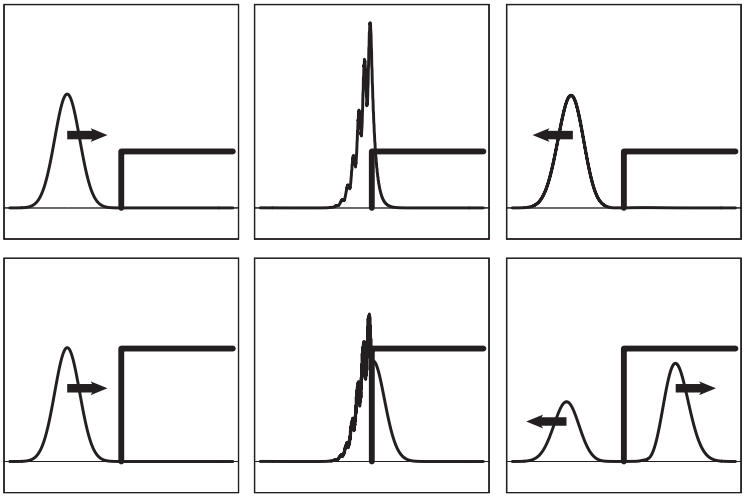


FIGURE 8.2. Illustration of the Klein paradox. The upper row of figures show the total reflection of a positive-energy wave packet at a potential step whose height is higher than all the energies of the wave packet. The lower row shows that when the same wave packet encounters a very high potential step, part of the packet is transmitted as a positron.

8.4.2. Relativistic scattering in one dimension: The Klein paradox

The Klein paradox occurs for a high one-dimensional electrostatic potential step. This step divides space into two regions with a different interpretation of particles and antiparticles. This is shown in Figure 8.2.

For the Dirac operator with a constant potential $V(\mathbf{x}) = V_0\,\mathbf{1}$, the energy spectrum consists of a positronic part $(-\infty, -\mathrm{m}\,c^2 + V_0]$ and an electronic part $[\mathrm{m}\,c^2 + V_0, \infty)$. The corresponding plane-wave solutions are $\underset{\rightarrow}{\omega}(E - V_0; x)$ and $\underset{\leftarrow}{\omega}(E - V_0; x)$ with the solutions defined in (7.138) and (7.141).

We consider the Dirac operator with a one-dimensional step potential

$$V(x) = \phi_{\mathrm{el}}(x)\,\mathbf{1}, \quad \phi_{\mathrm{el}}(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ V_0 & \text{for } x > 0. \end{cases} \tag{8.93}$$

The motion of wave packets in the region $x \leq 0$ is thus governed by the free-particle Dirac equation whereas the behavior in the region $x > 0$ is described by the Dirac equation in a constant potential. Whenever V_0 is bigger than $2\mathrm{m}c^2$, we see that the range of electronic energies according to the free-particle Dirac operator overlaps with the range of positronic energies

according to the Dirac operator with the constant potential. A particle with energies in $(\mathrm{m}c^2, -\mathrm{m}c^2+V_0)$ can thus propagate on both sides of the potential step, as an electron on the left side, and as a positron on the right side. The Dirac equation with the potential (8.93) now has the following plane-wave solution:

$$u(E;x) = \begin{cases} \underset{\rightarrow}{\omega}(E;x) + \underset{\leftarrow}{R}(E)\, \underset{\leftarrow}{\omega}(E;x) & \text{for } x \leq 0, \\ \underset{\rightarrow}{T}(E)\, \underset{\rightarrow}{\omega}(E-V_0;x) & \text{for } x > 0. \end{cases} \tag{8.94}$$

The coefficients $\underset{\leftarrow}{R}(E)$ (coefficient for reflection to the left) and $\underset{\rightarrow}{T}(E)$ (coefficient for transmission to the right) follow from a continuity condition at $x = 0$. The continuity condition reads

$$\underset{\rightarrow}{\omega}(E) + \underset{\leftarrow}{R}(E)\, \underset{\leftarrow}{\omega}(E) = \underset{\rightarrow}{T}(E)\, \underset{\rightarrow}{\omega}(E-V_0). \tag{8.95}$$

We insert the expressions (7.147) and (7.148) with the following abbreviations for the components,

$$n(E)\, b_+(E) = \eta, \qquad n(E-V_0)\, b_+(E-V_0) = \eta_0, \tag{8.96}$$

and note that, because of (7.149),

$$n(E)\, b_-(E) = \frac{1}{2\eta}. \tag{8.97}$$

Then, we can write the continuity condition as

$$\big(1 + \underset{\leftarrow}{R}(E)\big)\, \eta = \underset{\rightarrow}{T}(E)\, \eta_0, \quad \big(1 - \underset{\leftarrow}{R}(E)\big)\, \frac{1}{\eta} = \underset{\rightarrow}{T}(E)\, \frac{1}{\eta_0}. \tag{8.98}$$

From these equations, we determine

$$\underset{\leftarrow}{R}(E) = \frac{\eta_0^2 - \eta^2}{\eta_0^2 + \eta^2}, \qquad \underset{\rightarrow}{T}(E) = \frac{2\,\eta_0\,\eta}{\eta_0^2 + \eta^2}. \tag{8.99}$$

Let us consider the case $V_0 > 2\mathrm{m}c^2$. The transmission coefficient $\underset{\rightarrow}{T}(E)$ is seen to be real and positive for energies $\mathrm{m}c^2 < E < -\mathrm{m}c^2 + V_0$. Moreover, we have

$$\big|\underset{\leftarrow}{R}(E)\big|^2 + \big|\underset{\rightarrow}{T}(E)\big|^2 = 1. \tag{8.100}$$

A plot of $|\underset{\rightarrow}{T}(E)|^2$ for energies in $\mathrm{m}c^2 < E < -\mathrm{m}c^2 + V_0$ is shown in Figure 8.3.

We may use the plane wave solutions (8.94) to form a wave packet with energies in the range $\mathrm{m}c^2 < E < -\mathrm{m}c^2 + V_0$,

$$\psi(x,t) = \int_{\mathrm{m}c^2}^{-\mathrm{m}c^2+V_0} g(E)\, u(E;x)\, \mathrm{e}^{-\mathrm{i}Et}\, dE. \tag{8.101}$$

This describes a wave packet that is initially (for $t \ll 0$) located far to the left of the step, with positive velocities. At the step, the wave packet splits

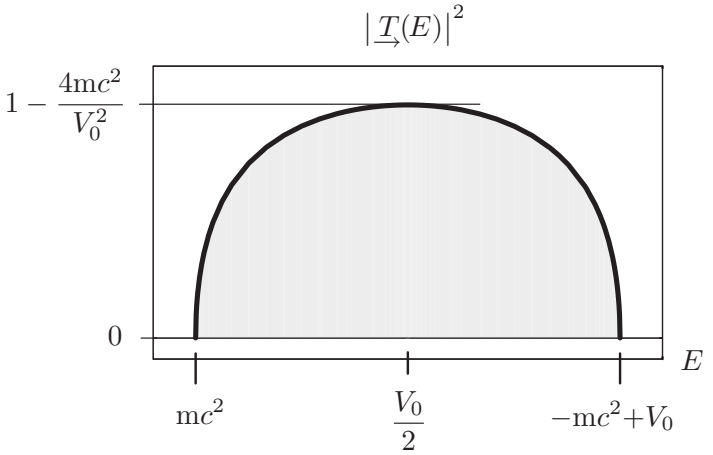
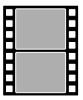


FIGURE 8.3. Square of the transmission coefficient for a high electrostatic potential step $V_0 > 2mc^2$.

into a reflected part and into a transmitted part. The reflected part (having positive energies) behaves like an electronic solution of the free-particle Dirac equation to the left of the potential step. The transmitted part belongs to the positronic energy region of the Dirac equation with a constant potential. It describes a wave packet that moves to the right and behaves like a positron. The existence of this solution is paradoxical, because it describes a wave packet that starts as an electron and has a certain chance to be finally detected as a positron. Hence, this solution violates the principle of charge conservation.



CD 6.20 shows the scattering of wave packets at potential steps of various heights. In case of a high potential step, we see Klein's paradox. The transmitted wave packet to the right of the step shows all characteristic features of negative-energy solutions. CD 6.21 discusses the scattering of Dirac-wave packets in momentum space. CD 6.22 shows variants of the Klein paradox in momentum space.

The Klein paradox remains unresolved in the framework of a single-particle theory with external fields. Here, it serves to illustrate the limitations of that theory. In reality, the field energy of a high potential step would presumably cause the generation of electron-positron pairs, which leads us into the range of quantum electrodynamics. Presently, quantum electrodynamics is still mathematically incomplete, and one may argue that there is no fully satisfactory and commonly accepted description of the Klein paradox at the time of this writing.

8.4.3. Physical Hilbert space and relativistic observables

The Hilbert space $\mathfrak{H} = L^2(\mathbb{R})^4$ contains states that are superpositions of positive and negative energy states. But, a single quantum system can hardly be imagined as a superposition of a particle and an antiparticle (although a similar phenomenon is known for neutral K-mesons). Therefore, one makes the assumption that physically allowed states should be restricted to either the positive-energy subspace or to the negative-energy subspace. This makes sense for free particles, because the time evolution leaves the subspaces with positive and negative energy invariant. An initial state with positive energy has positive energies for all times, that is,

$$\psi(t) \equiv \exp(-\mathrm{i}H_0 t)\,\psi \in \mathfrak{H}_{\mathrm{pos}} \quad \text{if and only if} \quad \psi \in \mathfrak{H}_{\mathrm{pos}}. \tag{8.102}$$

As we have seen, there are situations where the sign of the energy does not distinguish between electronic and positronic behavior (Klein's paradox). Consequently, transitions from electronic to positronic states cannot be excluded. In particular, the energy is not conserved if the potentials depend on time. Thus, an initial state with positive energy may turn into a superposition of electronic and positronic states.

Fortunately, these effects become important only if the potential energies are very high (of the order of $2mc^2$). These situations are not very likely to occur in atomic and molecular physics. For a sufficiently weak time-independent potential that vanishes, as $|\mathbf{x}| \to \infty$, one can still distinguish between particle and antiparticle solutions. Hence, there is a wide range of possible applications of the Dirac equation where the above-mentioned interpretational difficulties play no role, at least not in a practical sense.

The above reasoning suggests that in many cases of physical interest, the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^3)^4$ contains as a "physical" subspace $\mathfrak{H}_{\mathrm{electron}}$, the subspace containing electronic states. A Hilbert space of positronic states $\mathfrak{H}_{\mathrm{positron}}$ can be constructed via a charge conjugation from the orthogonal complement of $\mathfrak{H}_{\mathrm{electron}}$ in $\mathfrak{H}$. In case of free particles or weak (static) external fields, the $\mathfrak{H}_{\mathrm{electron}}$ coincides with the subspaces with positive energy $\mathfrak{H}_{\mathrm{pos}}$, and $\mathfrak{H}_{\mathrm{positron}}$ is given by $C\mathfrak{H}_{\mathrm{neg}}$. Superpositions of electronic and positronic states should be regarded as unphysical. Whenever these superpositions arise in an unavoidable way, one should apply the methods of quantum field theory. In these situations high energies are involved, pair creation is likely to happen, and the one-particle Dirac equation is an insufficient model of reality (as would be a "many-particle" version of the Dirac equation with a fixed number of particles).

From a pragmatic point of view, we believe that this interpretation of the Dirac equation is good for a certain set of physical situations (or range of energies) that encompasses the situations typically encountered in atomic

physics. Here, one should consider the positive-energy subspace $\mathfrak{H}_{\rm pos}$ as the physical Hilbert space for an electron. All calculations should be done within that Hilbert space. In particular, a minimization of the energy has to be carried out with the constraint that ψ be (approximately) orthogonal to $\mathfrak{H}_{\rm neg}$.[3] Doing otherwise would lead to the well-known "variational collapse," because the Dirac operator is not bounded from below on the space of all square-integrable functions.

One question, however, cannot be avoided completely. It is again that of relativistic observables. In view of this interpretation, only an observable that leaves the subspace of positive energy invariant, is a good observable. With $\psi \in \mathfrak{H}_{\rm pos}$ we should also require that $A\psi \in \mathfrak{H}_{\rm pos}$, otherwise a measurement of the observable A would throw the state out of the electronic Hilbert space.

This requirement rules out the standard position operator (multiplication by $\mathbf{x}$), because it does not commute with the sign of the energy. This is related to the fact that the subspace of positive energies does not contain strictly localized spinors. There is no wave function that vanishes everywhere in an open region of space. All positive-energy wave packets are essentially spread over all of space. Still, there are wave packets that are approximately localized in the same sense as a Gaussian wave packet. Examples of such Gaussian-type wave packets are, for example, in Figures 8.1 and 8.2. But, if we still maintain the interpretation of $|\psi(\mathbf{x})|^2$ as a density of the position probability, then we have to live with the fact that there are no strictly localized particles. Physical reasons may be given for this. If we want to prepare a particle whose wave function strictly vanishes outside some region, we would have to enclose the particle by an infinitely high potential barrier. But, this provides an infinite energy reservoir, and pair-creation processes or variants of the Klein paradox are to be expected.

8.4.4. Relativistic confinement

Our discussion of the electron-positron interpretation shows that the Dirac equation can have bound states for both attractive and repulsive electric potentials. Consider for example an attractive electrostatic potential well $eV(x) < 0$. It may support some bound states at energies E_n in the gap $(-\mathrm{m}c^2, \mathrm{m}c^2)$. A charge conjugation shows that the Dirac operator with a repulsive potential well $-eV(x)$ then has eigenvalues at the energies $-E_n$. According to our interpretation, the bound states in the attractive well are electrons, and the bound states in the "repulsive" well are, of course, positrons. As long as the well is not too deep, the electronic states and the positronic states remain well separated. Obviously, the situation becomes

[3] Griesemer, M., Siedentop, H., J. London Math. Soc. **60**, 490–500 (1999).

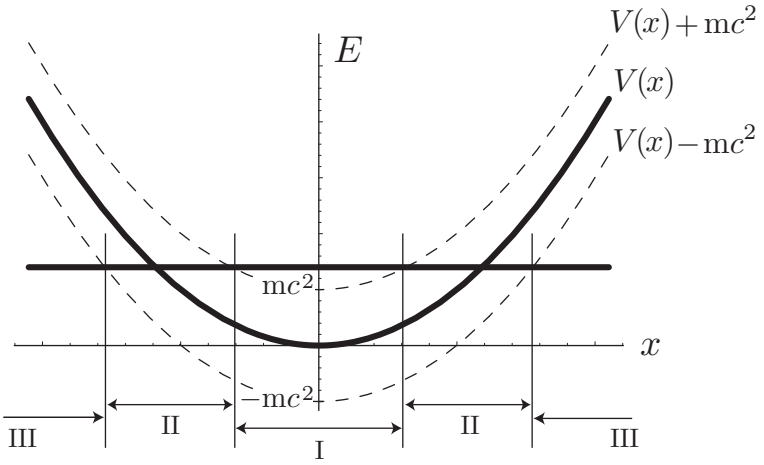
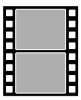


FIGURE 8.4. The (electrostatic) harmonic oscillator potential in the Dirac equation cannot bind particles. Any wave packet that is initially inside the potential (in the region I) can tunnel through the "classically forbidden" region II into the positronic region III.

more problematic if the ground state has a binding energy larger than $\mathrm{m}c^2$. In this case, the sign of the energy cannot separate the electronic from the positronic states any longer. If we continuously increase the strength of an attractive potential well, the bound state energies will move through the gap and finally dive into the positronic continuum where they turn into resonances. These problems have been investigated in the context of heavy ion collisions, where for some period of time a very heavy nucleus is formed. It is clear that in this context, the solutions of the Dirac equation have to be interpreted in the light of quantum field theory. Physically, as well as mathematically, this is partially an open problem.

Consider an electrostatic potential that goes to infinity, as $|\mathbf{x}| \to \infty$. A typical example would be the harmonic oscillator $V(\mathbf{x}) = x^2\, \mathbf{1}_4$. Because of the Klein paradox (see Fig. 8.2), this potential is not able to bind particles.



CD 6.24 shows how a wave packet escapes from a deep electrostatic potential well, because the wave packet tunnels into a region where it can escape as a positronic wave packet. This effect vanishes if the potential well is not so deep.

Figure 8.4 shows the harmonic oscillator potential $V(x)$ and the dashed curves $V(x) \pm \mathrm{m}c^2$. Electrons have a positive kinetic energy if the total

energy of the particle exceeds the potential energy plus the rest energy (in the region I). This is a region where a nonrelativistic wavepacket would be confined, with exponentially decaying "tails" into the classically forbidden region (where the total energy is less than $V(x) + \mathrm{m}c^2$). According to the Dirac equation, a positronic solution can exist in the region where the total energy is less than $V(x) - \mathrm{m}c^2$. Note that the "attractive" harmonic oscillator is actually repulsive for positronic wave packets. The Klein paradox shows that a wave packet can indeed tunnel through the forbidden region II from the electronic to the positronic region. Hence, any state that is initially in the region I will sooner or later tunnel through the region II and will finally escape as a positronic scattering state. Instead of a bound state, we can at best expect a resonance of positronic scattering. For this reason, the Dirac operator with an electric potential that goes to infinity has no bound states at all. The energy spectrum $\sigma(H)$ in this case is purely continuous and consists of the whole real line: $\sigma_(H) = \mathbb{R}$.

It should be stressed that the spectral properties of Dirac operators depend very much on the matrix structure of the potential. For a scalar potential $V(x) = \phi_{\mathrm{sc}}(\mathbf{x})\,\beta$, and after a unitary transformation with the constant matrix

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix}, \tag{8.103}$$

we can write the Dirac operator in the following form

$$H = \begin{pmatrix} 0 & c\,\boldsymbol{\sigma} \cdot \mathbf{p} + \mathrm{i}\big(\mathrm{m}c^2 + \phi_{\mathrm{sc}}(\mathbf{x})\big) \\ c\,\boldsymbol{\sigma} \cdot p - \mathrm{i}\big(\mathrm{m}c^2 + \phi_{\mathrm{sc}}(\mathbf{x})\big) & 0 \end{pmatrix}. \tag{8.104}$$

A charge conjugation does not change the sign of the scalar potential relative to the other terms in the Dirac operator. Hence, a scalar potential acts in the same way on electrons and positrons. If it is repulsive for electrons, so it is for positrons.

Next, we consider the case of a scalar potential that increases toward infinity. The square of this Dirac operator becomes

$$H^2 = \begin{pmatrix} A_+ & 0 \\ 0 & A_- \end{pmatrix}, \quad A_\pm = c^2 p^2 \pm c\boldsymbol{\sigma} \cdot \nabla \phi_{\mathrm{sc}}(\mathbf{x}) + (\mathrm{m}c^2 + \phi_{\mathrm{sc}}(\mathbf{x}))^2 \tag{8.105}$$

If $\phi_{\mathrm{sc}}(\mathbf{x})$ increases toward infinity (without oscillating too much), the square of the scalar potential will finally dominate the term $\boldsymbol{\sigma} \cdot \nabla \phi_{\mathrm{sc}}$. Hence, H^2 consists of Schrödinger operators with confining potentials. With H^2 also the Dirac operator H will have discrete eigenvalues. Therefore, a scalar harmonic oscillator potential $V(\mathbf{x}) = x^2\beta$ will produce a purely discrete energy spectrum.



CD 6.23 shows that there is no Klein paradox in case of a scalar potential step. CD 6.24.3 shows that a scalar potential well, no matter how deep, can bind particles.

8.5. Nonrelativistic Limit and Relativistic Corrections

8.5.1. The nonrelativistic limit

In a nonrelativistic theory, all velocities of the system are small in magnitude compared to c, the speed of light. Mathematically, the nonrelativistic limit of a relativistic theory is therefore described by the limit $c \to \infty$, which removes the relativistic bound on the propagation speed of signals. However, we cannot simply set $c = \infty$ in the Dirac equation, because this would just give infinity in all matrix elements of the free-particle Dirac operator (8.1).

Even the classical relativistic expression for the energy of a free particle goes to infinity, as c goes to infinity,

$$E(c) = \sqrt{c^2\mathbf{p}^2 + \mathrm{m}^2c^4} = \mathrm{m}c^2 + \frac{\mathbf{p}^2}{2\mathrm{m}} - \frac{\mathbf{p}^4}{8\mathrm{m}^3c^2} + O\Big(\frac{1}{c^4}\Big). \tag{8.106}$$

The leading term is the rest mass $\mathrm{m}c^2$, a purely relativistic object. It seems reasonable to subtract this term from the Dirac operator before taking the nonrelativistic limit. In the following, we consider the limit $c \to \infty$ of

$$\begin{aligned} H(c) - \mathrm{m}c^2\mathbf{1}_4 &= \begin{pmatrix} e\phi_{\mathrm{el}}(\mathbf{x}) + \phi_{\mathrm{sc}}(\mathbf{x}) & c\,\boldsymbol{\sigma}\cdot\big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\big) \\ c\,\boldsymbol{\sigma}\cdot\big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\big) & e\phi_{\mathrm{el}}(\mathbf{x}) - \phi_{\mathrm{sc}}(\mathbf{x}) - 2\,\mathrm{m}c^2 \end{pmatrix} \\ &= \begin{pmatrix} V_+ & c\,D^\dagger \\ c\,D & V_- - 2\,\mathrm{m}c^2 \end{pmatrix}. \end{aligned} \tag{8.107}$$

This expression refers to the standard representation. We have introduced the abbreviation

$$D = \boldsymbol{\sigma}\cdot\big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\big) = D^\dagger. \tag{8.108}$$

The distinction between D and its adjoint $D^\dagger$ in (8.107) has been made in view of possible generalizations.

The operator D still contains the parameter c. Nevertheless, we are going to keep D unchanged in the nonrelativistic limit. If we let the factor $1/c$ in front of the vector potential go to infinity, we would "turn off" the magnetic field. Actually, this factor is of an electrodynamic origin and has little to do with relativistic kinematics.

For the sake of simplicity, we assume that the potentials

$$V_\pm = e\phi_{\mathrm{el}}(\mathbf{x}) \pm \phi_{\mathrm{sc}}(\mathbf{x}) \tag{8.109}$$

are bounded functions. The treatment of the more general case, for example, the Coulomb potential, poses some technical complications (which can be dealt with by the use of bounded resolvents instead of unbounded operators).

Let us now consider the eigenvalue equation

$$\big(H(c) - \mathrm{m}c^2 \mathbf{1}_4\big)\,\psi = E\,\psi. \tag{8.110}$$

The eigenvalue E, like the eigenfunction ψ, will, of course, depend on the parameter c. We write

$$\begin{pmatrix} V_+ & c\,D^\dagger \\ c\,D & V_- - 2\,\mathrm{m}c^2 \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix} = E \begin{pmatrix} f \\ g \end{pmatrix} \tag{8.111}$$

with two-component spinors f and g. From this system, we may eliminate the two "lower components" g,

$$g = \Big(\mathbf{1} - \frac{1}{c^2}\frac{V_- - E}{2\mathrm{m}}\Big)^{-1} \frac{1}{2\mathrm{m}c}\,D\,f. \tag{8.112}$$

Inserting this expression into (8.111), we obtain the following equation for the "upper components" f:

$$\frac{1}{2\mathrm{m}}\,D^\dagger \Big(\mathbf{1} - \frac{1}{c^2}\frac{V_- - E}{2\mathrm{m}}\Big)^{-1} D\,f + V_+\,f = E\,f. \tag{8.113}$$

(This step is mathematically rigorous only for bounded V_- and large enough c, such that the operator $\mathbf{1} - (V_- - E)/(2\mathrm{m}c^2)$ is invertible.) The equation (8.113) contains the velocity of light only in the form of the small parameter $1/c^2$. In the nonrelativistic limit, we may set $1/c^2 = 0$, and the eigenvalue problem becomes

$$\frac{1}{2\mathrm{m}}\,D^\dagger D\,f^{(0)} + V_+\,f^{(0)} = E_0\,f^{(0)}. \tag{8.114}$$

Nonrelativistic limit of the Dirac operator:

The nonrelativistic limit of the Dirac operator

$$H(c) - \mathrm{m}c^2 = \begin{pmatrix} V_+ & c\,D^\dagger \\ c\,D & V_- - 2\,\mathrm{m}c^2 \end{pmatrix} \tag{8.115}$$

is the operator

$$H^{(0)} = \frac{D^\dagger D}{2\mathrm{m}} + V_+. \tag{8.116}$$

Using the above definition of D and $\mathbf{p} = -\mathrm{i}\,\nabla$, we obtain

$$\frac{1}{2\mathrm{m}}\,D^\dagger D + V_+ = \frac{1}{2\mathrm{m}}\Big(\boldsymbol{\sigma}\cdot\big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\big)\Big)^2$$
$$= \frac{1}{2\mathrm{m}}\big(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\big)^2 - \frac{e}{2\mathrm{m}c}\,\boldsymbol{\sigma}\cdot\mathbf{B}(\mathbf{x}). \tag{8.117}$$

This is precisely the nonrelativistic Hamiltonian for a particle with spin-1/2 in a magnetic field $\mathbf{B} = \nabla\times\mathbf{A}$. Hence, the energy due to the interaction of the spin with the magnetic field is

$$-g\boldsymbol{\mu}\cdot\mathbf{B}, \quad \text{with} \quad \boldsymbol{\mu} = \frac{e}{4\mathrm{m}c}\,\boldsymbol{\sigma}, \quad g = 2. \tag{8.118}$$

This shows that the Dirac theory of the electron has the correct g-factor $g = 2$ in the nonrelativistic limit.

Ψ The precise mathematical meaning of the statement in the box above is that the resolvent of the operator $H(c) - \mathrm{m}c^2$ converges (with respect to the norm of bounded operators) to the resolvent of the operator $H^{(0)}$ times a projection operator onto the upper components:

$$\lim_{c\to\infty}\big(H(c) - \mathrm{m}c^2\,\mathbf{1}_4 - z\,\mathbf{1}_4\big)^{-1} = \begin{pmatrix}\big(H^{(0)} - z\,\mathbf{1}_2\big)^{-1} & 0\\ 0 & 0\end{pmatrix} \tag{8.119}$$

holds for all $z\in\mathbb{C}$ with $\operatorname{Im} z \neq 0$.

8.5.2. Relativistic corrections

The eigenvalue problem, written in the form (8.113) as an equation for the upper components only, depends only on the parameter $1/c^2$. Hence, we assume[4] that the eigenvalues and eigenfunctions of (8.113) depend analytically on $1/c^2$. We are interested in an eigenvalue $E = E(c)$ close to the nth eigenvalue of the corresponding nonrelativistic problem and write

$$E(c) = E_n^{(0)} + \frac{1}{c^2}\,E^{(1)} + \mathcal{O}\Big(\frac{1}{c^4}\Big). \tag{8.120}$$

For simplicity, we assume that this eigenvalue is non-degenerate and that $f_n^{(0)}$ is the unique nonrelativistic eigenvector belonging to $E_n^{(0)}$,

$$\Big(\frac{D^\dagger D}{2\mathrm{m}} + V_+\Big)\,f_n^{(0)} = E_n^{(0)}\,f_n^{(0)}. \tag{8.121}$$

By the assumed analytic dependence of the relativistic eigenfunction $f = f(c)$ on $1/c^2$, we have

$$f(c) = f_n^{(0)} + \frac{1}{c^2}\,f^{(1)} + \mathcal{O}\Big(\frac{1}{c^4}\Big). \tag{8.122}$$

[4] A proof of the analyticity in $1/c^2$ for a large class of potentials including the Coulomb potential is given in [**10**].

We compute the lowest order relativistic corrections $E^{(1)}$ and $f^{(1)}$ to the nonrelativistic eigenvalue problem by the method presented in Appendix B.2. To that purpose, we assume that $f^{(1)}$ has an expansion in the eigenfunctions of the nonrelativistic problem

$$f^{(1)} = \sum_k a_k f_k^{(0)}, \qquad f(c) = \sum_k \big(\delta_{kn} + \frac{1}{c^2}\, a_k\big)\, f_k^{(0)} + \mathcal{O}\Big(\frac{1}{c^4}\Big). \tag{8.123}$$

Let us insert all this into the relativistic eigenvalue problem (8.113), ignoring terms of order $1/c^4$:

$$\begin{aligned} &\sum_k \Big(\frac{1}{2\mathrm{m}}\, D^\dagger \Big(\mathbf{1} - \frac{1}{c^2}\frac{V_- - E_n^{(0)}}{2\mathrm{m}}\Big)^{-1} D + V_+\Big)\, (\delta_{kn} + \frac{1}{c^2}\, a_k)\, f_k^{(0)} \\ &= \sum_k (E_n^{(0)} + \frac{1}{c^2}\, E^{(1)})\, (\delta_{kn} + \frac{1}{c^2}\, a_k)\, f_k^{(0)}. \end{aligned} \tag{8.124}$$

Taking the scalar product with $f_m^{(0)}$ and using the orthonormality of the eigenfunctions, we find

$$\begin{aligned} &\sum_k \langle f_m^{(0)}\,,\, \Big(\frac{1}{2\mathrm{m}}\, D^\dagger \Big(\mathbf{1} - \frac{1}{c^2}\frac{V_- - E_n^{(0)}}{2\mathrm{m}}\Big)^{-1} D + V_+\Big)\, f_k^{(0)} \rangle\, (\delta_{kn} + \frac{1}{c^2}\, a_k) \\ &= (E_n^{(0)} + \frac{1}{c^2}\, E^{(1)})\, (\delta_{mn} + \frac{1}{c^2}\, a_m). \end{aligned} \tag{8.125}$$

Next, we may expand[5]

$$\Big(\mathbf{1} - \frac{1}{c^2}\frac{V_- - E_n^{(0)}}{2\mathrm{m}}\Big)^{-1} = \mathbf{1} + \frac{1}{c^2}\frac{V_- - E_n^{(0)}}{2\mathrm{m}} + \mathrm{O}(1/c^4). \tag{8.126}$$

Using $\langle f_m^{(0)}, f_m^{(0)}\rangle = \delta_{mn}$ and keeping only terms up to the order $1/c^2$, we obtain

$$\begin{aligned} &E_m^{(0)}\, \delta_{mn} + \frac{1}{c^2}\frac{1}{4\mathrm{m}^2}\, \langle f_m^{(0)}\,,\, D^\dagger (V_- - E_n^{(0)})\, D\, f_k^{(0)}\rangle + \frac{1}{c^2}\, E_m^{(0)}\, a_m \\ &= E_n^{(0)}\, \delta_{mn} + \frac{1}{c^2}\, E^{(1)}\, \delta_{mn} + \frac{1}{c^2}\, E_n^{(0)}\, a_m. \end{aligned} \tag{8.127}$$

From this we obtain for $m = n$,

$$E^{(1)} = \frac{1}{4\mathrm{m}^2}\, \langle D\, f_n^{(0)}\,,\, (V_- - E_n^{(0)})\, D\, f_n^{(0)}\rangle \tag{8.128}$$

and for $m \neq n$,

$$a_m = \frac{1}{E_n^{(0)} - E_m^0}\, \langle D\, f_m^{(0)}\,,\, (V_- - E_n^{(0)})\, D\, f_n^{(0)}\rangle. \tag{8.129}$$

[5]Strictly speaking, the expansion formula (8.126) is only valid if the operator V_- is bounded (see Book One, Section 2.5.2).

We note that only the matrix element V_- of V is relevant for the calculation of the $1/c^2$-correction to the nonrelativistic eigenvalue. Equation (8.109) shows that V_- is different from V_+ in the presence of a scalar (non-electrostatic) potential.

$\boxed{\Psi}$ Using analytic perturbation theory (see Appendix C), one can extend the result above to potential matrices V that are relatively bounded with respect to the free-particle Dirac operator. The relative H_0-boundedness (Appendix C.1) is a regularity assumption that includes, for example, the Coulomb potential (with arbitrary coupling constant), but nothing with more severe singularities. In the degenerate case we have the following result:

Consider the Dirac operator

$$H = H_0 + V, \quad H_0 = \begin{pmatrix} \mathrm{m}c^2 & c\,D^\dagger \\ c\,D & -\mathrm{m}c^2 \end{pmatrix}, \quad V = \begin{pmatrix} V_+ & 0 \\ 0 & V_- \end{pmatrix}, \tag{8.130}$$

with a relatively H_0-bounded potential V. Let $E^{(0)}$ be an isolated, r-fold degenerate eigenvalue of the nonrelativistic problem,

$$\Big(\frac{D^\dagger D}{2\mathrm{m}} + V_+\Big)\, f_j^{(0)} = E^{(0)}\, f_j^{(0)}, \quad j = 1, \dots, r. \tag{8.131}$$

Then, there exists a neighborhood of $1/c^2 = 0$ where the Dirac operator $H - \mathrm{m}c^2$ has $k \leq r$ distinct eigenvalues E_j, $j = 1, \dots, r$. The eigenvalue E_j has the degree r_j of degeneracy such that $\sum_{j=1}^k r_j = r$. Moreover, E_j is analytic in $1/c^2$, and

$$E_j = E^{(0)} + \frac{1}{c^2}\, E_j^{(1)} + O\big(\frac{1}{c^4}\big), \tag{8.132}$$

where the numbers $E_j^{(1)}$, $j = 1, \dots, r$, are the eigenvalues of the Hermitian $r \times r$ matrix A with matrix elements

$$A_{ik} = \frac{1}{4\mathrm{m}^2}\, \Big(D\, f_i^{(0)}\,,\, (V_- - E^{(0)})\, D\, f_k^{(0)}\Big). \tag{8.133}$$

The r_j eigenvectors ψ_{j,k_j} of $H - \mathrm{m}c^2$ belonging to the eigenvalue E_j can be chosen in the form

$$\psi_{j,k_j}\big(\frac{1}{c}\big) = \begin{pmatrix} f_{j,k_j}(\frac{1}{c^2}) \\ \frac{1}{c}\, g_{j,k_j}(\frac{1}{c^2}) \end{pmatrix}, \qquad k_j = 1, \dots r_j, \tag{8.134}$$

where f_{j,k_j} and g_{j,k_j} are analytic in $1/c^2$, and

$$f_{j,k_j}(0) = f_j^{(0)}, \quad g_{j,k_j}(0) = \frac{1}{2\mathrm{m}}\, D\, f_j^{(0)}. \tag{8.135}$$

Therefore,

$$\psi_{j,k_j}(\frac{1}{c}) = \begin{pmatrix} f_j^{(0)} \\ 0 \end{pmatrix} + \frac{1}{2mc} \begin{pmatrix} 0 \\ D\, f_j^{(0)} \end{pmatrix} + \mathrm{O}(\frac{1}{c^2}). \tag{8.136}$$

Hence, the relativistic eigenvectors converge to the corresponding nonrelativistic eigenvectors, as $c \to \infty$. Convergence in the Hilbert space, however, does not imply pointwise convergence. Indeed, the eigenfunctions of the Dirac-Coulomb problem have a (square-integrable) singularity at the origin which is absent from the Coulomb eigenfunctions of the Schrödinger equation.

8.5.3. Spin-orbit interaction and the Darwin term

We consider the Dirac operator with an electrostatic potential

$$H(c) - \mathrm{m}c^2\, \mathbf{1} = c\boldsymbol{\alpha} \cdot \mathbf{p} + (\beta - \mathbf{1})\mathrm{m}c^2 + \phi_{\mathrm{el}}(\mathbf{x})\, \mathbf{1}. \tag{8.137}$$

In the standard representation, the matrix elements of (8.107) are given by

$$V_+(\mathbf{x}) = V_-(\mathbf{x}) = \phi_{\mathrm{el}}(\mathbf{x})\mathbf{1}_2, \qquad D = D^\dagger = \boldsymbol{\sigma} \cdot \mathbf{p} = -\mathrm{i}\, \boldsymbol{\sigma} \cdot \nabla. \tag{8.138}$$

The nonrelativistic limit of (8.137) is the Schrödinger operator

$$H^{(0)} = -\frac{1}{2\mathrm{m}}\, D^\dagger D + V_+(\mathbf{x}) = \Big(-\frac{1}{2\mathrm{m}}\Delta + \phi_{\mathrm{el}}(\mathbf{x})\Big)\, \mathbf{1}_2. \tag{8.139}$$

Acting on two-component wavefunctions, it describes the energy of particles with spin in an electrostatic field, without taking into account the spin-orbit interaction.

Relativistic perturbations:

Let $\phi_{\mathrm{el}}(\mathbf{x})$ be a bounded, twice continuously differentiable function of $\mathbf{x}$. Let $E^{(0)}$ be a non-degenerate eigenvalue of $H^{(0)}$,

$$H^{(0)}\, f^{(0)} = E_0\, f^{(0)}, \quad \|f^{(0)}\| = 1. \tag{8.140}$$

Then, a first approximation to the corresponding eigenvalue of the Dirac operator (8.137) is given by $E^{(0)} + E^{(1)}/c^2$, where

$$E^{(1)} = \frac{1}{4\mathrm{m}^2} \left\langle \boldsymbol{\sigma} \cdot \mathbf{p}\, f^{(0)}, (\phi_{\mathrm{el}} - E_0)\boldsymbol{\sigma} \cdot \mathbf{p}\, f^{(0)} \right\rangle \tag{8.141}$$

$$= -\frac{1}{8\mathrm{m}^3} \left\langle f^{(0)}\, ,\, \mathbf{p}^4\, f^{(0)} \right\rangle \tag{8.142}$$

$$+ \frac{1}{4\mathrm{m}^2} \left\langle f^{(0)}\, ,\, \big(\boldsymbol{\sigma} \cdot (\nabla\phi_{\mathrm{el}}) \times \mathbf{p}\big)\, f^{(0)} \right\rangle \tag{8.143}$$

$$+ \frac{1}{8\mathrm{m}^2} \left\langle f^{(0)}\, ,\, \Delta\phi_{\mathrm{el}}\, f^{(0)} \right\rangle. \tag{8.144}$$

PROOF. Equation (8.141) follows immediately from (8.128).

$$\begin{aligned} E^{(1)} &= \frac{1}{4\mathrm{m}^2}\,\langle f^{(0)}, \boldsymbol{\sigma}\cdot\mathbf{p}\,(\phi_{\mathrm{el}} - E^{(0)})\,\boldsymbol{\sigma}\cdot\mathbf{p}\,f^{(0)}\rangle \\ &= \frac{1}{4\mathrm{m}^2}\,\langle f^{(0)}, (\boldsymbol{\sigma}\cdot\mathbf{p}\,\phi_{\mathrm{el}}\,\boldsymbol{\sigma}\cdot\mathbf{p} - E^{(0)}(\boldsymbol{\sigma}\cdot\mathbf{p})^2)\,f^{(0)}\rangle \\ &= \frac{1}{4\mathrm{m}^2}\,\langle f^{(0)}, (\boldsymbol{\sigma}\cdot\mathbf{p}\,\phi_{\mathrm{el}}\,\boldsymbol{\sigma}\cdot\mathbf{p} - H^{(0)}\,\mathbf{p}^2) f^{(0)}\rangle \\ &= \frac{1}{4\mathrm{m}^2}\,\Big\langle f^{(0)}, \Big(-\frac{1}{2\mathrm{m}}\mathbf{p}^4 + [\,\boldsymbol{\sigma}\cdot\mathbf{p}\,,\phi_{\mathrm{el}}\,]\,\boldsymbol{\sigma}\cdot\mathbf{p}\Big)\,f^{(0)}\Big\rangle. \end{aligned} \tag{8.145}$$

With $\mathbf{p} = -\mathrm{i}\,\nabla$ we find

$$[\,\boldsymbol{\sigma}\cdot\mathbf{p}\,,\phi_{\mathrm{el}}]\,\boldsymbol{\sigma}\cdot\mathbf{p} = -\mathrm{i}\,(\boldsymbol{\sigma}\cdot\nabla\phi_{\mathrm{el}})\,\boldsymbol{\sigma}\cdot\mathbf{p} = -(\nabla\phi_{\mathrm{el}})\cdot\nabla + \boldsymbol{\sigma}\cdot(\nabla\phi_{\mathrm{el}})\times\mathbf{p}. \tag{8.146}$$

Next, we observe that

$$\begin{aligned} &\big\langle f^{(0)}, \frac{1}{2}\,(\Delta\phi_{\mathrm{el}})\,f^{(0)} + (\nabla\phi_{\mathrm{el}})\cdot(\nabla f^{(0)})\,\big\rangle \\ &= \frac{1}{2}\,\langle f^{(0)}, [\Delta,\phi_{\mathrm{el}}]f^{(0)}\rangle = \frac{1}{2}\,\langle f^{(0)}, 2\mathrm{m}[\phi_{\mathrm{el}} - E^{(0)}\,,\phi_{\mathrm{el}}]f^{(0)}\rangle = 0. \end{aligned} \tag{8.147}$$

Here, we just inserted $\Delta f^{(0)} = 2\mathrm{m}\,(\phi_{\mathrm{el}} - E^{(0)})\,f^{(0)}$ and used the self-adjointness of Δ. Finally, we obtain

$$\langle f^{(0)}\,,\,[\boldsymbol{\sigma}\cdot\mathbf{p}\,,\,\phi_{\mathrm{el}}\,]\,\boldsymbol{\sigma}\cdot\mathbf{p}\,f^{(0)}\rangle = \big\langle f^{(0)}\,,\,\big(\frac{1}{2}\,(\Delta\phi_{\mathrm{el}}) + \boldsymbol{\sigma}\cdot(\nabla\phi_{\mathrm{el}})\times\mathbf{p}\big)\,f^{(0)}\big\rangle, \tag{8.148}$$

from which (8.141) follows immediately. □

We note that the result (8.141) remains true for the Coulomb potential, while the derivation of (8.142)–(8.144) breaks down because of severe singularities at $\mathbf{x} = 0$, for example, a δ-function in (8.144). Nevertheless, it turns out that the scalar products (8.142)–(8.144) all remain finite even for the Coulomb potential.

The various terms (8.141) can be interpreted physically. The classical relativistic kinetic energy of electrons (with positive energy) is given by

$$\sqrt{c^2\mathbf{p}^2 + \mathrm{m}^2c^4} - \mathrm{m}c^2 = \frac{\mathbf{p}^2}{2\mathrm{m}} - \frac{\mathbf{p}^4}{8\mathrm{m}^3c^2} + O\Big(\frac{1}{c^4}\Big). \tag{8.149}$$

We see that (8.142) is just the expectation value of the operator corresponding to the term of order $1/c^2$. The operator

$$H_{\mathrm{kin}} = -\frac{1}{8\mathrm{m}^3}\,\mathbf{p}^4 = -\frac{1}{8\mathrm{m}^3}\,\Delta^2 \tag{8.150}$$

is therefore called the *relativistic kinetic-energy correction*.

The second summand (8.143) is the expectation value of

$$H_{\text{spin-orbit}} = \frac{1}{4\mathrm{m}^2}\,\boldsymbol{\sigma}\cdot(\nabla\phi_{\mathrm{el}})\times\mathbf{p} \tag{8.151}$$

This expression is called the *spin-orbit term*. It can be explained as follows. A particle with spin-1/2, charge q, and g-factor 2 has the magnetic moment

$$\boldsymbol{\mu} = \frac{q}{2\mathrm{m}}\,\boldsymbol{\sigma} \quad \text{(in units with } \hbar = 1\text{)}. \tag{8.152}$$

The magnetic moment interacts with an external magnetic field. As explained in Section 3.9.3, there is a magnetic field in the rest frame of the particle, and the corresponding interaction energy is, according to (3.133),

$$\frac{1}{c^2}\,\boldsymbol{\mu}\cdot\mathbf{v}\times\mathbf{E} = -\frac{1}{\mathrm{m}c^2}\,\boldsymbol{\mu}\cdot\mathbf{E}\times\mathbf{p} = \frac{1}{2\mathrm{m}^2c^2}\,\boldsymbol{\sigma}\cdot(\nabla\phi_{\mathrm{el}})\times\mathbf{p}. \tag{8.153}$$

Here, we used $q\mathbf{E} = -\nabla\phi_{\mathrm{el}}$. This expression is twice as large as $H_{\text{spin-orbit}}/c^2$. But, there is still another contribution due to the Thomas precession mentioned in Section 3.9.3. The term due to the Thomas precession combines with (8.153) to give the spin-orbit term $H_{\text{spin-orbit}}/c^2$.

The expression (8.144) is called the *Darwin term*,

$$H_{\mathrm{Darwin}} = \frac{1}{8\mathrm{m}^2}\,\Delta\phi_{\mathrm{el}}. \tag{8.154}$$

It is sometimes heuristically explained as an effect related to Zitterbewegung, but this is rather doubtful, because electronic bound states are stationary and hence perform no Zitterbewegung at all.

Ψ We note that the result (8.141) remains true for the Coulomb potential whereas the derivation of (8.142)–(8.144) breaks down because of severe singularities at $\mathbf{x} = 0$, for example, a δ-function in (8.144). Nevertheless, it turns out that the scalar products (8.142)–(8.144) all remain finite even for the Coulomb potential.

The perturbation formula (8.142)–(8.144) gives the wrong impression, that it is justified to apply first-order perturbation theory (as described in Appendix B.2, in particular, (B.28)) to a relativistically perturbed Hamiltonian

$$H_{\mathrm{rel}}(c) = H^{(0)} + \frac{1}{c^2}\big(H_{\mathrm{kin}} + H_{\text{spin-orbit}} + H_{\mathrm{Darwin}}\big). \tag{8.155}$$

Unfortunately, even in the absence of an external field, the term H_{kin} dominates in an operator-theoretic sense all other summands including $H^{(0)}$ so that for finite c, the spectrum of $H_{\mathrm{rel}}(c)$ is completely different from $H^{(0)}$ (for example, $H_{\mathrm{rel}}(c)$ has no negative eigenvalues at all). Hence, (8.155) should not be taken as the basis for perturbation theory.

However, an alternative approach to relativistic corrections due to Foldy and Wouthuysen leads in first order precisely to (8.155). This method consists in computing a sequence of approximate diagonalizations of the Dirac operator with an external field. At each step, the Foldy-Wouthuysen method introduces increasingly singular perturbations that cannot easily be justified

mathematically. In contrast, the method described in Section 8.5.2 uses only the mathematically well-established methods of analytic perturbation theory (see Appendix C).

8.6. Spherical Symmetry

8.6.1. Matrix potentials with spherical symmetry

In accordance with Section 1.4.1, we say that a Dirac operator H is spherically symmetric whenever it commutes with all rotations. But now, the potential $V(\mathbf{x})$ is a Hermitian 4×4 matrix, and hence the rotations (8.62) will affect both the argument $\mathbf{x}$ and the matrix structure of the potential. Therefore, it is not quite trivial to tell if $V(\mathbf{x})$ is spherically symmetric.

The free-particle Dirac operator is invariant under rotations, hence we find

$$\begin{aligned} \mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}}\, H\, \mathrm{e}^{\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}} &= H_0 + \mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}}\, V(\mathbf{x})\, \mathrm{e}^{\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{J}} \\ &= H_0 + \mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}}\, V(\mathbf{R}^{-1}\mathbf{x})\, \mathrm{e}^{\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}}. \end{aligned} \tag{8.156}$$

Therefore, the Dirac operator is invariant under rotations whenever

$$\mathrm{e}^{-\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}}\, V(\mathbf{R}^{-1}\mathbf{x})\, \mathrm{e}^{\mathrm{i}\,\varphi\,\mathbf{n}\cdot\mathbf{S}} = V(\mathbf{x}). \tag{8.157}$$

It is clear that an electrostatic potential of the form

$$V(\mathbf{x}) = \phi_{\mathrm{el}}(r)\, \mathbf{1}_4 \quad \text{with } r = |\mathbf{x}| \tag{8.158}$$

is invariant under rotations. Moreover, we find that the scalar potential (see Section 8.3.3)

$$V(\mathbf{x}) = \phi_{\mathrm{sc}}(r)\, \beta \quad \text{with } r = |\mathbf{x}| \tag{8.159}$$

is also invariant under rotations, because the Dirac matrix β commutes with all components of $\mathbf{S}$. Another example is given by the potential matrix

$$V(\mathbf{x}) = \mathrm{i}\,\beta\, \boldsymbol{\alpha} \cdot \mathbf{e}_r\, \phi_{\mathrm{am}}(r) \quad \text{with } \mathbf{e}_r = \mathbf{x}/r. \tag{8.160}$$

This type of potential matrix occurs for particles with an anomalous magnetic moment. In order to see that (8.160) is spherically symmetric, we use the relation (8.44). We insert $\gamma = (\beta, \beta\boldsymbol{\alpha})$, $x = (0, \mathbf{x})$, $\mathbf{T}(\Lambda) = \mathrm{e}^{-\mathrm{i}\,\varphi\mathbf{n}\cdot\mathbf{S}}$, and

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{R}(\varphi\mathbf{n}) \end{pmatrix} \tag{8.161}$$

into (8.44) and obtain

$$\mathrm{e}^{-\mathrm{i}\,\varphi\mathbf{n}\cdot\mathbf{S}}\, \beta\boldsymbol{\alpha} \cdot (\mathbf{R}^{-1}\mathbf{x})\, \mathrm{e}^{\mathrm{i}\,\varphi\mathbf{n}\cdot\mathbf{S}} = \beta\boldsymbol{\alpha} \cdot \mathbf{x}. \tag{8.162}$$

From this (8.157) follows immediately.

Dirac operators with spherical symmetry:
The Dirac operator $H = H_0 + V$ with a potential matrix of the form

$$V(\mathbf{x}) = \phi_{\rm el}(r)\,\mathbf{1}_4 + \phi_{\rm sc}(r)\,\beta + \mathrm{i}\,\phi_{\rm am}(r)\,\beta\,\boldsymbol{\alpha}\cdot\mathbf{x}/r \tag{8.163}$$

is spherically symmetric. It commutes with all rotations $\exp(-\mathrm{i}\varphi\,\mathbf{n}\cdot J)$.

As shown in (8.85), the potential matrix (8.160) can be used to describe a particle with an anomalous magnetic moment. For example, consider an electron with an anomalous moment $\mu_{\rm a}$. In a spherically symmetric electric field

$$\mathbf{E}(\mathbf{x}) = -\nabla\phi_{\rm el}(r) = -\frac{1}{r}\,\frac{d}{dr}\,\phi_{\rm el}(r) \tag{8.164}$$

it has the potential energy

$$V(\mathbf{x}) = e\,\phi_{\rm el}(r)\,\mathbf{1}_4 - \mathrm{i}\,\mu_{\rm a}\beta\,\frac{\boldsymbol{\alpha}\cdot\mathbf{x}}{r}\,\frac{d\phi_{\rm el}(r)}{dr}. \tag{8.165}$$

The Dirac operator with this potential matrix is spherically symmetric.

8.6.2. Operators that commute with the Dirac operator

In order to write the Dirac operator in spherical coordinates, we use the formula

$$-\mathrm{i}\,\nabla = -\mathrm{i}\,\mathbf{e}_r\,\frac{\partial}{\partial r} - \frac{1}{r}\,(\mathbf{e}_r\times\mathbf{L}). \tag{8.166}$$

This expression for the gradient in spherical coordinates can be obtained by combining the equations (1.87) and (1.90). Next, we use the algebraic relation

$$(\boldsymbol{\alpha}\cdot\mathbf{A})\,(2\,\mathbf{S}\cdot\mathbf{B}) = \mathrm{i}\,\gamma_5\mathbf{A}\cdot\mathbf{B} - \mathrm{i}\,\boldsymbol{\alpha}\cdot(\mathbf{A}\times\mathbf{B}) \tag{8.167}$$

which holds for any two vectors $\mathbf{A}$ and $\mathbf{B}$. Here, $\gamma_5 = -\mathrm{i}\alpha_1\alpha_2\alpha_3$. This implies that

$$-\frac{1}{r}\,\boldsymbol{\alpha}\cdot(\mathbf{e}_r\times\mathbf{L}) = \frac{\mathrm{i}}{r}\,(\boldsymbol{\alpha}\cdot\mathbf{e}_r)\,(2\mathbf{S}\cdot\mathbf{L}). \tag{8.168}$$

Hence, we find

$$-\mathrm{i}\,\boldsymbol{\alpha}\cdot\nabla = -\mathrm{i}\,(\boldsymbol{\alpha}\cdot\mathbf{e}_r)\,\frac{\partial}{\partial r} + \frac{\mathrm{i}}{r}\,(\boldsymbol{\alpha}\cdot\mathbf{e}_r)\,(2\,\mathbf{S}\cdot\mathbf{L}) \tag{8.169}$$

and, putting everything together, we arrive at the following expression for the free-particle Dirac operator in spherical coordinates:

$$H_0 = -\mathrm{i}\,c(\boldsymbol{\alpha}\cdot\mathbf{e}_r)\Big(\frac{\partial}{\partial r} + \frac{1}{r} - \frac{1}{r}\beta K\Big) + \beta\mathrm{m}c^2. \tag{8.170}$$

Here, we have introduced the *relativistic spin-orbit operator*

$$K = \beta(2\,\mathbf{S}\cdot\mathbf{L} + 1) = \beta(J^2 - L^2 + \tfrac{1}{4}). \tag{8.171}$$

The operator K describes the part of the Dirac operator that acts on the angular variables. All other summands in (8.170) obviously only affect the radial variable r of a wave function. Clearly, the spin-orbit operator (8.171) is the relativistic analog of the spin-orbit operator defined in (3.143).

The reason for introducing the Dirac matrix β in the definition of the relativistic spin-orbit operator K is that this operator commutes with H_0. First, we note that $2\,\mathbf{S}\cdot\mathbf{L}+1$ anticommutes with $\boldsymbol{\alpha}\cdot\mathbf{e}_r$,

$$(2\,\mathbf{S}\cdot\mathbf{L}+1)\,\boldsymbol{\alpha}\cdot\mathbf{e}_r = -\boldsymbol{\alpha}\cdot\mathbf{e}_r\,(2\,\mathbf{S}\cdot\mathbf{L}+1). \tag{8.172}$$

As the Dirac matrix β also anticommutes with $\boldsymbol{\alpha}\cdot\mathbf{e}_r$, we conclude that

$$[K, \boldsymbol{\alpha}\cdot\mathbf{e}_r] = 0. \tag{8.173}$$

Moreover, β commutes with $\mathbf{S}\cdot\mathbf{L}$, and therefore

$$[K, \beta] = 0. \tag{8.174}$$

Hence, K commutes with the free-particle Dirac operator H_0 in (8.170),

$$[K, H_0] = 0. \tag{8.175}$$

We know already that H_0 commutes with each of the angular-momentum operators J_k because H_0 and J_k are generators of a representation of the Poincaré group.

$$[H_0, J_k] = 0, \quad \text{for } k = 1,\, 2,\, 3. \tag{8.176}$$

Also, K commutes with rotations and hence with J^2 and J_k. We conclude that H_0, K, J^2, and J_3 form a system of commuting operators. J^2 is, however, not independent of K because we have, as in (3.144),

$$J^2 = K^2 - \frac{1}{4}. \tag{8.177}$$

Moreover, the relation

$$L^2 = K^2 - \beta\, K \tag{8.178}$$

shows that L^2 commutes with K and J_3 but not with H_0, because the matrix β anticommutes with $\boldsymbol{\alpha}$.

Moreover, because K commutes with β and $\beta\boldsymbol{\alpha}\cdot\mathbf{e}_r$, and because $\mathbf{L}$ performs only angular derivatives, we find that K commutes with the spherically symmetric potential matrix (8.163). Therefore, if H is a Dirac operator with the spherically symmetric potential (8.163), then H, K, and J_3 are a system of commuting operators.

EXERCISE 8.1. *Verify that each component J_k of the total angular momentum commutes with $\mathbf{S}\cdot\mathbf{L}$ and hence with K.*

8.6.3. Angular-momentum eigenfunctions

The operators K and J_3 act only on the angular coordinates of a wave function $\psi(r, \vartheta, \varphi)$ in spherical coordinates. Hence, we may consider these expressions as operators acting only in the Hilbert space $L^2(S^2)^4$. This is a Hilbert space of spinor-wave functions defined on the surface of the unit sphere S^2. In this section, we describe the eigenvalues and eigenfunctions of these operators, which are easily constructed from the results of the nonrelativistic theory in Section 3.9.4.

There is a complete system of common orthonormal eigenvectors $\Phi_{\kappa,m_j} \in L^2(S^2)^4$ and eigenvalue parameters κ and m_j, such that

$$K\,\Phi_{\kappa,m_j} = \kappa\,\Phi_{\kappa,m_j}, \qquad \kappa = \pm 1, \pm 2, \pm 3, \ldots, \tag{8.179}$$

$$J_3\,\Phi_{\kappa,m_j} = m_j\,\Phi_{\kappa,m_j}, \qquad m_j = -j, -j+1, \ldots, +j, \tag{8.180}$$

where

$$j = |\kappa| - \frac{1}{2} \tag{8.181}$$

is the angular-momentum quantum number describing the eigenvalue of J^2,

$$J^2\,\Phi_{\kappa,m_j} = j(j+1)\,\Phi_{\kappa,m_j}, \quad j = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \ldots, \tag{8.182}$$

The eigenvectors Φ_{κ,m_j} are functions of ϑ and φ (that is, functions on the unit-sphere S^2). For each pair of eigenvalues κ, m_j, one finds in fact two orthogonal eigenfunctions $\Phi^{\pm}_{\kappa,m_j}$ in $L^2(S^2)^4$, hence the corresponding common eigenspace of the operators K, J_3 is two-dimensional. We may choose a basis and write any eigenvector as a linear combination of these basis vectors,

$$\Phi_{\kappa,m_j} = c^+ \Phi^+_{\kappa,m_j} + c^- \Phi^-_{\kappa,m_j}, \tag{8.183}$$

with suitable complex constants c^+ and c^-. Here, we define the following basis vectors with respect to the standard representation

$$\Phi^+_{\kappa,m_j} = \begin{pmatrix} \mathrm{i}\,\mathcal{Y}_{\kappa,m_j} \\ 0 \\ 0 \end{pmatrix}, \quad \Phi^-_{\kappa,m_j} = \begin{pmatrix} 0 \\ 0 \\ \mathcal{Y}_{-\kappa,m_j} \end{pmatrix}, \tag{8.184}$$

where $\mathcal{Y}_{\kappa,m_j}$ are the spinor harmonics introduced in (3.151) and (3.152). Note that the spinor harmonics are two-component spinors, hence the spinors $\Phi^{\pm}_{\kappa,m_j}$ have four components. Obviously, they are also eigenfunctions of the Dirac matrix β (which is diagonal in the standard representation),

$$\beta\,\Phi^{\pm}_{\kappa,m_j} = \pm\Phi^{\pm}_{\kappa,m_j}. \tag{8.185}$$

The functions $\Phi^{\pm}_{\kappa,m_j}$ are also eigenfunctions of $L^2 = K^2 - \beta K$ and of $J^2 = K^2 - 1/4$ (because they are eigenfunctions of β and K). We have

$$L^2\,\Phi^{\pm}_{\kappa,m_j} = (\kappa^2 \mp \kappa)\,\Phi^{\pm}_{\kappa,m_j}, \quad J^2\,\Phi^{\pm}_{\kappa,m_j} = \left(\kappa^2 - \frac{1}{4}\right)\Phi^{\pm}_{\kappa,m_j}. \tag{8.186}$$

As usual, we denote the eigenvalues of L^2 by $\ell(\ell+1)$ and the eigenvalues of J^2 by $j(j+1)$. We conclude that both eigenfunctions have the total angular-momentum quantum number $j = |\kappa| - 1/2$. But, we find that Φ^+_{κ,m_j} has the orbital angular-momentum quantum number $\ell = \ell^+$, where

$$\ell^+ = \begin{cases} \kappa - 1 = j - 1/2, & \text{if } \kappa > 0, \\ -\kappa = j + 1/2, & \text{if } \kappa < 0, \end{cases} \tag{8.187}$$

whereas the second basis state Φ^-_{κ,m_j} belongs to $\ell = \ell^- = \ell^+ + 1$. As in nonrelativistic quantum mechanics, the quantum number ℓ has integer values. We have $\ell = j - 1/2 = 0, 1, 2, \ldots$ for Φ^+_{κ,m_j} if $\kappa > 0$ and for Φ^-_{κ,m_j} if $\kappa < 0$. In these cases, the orbital angular momentum ℓ and the spin $1/2$ add up to the total angular momentum $j = \ell + 1/2$. We have $\ell = j + 1/2 = 1, 2, 3, \ldots$ for Φ^+_{κ,m_j} if $\kappa > 0$ and for Φ^-_{κ,m_j} if $\kappa < 0$. In these cases, the spin $1/2$ is subtracted from the orbital angular momentum to give $j = \ell - 1/2$.

Finally, we need the formula

$$\mathrm{i}\boldsymbol{\alpha} \cdot \mathbf{e}_r \, \Phi^\pm_{\kappa,m_j} = \mp \Phi^\mp_{\kappa,m_j}. \tag{8.188}$$

This follows from a property of the spinor harmonics $\mathcal{Y}_{\kappa,m_j}$ which we state here without proof:

$$\boldsymbol{\sigma} \cdot \mathbf{e}_r \, \mathcal{Y}_{\kappa,m_j} = \mathcal{Y}_{-\kappa,m_j}. \tag{8.189}$$

8.6.4. The angular-momentum subspaces

We denote the two-dimensional subspace of the Hilbert space $L^2(S^2)^4$ that contains the linear combinations of the basis vectors $\Phi^\pm_{\kappa,m_j}$ by

$$\mathcal{K}_{\kappa,m_j} = \{\, c^+\Phi^+_{\kappa,m_j} + c^-\Phi^-_{\kappa,m_j} \mid (c^+, c^-) \in \mathbb{C}^2 \,\}. \tag{8.190}$$

This is the simultaneous eigenspace of the operators J_3, and K. From (8.185) and (8.188), we see that the subspaces $\mathcal{K}_{\kappa,m_j}$ are left invariant by the operators β and $\boldsymbol{\alpha} \cdot \mathbf{e}_r$. With respect to the basis $\{\Phi^+_{\kappa,m_j}, \Phi^-_{\kappa,m_j}\}$, these operators are represented by two-by-two matrices,

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad -\mathrm{i}\boldsymbol{\alpha} \cdot \mathbf{e}_r = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{8.191}$$

The completeness of the eigenfunctions means that any Dirac spinor $\psi(\vartheta, \varphi)$ that is square-integrable in the angular variables ϑ and φ is a linear combination

$$\psi(\vartheta, \varphi) = \sum_{\kappa,m_j} \left(c^+_{\kappa,m_j}\Phi^+_{\kappa,m_j} + c^-_{\kappa,m_j}\Phi^-_{\kappa,m_j} \right) \tag{8.192}$$

of the basis eigenfunctions with square-summable coefficients,

$$\sum_{\kappa,m_j} \big(|c^+_{\kappa,m_j}|^2 + |c^-_{\kappa,m_j}|^2\big) = \|\psi\|^2. \tag{8.193}$$

Consider two square-integrable functions $f^+(r)$ and $f^-(r)$ on $0 \leq r < \infty$. Define

$$\begin{aligned}\psi(r,\vartheta,\varphi) &= \frac{1}{r}\, f^+(r)\, \Phi^+_{\kappa,m_j}(\vartheta,\varphi) + \frac{1}{r}\, f^-(r)\, \Phi^-_{\kappa,m_j}(\vartheta,\varphi) \\ &= \frac{1}{r} \begin{pmatrix} \mathrm{i}\, f^+(r)\, \mathcal{Y}_{\kappa,m_j}(\vartheta,\varphi) \\ f^-(r)\, \mathcal{Y}_{-\kappa,m_j}(\vartheta,\varphi) \end{pmatrix}. \end{aligned} \tag{8.194}$$

This is a function in the *angular-momentum subspace* $L^2((0,\infty), dr) \otimes \mathcal{K}_{\kappa,m_j}$. Its norm is given by

$$\|\psi\|^2 = \int_0^\infty \big(|f^+(r)|^2 + |f^-(r)|^2\big)\, dr. \tag{8.195}$$

Making the transition back to Cartesian coordinates, we obtain a square integrable function $\Psi(\mathbf{x}) = \psi(r,\vartheta,\varphi)$ in $L^2(\mathbb{R}^3)^4$ whose norm is given by the above expression. An arbitrary Dirac spinor is a linear combination of the form

$$\Psi(\mathbf{x}) = \frac{1}{r} \sum_{\kappa,m_j} \Big(f^+_{\kappa,m_j}(r)\, \Phi^+_{\kappa,m_j}(\vartheta,\varphi) + f^-_{\kappa,m_j}(r)\, \Phi^-_{\kappa,m_j}(\vartheta,\varphi) \Big), \tag{8.196}$$

with square-integrable coefficient functions $f^\pm \in L^2(0,\infty)$.

The action of the Dirac operator H_0 on a function like (8.194) in a particular angular-momentum subspace can be obtained in a straightforward manner, because we know already the action of all parts of H_0 that appear in the expression (8.170). For example, the action of K is just multiplication by the eigenvalue κ. The action of the Dirac matrices β and $\boldsymbol{\alpha}\cdot\mathbf{e}_r$ in the angular-momentum subspace is described by (8.191). Likewise, we can compute the action of a spherically symmetric potential in one of the angular-momentum subspaces. We finally note that the factor $1/r$ in the definition of the wave functions in spherical coordinates (8.194) simplifies the operator $d/dr + 1/r$, which is part of expression for the Dirac operator in polar coordinates, because

$$\Big(\frac{d}{dr} + \frac{1}{r}\Big)\, \frac{1}{r}\, f(r) = \frac{1}{r}\, \frac{d}{dr}\, f(r). \tag{8.197}$$

We put everything together and obtain the following result:

The radial Dirac operator:

The action of the Dirac operator $H = -\mathrm{i}c\,\boldsymbol{\alpha}\cdot\nabla + \beta\,\mathrm{m}c^2 + V$ with a spherically symmetric potential

$$V(\mathbf{x}) = \phi_{\mathrm{sc}}(r)\beta + \phi_{\mathrm{el}}(r)\mathbf{1}_4 + \mathrm{i}\,\beta\,\boldsymbol{\alpha}\cdot\mathbf{e}_r\,\phi_{\mathrm{am}}(r) \tag{8.198}$$

on a spinor of the form

$$\Psi(\mathbf{x}) = \frac{1}{r}\big(f^+(r)\,\Phi^+_{\kappa,m_j}(\vartheta,\varphi) + f^-(r)\,\Phi^-_{\kappa,m_j}(\vartheta,\varphi)\big) \tag{8.199}$$

is described by the radial Dirac operator

$$h_\kappa = \begin{pmatrix} \mathrm{m}c^2 + \phi_{\mathrm{sc}}(r) + \phi_{\mathrm{el}}(r) & c\big(-\frac{d}{dr} - \frac{\kappa}{r}\big) + \phi_{\mathrm{am}}(r) \\ c\big(\frac{d}{dr} - \frac{\kappa}{r}\big) + \phi_{\mathrm{am}}(r) & -\mathrm{m}c^2 - \phi_{\mathrm{sc}}(r) + \phi_{\mathrm{el}}(r) \end{pmatrix} \tag{8.200}$$

acting on the coefficient functions $\begin{pmatrix} f^+(r) \\ f^-(r) \end{pmatrix}$.

This means that

$$(H_0 + V)\Psi(\mathbf{x}) = \frac{1}{r}\big(g^+(r)\,\Phi^+_{\kappa,m_j}(\vartheta,\varphi) + g^-(r)\,\Phi^-_{\kappa,m_j}(\vartheta,\varphi)\big) \tag{8.201}$$

where

$$\begin{pmatrix} g^+(r) \\ g^-(r) \end{pmatrix} = h_\kappa \begin{pmatrix} f^+(r) \\ f^-(r) \end{pmatrix}. \tag{8.202}$$

In particular, the Dirac operator leaves each of the angular-momentum subspaces invariant, that is, the result of its action is a wave function in the same angular-momentum subspace.

It is therefore sufficient to solve the eigenvalue problem $H\Psi = E\Psi$ in each of the angular-momentum subspaces. Whenever we have a solution of

$$h_\kappa \begin{pmatrix} f^+(r) \\ f^-(r) \end{pmatrix} = E \begin{pmatrix} f^+(r) \\ f^-(r) \end{pmatrix} \tag{8.203}$$

(this is just a two-dimensional system of ordinary differential equations), then E is an eigenvalue of the Dirac operator H, and Ψ defined as in (8.199) is the corresponding eigenvector. Note that the energy eigenvector Ψ is a superposition of a vector with orbital angular momentum ℓ^+ and another vector with $\ell^- = \ell^+ + 1$, hence it is not an eigenvector of L^2. This is, of course, due to the fact that H and L^2 do not commute and is often expressed by saying that the orbital angular momentum is "not a good quantum number" in relativistic physics.

The radial Dirac operator h_κ does not depend on the magnetic quantum number m_j. Hence, there are $2j+1$ orthogonal eigenfunctions of H belonging to the same eigenvalue E. These eigenfunctions are labeled by the quantum

number $m_j = -j, -j+1, \ldots, +j$, with $j = |\kappa| - 1/2$. This is the degeneracy due to the spherical symmetry of the system. In nonrelativistic quantum mechanics, the degeneracy due to spherical symmetry of the Schrödinger equation with spin is $2(2\ell+1)$, as discussed in Section 3.9.1. A part of this degeneracy has been removed by relativity.

In general, the energy depends on the sign of κ. This sign describes the two possible orientations of the spin with respect to the orbital angular momentum. We can understand the splitting of the energy levels heuristically: From the point of view of a particle that moves in an electrostatic field, it's the field-generating charges that are moving. Hence, the particle "sees" a magnetic field that causes the two spin-states to be energetically different.

8.7. The Dirac-Coulomb Problem

8.7.1. The radial Dirac-Coulomb equation

The stationary Dirac equation for a hydrogenic atom reads (in SI units)

$$\left(-\mathrm{i}\hbar c\,\boldsymbol{\alpha}\cdot\nabla + \beta\,\mu c^2 - \frac{Ze^2}{4\pi\epsilon_0}\,\frac{1}{|\mathbf{x}|}\right)\psi = E\psi. \tag{8.204}$$

Here, μ is the reduced mass of the electron-nucleus system. We divide this equation by $\hbar c$ and consider the eigenvalue equation

$$\left(-\mathrm{i}\boldsymbol{\alpha}\cdot\nabla + \beta\,\mathrm{m} - \frac{\gamma_\mathrm{c}}{|\mathbf{x}|}\right)\psi = \epsilon\,\psi \tag{8.205}$$

with the abbreviations

$$\mathrm{m} = \frac{\mu c}{\hbar}, \quad \gamma_\mathrm{c} = \frac{Ze^2}{4\pi\epsilon_0\,\hbar c}, \quad \epsilon = \frac{E}{\hbar c}. \tag{8.206}$$

The atomic number Z is an integer, but we treat γ_c simply as a real parameter. We note that

$$\frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137.036}. \tag{8.207}$$

is the dimensionless fine structure constant.

In this section, we assume that γ_c is positive. A positive coupling constant $\gamma_\mathrm{c} > 0$ describes an attractive Coulomb potential. The Schrödinger equation has no bound states for "repulsive" potentials, that is, for negative γ_c. In case of the Dirac equation, however, a potential that is repulsive for electrons is attractive for positrons (negative-energy wave packets). Hence, we expect that the Dirac equation has bound states also for negative coupling constants γ_c. It is not necessary to discuss this separately, because the positronic bound states for $\gamma_\mathrm{c} < 0$ can be obtained from the electronic bound states by a charge conjugation.

Using the spherical symmetry of the Coulomb potential, we introduce the radial operators

$$h_\kappa = \begin{pmatrix} \mathrm{m} - \frac{\gamma_\mathrm{c}}{r} & -\frac{d}{dr} - \frac{\kappa}{r} \\ \frac{d}{dr} - \frac{\kappa}{r} & -\mathrm{m} - \frac{\gamma_\mathrm{c}}{r} \end{pmatrix} = -\mathrm{i}\sigma_2 \frac{d}{dr} - \sigma_1 \frac{\kappa}{r} + \sigma_3 \mathrm{m} - \frac{\gamma_\mathrm{c}}{r} \mathbf{1}. \tag{8.208}$$

The eigenvalue problem (8.205) can now be replaced by the eigenvalue problems

$$(h_\kappa - \epsilon) \begin{pmatrix} f^+ \\ f^- \end{pmatrix} = 0, \quad \kappa = \pm 1, \pm 2, \ldots, \tag{8.209}$$

The parameter κ is the eigenvalue of the spin-orbit operator K defined in (8.171).

8.7.2. A useful similarity transformation

As on several occasions before, we use a method of solving the radial Dirac equation that works in an essentially algebraic way by the method of factorization. First, we multiply the eigenvalue equation (8.209) by the matrix $\mathrm{i}\sigma_2$ in order to bring the derivative to the main diagonal,

$$\mathrm{i}\sigma_2 (h_\kappa - \epsilon) = \mathbf{1} \frac{d}{dr} + \begin{pmatrix} -\kappa & -\gamma_\mathrm{c} \\ \gamma_\mathrm{c} & \kappa \end{pmatrix} \frac{1}{r} + \begin{pmatrix} 0 & -\epsilon - \mathrm{m} \\ \epsilon - \mathrm{m} & 0 \end{pmatrix}. \tag{8.210}$$

We can now diagonalize the matrix in front of $1/r$. We find

$$\mathrm{M}^{-1} \begin{pmatrix} -\kappa & -\gamma_\mathrm{c} \\ \gamma_\mathrm{c} & \kappa \end{pmatrix} \mathrm{M} = \begin{pmatrix} s & 0 \\ 0 & -s \end{pmatrix}, \quad s = \sqrt{\kappa^2 - {\gamma_\mathrm{c}}^2}. \tag{8.211}$$

with

$$\mathrm{M} = \begin{pmatrix} s - \kappa & \gamma_\mathrm{c} \\ \gamma_\mathrm{c} & s - \kappa \end{pmatrix}. \tag{8.212}$$

In terms of the functions

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathrm{M}^{-1} \begin{pmatrix} f^+ \\ f^- \end{pmatrix} \tag{8.213}$$

the eigenvalue equation (8.209) is thus equivalent to

$$\mathrm{M}^{-1} \mathrm{i}\sigma_2 (h_\kappa - \epsilon) \mathrm{M} \begin{pmatrix} u \\ v \end{pmatrix} = 0. \tag{8.214}$$

Multiplying this equation by $-\mathrm{i}\sigma_2$, we obtain the following result.

An equivalent eigenvalue problem:

The radial Dirac-Coulomb equation

$$\begin{pmatrix} \mathrm{m} - \frac{\gamma_c}{r} & -\frac{d}{dr} - \frac{\kappa}{r} \\ \frac{d}{dr} - \frac{\kappa}{r} & -\mathrm{m} - \frac{\gamma_c}{r} \end{pmatrix} \begin{pmatrix} f^+ \\ f^- \end{pmatrix} = \epsilon \begin{pmatrix} f^+ \\ f^- \end{pmatrix} \tag{8.215}$$

is equivalent to the equation

$$\begin{pmatrix} \mathrm{m} & -\frac{d}{dr} + \frac{s}{r} - \frac{\gamma_c \epsilon}{s} \\ \frac{d}{dr} + \frac{s}{r} - \frac{\gamma_c \epsilon}{s} & -\mathrm{m} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = -\frac{\kappa\epsilon}{s} \begin{pmatrix} u \\ v \end{pmatrix}. \tag{8.216}$$

In the new eigenvalue equation (8.216), all r-dependent parts appear as off-diagonal matrix elements. Note that the new matrix explicitly depends on the eigenvalue ϵ to be determined. We have to find those values of ϵ, for which the equation (8.216) admits square-integrable solutions.

8.7.3. The second-order equations

We introduce the abbreviation

$$D_0(\epsilon) \equiv \frac{d}{dr} + \frac{s}{r} - \frac{\gamma_c \epsilon}{s}, \tag{8.217}$$

and note that the adjoint of $D_0(\epsilon)$ is simply given by

$$D_0(\epsilon)^\dagger \equiv -\frac{d}{dr} + \frac{s}{r} - \frac{\gamma_c \epsilon}{s}. \tag{8.218}$$

Inserting $D_0(\epsilon)$, Eq. (8.216) becomes

$$\begin{pmatrix} \mathrm{m} & D_0(\epsilon)^\dagger \\ D_0(\epsilon) & \mathrm{m} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = -\frac{\kappa\epsilon}{s} \begin{pmatrix} u \\ v \end{pmatrix}. \tag{8.219}$$

The structure of the system (8.219) is very special. It is again a system of two ordinary differential equations of first order, but the coupling between u and v is now rather simple. Applying the matrix operator in (8.219) once again to both sides of that equation immediately leads to the two uncoupled equations

$$D_0(\epsilon)^\dagger D_0(\epsilon)\, u(r) + \mathrm{m}^2\, u(r) = \frac{\kappa^2\epsilon^2}{s^2}\, u(r), \tag{8.220}$$

and

$$D_0(\epsilon) D_0(\epsilon)^\dagger \, v(r) + \mathrm{m}^2 \, v(r) = \frac{\kappa^2 \epsilon^2}{s^2} \, v(r). \tag{8.221}$$

Using the definition of $D_0(\epsilon)$ in (8.217), we rewrite (8.220) as

$$\Big(-\frac{d^2}{dr^2} + \frac{s(s+1)}{r^2} - \frac{2\gamma_{\mathrm{c}}\epsilon}{r} + \mathrm{m}^2 \Big) \, \psi = \epsilon^2 \, \psi, \tag{8.222}$$

and (8.221) as

$$\Big(-\frac{d^2}{dr^2} + \frac{s(s-1)}{r^2} - \frac{2\gamma_{\mathrm{c}}\epsilon}{r} + \mathrm{m}^2 \Big) \, \psi = \epsilon^2 \, \psi. \tag{8.223}$$

(Here, we used $\kappa^2 = s^2 + {\gamma_{\mathrm{c}}}^2$.)

For later use, we define the sequence of operators

$$D_{n_r}(\epsilon) = \frac{d}{dr} + \frac{s+n_r}{r} - \frac{\gamma_{\mathrm{c}}\epsilon}{s+n_r} \qquad \text{for } n_r = 0, 1, 2, \ldots. \tag{8.224}$$

The operator $D_{n_r}(\epsilon)$ is obtained from $D_0(\epsilon)$ in (8.217) through the replacement $s \to s + n_r$. For each n_r, the adjoint of $D_{n_r}(\epsilon)$ is the operator

$$D_{n_r}(\epsilon)^\dagger = -\frac{d}{dr} + \frac{s+n_r}{r} - \frac{\gamma_{\mathrm{c}}\epsilon}{s+n_r} \, . \tag{8.225}$$

The crucial property of the operators $D_{n_r} = D_{n_r}(\epsilon)$ and $D_{n_r}^\dagger$ is the relation

$$D_{n_r-1}^\dagger D_{n_r-1} - \frac{{\gamma_{\mathrm{c}}}^2 \epsilon^2}{(s+n_r-1)^2} = D_{n_r} D_{n_r}^\dagger - \frac{{\gamma_{\mathrm{c}}}^2 \epsilon^2}{(s+n_r)^2}, \tag{8.226}$$

which holds for all $n_r \geq 1$.

8.7.4. Supersymmetry

We note that the two "second-order equations" (8.220) and (8.221) are related by the spectral supersymmetry described in Section 2.4.1 (and in Book One, Section 7.7.1). It is also important that a solution of one of the second order equations in turn determines a solution of the "first-order system" (8.219). We state this result in the following box:

Supersymmetric structure:

Let D be a suitable operator and let u be a solution of

$$D^\dagger D u + \mathrm{m}^2 u = \lambda^2 u, \tag{8.227}$$

with $Du \neq 0$. Then

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ \frac{1}{\lambda+\mathrm{m}} Du \end{pmatrix} \tag{8.228}$$

is a solution of

$$\begin{pmatrix} \mathrm{m} & D^\dagger \\ D & -\mathrm{m} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \end{pmatrix}. \tag{8.229}$$

Moreover, v is a solution of

$$D D^\dagger v + \mathrm{m}^2 v = \lambda^2 v. \tag{8.230}$$

Similarly, for any solution v of (8.230) such that $D^\dagger v \neq 0$ we find that

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda-\mathrm{m}} D^\dagger v \\ v \end{pmatrix} \tag{8.231}$$

is a solution of (8.229) and that u is a solution of (8.227).

You can verify this result by inserting (8.228) or (8.231) into the first-order equation (8.229). We note that the condition $Du \neq 0$ is equivalent with $D^\dagger D u \neq 0$. Therefore, (8.227) implies $\lambda^2 \neq \mathrm{m}^2$ and the solution (8.228) is well defined. A similar observation holds for (8.231).

Next, consider a nonzero solution u of (8.227) with $Du = 0$. Then, supersymmetry tells us nothing about the solutions of the other second-order equation (8.230). But, we can learn something about the matrix equation (8.229). Inserting $Du = 0$ into this equation gives

$$\mathrm{m}\, u + D^\dagger v = \lambda\, u, \qquad -\mathrm{m}\, v = \lambda\, v. \tag{8.232}$$

So either we have $v = 0$ and $\lambda = \mathrm{m}$, or we have $v \neq 0$, $\lambda = -\mathrm{m}$ and $D^\dagger v = -2\mathrm{m}\, u$. In the second case, we find $D D^\dagger v = -2\mathrm{m}\, D u = 0$. Hence, the first equation in (8.232) leads to $\lambda = \mathrm{m}$, because u is assumed to be nonzero and because $D D^\dagger v = 0$ is equivalent to $D^\dagger v = 0$. This is an obvious contradiction. So the only remaining possibility is that $v = 0$ and $\lambda = \mathrm{m}$.

A quite similar statement can be proved for a nonzero solution v of (8.230) with $D^\dagger v \neq 0$. We summarize these results as follows.

The solutions of the equation

$$\begin{pmatrix} \mathrm{m} & D^\dagger \\ D & -\mathrm{m} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \end{pmatrix} \quad (\text{with } \mathrm{m} > 0) \tag{8.233}$$

have the following property:

$$u \neq 0 \quad \text{and} \quad Du = 0 \qquad \text{implies} \qquad v = 0 \quad \text{and} \quad \lambda = \mathrm{m}. \tag{8.234}$$

Similarly,

$$v \neq 0 \quad \text{and} \quad D^\dagger v = 0 \qquad \text{implies} \qquad u = 0 \quad \text{and} \quad \lambda = -\mathrm{m}. \tag{8.235}$$

Thus, we arrive at the following conclusion:

An operator of the form

$$\begin{pmatrix} \mathrm{m} & D^\dagger \\ D & -\mathrm{m} \end{pmatrix} \tag{8.236}$$

has the eigenvalue $\lambda = \mathrm{m}$ if and only if u is a nonzero solution of $Du = 0$. The corresponding eigenvector is $\binom{u}{0}$.

Similarly, (8.236) has the eigenvalue $\lambda = -\mathrm{m}$ if and only if v is a nonzero solution of $D^\dagger v = 0$. In this case, the corresponding eigenvector is $\binom{0}{v}$.

In the following sections, we are going to apply these results in order to determine (non-normalized) eigenfunctions and eigenvalues of the Dirac-Coulomb problem.

8.7.5. The ground state

The differential equation $D_0(\epsilon)^\dagger\, v(r) = 0$, or

$$\left(-\frac{d}{dr} + \frac{s}{r} - \frac{\gamma_\mathrm{c}\epsilon}{s}\right) v(r) = 0, \tag{8.237}$$

has for arbitrary values of the parameters the solution

$$v(r) = r^s \mathrm{e}^{-(\gamma_\mathrm{c}\epsilon/s)r}. \tag{8.238}$$

This solution is square-integrable if and only if $\gamma_\mathrm{c}\epsilon/s > 0$. By definition, we have $s > 0$, and by assumption $\gamma_\mathrm{c} > 0$, hence we must have $\epsilon > 0$. The first-order equation (8.216) has the form of an eigenvalue equation for an operator of the form (8.236) and with $\lambda = -\kappa\epsilon/s$. Now the result at the end of the previous section tells us that $D_0(\epsilon)^\dagger v = 0$ implies $\lambda = -\mathrm{m}$, that is,

$$\frac{\kappa\epsilon}{s} = \mathrm{m}, \quad \text{or} \quad \epsilon = \epsilon_0 = \frac{\mathrm{m}s}{\kappa}. \tag{8.239}$$

Because of $\epsilon > 0$, this condition can only be satisfied for $\kappa > 0$.

We note that the square-integrable solution of (8.216) belonging to $\epsilon = \epsilon_0$ is

$$\begin{pmatrix} u_0(r) \\ v_0(r) \end{pmatrix} = \begin{pmatrix} 0 \\ r^2\, \exp\bigl(-(\gamma_\mathrm{c}\epsilon_0/s)\, r\bigr) \end{pmatrix}. \tag{8.240}$$

Automatically, the second component v_0 is also a solution of the second-order equation (8.221). The non-negativity of the operator $D_0(\epsilon)D_0(\epsilon)^\dagger$ implies that ϵ_0 is the smallest possible value of the energy parameter ϵ. This can be seen as follows. Let v be any square-integrable solution of (8.221). Take the scalar product of (8.221) with v to obtain

$$\langle v, D_0(\epsilon)D_0(\epsilon)^\dagger\, v\rangle + \mathrm{m}^2\,\|v\|^2 = \frac{\kappa^2\epsilon^2}{s^2}\,\|v\|^2. \tag{8.241}$$

Here, the first term is non-negative:

$$\big\langle v, D_0(\epsilon)D_0(\epsilon)^\dagger\, v\big\rangle = \|D_0(\epsilon)^\dagger\, v\|^2 \geq 0. \tag{8.242}$$

This implies immediately

$$\mathrm{m}^2 \leq \frac{\kappa^2\epsilon^2}{s^2}. \tag{8.243}$$

With ϵ_0 as in (8.244), we obtain the smallest possible value ($= \mathrm{m}^2$) on the right-hand side of (8.243).

Equation (8.216) is equivalent with the radial Dirac-Coulomb equation. According to (8.213), we obtain the corresponding eigenfunction by applying the matrix M to the solution (8.240). We summarize our findings in the following box.

Ground-state of the Dirac-Coulomb problem:

The lowest possible eigenvalue of the radial Dirac equation (8.215) is

$$\epsilon_0 = \frac{\mathrm{m}s}{\kappa} = \mathrm{m}\left(1 + \frac{{\gamma_\mathrm{c}}^2}{\kappa^2 - {\gamma_\mathrm{c}}^2}\right)^{-\frac{1}{2}} \qquad (\kappa > 0). \tag{8.244}$$

This eigenvalue is only possible for $\kappa > 0$. The corresponding square-integrable solution of the radial Dirac-Coulomb equation is

$$\begin{pmatrix} f_0^+(r) \\ f_0^-(r) \end{pmatrix} = N_0 \mathrm{M} \begin{pmatrix} 0 \\ v_0(r) \end{pmatrix} = n_0 \begin{pmatrix} \gamma_\mathrm{c} v_0 \\ (s-\kappa)v_0 \end{pmatrix}, \tag{8.245}$$

where N_0 is an appropriate normalization constant, and

$$v_0(r) = r^s\, \exp\Big(-\frac{\gamma_\mathrm{c}\epsilon_0}{s}\, r\Big). \tag{8.246}$$

A second possibility to obtain a solution of the Dirac-Coulomb problem would be to solve $D_0(\epsilon)u(r) = 0$. This differential equation is solved (for any ϵ) by

$$u(r) = r^{-s} e^{(\gamma_\mathrm{c}\epsilon/s)r}. \tag{8.247}$$

This solution has to be excluded because of the singularity at $r = 0$.

Ψ The function $u(r)$ defined in (8.247) is not square integrable whenever $s > 1/2$, that is, whenever $|\gamma_c| < (\kappa^2 - 1/4)^{1/2}$. But, for $s < 1/2$ (and $\gamma_c \epsilon < 0$), u is square integrable despite being singular at $r = 0$. In this case, the solution can only be excluded by specifying a boundary condition of the form

$$\lim_{r \to 0} u(r) = 0. \tag{8.248}$$

From a mathematical point of view, the choice of a boundary condition (a condition for a reflection at the origin) amounts to the choice of a distinguished self-adjoint extension of the operator h_κ for values of γ_c for which h_κ is not essentially self-adjoint. As long as $s < 1$, this method of obtaining a self-adjoint Dirac operator is fairly unambiguous.

8.7.6. The first excited state

We observe that (8.222) can be obtained from (8.223) simply by replacing the parameter s with $s + 1$. The same must be true for (8.220) and (8.221). Hence, if we replace in (8.221) the operator $D_0(\epsilon)$ by $D_1(\epsilon)$ and $\kappa^2 = s^2 + \gamma_c^2$ by $(s+1)^2 + \gamma_c^2$, then we obtain (8.220) in the form

$$\big(D_1(\epsilon)\, D_1(\epsilon)^\dagger + \mathrm{m}^2\big) u = \left(\epsilon^2 + \frac{{\gamma_c}^2 \epsilon^2}{(s+1)^2}\right) u. \tag{8.249}$$

We can easily find a solution for this equation if we only know a solution of

$$D_1(\epsilon)^\dagger u = 0. \tag{8.250}$$

From the definition (8.225) of $D_1(\epsilon)^\dagger$, we see that this equation is of the same type as (8.237). Therefore, for any ϵ, we obtain the solution

$$u_1(r) = r^{s+1} \exp\left(-\frac{\gamma_c \epsilon}{s+1} r\right), \tag{8.251}$$

just by replacing s with $s + 1$ in (8.238). For $\epsilon > 0$, $\gamma_c > 0$, the function u_1 is even square-integrable. Inserting u_1 into (8.249), we find a solution for $\epsilon = \epsilon_1$, where

$$\epsilon_1 = \mathrm{m}^2 \left\{1 + \frac{{\gamma_c}^2}{(s+1)^2}\right\}^{-1/2}. \tag{8.252}$$

Having found a solution u_1 of (8.220) with $D_0(\epsilon_1) u_1(r) \neq 0$, we obtain a solution of the first-order equation (8.219) according to (8.228). Finally, an eigenvector belonging to the eigenvalue ϵ_1 of the radial Dirac operator (8.208) is obtained by

$$\begin{pmatrix} f_1^+ \\ f_1^- \end{pmatrix} = N_1 \, \mathrm{M} \begin{pmatrix} u_1 \\ \big(\mathrm{m} - \frac{\kappa \epsilon_1}{s}\big)^{-1} D_0(\epsilon_1)\, u_1 \end{pmatrix}. \tag{8.253}$$

with the matrix M as defined in (8.212) and an appropriate normalization constant N_1.

The eigenvalue ϵ_1 is the smallest parameter for which (8.249) (or, equivalently, (8.220)) has a square-integrable solution, because $D_1 D_1^\dagger$ is a nonnegative operator. As the solutions of the Dirac equation with $\epsilon > \epsilon_0$ are in one-to-one correspondence to the solutions of (8.220), we find that there is no further eigenvalue between ϵ_0 and ϵ_1.

8.7.7. Further eigenfunctions

For the operators $D_{n_r} = D_{n_r}(\epsilon)$ defined in (8.224), we can make the following observation:

For $n_r \geq 1$, u is a solution of the equation

$$(D_{n_r} D_{n_r}^\dagger + \mathrm{m}^2)u = \left(\epsilon^2 + \frac{{\gamma_\mathrm{c}}^2 \epsilon^2}{(s+n_r)^2}\right)u \tag{8.254}$$

if and only if $v = D_{n_r} u$ is a solution of

$$(D_{n_r-1} D_{n_r-1}^\dagger + \mathrm{m}^2)v = \left(\epsilon^2 + \frac{{\gamma_\mathrm{c}}^2 \epsilon^2}{(s+n_r-1)^2}\right)v \tag{8.255}$$

with $D_{n_r-1}^\dagger v \neq 0$.

This can be proved as follows: Consider the equation

$$(D_{n_r-1}^\dagger D_{n_r-1} + \mathrm{m}^2)u = \left(\epsilon^2 + \frac{{\gamma_\mathrm{c}}^2 \epsilon^2}{(s+n_r-1)^2}\right)u. \tag{8.256}$$

The relation (8.226) shows that (8.256) is the same as (8.254). Moreover, for arbitrary $n_r \geq 1$, the equation $D_{n_r-1} u = 0$ has no suitable square-integrable solution at all (see the discussion at the end of Section 8.7.5). This shows that any square integrable solution u of (8.256) satisfies $D_{n_r-1} u \neq 0$ and is hence an eigenvector of $D_{n_r-1}^\dagger D_{n_r-1}$ belonging to a nonzero eigenvalue. Hence, by the spectral supersymmetry described in Exercise 2.5, the operator $D_{n_r-1} D_{n_r-1}^\dagger$ in (8.255) has the same eigenvalue, and the corresponding eigenvector is given by $v = D_{n_r} u$. The condition $D_{n_r-1}^\dagger v \neq 0$ is equivalent to $D_{n_r-1} D_{n_r-1}^\dagger v \neq 0$. The spectral supersymmetry holds for all nonzero eigenvalue of $D_{n_r-1}^\dagger D_{n_r-1}$ and $D_{n_r-1} D_{n_r-1}^\dagger$, and therefore all solutions of (8.255) with $D_{n_r-1}^\dagger v \neq 0$ can be obtained in this way from solutions of (8.254).

Obviously, if v is a solution of (8.254) with $n_r \geq 2$, then

$$u = D_1 D_2 \cdots D_{n_r-1} v \tag{8.257}$$

is a solution of the equation with $n_r = 1$. Equation (8.254) with $n_r = 1$ is the same as Eq. (8.249), which in turn is identical with (8.220).

We can now determine successively all solutions of (8.249) by considering only the ground-state solutions for the problems (8.254) with $n_r = 0, 1, 2, \ldots$. For example, the ground state of the problem with $n = 3$ gives the first excited state of the problem with $n_r = 2$, hence the second excited state of the problem with $n_r = 1$ (which is just (8.249)).

We find the ground-state solution of (8.254) for arbitrary n_r by solving the by now well-known equation

$$(D_{n_r})^\dagger v = \Big(-\frac{d}{dr} + \frac{s+n_r}{r} - \frac{\gamma_c \epsilon_n}{(s+n_r)}\Big) v = 0, \tag{8.258}$$

which gives

$$v_{n_r}(r) = r^{s+n_r} \exp\Big(-\frac{\gamma_c \epsilon_{n_r}}{s+n_r} r\Big). \tag{8.259}$$

Again, for $\gamma_c > 0$, v_{n_r} is square integrable if and only if $\epsilon_{n_r} > 0$. By inserting v_{n_r} into (8.254) we find that the energy eigenvalue is

$$\epsilon_{n_r} = \mathrm{m}\left(1 + \frac{{\gamma_c}^2}{(s+n_r)^2}\right)^{-\frac{1}{2}}. \tag{8.260}$$

We note that the factor describing the exponential decay of the eigenfunctions can be written as

$$\frac{\gamma_c \epsilon_n}{s+n_r} = \frac{\mathrm{m}\gamma_c}{\sqrt{(s+n_r)^2 + {\gamma_c}^2}} = \sqrt{\mathrm{m}^2 - \epsilon_{n_r}^2}. \tag{8.261}$$

For all $n_r \geq 2$, we obtain the solution of the Dirac-Coulomb equation belonging to the energy ϵ_{n_r} as

$$\begin{pmatrix} f^+_{n_r}(r) \\ f^-_{n_r}(r) \end{pmatrix} = N_{n_r}\, \mathrm{M} \begin{pmatrix} u_{n_r}(r) \\ \big(\mathrm{m} - \frac{\kappa \epsilon_{n_r}}{s}\big)^{-1} D_0 u_{n_r}(r) \end{pmatrix}, \tag{8.262}$$

where

$$u_{n_r}(r) = D_1(\epsilon_n)\, D_2(\epsilon_n) \cdots D_{n_r-1}(\epsilon_{n_r})\, v_{n_r}(r). \tag{8.263}$$

Again, N_{n_r} is an appropriate normalization constant. Determining the norm by computing the integral (8.195) is a long and tedious calculation, which is not done here.

8.8. Relativistic Hydrogen Atom

8.8.1. Eigenvalues and eigenfunctions

It is possible to express the radial eigenfunctions (8.262) in terms of hypergeometric functions. Without going into the details of this calculation, we present the results in the following box.

Solution of the Dirac-Coulomb problem:

For $0 < \gamma_c < 1$ the radial Dirac-Coulomb operator

$$h_\kappa = \begin{pmatrix} \mathrm{m} - \frac{\gamma_c}{r} & -\frac{d}{dr} - \frac{\kappa}{r} \\ \frac{d}{dr} - \frac{\kappa}{r} & -\mathrm{m} - \frac{\gamma_c}{r} \end{pmatrix} \tag{8.264}$$

has infinitely many eigenvalues

$$\epsilon_{n_r} = \mathrm{m}\left(1 + \frac{{\gamma_c}^2}{\left(n_r + \sqrt{\kappa^2 - {\gamma_c}^2}\right)^2}\right)^{-1/2}, \quad n_r = 0, 1, 2, 3, \ldots . \tag{8.265}$$

The corresponding eigenvectors are

$$\begin{aligned} \begin{pmatrix} f^+_{n_r}(r) \\ f^-_{n_r}(r) \end{pmatrix} &= r^s \, \mathrm{e}^{-k_{n_r} r} \left(\frac{\gamma_c \mathrm{m}}{k_{n_r}} + \kappa\right) {}_1F_1(-n_r, 2s+1, 2k_{n_r} r) \begin{pmatrix} N^+_{n_r} \\ -N^-_{n_r} \end{pmatrix} \\ &\quad - r^s \, \mathrm{e}^{-k_{n_r} r} \, n_r \, {}_1F_1(1 - n_r, 2s+1, 2k_{n_r} r) \begin{pmatrix} N^+_{n_r} \\ N^-_{n_r} \end{pmatrix}. \end{aligned} \tag{8.266}$$

Here, $k_{n_r} = \sqrt{\mathrm{m}^2 - \epsilon^2_{n_r}}$, $s = \sqrt{\kappa^2 - {\gamma_c}^2}$.

With the constants

$$N^\pm_{n_r} = \frac{(2k_{n_r})^{s+3/2}}{4\mathrm{m}\,\Gamma(2s+1)} \sqrt{\frac{(\mathrm{m} \pm \epsilon_{n_r})\,\Gamma(2s + n_r + 1)}{\gamma_c \left(\gamma_c \mathrm{m} + \kappa k_{n_r}\right) n_r!}}, \tag{8.267}$$

these solutions are normalized,

$$\int_0^\infty \left(|f^+_{n_r}(r)|^2 + |f^-_{n_r}(r)|^2\right) dr = 1. \tag{8.268}$$

All eigenvalues of the Dirac-Coulomb problem are in the interval $0 < \epsilon < \mathrm{m}$.

The eigenvalues and eigenfunctions of the radial problem determine the eigenvalues and eigenfunctions of the Dirac-Coulomb operator

$$H_{\mathrm{DC}} = c\,\boldsymbol{\alpha} \cdot \mathbf{p} + \beta\,\mu c^2 - \frac{\gamma}{|\mathbf{x}|} \tag{8.269}$$

(here $\gamma = \gamma_c \hbar c = Ze^2/4\pi\epsilon_0$). The eigenvalues of H_{DC} are related to the eigenvalues (8.265) of the radial operator h_κ by $E_{n_r} = \epsilon_{n_r} \hbar c$. In (8.265), we replace γ_c by $\gamma/\hbar c$ and m by $\mu c/\hbar$ and obtain

$$E_{n_r} = \mu c^2 \left(1 + \frac{\gamma^2/\hbar^2 c^2}{\left(n_r + \sqrt{\kappa^2 - \gamma^2/\hbar^2 c^2}\right)^2}\right)^{-1/2}. \tag{8.270}$$

All energies E_{n_r} are in the interval $0 < E < \mu c^2$. The eigenfunctions (8.266) of the radial Dirac-Coulomb problem determine the eigenfunctions of H_{DC}

by

$$\Psi(\mathbf{x}) = \frac{1}{r}\big(f^+_{n_r}(r)\,\Phi^+_{\kappa,m_j}(\vartheta,\varphi) + f^-_{n_r}(r)\,\Phi^-_{\kappa,m_j}(\vartheta,\varphi)\big), \tag{8.271}$$

where (r,ϑ,φ) are the spherical coordinates of $\mathbf{x}\in\mathbb{R}^3$.

8.8.2. Degeneracy and higher symmetry

The eigenvalues do not depend on the sign of the spin-orbit quantum number κ. The degeneracy of an eigenvalue E_{n_r} with angular momentum j is $2(2j+1)$, except if $n_r = 0$. The degeneracy is thus higher than one would expect from spherical symmetry alone (this is the "higher symmetry" of the relativistic Coulomb problem). For $n_r = 0$, only the radial eigenfunction with $\kappa > 0$ is square-integrable. Indeed, the normalization constants $N_0^\pm$ are not defined for $\kappa < 0$, because the factor $\gamma_{\mathrm{c}}\mathrm{m} + \kappa k_0 = \gamma_{\mathrm{c}}\mathrm{m}(1+\operatorname{sgn}\kappa)$ in the denominator of the square root becomes zero.

The higher symmetry of the relativistic Coulomb problem is related to the existence of an additional conserved quantity. One can show that the Dirac-Coulomb operator H_{DC} commutes with

$$B = \frac{2}{\mathrm{mc}}\,\mathbf{S}\cdot\left(\frac{\beta}{2}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}) + \mathrm{m}\gamma\,\frac{\mathbf{x}}{|\mathbf{x}|}\right) + \frac{\gamma}{\mathrm{mc}^2}\frac{1}{|\mathbf{x}|}K\alpha_1\alpha_2\alpha_3. \tag{8.272}$$

The operator B is called the *Biedenharn-Johnson-Lippmann operator*. It is the relativistic counterpart of the Runge-Lenz vector (see (2.14) and (2.25))

$$\frac{1}{2}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}) + \mathrm{m}\gamma\frac{\mathbf{x}}{|\mathbf{x}|}, \tag{8.273}$$

which is a conserved quantity for the nonrelativistic Coulomb problem. Unfortunately, the operator B does not commute with the relativistic spin-orbit operator K and hence cannot be diagonalized simultaneously with the angular momentum. Instead, B anticommutes with the spin-orbit operator K,

$$\{B,K\} = BK + KB = 0. \tag{8.274}$$

If Ψ is a common eigenfunction of the energy H_{DC} and the spin-orbit operator K with eigenvalues E_n and κ, then $B\Psi$ belongs to the eigenvalues E_n and $-\kappa$. This explains the degeneracy of the Dirac-Coulomb eigenvalues with respect to the sign of κ.

8.8.3. Angular momentum and spectroscopical notation

In order to compute the nonrelativistic limit of the Coulomb energies, we subtract the rest-mass μc^2 and consider the limit, as c goes to infinity, of

$$E_{n_r} - \mu c^2 \;\overset{c\to\infty}{\longrightarrow}\; E_{\mathrm{nonrel}} = -\frac{\mu\gamma^2}{\hbar^2}\,\frac{1}{2\,(n_r + |\kappa|)^2}. \tag{8.275}$$

In the nonrelativistic limit, the eigenvalue depends only on the "principal quantum number"

$$n = n_r + |\kappa|, \tag{8.276}$$

which leads to the much higher degree of degeneracy in the nonrelativistic Coulomb problem.

The angular-momentum operators $\mathbf{S}$ and $\mathbf{L}$ commute with each other, but not with the Dirac-Coulomb Hamiltonian. Only the total angular momentum is conserved, but the spin-angular momentum and the orbital angular momentum depend on time. In the nonrelativistic limit, the Dirac-Coulomb operator becomes the two-component Schrödinger operator (3.125), which commutes with both $\mathbf{S}$ and $\mathbf{L}$, as described in Section 3.9.

Furthermore, in the nonrelativistic limit, the lower components of the bound-state wave functions vanish. Therefore, Dirac's matrix β becomes the identity operator on eigenstates. The operator $K = \beta(2\mathbf{S}\cdot\mathbf{L}+\mathbf{1})$ simply becomes the nonrelativistic spin-orbit operator (3.143). If K is positive on some eigenstate (that is, $\kappa > 0$), then the spin is parallel to the orbital angular momentum in the nonrelativistic limit, that is, $j = \ell + 1/2$. On the other hand, $\kappa < 0$ implies for the nonrelativistic limit that $j = \ell - 1/2$ For all common eigenstates of $H_{\rm DC}$, K, and J^2, the values of the spin-orbit quantum number κ and of the total angular momentum j are unchanged in the nonrelativistic limit. Hence, κ is uniquely determined by j and the value of the orbital angular momentum in the nonrelativistic limit,

$$\kappa = \begin{cases} -\ell & \text{if } j = \ell - 1/2, \\ \ell + 1 & \text{if } j = \ell + 1/2. \end{cases} \tag{8.277}$$

The principal quantum number n, the total angular momentum j, and the orbital angular momentum ℓ can therefore be used to label the relativistic Coulomb eigenstates, even if the orbital angular momentum is not conserved in the relativistic theory. In the spectroscopic notation one writes nx_j, with x denoting the nonrelativistic orbital angular momentum as usual:

$$\ell = 0, 1, 2, 3, 4, \ldots \quad \text{corresponds to} \quad x = s, p, d, f, g, \ldots . \tag{8.278}$$

Hence, the eigenstates of $H_{\rm DC}$ are denoted by

$$\begin{array}{lllll} 1s_{1/2} & & & & \\ 2s_{1/2} & 2p_{1/2} & 2p_{3/2} & & \\ 3s_{1/2} & 3p_{1/2} & 3p_{3/2} & 3d_{3/2} & 3d_{5/2} \\ \text{etc.} & & & & \end{array} \tag{8.279}$$

This notation does not distinguish between states that belong to different eigenvalues m_j of J_3. Hence, there are $2j + 1$ orthogonal states denoted by nx_j, and the degree of degeneracy of the corresponding eigenvalue is

$2j+1$. The higher symmetry of the relativistic Coulomb problem (that is, the degeneracy with respect to the sign of κ) implies that the following states have the same energy, respectively:

$$\begin{aligned} &np_{1/2} \quad (\kappa=-1) \quad \text{and} \quad ns_{1/2} \quad (\kappa=1), \quad \text{for } n\geq 2, \\ &nd_{3/2} \quad (\kappa=-2) \quad \text{and} \quad np_{3/2} \quad (\kappa=2), \quad \text{for } n\geq 3, \quad \text{etc.} \end{aligned} \tag{8.280}$$

The states $1s_{1/2}$, $2p_{3/2}$, $3d_{5/2}$, and so forth, corresponding to $n_r = 0$ are all nondegenerate. All states with the same n have the same energy in the nonrelativistic limit. We also note that the relativistic electron bound state energies are always below the corresponding nonrelativistic eigenvalues.

8.8.4. Fall to the center

Consider the eigenvalues (8.265) as a function of the coupling constant. As soon as $\gamma_c > 0$, all eigenvalues are strictly below the threshold $\epsilon = \mathrm{m}$ (which corresponds to the energy μc^2). If the coupling strength is increased, the eigenvalues decrease continuously (that is, the *binding energy* $\mu c^2 - \epsilon_{n_r}$ increases). In the angular-momentum subspace $\kappa = 1$, the lowest eigenvalue ϵ_0 (which belongs to the ground state of the Coulomb problem) approaches 0 for $\gamma_c \to 1$,

$$\lim_{\gamma_c \to 1} \epsilon_0(\gamma_c) = 0, \tag{8.281}$$

with

$$\frac{d\epsilon_0(\gamma_c)}{d\gamma_c} \to \infty, \quad \text{as } \gamma_c \to 1 \text{ (for } \kappa = 1). \tag{8.282}$$

Here, the binding energy approaches the rest-energy μc^2. All eigenvalues $|\kappa| = 1$ become complex for and $\gamma_c > 1$. This indicates that the operator $h_\kappa = h_1$ (and hence the Dirac-Coulomb operator H_{DC}) ceases to be self-adjoint. No Coulomb-eigenvalue dives below $\epsilon = 0$.

The formula for the bound-state energies obviously looses its validity as soon as the coupling constant γ_c gets too large. The reason for this limitation lies in relativistic kinematics. Because of the relativistic energy-momentum relation, the kinetic energy increases linearly with $|\mathbf{p}|$ for large momenta. (In the nonrelativistic case, E increases like $|\mathbf{p}|^2$.) The kinetic energy combines with the potential energy $-\gamma_c/|\mathbf{x}|$ to a constant total energy. With increasing coupling strength, the wave function gets closer to the origin where both the kinetic and the potential energy tend to infinity. In the relativistic case, the slow increase of the kinetic energy cannot balance the more rapid increase of the potential energy, and the problem becomes ill defined. Physically, one may think that the particle is drawn into the singularity and hits the origin within a finite time.

In nonrelativistic quantum mechanics, the angular-momentum barrier prevents the collapse to the center. The angular-momentum barrier is an

effective potential of the form $\ell(\ell+1)/2r^2$ that has its origin in the expression for the kinetic energy in polar coordinates. In the radial Dirac equation, the angular-momentum barrier is given by the off-diagonal term κ/r. For both signs of κ, this term acts like a repulsive potential: For large $|\kappa|$, the wave function is driven away from the origin by the factor r^s in (8.266). We run into problems as soon as the Coulomb attraction dominates the angular-momentum barrier, that is, as soon as $|\gamma_c| > |\kappa|$.

Even in classical physics, the time evolution and the particle's trajectory are not defined after the particle hits a singularity of the potential, unless one specifies a "reflection condition" that allows a unique continuation of the trajectory. In quantum mechanics, such a condition is specified by a boundary condition at the origin. Indeed, the boundary condition $\lim_{r\to 0} f^{\pm}(r) = 0$ was necessary to define a unique square-integrable eigenfunction (see Section 8.7.5) for $\gamma_c^2 > \kappa^2 - 1/4$. For $\kappa = 1$ and $\gamma_c > \sqrt{3}/2$ (corresponding to atomic numbers $Z > 118$), this boundary condition defines a self-adjoint extension of the Dirac operator which is physically distinguished by the property that the kinetic and the potential energy both remain finite for all bound-state wave functions.

For $\gamma_c > 1$, the eigenvalues (8.265) with $|\kappa| = 1$ become complex. A complex eigenvalue means that the time evolution is not unitary, hence the generator (that is, the Dirac operator) cannot be self-adjoint any longer. We may define the complex square-root in such a way that ϵ_{n_r} has a negative imaginary part $\operatorname{Im}\epsilon_{n_r} < 0$ for $\gamma_c > 1$ (corresponding to the definition of a non-self-adjoint extension of the Dirac operator). Then, the wave function gets sucked away with time according to

$$\exp(-\mathrm{i}ct\epsilon_{n_r}) = \exp(-\mathrm{i}ct \operatorname{Re}\epsilon_{n_r}) \exp(ct \operatorname{Im}\epsilon_{n_r}) \to 0 \quad \text{as } t \to \infty. \tag{8.283}$$

This would correspond to the heuristic image of a particle falling into the singularity.

The condition $\gamma_c < 1$ is not a serious limitation. It means that the number of protons in the nucleus must be less than 137. This is rather far away in the transuranic region. When Z approaches 137, the ground-state wave function is drawn very close to the origin, and the finite size of the nucleus will become important. For large atomic numbers, the Coulomb potential with a point-like singularity (which is the origin of all troubles) is certainly not a good model of the physical reality. A potential model for an extended nucleus would rather resemble a smooth potential well. In this case, all problems with self-adjointness would disappear. A problem would still arise at much higher coupling constants, when the binding energy becomes $-2\mu c^2$. In this case, the lowest eigenvalue dives into the negative-energy continuum and becomes a resonance for positronic scattering states—a situation similar to the one discussed in Section 8.4.4. It is not quite clear

how this should be interpreted. One expects some sort of spontaneous pair creation—something that the Dirac equation alone cannot handle. Quantum field theory is needed to describe these phenomena, but the mathematics is tricky, and a final word has not yet been spoken.

Introducing a cut-off of the Coulomb potential at short distances introduces some arbitrariness. The anomalous magnetic moment of the electron also regularizes the Dirac-Coulomb Hamiltonian. One can show that the additional term describing the anomalous magnetic moment, though highly singular, is repulsive and forces the wave function away from the singularity. Hence, the Dirac-Coulomb operator with an anomalous magnetic moment is well defined and self-adjoint for all values of γ_c and κ. The behavior of the eigenvalues is similar to the case of a potential with cut-off singularity.

Appendix A

Synopsis of Quantum Mechanics

This book assumes that the reader is familiar with the basic notions of quantum theory. Here, we give a brief outline of the quantum-mechanical formalism to the extent that is needed for this book. Details, explanations, and examples can be found in many introductory books, for example, in Book One [**11**].

A.1. The Hilbert Space of Quantum Mechanics

Quantum mechanics describes the physical reality in terms of rather abstract mathematical objects. From these, any predictions about the results of measurements have to be derived with the help of interpretation rules. The mathematical framework for the quantum-mechanical formalism is defined by associating a suitable Hilbert space with a given physical system.[1]

A *Hilbert space* $\mathfrak{H}$ is a linear space over the field of complex numbers $\mathbb{C}$ with a scalar product $\langle \cdot, \cdot \rangle$ (see Book One, Section 2.2). A *scalar product* is a mapping that associates a complex number with an ordered pair of vectors ϕ and ψ such that

- $\langle \psi, \psi \rangle > 0$ for $\psi \neq 0$ (positive definiteness)
- $\langle \phi, a\psi_1 + b\psi_2 \rangle = a \langle \phi, \psi_1 \rangle + b \langle \phi, \psi_2 \rangle$ (linearity in 2nd argument)
- $\langle \phi, \psi \rangle = \overline{\langle \psi, \phi \rangle}$ (skew-symmetry)

The scalar product induces a *norm*

$$\|\psi\| = \sqrt{\langle \psi, \psi \rangle} \tag{A.1}$$

and Hilbert spaces are by definition *complete* with respect to this norm (see Book One, Section 2.2.3).

[1] Here, we do not attempt to give a formal definition of the expression "physical system." It has the usual intuitive meaning of something that can be described in physical terms and that is sufficiently distinguished so that it is clear what one is talking about. In this book we deal with rather simple systems—an electron, or just the spin of an electron, an atom, or a measuring apparatus.

Examples: A (structureless) particle in an n-dimensional space is associated with the infinite dimensional Hilbert space $L^2(\mathbb{R}^n)$. This is the space of square-integrable complex-valued functions with the scalar product

$$\langle \phi \,,\, \psi \rangle = \int_{\mathbb{R}^3} \overline{\phi(\mathbf{x})}\, \psi(\mathbf{x})\, d^3x. \tag{A.2}$$

If the motion is restricted, for example to an interval $[a, b]$ in $\mathbb{R}$, then the Hilbert space is $L^2([a, b])$. The states of a relativistic particle with spin-1/2 form the Hilbert space $L^2(\mathbb{R}^3)^4$ of $\mathbb{C}^4$-valued square-integrable functions (see Section 7.3.1).

If we are only interested in the internal states of a spin-1/2 particle, we choose the two-dimensional Hilbert space $\mathbb{C}^2$ (see Chapter 4). A general example of a finite-dimensional Hilbert space is the n-dimensional complex vector space $\mathbb{C}^n$, consisting of vectors $\psi = (\psi_1, \ldots, \psi_n)$, with $\psi_i \in \mathbb{C}$, and with the scalar product defined by

$$\langle \phi \,,\, \psi \rangle = \sum_{i=1}^{n} \overline{\phi_i}\, \psi_i \,. \tag{A.3}$$

A.2. States of a Physical System

The *state* of a physical system is a usually well-defined collection of information about the system, for example, the values of certain physical quantities at some given time (see Book One, Section 4.1.1). This set of information should be as complete as possible in the sense that it allows one to make statements about physical quantities of interest. Moreover, the information should be precise enough so as to determine the future behavior of the state. In the mathematical framework of quantum mechanics, states are described in terms of vectors in the Hilbert space associated with the system, but the mapping between vectors and states is not one-to-one (see Book One, Section 4.1.1).

Postulate A.1. *The (pure) states of the physical system at any instant of time are uniquely given by the one-dimensional subspaces (rays)*

$$\hat{\psi} = \{\phi = \lambda\psi \mid \lambda \in \mathbb{C}\} \tag{A.4}$$

of the Hilbert space $\mathfrak{H}$ of the system.

Here, the word "pure" distinguishes this type of state from the mixed states discussed in Section 5.6.1. By choosing a representative $\psi \in \hat{\psi}$ with $\|\psi\| = 1$, one can say that the state is described by a normalized vector in the Hilbert space. This is usually assumed in the formulas that describe the interpretation rules. The *projection operator*

$$\rho_\psi = \psi\, \langle \psi, \cdot \rangle. \tag{A.5}$$

projects onto the one-dimensional subspace spanned by ψ and thus can also serve to characterize the state of the system.

Superposition principle: The linear structure of the Hilbert space allows us to form arbitrary linear combination of state vectors. The result of the linear combination is again a vector in the Hilbert space. The corresponding one-dimensional subspace defines a new state, which is called a *superposition* of the given states. Obviously, any superposition of states is another possible state of the quantum system. This statement is called the superposition principle.

A.3. Observables

An *observable* of a physical system is a physical quantity for that a numerical value can be obtained in a well-defined process. This process is usually called a measurement.[2]

POSTULATE A.2. *The physical observables are represented by linear self-adjoint operators acting in the Hilbert space of the system.*

In a finite-dimensional Hilbert space $\mathfrak{H}$, a linear operator A is a mapping $\mathfrak{H} \to \mathfrak{H}$ with the property

$$A(a\psi_1 + b\psi_2) = aA\psi_1 + bA\psi_2 \tag{A.6}$$

and its adjoint is a linear operator $A^\dagger$ defined by

$$\langle \phi, A\psi \rangle = \langle A^\dagger \phi, \psi \rangle. \tag{A.7}$$

An operator is *self-adjoint* if $A = A^\dagger$. For a self-adjoint operator, the quantity $\langle \psi, A\psi \rangle$ is always a real number. In the case of infinite-dimensional Hilbert spaces, one has to take into account the domain of definition of the linear operators, as discussed in Book One (Section 6.6).

Examples: The components of position are represented by the operators of multiplication with the variables $\mathbf{x} = (x_1, \dots, x_n)$. The momentum operator is given by the differential operator $-\mathrm{i}\nabla$ (where i is the imaginary unit and ∇ is the gradient). A particular role is played by the energy operator (*Hamiltonian*) H. For a particle with mass m and charge q moving under the influence of an electromagnetic field, the energy operator is

$$H = -\frac{1}{2m}\left(-\mathrm{i}\nabla - \frac{q}{c}\mathbf{A}(\mathbf{x})\right) + qV(\mathbf{x}). \tag{A.8}$$

Here, $(V, \mathbf{A})$ are the electromagnetic potentials that describe the electric and magnetic fields by $\mathbf{E} = -\nabla V$ and $\mathbf{B} = \nabla \times \mathbf{A}$.

[2]We do not claim that an observable has a value independently of the measurement.

A.4. Interpretation Rule

In general, quantum mechanics makes only probabilistic predictions about the outcomes of measurements, even if the state of the system is known exactly. The measurement of an observable A is thus a random experiment. It allows one to define a random variable describing the numerical values associated with the observable in each run of the experiment.[3]

If, at any particular moment, the state of a physical system is represented by a normalized state ψ, then

$$\langle A\rangle_\psi = \langle \psi\,, A\psi\rangle \tag{A.9}$$

is the *expectation value* (*mean value*) of the observable A.

The *uncertainty* or *standard deviation* $\Delta_\psi A$

$$\Delta_\psi A \equiv \|(A - \langle A\rangle_\psi)\psi\| = \sqrt{\langle (A - \langle A\rangle_\psi)^2\rangle_\psi} \tag{A.10}$$

of an observable A in the state ψ (which is assumed to be normalized) describes the dispersion of the values of A around the mean value $\langle A\rangle_\psi$.

If the state ψ is an eigenvector of the linear operator A, that is,

$$A\psi = a\psi, \quad \text{for some real number } a, \tag{A.11}$$

then the result of a measurement of A gives a with certainty, that is, $\Delta_\psi A = 0$. Note, however, that there are self-adjoint operators (for example, position and momentum) that have no eigenvalues at all (they have a purely continuous spectrum).

For an observable that has only eigenvalues (no continuous spectrum), these eigenvalues are the only possible measurement results.

Two observables are called incompatible if the corresponding operators A and B do not commute, that is $[A, B] = AB - BA \neq 0$. One cannot obtain numerical values for two incompatible observables within a single experiment. The product of their uncertainties in a given state ψ obeys the inequality

$$\Delta_\psi A\ \Delta_\psi B \geq \tfrac{1}{2}\,\big|\,\langle [A, B]\,\rangle_\psi\,\big|, \tag{A.12}$$

hence the uncertainties of incompatible observables A and B cannot both be small.

A.5. Projections and Properties

An orthogonal projection operator is a bounded self-adjoint operator P that has only the eigenvalues 0 and 1. Hence, it is characterized by $P^2 = P = P^\dagger$. The physical observable represented by a projection operator describes whether the system has a certain property or not. Measuring a property can

[3]The observable A, the self-adjoint operator representing A, and the random variable describing the values of A are usually denoted by the same letter.

only give the results "yes" (eigenvalue 1) or "no" (eigenvalue 0). The range of P, that is, the eigenspace belonging to the eigenvalue 1, is the subspace of states with that property.

For every observable A, there is a unique family of projection operators $P_B(A)$, labeled by (measurable) subsets B of $\mathbb{R}$. The projection operator $P_B(A)$ describes whether the observable A has a value in the set $B \subset \mathbb{R}$ or not. The mapping $B \to P_B(A)$ is called an orthogonal projection operator-valued measure, or spectral measure associated with the observable A. The famous spectral theorem states that this mapping between self-adjoint operators and spectral measures is one-to-one.

For example, the operator of multiplication by the characteristic function $\chi_B(x)$ of a region B is a projection operator. Its expectation value is just the probability of finding the particle in the region B. The operators $\chi_B(x)$ define the spectral measure of the position operator x.

Assume that a particle in $\mathbb{R}$ is described by a normalized wave function $\psi \in L^2(\mathbb{R})$. Then, the quantity $\int_B |\psi(x)|^2\,dx$ is the probability of finding the particle in the region $B \subset \mathbb{R}$. Similarly: Let $\hat{\psi}$ be the Fourier transform of ψ. Then $\int_B |\hat{\psi}(k)|^2\,dk$ is the probability of finding the momentum of the particle in $B \subset \mathbb{R}$.

A one-dimensional projection operator ρ_ϕ is given by

$$\rho_\phi \psi = \phi \,\langle \phi \,, \psi \rangle, \quad \text{also written as} \quad \rho_\phi = \phi \,\langle \phi, \cdot \,\rangle = |\phi\rangle\langle\phi|. \tag{A.13}$$

It projects onto the one-dimensional subspace $\{\lambda\phi \mid \lambda \in \mathbb{C}\}$. The observable ρ_ϕ describes the property of being in the state described by the vector ϕ. The expectation value $\langle \rho_\phi \rangle_\psi$ is the probability that a system in the state ψ is found to be in the state ϕ,

$$p_{\psi \to \phi} = \langle \rho_\phi \rangle_\psi = |\langle \phi, \psi \rangle|^2. \tag{A.14}$$

This is called the *transition probability* from ψ to ϕ.

A.6. Time Evolution

The third postulate describes the time evolution of a state. An operator is called unitary if $U^\dagger = U^{-1}$.

Postulate A.3. *The time evolution of a physical system is described by the unitary operator* $\exp(-\mathrm{i}Ht)$*, where* H *is the energy operator of the system. If the state at time* t_0 *is given by* ψ_0*, then the state at time* t *is given by*

$$\psi(t) = \mathrm{e}^{-\mathrm{i}H(t-t_0)}\,\psi_0. \tag{A.15}$$

In particular, if the initial state ψ_0 is in the domain $\mathfrak{D}(H)$ of the Hamiltonian H, then $\psi(t)$ is a solution of the Schrödinger equation

$$\mathrm{i}\,\frac{d}{dt}\,\psi(t) = H\,\psi(t). \tag{A.16}$$

If the initial state ψ_0 is an eigenstate of the Hamiltonian, $H\psi_0 = E\psi_0$ (with some real number E), then (A.15) specializes to

$$\psi(t) = \mathrm{e}^{-\mathrm{i}E(t-t_0)}\,\psi_0. \tag{A.17}$$

The time evolution is thus reduced to a simple multiplication by a time-dependent phase factor, something that does not change the state at all. Hence, the eigenstates are also called *stationary states*.

Unitary Groups: An important mathematical theorem (*Stone's Theorem*) states that any self-adjoint operator H generates a group of unitary operators $\exp(-\mathrm{i}Ht)$ such that Eqs. (A.16) and (A.15) hold. The operator-valued exponential function satisfies in particular

$$\mathrm{e}^{-\mathrm{i}H0} = \mathbf{1}, \quad \mathrm{e}^{-\mathrm{i}Ht}\,\mathrm{e}^{-\mathrm{i}Hs} = \mathrm{e}^{-\mathrm{i}H(t+s)}, \tag{A.18}$$

and the mapping $\psi \to \psi(t) = \mathrm{e}^{-\mathrm{i}Ht}\psi$ is continuous with respect to t for all ψ in the Hilbert space.[4]

Among the unitary groups that are important in quantum mechanics, we mention the translations, which are generated by the momentum operator, and the rotations, which are generated by the angular-momentum operators (see Section 1.2.2).

Time-dependence of observables: The time evolution of an observable A is defined by

$$A(t) = \mathrm{e}^{\mathrm{i}Ht}\,A\,\mathrm{e}^{-\mathrm{i}Ht}. \tag{A.19}$$

Hence, the expectation value of $A(t)$ fulfills the relation

$$\langle A(t)\rangle_\psi = \langle A\rangle_{\psi(t)}. \tag{A.20}$$

The function $t \to A(t)$ is a solution of Heisenberg's equation of motion,

$$\frac{d}{dt}\,A(t) = \mathrm{i}[H, A(t)], \tag{A.21}$$

with the initial condition $A(0) = A$. Observables A that are conserved under the time evolution, $A(t) = A$, are called constants of motion. They generate symmetry transformations of the system. The constants of motion are self-adjoint operators that commute with the Hamiltonian, $[H, A] = 0$.

[4]For a fixed t, the mapping $\psi \to \psi(t)$ is also continuous with respect to ψ, because a unitary operator is continuous. This means that the solution depends continuously on the initial data.

A.7. Measurements

A measurement is usually a complicated interaction between an object and a measuring device. For a surprisingly large variety of situations, it is sufficient to use one of the following simple models of measurement processes.

(1) Determinative measurement or state verification. It is the goal to determine the state of a system or to verify a quantum-mechanical prediction. One has to repeat the measurement very often on identically prepared copies of the system, because one has to verify a probability distribution predicted by the quantum mechanical formalism. The collection of identically prepared systems is called a statistical ensemble. In order to be explicit, we sometimes call the measurement on an individual system (that is, a single run of the random experiment) an elementary experiment. With a finite number of elementary experiments, it is only possible to obtain approximate information about the state (*state estimation*).

(2) Preparative measurements or state preparation. This can be interpreted as an operation performed on an individual system. The measurement of a physical quantity usually gives a random result, but once the result is determined, immediately after the measurement, the system is in a state (or in a subspace of states) that is characterized by the outcome of the measurement. A second measurement would then give the same result with certainty. The state preparation procedure is described by a projection operator (projection postulate). As a special case, consider an observable A with discrete eigenvalues. If the measurement of A gives the value a (that is, one of the eigenvalues), then, immediately after the measurement, the system is in the eigenspace belonging to the eigenvalue a.

The description of a preparative measurement as a (non-unitary) projection is also called a *collapse* of the state. The projection postulate is a useful and simple mathematical model of a complicated physical process. It describes how the information content of the system changes as a result of the interaction between the system and an (often macroscopic) object (measurement device). Much more could be said about this in a more detailed quantum mechanical analysis (*decoherence theory*), which is, unfortunately, beyond the scope of this text.

A.8. Dirac's Formalism

In this book, I make use of Dirac's bra-ket formalism only for finite-dimensional systems (in the context of quantum information theory). The Dirac formalism provides some mathematical pitfalls, is difficult to justify rigorously, and is completely unusual in the mathematical literature about the subject. For these reasons, I avoid the Dirac formalism in the quantum theory of systems with infinite dimensional Hilbert spaces. Nevertheless, in

order to be able to read other books on quantum mechanics, one should know some basic facts about this formalism.

A vector ψ in Hilbert space is denoted by a *ket* symbol $|\psi\rangle$. Very often, if an eigenstate of the Hamiltonian is characterized by certain quantum numbers, then the list of these quantum numbers is written instead of the ψ. For example, the eigenstates of the hydrogen atom $\psi_{n,\ell,m}$ would be written as $|n,\ell,m\rangle$ (it is more common to use the principal quantum number N instead of n). The linear forms in the Hilbert space are denoted by a *bra* symbol $\langle\phi|$. Linear forms (sometimes also called linear functionals) are continuous linear maps from the Hilbert space into the complex numbers. The vector space of linear forms on a Hilbert space can be identified in a natural way with the Hilbert space itself (Riesz lemma). Hence, there is actually no need to distinguish in notation between the linear forms and the vectors in the Hilbert space. As a linear form, a vector ϕ acts on the Hilbert space by associating to each vector ψ the scalar product $\langle\phi,\psi\rangle$. In Dirac's formalism, the application of a bra $\langle\phi|$ to a ket $|\psi\rangle$ is written as $\langle\phi|\psi\rangle$ and gives again just the scalar product of two vectors ϕ and ψ.

Let us list the entries in our dictionary:

$$\psi \quad\longleftrightarrow\quad |\psi\rangle \qquad \text{(vector in Hilbert space)} \tag{A.22}$$

$$\langle\psi\,,\cdot\,\rangle \quad\longleftrightarrow\quad \langle\psi| \qquad \text{(linear form in Hilbert space)} \tag{A.23}$$

$$\langle\phi\,,\psi\rangle \quad\longleftrightarrow\quad \langle\phi\,|\,\psi\rangle \qquad \text{(scalar product)} \tag{A.24}$$

The action of a linear operator A on a vector ψ is written as $A|\psi\rangle$ in Dirac's formalism. Moreover, one introduces an action to the left:

$$A\,\psi \quad\longleftrightarrow\quad A\,|\psi\rangle \qquad \text{(applying an operator)}, \tag{A.25}$$

$$A^\dagger\,\psi \quad\longleftrightarrow\quad \langle\psi|\,A \qquad \text{(adjoint operator)}. \tag{A.26}$$

Here is a little problem with Dirac's formalism: We have the identification

$$\langle\phi\,,A\psi\rangle \quad\longleftrightarrow\quad \langle\phi|\,A\,|\psi\rangle \tag{A.27}$$

for all $\phi\in\mathfrak{H}$ and $\psi\in\mathfrak{D}(A)$. On the other hand, one also has

$$\langle A^\dagger\phi\,,\psi\rangle \quad\longleftrightarrow\quad \langle\phi|\,A\,|\psi\rangle, \tag{A.28}$$

but this time for $\phi\in\mathfrak{D}(A^\dagger)$ and all $\psi\in\mathfrak{H}$. Hence, Dirac's formalism does not encourage a correct treatment of the domains of operators. But, a careless manipulation of expressions is dangerous as physical properties depend on the domains of the operators.

There is, however, no problem with bounded operators (in particular, with operators in finite-dimensional Hilbert spaces). In this case, the domain

can always be assumed to be the whole Hilbert space. For example, a one-dimensional projection operator can be written in a very intuitive form

$$\psi \langle \psi \, , \cdot \rangle \quad \longleftrightarrow \quad |\psi\rangle\langle\psi| \quad \text{(projection operator)}. \tag{A.29}$$

Hence, for example, the expectation value of a projection operator $P = |\psi\rangle\langle\psi|$ in a state $|\phi\rangle$ is calculated as

$$\langle P \rangle_\phi = \langle\phi|P|\phi\rangle = \langle\phi|\psi\rangle\langle\psi|\phi\rangle = |\langle\phi|\psi\rangle|^2. \tag{A.30}$$

The orthonormality and completeness of an orthonormal basis $\{\psi_j\}$ just becomes

$$\langle \psi_j \,|\, \psi_k \rangle = \delta_{jk}, \qquad \sum_k |\psi_k\rangle\langle\psi_k| = \mathbf{1}, \tag{A.31}$$

and the expansion of a vector ϕ in the orthonormal basis can be done as follows

$$|\phi\rangle = \mathbf{1}\,|\phi\rangle = \sum_k |\psi_k\rangle\langle\psi_k|\phi\rangle. \tag{A.32}$$

The Dirac notation comes in handy whenever we want to specify a state in terms of a complete set of quantum numbers or in terms of a graphical symbol. It is notationally more convenient to write $|n, \ell, m_\ell, m_s\rangle$ for an eigenvector of the Coulomb problem with spin, than to attach the quantum numbers as indices, as in ψ_{n,ℓ,m_ℓ,m_s}.

Another feature of Dirac's formalism is that it introduces the symbols $|x\rangle$ and $|k\rangle$ for the generalized eigenvectors of the position and momentum operators, respectively. The object $|x\rangle$ corresponds to Dirac's delta function $\delta(\cdot - x)$, and $|k\rangle$ is just the plain wave $\exp(\mathrm{i}kx)$. Both do not belong to the particle's Hilbert space. One often writes

$$\psi(x) \quad \longleftrightarrow \quad \langle x|\psi\rangle \quad \text{(values in position space)} \tag{A.33}$$

$$\hat{\psi}(k) \quad \longleftrightarrow \quad \langle k|\psi\rangle \quad \text{(values in momentum space)}. \tag{A.34}$$

The ket-vector $|x\rangle$ is in fact a distribution. Like the plane wave $|k\rangle$, it cannot be interpreted in terms of position probabilities. A mathematically rigorous justification of these elements of Dirac's formalism can be done in the framework of Gelfand triples. This is, however, highly nontrivial.

It is considered an advantage of the Dirac formalism that it deals with objects $|\psi\rangle$ in an abstract Hilbert space, and position and momentum representation are just different concrete realizations. In that sense, the values $\psi(x)$ are just the components of ψ in the basis of position eigenfunctions, in the same sense as $\psi_{n\ell m} = \langle n, \ell, m|\psi\rangle$ are the components of ψ in the eigenbasis of the hydrogen Hamiltonian. However, you should note that all Hilbert spaces are isomorphic to L^2 in the infinite dimensional case and to $\mathbb{C}^n$ in the finite-dimensional case. Hence, from a mathematical point of view, the Dirac formalism is by no means more general than the usual formalism.

Appendix B

Perturbation of Eigenvalues

Very often in physics and mathematics, we need to solve a problem that differs only slightly from a problem that has already been solved. A typical situation is the following: Assume that we know the eigenvalues and eigenfunctions of a Hamiltonian with potential V. Now, we change the potential a bit by adding a small perturbation W. What can we say about the eigenvalues and eigenfunctions of the Hamiltonian with potential $V + W$?

B.1. Introduction

Consider a self-adjoint operator $H^{(0)}$ (the *unperturbed operator*) in some Hilbert space $\mathfrak{H}$. Usually, $H^{(0)}$ is the Hamiltonian of a physical system whose solutions are well-known. In the following, the superscript "(0)" will always refer to this unperturbed system.

We are interested in the *perturbed operator*

$$H(\gamma) = H^{(0)} + \gamma\, H^{(1)}, \tag{B.1}$$

where $H^{(1)}$ is some other self-adjoint operator, called the *perturbation*, and γ is some (small) parameter, called the *coupling constant* of the perturbation, or the *perturbation parameter*.

Let $\psi_n^{(0)}$ be an eigenvector belonging to the nth discrete[1] eigenvalue $E_n^{(0)}$ of $H^{(0)}$

$$H^{(0)}\,\psi_n^{(0)} = E_n^{(0)}\,\psi_n^{(0)}. \tag{B.2}$$

We look for a solution of the *perturbed eigenvalue problem*

$$H(\gamma)\,\psi(\gamma) = E(\gamma)\,\psi(\gamma). \tag{B.3}$$

Our notation indicates that the perturbed eigenvalues and eigenvectors (if they exist) will somehow depend on γ. We want to address the following questions:

[1] An eigenvalue is *isolated* if it is contained in an interval that contains no other part of the spectrum of the operator. An eigenvalue is *discrete* if it is isolated and has a finite degree of degeneracy.

- Can we find a solution $E_n(\gamma)$ and $\psi_n(\gamma)$ of the perturbed eigenvalue problem, such that $E_n(0) = E_n^{(0)}$ and $\psi_n(0) = \psi_n^{(0)}$?
- Will the perturbed eigenvalues and eigenvectors depend smoothly on γ?

The answers to these questions will depend on the "smallness" of the perturbation $\gamma H^{(1)}$, which is not only determined by the smallness of γ but also by the mathematical properties of $H^{(1)}$ and $H^{(0)}$.

EXAMPLE B.1. As a simple example, consider the Hamiltonian of a qubit with a magnetic field in the z-direction. In the standard representation, it can be written as $H^{(0)} = \omega\sigma_3$, with some constant ω describing the strength of the magnetic field and of the magnetic moment (for electrons, $\omega = \mu_B B_z$). Its eigenvalues are $\pm\omega$, and the eigenvectors are the spinors of the standard basis, ψ_+ and ψ_-. The perturbation is a small magnetic field in the x-direction. We assume that the potential energy due to the perturbation is $\gamma\sigma_1$. Hence, the perturbed Hamiltonian (in the standard representation) is of the form

$$H(\gamma) = \omega\sigma_3 + \gamma\sigma_1 = \begin{pmatrix} \omega & \gamma \\ \gamma & -\omega \end{pmatrix}. \tag{B.4}$$

This matrix has the eigenvalues $\pm\sqrt{\omega^2+\gamma^2}$. We can expand these eigenvalues in power series with respect to the perturbation parameter γ. For example, the positive eigenvalue is (assuming $\omega > 0$)

$$E(\gamma) = \sqrt{\omega^2+\gamma^2} = \omega + \frac{1}{2\omega}\,\gamma^2 - \frac{1}{8\omega^3}\,\gamma^4 + \dots . \tag{B.5}$$

Due to the branch-cut singularities of the square-root (for complex $\gamma = \pm i\omega$), this power series has a finite radius of convergence even for real γ (in this example, it is actually a power series in γ^2, because the coefficients of the odd powers of γ vanish). For the corresponding normalized eigenvector,

$$\begin{aligned}\psi(\gamma) &= \frac{1}{\sqrt{2\left(\omega^2+\omega\sqrt{\omega^2+\gamma^2}+\gamma^2\right)}} \begin{pmatrix} \omega+\sqrt{\omega^2+\gamma^2} \\ \gamma \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 1/(2\omega) \end{pmatrix} - \gamma^2 \begin{pmatrix} 1/(8\omega^2) \\ 0 \end{pmatrix} - \gamma^3 \begin{pmatrix} 0 \\ 3/(16\omega^3) \end{pmatrix} + \dots ,\end{aligned} \tag{B.6}$$

we also obtain a Taylor series in γ that converges in some neighborhood of $\gamma = 0$.

In the next example, the addition of the perturbation changes the domain of the unperturbed operator. In this case, perturbation theory becomes a delicate task.

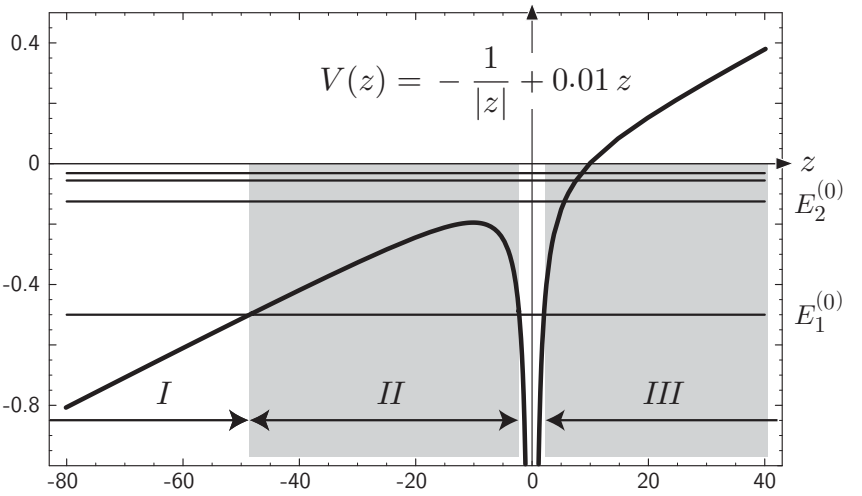


FIGURE B.1. The sum of a Coulomb potential and a linear potential has no bound states, because a particle would tunnel through a classically forbidden region as long as the region has a finite diameter. The gray regions II and III are the classically forbidden regions for the ground-state energy $E_1^{(0)}$ of the unperturbed Coulomb problem.

EXAMPLE B.2. **The Stark effect**: We consider a hydrogen atom with Hamiltonian (in atomic units)

$$H^{(0)} = -\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|} \tag{B.7}$$

and perturb it with a small electrostatic field. The external field is a constant homogeneous field in the z-direction, $\mathbf{E} = \gamma(0, 0, 1)$. This adds a potential energy that depends linearly on the z-coordinate.

$$H(\gamma) = -\frac{1}{2}\,\Delta - \frac{1}{|\mathbf{x}|} + \gamma\, z. \tag{B.8}$$

We know that the unperturbed hydrogen atom has infinitely many isolated eigenvalues. The operator (B.8), however, has no eigenvalues at all, as soon as $\gamma \neq 0$.

We want to explain the vanishing of bound states with the help of Figure B.1. It shows the potential $-1/|\mathbf{x}| + 0.01z$ along the z-axis (where $|\mathbf{x}| = |z|$). Because of the linear perturbation, the Coulomb potential essentially becomes a potential well with a finite potential barrier on one side. The figure shows a local maximum of the potential energy $V(z)$ at $z = -10$ (in three dimensions, this is actually a saddle point).

The horizontal lines in Figure B.1 are the energy levels of the unperturbed system (Coulomb problem in atomic units). We see that the excited levels $E_2^{(0)}, E_3^{(0)}, \ldots$ are all above the local maximum of the perturbed potential $V(z)$ at $z = -10a_0$. Electrons with energies higher than the maximal value of the potential barrier will escape to the region behind the barrier, that is, toward $z \to -\infty$. This is true for quantum mechanics as well as for classical mechanics.

Now, consider an electron in the ground state of the unperturbed system. It has the energy $E_1^{(0)} = -1/2$, and its wave function is centered around the origin (essentially in the white region between II and III in Figure B.1). This wave function decreases exponentially for $|\mathbf{x}| \to \infty$, but it is nonzero everywhere in space. As soon as the perturbation is "turned on," the electron is separated by a potential barrier of finite width (region II) from the region I. Given enough time, the electron will tunnel through this barrier (a classically forbidden region) into the region I where it accelerates toward $-\infty$ (this is essentially a free fall in a linear potential).

B.2. Rayleigh-Schrödinger series

In this section, we treat a special case, where the perturbed eigenvalue depends analytically on the perturbation parameter γ. The assumption of analyticity means that the perturbed quantities have a power series expansion in a neighborhood of $\gamma = 0$.

Assumption 1: For small γ, the perturbed eigenvalue and the perturbed eigenvector are given by power series in γ,

$$E_n(\gamma) = E_n^{(0)} + \gamma\, E_n^{(1)} + \gamma^2\, E_n^{(2)} + \ldots\,, \tag{B.9}$$

$$\psi_n(\gamma) = \psi_n^{(0)} + \gamma\, \psi_n^{(1)} + \gamma^2\, \psi_n^{(2)} + \ldots\,. \tag{B.10}$$

Actually, this is a rather strong assumption that is not justified in some physically interesting situations, like in Example B.2 above. Under this assumption, we may insert the expressions (B.9) and (B.10) into the eigenvalue equation (B.3),

$$\begin{aligned}(H^{(0)} + \gamma\, H^{(1)})\,(\psi_n^{(0)} + \gamma\, \psi_n^{(1)} + \ldots) \\ = (E_n^{(0)} + \gamma\, E_n^{(1)} + \ldots)\,(\psi_n^{(0)} + \gamma\, \psi_n^{(1)} + \ldots),\end{aligned} \tag{B.11}$$

and compare the coefficients of the equal powers of γ on both sides of this equation. This gives

$$\gamma^0: \qquad H^{(0)}\,\psi_n^{(0)} = E_n^{(0)}\,\psi_n^{(0)}, \tag{B.12}$$

$$\gamma^1: \qquad H^{(0)}\,\psi_n^{(1)} + H^{(1)}\,\psi_n^{(0)} = E_n^{(0)}\,\psi_n^{(1)} + E_n^{(1)}\,\psi_n^{(0)}, \tag{B.13}$$

$$\gamma^2: \qquad H^{(0)}\,\psi_n^{(2)} + H^{(1)}\,\psi_n^{(1)} = E^{(0)}\,\psi_n^{(2)} + E^{(1)}\,\psi_n^{(1)} + E^{(2)}\,\psi_n^{(0)} \tag{B.14}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

The first equation (B.12) is identical with the unperturbed eigenvalue equation (B.2).

Let us now consider the special case that the unperturbed eigenvalue $E_n^{(0)}$ is non-degenerate. Hence, $\psi_n^{(0)}$ is the only eigenvector belonging to this eigenvalue. We may form the scalar product of $\psi_n^{(0)}$ with both sides of (B.13). For the first term on the left side of this equation we obtain

$$\langle\, \psi_n^{(0)}\,,\, H^{(0)}\,\psi_n^{(1)}\,\rangle = \langle\, H^{(0)}\,\psi_n^{(0)}\,,\, \psi_n^{(1)}\,\rangle = E_n^{(0)}\,\langle\, \psi_n^{(0)}\,,\, \psi_n^{(1)}\,\rangle. \tag{B.15}$$

The first term on the right side of (B.13) gives the same expression,

$$\langle\, \psi_n^{(0)}\,,\, E_n^{(0)}\,\psi_n^{(1)}\,\rangle = E_n^{(0)}\,\langle\, \psi_n^{(0)}\,,\, \psi_n^{(1)}\,\rangle. \tag{B.16}$$

Therefore, (B.13) simplifies to

$$E_n^{(1)} = \langle\, \psi_n^{(0)}\,,\, H^{(1)}\,\psi_n^{(0)}\,\rangle. \tag{B.17}$$

This is the most important formula in this section. We conclude that up to terms of second order in γ, the perturbed eigenvalue is just the expectation value of $H(\gamma)$ in the unperturbed eigenstate:

$$E_n(\gamma) = \langle\, \psi_n^{(0)}\,,\, H(\gamma)\,\psi_n^{(0)}\,\rangle + \mathcal{O}(\gamma^2). \tag{B.18}$$

Here, the expression $\mathcal{O}(\gamma^2)$ (*Landau symbol*) is a handy abbreviation for "a term that behaves like γ^2 in the limit under consideration." It means that $\lim_{\gamma\to 0}\mathcal{O}(\gamma^2)/\gamma^2$ is a finite nonzero constant.

In order to compute the first-order perturbation of the eigenvector, we make the following assumption:

Assumption 2: The first-order perturbation $\psi_n^{(1)}$ is contained in the subspace spanned by the eigenvectors of $H^{(0)}$.

Given this assumption, $\psi_n^{(1)}$ can be written as

$$\psi_n^{(1)} = \sum_k a_{n,k}\,\psi_k^{(0)} \tag{B.19}$$

where $\{\psi_k^{(0)} \mid k = 1, 2, \ldots\}$ is the orthonormal set of eigenvectors of the unperturbed operator $H^{(0)}$, which may be finite or infinite. We insert the

expression (B.19) into (B.13). A scalar multiplication of this equation by $\psi_m^{(0)}$ with $m \neq n$ then gives

$$E_m^{(0)}\, a_{n,m} + \left\langle\, \psi_m^{(0)}\,,\, H^{(1)}\, \psi_n^{(0)} \,\right\rangle = E_n^{(0)}\, a_{n,m}. \tag{B.20}$$

The term with $E_n^{(1)}$ vanishes, because $\langle \psi_m^{(0)}, \psi_n^{(0)} \rangle = 0$. From $m \neq n$ and the fact that $E_n^{(0)}$ is non-degenerate it follows that $E_n^{(0)} \neq E_m^{(0)}$. Hence, we find

$$a_{n,m} = \frac{\left\langle\, \psi_m^{(0)}\,,\, H^{(1)}\, \psi_n^{(0)} \,\right\rangle}{E_n^{(0)} - E_m^{(0)}} \qquad (\text{for } m \neq n). \tag{B.21}$$

This finally gives the first-order correction of the eigenvector:

$$\psi_n^{(1)} = \sum_k a_{n,k}\, \psi_k^{(0)} = a_{n,n}\, \psi_n^{(0)} + \sum_{k \neq n} \frac{\left\langle\, \psi_k^{(0)}\,,\, H^{(1)}\, \psi_n^{(0)} \,\right\rangle}{E_n^{(0)} - E_k^{(0)}}\, \psi_k^{(0)}. \tag{B.22}$$

The still unknown coefficient $a_{n,n}$ can be determined from a normalization condition. It is most convenient to require that

$$\langle \psi_n^{(0)}, \psi_n(\gamma) \rangle = 1. \tag{B.23}$$

(Hence, in general, $\|\psi_n(\gamma)\| \neq 1$). This requirement implies that for all λ in a neighborhood of $\lambda = 0$,

$$\langle \psi_n^{(0)}, \psi_n(\gamma) \rangle = \langle \psi_n^{(0)}, \psi_n^{(0)} \rangle + \gamma \langle \psi_n^{(0)}, \psi_n^{(1)} \rangle + \gamma^2 \langle \psi_n^{(0)}, \psi_n^{(2)} \rangle + \ldots = 1, \tag{B.24}$$

and from $\langle \psi_n^{(0)}, \psi_n^{(0)} \rangle = 1$ we find

$$\langle \psi_n^{(0)}, \psi_n^{(1)} \rangle = \langle \psi_n^{(0)}, \psi_n^{(2)} \rangle = \ldots = 0. \tag{B.25}$$

This result together with (B.22) implies immediately that $a_{n,n} = 0$. We collect out results in the following box:

First order perturbation (non-degenerate eigenvalue):

Let $E_n^{(0)}$ be an isolated, non-degenerate eigenvalue of $H^{(0)}$,

$$H^{(0)}\,\psi_n^{(0)} = E_n^{(0)}\,\psi_n^{(0)}. \tag{B.26}$$

Under the assumptions stated above, the perturbed operator $H(\gamma) = H^{(0)} + \gamma\,H^{(1)}$ has for small γ an eigenvalue

$$E_n(\gamma) = E_n^{(0)} + \gamma\,E_n^{(1)} + \gamma^2\,E_n^{(2)} + \ldots \tag{B.27}$$

with

$$E_n^{(1)} = \left\langle\, \psi_n^{(0)}\,,\, H^{(1)}\,\psi_n^{(0)} \,\right\rangle. \tag{B.28}$$

Assuming that the corresponding eigenvector $\psi_n(\gamma)$ is normalized according to

$$\langle \psi_n^{(0)}, \psi_n(\gamma)\rangle = 1, \tag{B.29}$$

we find the first-order correction

$$\psi_n^{(1)} = \sum_{k\neq n} \frac{\left\langle\, \psi_k^{(0)}\,,\, H^{(1)}\,\psi_n^{(0)} \,\right\rangle}{E_n^{(0)} - E_k^{(0)}}\,\psi_k^{(0)}. \tag{B.30}$$

We note that formulas can also be obtained for the higher-order perturbations, for example,

$$E_n^{(2)} = \sum_{k\neq n} \frac{\left|\left\langle\, \psi_k^{(0)}\,,\, H^{(1)}\,\psi_n^{(0)} \,\right\rangle\right|^2}{E_n^{(0)} - E_k^{(0)}}. \tag{B.31}$$

These formulas become more nasty with increasing order, and today they have less practical importance due to the availability of efficient numerical algorithms for computing eigenvalues and eigenfunctions.

B.3. Degenerate eigenvalues

Here, we treat perturbation theory for degenerate energy levels. We assume that the nth discrete eigenvalue $E_n^{(0)}$ of the unperturbed system is d-fold degenerate. The corresponding orthonormal eigenvectors are denoted by $\psi_{n,j}^{(0)}$, $j = 1, 2, \ldots, d$. In order to compute the first-order corrections according to the perturbation $H^{(1)}$, we choose an unperturbed eigenvector $\psi_n^{(0)}$ in the eigenspace belonging to $E_n^{(0)}$. This eigenvector may be expanded in the given orthonormal basis,

$$\psi_n^{(0)} = \sum_{j=1}^{d} c_j\,\psi_{n,j}^{(0)}. \tag{B.32}$$

We may insert this expression into the first-order equation (B.13) and take the inner product with the vector $\psi^{(0)}_{n,k}$. The terms involving $\psi^{(1)}_n$ on each side cancel and, exploiting orthonormality, we obtain

$$\sum_j c_j \left\langle \psi^{(0)}_{n,k} , H^{(1)} \psi^{(0)}_{n,j} \right\rangle = E^{(1)}_n c_k. \tag{B.33}$$

The $d \times d$ matrix A with the components

$$A_{kj} = \left\langle \psi^{(0)}_{n,k} , H^{(1)} \psi^{(0)}_{n,j} \right\rangle \tag{B.34}$$

is Hermitian if, as it is normally the case, the perturbation $H^{(1)}$ is a symmetric operator. This follows from

$$\left\langle \psi^{(0)}_{n,k} , H^{(1)} \psi^{(0)}_{n,j} \right\rangle = \left\langle H^{(1)} \psi^{(0)}_{n,k} , \psi^{(0)}_{n,j} \right\rangle = \overline{\left\langle \psi^{(0)}_{n,j} , H^{(1)} \psi^{(0)}_{n,k} \right\rangle}. \tag{B.35}$$

(Here, as usual, the overbar signifies complex conjugation.) Note that (B.33) is just the eigenvalue equation with the matrix A,

$$\sum_j A_{kj} c_j = E^{(1)}_n c_k. \tag{B.36}$$

The column vector of the coefficients $(c_1, c_2, \ldots, c_d)$ thus has to be chosen as an eigenvector of the matrix A. The Hermitian $d \times d$ matrix A has exactly d orthogonal eigenvectors $\mathbf{c}_\nu$, $\nu = 1, \ldots d$. For each eigenvector, we obtain the corresponding eigenvalue $E^{(1)}_{n,\nu}$ as a possible solution of (B.33). A Hermitian $d \times d$ matrix has only real eigenvalues, some of which may be equal.

Perturbation of a degenerate eigenvalue:

Assume that the nth discrete eigenvalue $E^{(0)}_n$ is d-fold degenerate. Then, there are d perturbed eigenvalues

$$E_{n,\nu}(\gamma) = E^{(0)}_n + \gamma E^{(1)}_{n,\nu} + \ldots, \quad \nu = 1, 2, \ldots, d, \tag{B.37}$$

all analytic in γ in a neighborhood of $\gamma = 0$, but not necessarily different from another. The first-order corrections $E^{(1)}_{n,\nu}$ are obtained as the eigenvalues of the Hermitian $d \times d$ matrix with entries

$$\left\langle \psi^{(0)}_{n,k} , H^{(1)} \psi^{(0)}_{n,j} \right\rangle. \tag{B.38}$$

B.4. Alkali atoms

A canonical example is the following model of alkali atoms. The spectra of alkali atoms (Li, Na, K, ...) are very similar to the spectra of hydrogen atoms. Alkali atoms have a single *valence electron* and a *core* consisting of the nucleus and $Z-1$ *inner electrons* (where Z is the atomic number). The core is a stable configuration that is little perturbed by the presence of the valence electron. Hence, one may assume that the valence electron moves in an electrostatic field that consists of the Coulomb field of the nucleus and an averaged field of the core electrons (which is repulsive). The field of the core electrons shields the field of the nucleus. Outside the core, where the valence electron is found with the highest probability, the potential energy is (using atomic units)

$$V(\mathbf{x}) \approx -1/|\mathbf{x}|, \quad \text{for } |\mathbf{x}| \text{ large.} \tag{B.39}$$

Very close to the nucleus, we expect

$$V(\mathbf{x}) \approx -Z/|\mathbf{x}|, \quad \text{for } |\mathbf{x}| \text{ small.} \tag{B.40}$$

A simple model for the potential energy of the valence electron is

$$V(\mathbf{x}) = -\frac{1}{|\mathbf{x}|} + V_c(\mathbf{x}), \tag{B.41}$$

where V_c is only nonzero in the neighborhood of the nucleus, and negative.

$$V_c(\mathbf{x}) \to \begin{cases} \dfrac{1-Z}{|\mathbf{x}|}, & \text{as } |\mathbf{x}| \to 0, \\ 0, & \text{as } |\mathbf{x}| \to \infty. \end{cases} \tag{B.42}$$

The contribution of the potential V_c is only significant if the valence electron enters deep into the core. So we can treat $V_c = H^{(1)}$ as a perturbation of the Coulomb Hamiltonian

$$H(\gamma) = -\frac{1}{2}\Delta - \frac{1}{|\mathbf{x}|} + \gamma V_c(\mathbf{x}). \tag{B.43}$$

Actually, we are interested in the Hamiltonian with $\gamma = 1$, which is not small. Therefore, one should not expect quantitatively accurate results from first-order perturbation theory. But we can expect some qualitative results. In an eigenspace belonging to the eigenvalue $E_n^{(0)}$, we choose the basis consisting of the eigenvectors $\psi_{n,\ell,m}^{(0)}$ of the unperturbed Coulomb problem. The eigenvalues and eigenvectors of the Coulomb problem have been described before, see Chapter 2, Eqs. (2.67) and (2.69). Here, we attached the superscript "(0)" to indicate that these quantities refer to the unperturbed

problem. According to (B.38), we consider the matrix

$$\langle\, \psi^{(0)}_{n,\ell,m}\,,\, V_c\, \psi^{(0)}_{n,\ell',m'}\,\rangle = \int_{\mathbb{R}^3} \overline{\psi^{(0)}_{n,\ell,m}(\mathbf{x})}\, V_c(\mathbf{x})\, \psi^{(0)}_{n,\ell',m'}(\mathbf{x})\, d^3x. \tag{B.44}$$

In a first approximation, we may assume that V_c is spherically symmetric, $V_c(\mathbf{x}) = V_c(r)$. Then, we may perform the integration in (B.44) using spherical coordinates. Using the expression $\psi^{(0)}_{n,\ell,m} = (1/r)\, f_{n,\ell}(r)\, Y^m_\ell(\vartheta,\varphi)$ for the Coulomb eigenfunction in spherical coordinates, the matrix element above becomes

$$\int_0^\infty V_c(r)\, f_{n,\ell}(r)\, f_{n,\ell'}(r)\, dr \int_{S^2} \overline{Y^m_\ell(\vartheta,\varphi)}\, Y^{m'}_{\ell'}(\vartheta,\varphi)\, d\Omega \tag{B.45}$$

(with $d\Omega = \sin\vartheta\, d\vartheta\, d\varphi$). The spherical harmonics form an orthonormal set of functions on the unit sphere S^2, hence

$$\langle\, \psi^{(0)}_{n,\ell,m}\,,\, V_c\, \psi^{(0)}_{n,\ell',m'}\,\rangle = \delta_{\ell\ell'}\, \delta_{mm'} \int_0^\infty V_c(r)\, |f_{n,\ell}(r)|^2\, dr\,. \tag{B.46}$$

This defines a diagonal matrix. We conclude that the eigenvalue $E^{(0)}_n$ splits into at most n different eigenvalues $E_{n,\ell}(\gamma)$, with $\ell = 0, 1, \ldots, n-1$. These eigenvalues are still degenerate, because a spherically symmetric potential always leads to energy eigenvalues that are degenerate with respect to the quantum number m (see Section 1.10.1), and the degree of degeneracy is given by $2\ell + 1$.

The radial integral in (B.46) measures the shift of energy in first-order. Because V_c is everywhere negative, the integral is negative, hence all levels of the alkali atom are below the corresponding level of the hydrogen atom. Moreover, V_c is small for large r and gives significant contributions only for small r. The radial position probability amplitude $f_{n,\ell}$ is concentrated at small values of r only if ℓ is small. For larger values of ℓ, the integral must be smaller. Hence, we expect that in first-order perturbation theory,

$$E^{(1)}_{n,\ell} > E^{(1)}_{n,\ell'}, \quad \text{if } \ell < \ell'. \tag{B.47}$$

B.5. Ground state of helium

The helium atom consists of a nucleus with charge $2e$ surrounded by two electrons. Assuming that the nucleus has infinite mass, we obtain the Hamiltonian

$$H = \frac{1}{2\mathrm{m}_e}\,(\mathbf{p}_1^2 + \mathbf{p}_2^2) - 2\gamma_0 \left(\frac{1}{|\mathbf{x}_1|} + \frac{1}{|\mathbf{x}_2|}\right) + \gamma\, \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|}. \tag{B.48}$$

Here, $\mathbf{p}_j$ are the momenta and $\mathbf{x}_j$ are the position vectors for the two electrons, and γ_0 is the hydrogen coupling constant

$$\gamma_0 = \frac{e^2}{4\pi\epsilon_0}. \tag{B.49}$$

The last summand in (B.48) is interpreted as a perturbation,

$$H^{(1)} = \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|}. \tag{B.50}$$

It describes the repulsion between the two electrons. γ is the perturbation parameter and the physical helium atom is obtained for $\gamma = \gamma_0$.

The unperturbed Hamiltonian

$$H^{(0)} = \frac{1}{2\mathrm{m}_e}\,\mathbf{p}_1^2 - \frac{2\gamma_0}{|\mathbf{x}_1|} + \frac{1}{2\mathrm{m}_e}\,\mathbf{p}_2^2 - \frac{2\gamma_0}{|\mathbf{x}_2|} \tag{B.51}$$

is a sum of two independent single-electron Coulomb Hamiltonians. Its ground state is the product of two hydrogenic ground states,

$$\psi^{(0)}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi}\left(\frac{2}{a_0}\right)^3 \exp\Big(-\frac{2}{a_0}\,(|\mathbf{x}_1| + |\mathbf{x}_2|)\Big). \tag{B.52}$$

Here, $a_0/2$ is the Bohr radius of a hydrogen-like atom with atomic number $Z = 2$.

In our consideration, we neglect contributions to the energy due to the spin of the electrons. Hence, the required antisymmetry of the two-electron wave function can be taken into account by an antisymmetric spin-part. The fermionic ground state of the Hamiltonian (B.51) in the Hilbert space of antisymmetric spinor-valued two-particle wave functions is

$$\begin{aligned}\psi(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{\sqrt{2}}\left((\phi_1(\mathbf{x}_1)\psi_+) \otimes (\phi_2(\mathbf{x}_2)\psi_-) - (\phi_1(\mathbf{x}_1)\psi_-) \otimes (\phi_2(\mathbf{x}_2)\psi_+)\right) \\ &= \phi_1(\mathbf{x}_1)\,\phi_2(\mathbf{x}_2)\frac{1}{\sqrt{2}}\,(\psi_+ \otimes \psi_- - \psi_- \otimes \psi_+) \\ &= \phi_1(\mathbf{x}_1)\,\phi_2(\mathbf{x}_2)\,\psi_o^-. \end{aligned} \tag{B.53}$$

Here, ψ_o^- is the odd-parity Bell state defined in Eq. (5.30).

The ground-state energy of the Hamiltonian (B.51) is the sum of the ground-state energies of two independent Coulomb Hamiltonians with atomic number $Z = 2$,

$$E^{(0)} = -2\,\frac{Z^2\gamma_0}{2a_0} = -4\,\frac{\gamma_0}{a_0}. \tag{B.54}$$

If we compare this with the experimental value

$$E_{\mathrm{exp}} = -2.92\,\frac{\gamma_0}{a_0}, \tag{B.55}$$

we see that the zero-order approximation is not very good.

In order to compute the first-order approximation $\langle \psi^{(0)}, H^{(1)}\, \psi^{(0)} \rangle$, we need the integral

$$\int_{\mathbb{R}^6} \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|}\, \exp\bigl(-c\,(|\mathbf{x}_1| + |\mathbf{x}_2|)\bigr)\, d^3x_1\, d^3x_2 = \frac{20\pi^2}{c^5}. \tag{B.56}$$

Using this formula, one finally obtains

$$\langle \psi^{(0)}, H^{(1)}\, \psi^{(0)} \rangle = \frac{5}{4a_0}. \tag{B.57}$$

Hence,

$$E(\gamma) \approx E^{(0)} + \gamma E^{(1)} = -4\,\frac{\gamma_0}{a_0} + \frac{5}{4}\,\frac{\gamma}{a_0}. \tag{B.58}$$

Setting $\gamma = \gamma_0$, we obtain the result

$$E(\gamma) \approx -2.75\,\frac{\gamma_0}{a_0}, \tag{B.59}$$

which is already rather close to the observed value (B.55).

Appendix C

Special Topic: Analytic Perturbation Theory

In this appendix, we sketch some ideas of analytic perturbation theory, which is the mathematical basis of the results in Appendix B. In particular, we show how the analyticity of the eigenvalues is related to the analyticity of the resolvent of the perturbed Hamiltonian. More mathematical details and proofs of the quoted results can be found, for example, in the books [**4**] or [**7**].

C.1. Relative boundedness

In Section B.2, we needed some assumptions in order to compute perturbation series for eigenvalues. The mathematical condition that is implicit in these calculations is stated here for the sake of completeness:

Assumption on operator domains: The domain of the self-adjoint perturbation $H^{(1)}$ is larger than the domain of the unperturbed operator $H^{(0)}$.

This assumption guarantees that the perturbed operator $H(\gamma) = H^{(0)} + \gamma\, H^{(1)}$ is self-adjoint on the domain of the unperturbed operator for small enough values of γ. Hence, the perturbed eigenvector and its approximations are all in the domain of $H^{(0)}$, $H^{(1)}$, and $H(\gamma)$. The assumption on the domains implies, in particular, the following statement:

Relative boundedness condition: There are non-negative constants a and b such that for all vectors in the domain of $H^{(0)}$ the following inequality holds:

$$\|H^{(1)}\,\psi\| \leq a\,\|H^{(0)}\,\psi\| + b\,\|\psi\| \quad \text{for all } \psi \in \mathfrak{D}(H^{(0)}). \tag{C.1}$$

The analytic dependence of isolated eigenvalues on γ (Assumption 1) follows from this condition. For Schrödinger operators $H^{(0)} = H_0 + V(\mathbf{x})$ and $H^{(1)} = V_1(\mathbf{x})$, the relative boundedness condition is fulfilled (as a rule of thumb) whenever V_1 is less singular than V. That is, whenever $V_1(\mathbf{x})$ tends to infinity (for $\mathbf{x} \to \mathbf{x}_0$ or for $\mathbf{x} \to \infty$), then the unperturbed potential V should go to infinity at least as fast as V_1. If $H^{(0)}$ is a Coulomb Hamiltonian,

then $V_1(\mathbf{x}) = \gamma/|\mathbf{x}|$ is fine but $V_1(\mathbf{x}) = \gamma\,\mathbf{S}\cdot\mathbf{L}/|\mathbf{x}|^3$ (spin-orbit term) is too singular for analytic perturbation theory. If $H^{(0)}$ is a harmonic oscillator Hamiltonian, then $V_1(\mathbf{x}) = \gamma|\mathbf{x}|$ is fine but $V_1(\mathbf{x}) = \gamma\,|\mathbf{x}|^4$ (anharmonic oscillator) diverges too fast, as $|\mathbf{x}| \to \infty$. Analytic perturbation theory does not work whenever the perturbation is more singular than the unperturbed potential. Unfortunately, this is the case for many well-known examples including the Stark effect and the anharmonic oscillator. It is interesting that perturbation theory is not quite useless in these cases and that the perturbation formulas do retain a certain physical and mathematical meaning.

We note that the relative boundedness condition is central to many results in perturbation theory. For example, the *Kato-Rellich theorem* on self-adjointness: Let $H^{(0)}$ be self-adjoint. Assume that we can find $b > 0$ and $a < 1$ such that (C.1) holds. Then, $H(\gamma)$ is self-adjoint on the domain of $H^{(0)}$ for all γ with $\gamma a < 1$.

C.2. The resolvent

Today, analytic perturbation theory is less important as a practical tool for computing eigenvalues, but it is essential as a method to decide whether there are perturbed eigenvalues and if they depend on γ in an analytic way (power series).

A useful tool for perturbation theory is the *resolvent* of a self-adjoint operator H. The resolvent R is an operator valued function of a complex variable z. It is defined as the inverse of the operator $H - z\mathbf{1}$ for those $z \in \mathbb{C}$ for which the inverse is defined as a bounded operator. It is usually simply written as

$$R(z) = (H - z)^{-1}. \tag{C.2}$$

Note that $H - z\mathbf{1}$ is not invertible if z is an eigenvalue of H, because then $(H - z\mathbf{1})f = 0$ for some nonzero vector, that is, $H - z\mathbf{1}$ is not injective. The set of all $z \in \mathbb{C}$ for which the resolvent either does not exist, or is an unbounded operator, is called the *spectrum* $\sigma(H)$ of the operator H. In case of a self-adjoint operator, the spectrum is always subset of the real axis. The *resolvent set* $\rho(H)$ consists of all z that are not in the spectrum of H. For $z \in \rho(H)$, the operator $R(z)$ is defined everywhere on the Hilbert space and maps onto the domain of H. The resolvent set is open in $\mathbb{C}$, that is, for every point $z_0 \in \rho(H)$, there is a disk of radius ϵ around z_0 which also belongs to the resolvent set.

Whenever H has an eigenvalue E, then the eigenvector ψ is also an eigenvalue of $R(z)$,

$$R(z)\,\psi = \frac{1}{E - z}\,\psi \quad \text{if and only if } H\psi = E\psi. \tag{C.3}$$

The resolvent contains all information about the eigenvalues of the operator H. Consideration of the resolvent frees us from tedious technicalities (like domain questions), because it is a bounded operator that is everywhere defined.

The resolvent is an analytic function of z. This follows from the *first resolvent equation*

$$\frac{R(z) - R(z_0)}{z - z_0} = R(z)\, R(z_0). \tag{C.4}$$

(You can verify this equation by multiplying both sides from the left by $H - z$ and from the right by $H - z_0$.) Taking the limit as $z \to z_0$ on both sides of (C.4), we obtain

$$\frac{d}{dz}\, R(z)\Big|_{z_0} = R(z_0)^2. \tag{C.5}$$

Thus, the resolvent is differentiable, hence analytic. Analyticity means that there is a power series expansion for z in a neighborhood of z_0,

$$R(z) = \sum_{n=0}^{\infty} (z - z_0)^n\, R(z_0)^{n+1}. \tag{C.6}$$

This is called the *Neumann series* of the resolvent. As an analytic function, the resolvent fulfills an operator analog of Cauchy's formula,

$$\oint_\Gamma R(z)\, dz = 0 \tag{C.7}$$

whenever Γ is a closed curve that lies, together with its interior, entirely within the resolvent set of H. Now, assume that Γ encircles an isolated eigenvalue E of H, such that no other part of the spectrum is inside the curve. Then it can be shown that

$$P = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma R(z)\, dz \tag{C.8}$$

is the projection operator onto the eigenspace belonging to the eigenvalue E. The projection of H onto that eigenspace is

$$PHP = EP = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma z\, R(z)\, dz. \tag{C.9}$$

This operator can be represented by a $m \times m$ matrix if the degree of degeneracy of the eigenvalue E is m (the degree of degeneracy is the dimension of the eigenspace). If Γ encircles more than one eigenvalue, then P is the projection operator onto the subspace spanned by all eigenvectors belonging to the eigenvalues inside Γ.

C.3. Analytic dependence on the perturbation parameter

Next, we consider the behavior of the resolvent under a perturbation of the self-adjoint operator. Let $H(\gamma) = H^{(0)} + \gamma\, H^{(1)}$, and write

$$R^{(0)}(z) = (H^{(0)} - z)^{-1}, \qquad R(\gamma, z) = (H^{(0)} + \gamma\, H^{(1)} - z)^{-1}. \tag{C.10}$$

Then, we find the second resolvent equation

$$R(\gamma, z) = R^{(0)}(z) - \gamma\, R(\gamma, z)\, H^{(1)}\, R^{(0)}(z) \tag{C.11}$$

(which can be verified by multiplying from the left by $H - z$ and from the right by $H^{(0)} - z$.) One usually tries to solve this equation by iteration, writing

$$\begin{aligned} R(\gamma, z) &= R^{(0)}(z) - \gamma\, R^{(0)}(z)\, H^{(1)}\, R^{(0)}(z) \\ &\quad + \gamma^2\, R^{(0)}(z)\, H^{(1)}\, R^{(0)}(z)\, H^{(1)}\, R^{(0)}(z) + \dots \\ &= \sum_{n=0}^{\infty} \gamma^n\, R^{(n)}(z) \end{aligned} \tag{C.12}$$

where

$$R^{(n)}(z) = R^{(0)}(z)\, \big(H^{(1)}\, R^{(0)}(z)\big)^n, \quad n = 0, 1, 2 \dots . \tag{C.13}$$

These expressions are well defined whenever $H^{(1)}$ is well defined on the domain of $H^{(0)}$. For this, the relative boundedness condition (C.1) is needed. It implies that the operator $H^{(1)}\, R^{(0)}(z)$ has a finite norm. Hence, the power series (C.12) converges, whenever the absolute value of γ is small enough. The fact that there is a power series expansion of the resolvent in the perturbation parameter γ around $\gamma = 0$ shows that the resolvent is an analytic operator-valued function of γ. (In the mathematical literature, the family $H(\gamma)$ described here is called an analytic family of type A.) One can show that if z is in the resolvent set of $H(\gamma_0)$, then it is in the resolvent set of $H(\gamma)$ for γ in some neighborhood of γ_0.

Again, let Γ be a small circle around an isolated eigenvalue $E^{(0)}$ of $H^{(0)}$, and consider the m-dimensional projection operator

$$P^{(0)} = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma R^{(0)}(z)\, dz. \tag{C.14}$$

For γ small, the points on Γ are still in the resolvent set of $H(\gamma)$ and hence

$$P(\gamma) = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma R(\gamma, z)\, dz. \tag{C.15}$$

is the projection operator onto the eigenspaces of all eigenvalues of $H(\gamma)$ inside Γ. The total degree of degeneracy of these eigenvalues is m. From the

analyticity of $R(\gamma, z)$ in the parameter γ it follows that also $P(\gamma)$ is analytic in γ near $\gamma = 0$.

$$P(\gamma) = \sum_{n=0}^{\infty} \gamma^n P^{(n)}, \qquad P^{(n)} = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma R^{(n)}\, dz. \tag{C.16}$$

With the help of the projection operators $P(\gamma)$ and $P^{(0)} = P(0)$, we define the unitary operator

$$U(\gamma) = \big(\mathbf{1} - (P(\gamma) - P^{(0)})^2\big)^{-1/2} \big(P(\gamma)P^{(0)} + (\mathbf{1} - P(\gamma))(\mathbf{1} - P^{(0)})\big) \tag{C.17}$$

Note that $A \equiv (P(\gamma) - P^{(0)})^2$ commutes with $P(\gamma)$ and $P^{(0)}$. The operator A is bounded, and we have $\|A\| < 1$ if $|\gamma|$ is sufficiently small, because $\lim_{\gamma\to 0} P(\gamma) = P^{(0)}$ with respect to the operator norm. For $\|A\| < 1$ can define the inverse square root in Eq. (C.17) by the power series

$$(\mathbf{1} - A)^{-1/2} = \mathbf{1} + \tfrac{1}{2}\, A + \tfrac{3}{8}\, A^2 + \dots\,. \tag{C.18}$$

The operator valued function $U(\gamma)$ is called transformation function for $P(\gamma)$, because

$$U(\gamma)\, P^{(0)} = P(\gamma)\, U(\gamma), \tag{C.19}$$

that is, $U(\gamma)$ intertwines between the projection operators $P(\gamma)$ and $P^{(0)}$. Clearly, $U(\gamma)$ is analytic in γ around $\gamma = 0$, from the expansion of $P(\gamma)$ and Eq. (C.18), we obtain

$$U(\gamma)\, P^{(0)} = P^{(0)} + \gamma P_1 P^{(0)} + \gamma^2 (P_2 P^{(0)} + \tfrac{1}{2}\, P_1^2 P^{(0)}) + \dots \tag{C.20}$$

Similarly as in Eq. (C.9), we can define an analytic family of bounded self-adjoint operators

$$P(\gamma) H(\gamma) P(\gamma) = -\frac{1}{2\pi \mathrm{i}} \oint_\Gamma z\, R(\gamma, z)\, dz. \tag{C.21}$$

The range of the projection operator $P(\gamma)$ is generally different for each γ. But with the help of the transformation function, we can define an analytic family of bounded self-adjoint operators which is unitarily equivalent to (C.21) and which acts on one fixed subspace $P^{(0)}\mathfrak{H}$.

$$\begin{aligned} \tilde{H}(\gamma) &= U(\gamma)^{-1} P(\gamma) H(\gamma) P(\gamma) U(\gamma) \\ &= -\frac{1}{2\pi \mathrm{i}}\, P^{(0)} U(\gamma)^{-1} \oint_\Gamma z\, R(\gamma, z)\, dz\, U(\gamma) P^{(0)} \\ &= E^{(0)} P^{(0)} - \frac{1}{2\pi \mathrm{i}}\, P^{(0)} U(\gamma)^{-1} \oint_\Gamma (z - E^{(0)})\, R(\gamma, z)\, dz\, U(\gamma) P^{(0)}. \end{aligned} \tag{C.22}$$

Hence, the problem of finding the eigenvalues of $H(\gamma)$ near the eigenvalue $E^{(0)}$ of $H(0)$ is completely reduced to the diagonalization of a self-adjoint

operator in a finite dimensional Hilbert space. Hence, we can apply the following theorem of Rellich.

THEOREM C.1. *Let $\tilde{H}(\gamma)$ be a family of operators in a finite dimensional Hilbert space, such that $\tilde{H}(\gamma)$ is analytic in a neighborhood of $\gamma = 0$. Suppose that $\tilde{H}(\gamma)$ is self-adjoint for real γ and let $E^{(0)}$ be an eigenvalue of $\tilde{H}(0)$ whose degree of degeneracy is m. Then, there are $k < m$ distinct functions, $E_1(\gamma), \ldots, E_k(\gamma)$, which are analytic in γ in a neighborhood of $\gamma = 0$, and which are all eigenvalues of $\tilde{H}(\gamma)$. The degree of degeneracy of the eigenvalue $E_j(\gamma)$ is m_j, such that $\sum_{j=1}^{k} m_j = m$.*

We also note that the eigenvectors of $\tilde{H}(\gamma)$ are simply given by

$$\phi_l(\gamma) = U(\gamma)\phi_l^{(0)}, \quad l = 1, \ldots, m, \tag{C.23}$$

where the vectors $\phi_l^{(0)}$ form a basis of the eigenspace of $H^{(0)}$ belonging to the eigenvalue $E^{(0)}$. This result follows immediately from the fact that the unitary operator $U(\gamma)$ maps $P^{(0)}\mathfrak{H}$ onto $P(\gamma)\mathfrak{H}$. Analyticity of $\tilde{H}$ and its eigenvalues implies that there is an expansion of the form

$$\tilde{H}(\gamma) = \sum_{n=0}^{\infty} \gamma^n \, \tilde{H}^{(n)}, \qquad E_j(\gamma) = \sum_{n=0}^{\infty} \gamma^n \, E_j^{(n)}. \tag{C.24}$$

If $E^{(0)}$ is nondegenerate, then the first-order term in the expansion of $E(\gamma)$ is given by

$$E^{(1)} = \langle \phi^{(0)}, \tilde{H}^{(1)} \phi^{(0)} \rangle. \tag{C.25}$$

If the degree of degeneracy of $E^{(0)}$ is m, then $E_j^{(1)}$, $j = 1, \ldots, k$, are the eigenvalues of the self-adjoint $m \times m$ matrix

$$A_{jk} = \langle \phi_j^{(0)}, \tilde{H}^{(1)} \phi_k^{(0)} \rangle. \tag{C.26}$$

Appendix D

Variational Method

The ground state can be obtained by minimizing the energy as a functional of the state vector.

D.1. Critical points of the energy functional

Here, we consider the expectation value of a Hamiltonian H as a functional of the state vector ψ. For a vector ψ of arbitrary normalization in the Hilbert space $\mathfrak{H}$, we define the energy functional

$$W(\psi) = \frac{\langle \psi, H\psi \rangle}{\|\psi\|^2} \tag{D.1}$$

for all ψ in the domain $\mathfrak{D}(H)$ of the operator H. A functional is a mapping from the Hilbert space into the complex numbers. As H is assumed to be self-adjoint, W is a real-valued functional.

A vector ψ in $\mathfrak{D}(H)$ is called a *critical point* or *stationary point* if

$$\frac{d}{d\lambda} W(\psi + \lambda\phi)\Big|_{\lambda=0} = 0 \quad \text{for all } \phi \in \mathfrak{D}(H). \tag{D.2}$$

Examples for critical points are maxima, minima, and saddle points.

Theorem:

ψ is a critical point of W if and only if ψ is an eigenvector of H with eigenvalue $W(\psi)$.

Proof. From the definition (D.1),

$$W(\psi + \lambda\phi)\, \|\psi + \lambda\phi\|^2 = \langle \psi + \lambda\phi, H(\psi + \lambda\phi) \rangle. \tag{D.3}$$

Expanding both sides we obtain

$$\begin{aligned} W(\psi + \lambda\phi)\, \big(\|\psi\|^2 + 2\lambda \operatorname{Re} \langle \phi, \psi \rangle + \lambda^2 \|\phi\|^2\big) \\ &= \langle \psi, H\psi \rangle + \lambda \langle \psi, H\phi \rangle + \lambda \langle \phi, H\psi \rangle + \lambda^2 \langle \phi, H\phi \rangle \\ &= \langle \psi, H\psi \rangle + 2\lambda \operatorname{Re} \langle \phi, H\psi \rangle + \lambda^2 \langle \phi, H\phi \rangle. \end{aligned} \tag{D.4}$$

Forming the derivative at $\lambda = 0$ gives

$$\frac{d}{d\lambda} W(\psi + \lambda\phi)\Big|_{\lambda=0} \|\psi\|^2 + 2\, W(\psi)\, \mathrm{Re}\, \langle \phi, \psi \rangle = 2\, \mathrm{Re}\, \langle \phi, H\psi \rangle, \tag{D.5}$$

or

$$\frac{d}{d\lambda} W(\psi + \lambda\phi)\Big|_{\lambda=0} = 2\, \mathrm{Re}\, \frac{\langle \phi, H\psi \rangle - W(\psi)\, \langle \phi, \psi \rangle}{\|\psi\|^2}. \tag{D.6}$$

If $H\psi = E\psi$, then $W(\psi) = E$, and (D.6) evaluates to zero for all ϕ in the domain of H. Conversely, if (D.6) is zero for some $\psi \neq 0$ and all $\phi \in \mathfrak{D}(H)$, then it is zero also for $\mathrm{i}\phi$, that is, we can omit taking the real part in (D.6)

$$0 = \langle \phi, H\psi \rangle - W(\psi)\, \langle \phi, \psi \rangle = \langle \phi, (H - W(\psi))\, \psi \rangle. \tag{D.7}$$

Hence, $(H - W(\psi))\, \psi$ is orthogonal to all vectors in $\mathfrak{D}(H)$ (a subspace that is dense in $\mathfrak{H}$). But, the only vector that is orthogonal to a dense subspace is the zero vector, and we conclude

$$H\psi = W(\psi)\, \psi. \tag{D.8}$$

This proves the theorem. □

D.2. Semibounded Hamiltonians and the minimal energy

The energy spectra of the Hamiltonians used in atomic physics have in most cases the following structure. There is a finite or infinite number of *discrete* eigenvalues at the bottom of the energy spectrum. A discrete eigenvalue is, by definition, an isolated eigenvalue with a finite degree of degeneracy. The ground-state energy E_1 is usually non-degenerate (or has a "trivial" degree of degeneracy due to the spin of the electron). The discrete eigenvalues can be ordered according to their size,

$$E_1 \leq E_2 \leq E_3 \leq \ldots \tag{D.9}$$

Then there is the *essential spectrum*, usually a semi-infinite interval $[E_\infty, \infty)$. It consists of scattering energies forming the continuous spectrum and in addition it may contain embedded (non-isolated) eigenvalues, or accumulation points of eigenvalues, or eigenvalues with infinite degree of degeneracy. The bottom of the essential spectrum, E_∞, may or may not be an eigenvalue. There is no spectrum (no eigenvalues and no essential spectrum) in the interval $(-\infty, E_1)$ to the left of the ground-state energy. A Hamiltonian with this property is called *semibounded*. An important counter-example is the Dirac operator (see Section 8.1.2).

In the subspace of bound states, the eigenstates form an orthonormal basis. Assume that ψ is in that subspace, then

$$\psi = \sum_n c_n \psi_n, \quad \text{with } c_n = \langle \psi_n, \psi \rangle, \qquad \sum_n |c_n|^2 = \|\psi\|^2. \tag{D.10}$$

We find

$$\begin{aligned}\langle\psi, H\psi\rangle &= \sum_n \langle\psi,\psi_n\rangle\langle\psi_n, H\,\psi\rangle = \sum_n \langle\psi,\psi_n\rangle\langle H\,\psi_n,\psi\rangle \\ &= \sum_n E_n\,\langle\psi,\psi_n\rangle\langle\psi_n,\psi\rangle = \sum_n E_n\,|\langle\psi,\psi_n\rangle|^2 \\ &\geq E_1\,\sum_n |\langle\psi,\psi_n\rangle|^2 = E_1\,\|\psi\|^2. \end{aligned} \tag{D.11}$$

Hence,

$$W(\psi) \geq E_1 \quad \text{for all } \psi \in \mathfrak{D}(H) \tag{D.12}$$

and

$$W(\psi) = E_1 \quad \text{if and only if } H\psi = E_0\psi. \tag{D.13}$$

This observation can be turned into a practical method to estimate the ground state energy from above. We consider a Hamiltonian H with eigenstates ψ_n and eigenvalues E_n at the bottom of its energy spectrum. In order to find the ground-state energy, we have to find the minimum of the functional W, that is,

$$E_1 = \min_{\psi\in\mathfrak{D}(H)} W(\psi). \tag{D.14}$$

As it might be difficult or impossible to minimize W over all $\psi \in \mathfrak{D}(H)$, one often chooses vectors ψ_λ depending on one or more parameters λ, and considers the function

$$f(\lambda) = W(\psi_\lambda) \geq E_1. \tag{D.15}$$

Then, the minimal value of f is an upper bound for the ground-state energy:

$$\min_\lambda f(\lambda) \geq E_1. \tag{D.16}$$

This method is useful, when one has already a rough idea how the ground state wave function looks. Then, we set up functions with a similar shape and use the parameter λ for fine tuning.



In general, it may be necessary to write "infimum" instead of "minimum" in (D.14) and (D.16), because the minimal value could be achieved for $\lambda \to \infty$, or for some $\lambda \to \lambda_0$ for which ψ_{λ_0} is not in the domain of H (consequently, f is not defined for λ_0). The infimum is the largest lower bound of a set, but it need not belong to that set. A minimum is an infimum that is actually attained by some element within that set. In particular, if there are no eigenvalues at the bottom of the energy spectrum, then $\min W(\psi)$ does not exist. As an example, consider the free-particle Hamiltonian $H_0 = -\Delta/2\mathrm{m}$. This operator has no eigenvalues at all, but it is semibounded, which means that

$$W_0(\psi) = \frac{\langle\psi, H_0\psi\rangle}{\|\psi\|^2} > 0 \quad \text{for all } \psi \in \mathfrak{D}(H_0). \tag{D.17}$$

The infimum of $W(\psi)$ over all $\psi \in \mathfrak{D}(H)$ exists and is equal to zero (the bottom of the essential spectrum),

$$\inf_{\psi\in\mathfrak{D}(H_0)} W_0(\psi) = 0, \tag{D.18}$$

but this value is not achieved for any square-integrable function ψ.

D.3. The ground state of helium (again)

Again we take the Hamiltonian

$$H = \frac{1}{2\mathrm{m}_e}(\mathbf{p}_1^2 + \mathbf{p}_2^2) - 2\gamma_0 \Big(\frac{1}{|\mathbf{x}_1|} + \frac{1}{|\mathbf{x}_2|}\Big) + \frac{\gamma_0}{|\mathbf{x}_1 - \mathbf{x}_2|}. \tag{D.19}$$

We consider the family of functions $\psi_\lambda \in L^2(\mathbb{R}^6)$, defined by

$$\psi_\lambda(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi}\Big(\frac{\lambda}{a_0}\Big)\, \exp\Big(-\frac{\lambda}{a_0}\big(|\mathbf{x}_1| + |\mathbf{x}_2|\big)\Big). \tag{D.20}$$

Hence, this wave function describes the ground state of

$$H_\lambda = \frac{1}{2\mathrm{m}_e}(\mathbf{p}_1^2 + \mathbf{p}_2^2) - \lambda\gamma_0 \Big(\frac{1}{|\mathbf{x}_1|} + \frac{1}{|\mathbf{x}_2|}\Big) + \frac{\gamma_0}{|\mathbf{x}_1 - \mathbf{x}_2|}, \tag{D.21}$$

belonging to the ground-state energy

$$E_\lambda = -\lambda^2\, \frac{\gamma_0}{a_0}. \tag{D.22}$$

From the virial theorem, we conclude immediately that this energy is minus the average kinetic energy and half of the average potential energy. Hence,

$$\big\langle \psi_\lambda\,,\, \frac{1}{2\mathrm{m}_e}(\mathbf{p}_1^2 + \mathbf{p}_2^2)\, \psi_\lambda \big\rangle = -E_\lambda, \tag{D.23}$$

$$\big\langle \psi_\lambda\,,\, -\lambda\gamma_0 \Big(\frac{1}{|\mathbf{x}_1|} + \frac{1}{|\mathbf{x}_2|}\Big)\, \psi_\lambda \big\rangle = 2\, E_\lambda. \tag{D.24}$$

With the help of (B.56), we compute

$$\big\langle \psi_\lambda\,,\, \frac{\gamma_0}{|\mathbf{x}_1 - \mathbf{x}_2|}\, \psi_\lambda \big\rangle = \frac{5}{8}\, \lambda\, \frac{\gamma_0}{a_0} = -\frac{5}{8}\, \frac{E_\lambda}{\lambda}. \tag{D.25}$$

Putting all these pieces together, we obtain

$$f(\lambda) = \langle \psi_\lambda, H\, \psi_\lambda \rangle = -E_\lambda + \frac{2}{\lambda}\, E_\lambda - \frac{5}{8\lambda} E_\lambda = \big(\lambda^2 - \frac{27}{8}\, \lambda\big)\, \frac{\gamma_0}{a_0}. \tag{D.26}$$

This expression achieves its minimum value at $\lambda = 27/16$. We obtain

$$\min f(\lambda) = -\Big(\frac{27}{16}\Big)^2\, \frac{\gamma_0}{a_0} \approx 2.85\, \frac{\gamma_0}{a_0}. \tag{D.27}$$

The constant 2.85 is closer to the observed value 2.92 than the value 2.75 obtained by first-order perturbation theory. This is due to the fact that the

unperturbed eigenvector (B.52) is among the functions ψ_λ (indeed, $\psi^{(0)} = \psi_{\lambda_0}$ with $\lambda_0 = 2$). Hence,

$$E^{(0)} + \gamma_0 E^{(1)} = \langle \psi^{(0)}, (H^{(0)} + \gamma_0 H^{(1)})\psi^{(0)} \rangle = f(\lambda_0) \geq \min f(\lambda). \quad \text{(D.28)}$$

D.4. Finding excited states by variational methods

We have seen that the ground-state energy E_1 is characterized by

$$E_1 = \min_{\phi \in \mathfrak{D}(H)} W(\phi), \quad \text{with} \quad W(\phi) = \frac{\langle \phi, H\phi \rangle}{\|\phi\|^2}. \quad \text{(D.29)}$$

Now suppose that we know the ground state ψ_1, for which $W(\psi_1) = E_1$. We can restrict the Hamiltonian H to the subspace of the Hilbert space that is orthogonal to ψ_1. In this subspace, the lowest energy eigenvalue is E_2. Thus, we can find the eigenvalue E_2 by minimizing $W(\phi)$ over all ϕ that are orthogonal to ψ_1,

$$E_2 = \min_{\substack{\phi \in \mathfrak{D}(H) \\ \phi \perp \psi_1}} W(\phi). \quad \text{(D.30)}$$

In general, we know only an approximation φ to the true eigenvector ψ_1. We assume that φ can be written as a linear combination of eigenvectors,

$$\varphi = \sum_n a_n \psi_n. \quad \text{(D.31)}$$

Choose, for example, $\phi = \overline{a_2}\,\psi_1 - \overline{a_1}\,\psi_2$. Then, it is easy to see that ϕ is orthogonal to φ and

$$W(\phi) = \frac{|a_2|^2 E_1 + |a_1|^2 E_2}{|a_1|^2 + |a_2|^2} \leq E_2, \quad \text{(D.32)}$$

because $E_1 \leq E_2$. Hence, taking the minimum of W over all ϕ that are orthogonal to φ gives something that is smaller than E_2! Therefore,

$$E_2 = \max_{\psi} \min_{\substack{\phi \in \mathfrak{D}(H) \\ \phi \perp \psi}} W(\phi). \quad \text{(D.33)}$$

Note that this characterization of E_2 is independent of ψ_1.

Assuming that H has at least n eigenvalues at the bottom of its spectrum, we find that we can characterize the nth eigenvalue in an analogous way. Denoting the subspace spanned by m vectors $\varphi_1, \ldots \varphi_m$ by $[\varphi_1, \ldots \varphi_m]$, we find

$$E_n = \max_{\varphi_1, \ldots \varphi_{n-1}} \min_{\substack{\phi \in \mathfrak{D}(H) \\ \phi \perp [\varphi_1, \ldots \varphi_{n-1}]}} W(\phi). \quad \text{(D.34)}$$

D.5. The minimax principle and the Rayleigh-Ritz technique

Assuming that we are given a semibounded Hamiltonian H that has eigenvalues at the bottom of its spectrum, we want to determine the first n of these eigenvalues. Each of these eigenvalues is given by a formula like (D.34).

We can set up a system of n trial functions $\phi_1, \ldots, \phi_n$ in the domain of H, and we assume that the trial functions have been chosen to be orthonormal. Hence, these functions form the basis of an n-dimensional subspace $\mathfrak{H}_n$ of the Hilbert space $\mathfrak{H}$. The restriction of the Hamiltonian to this subspace can be represented by the Hermitian matrix $\hat{H} = (H_{ij})$ with matrix elements $H_{ij} = \langle \phi_i, H\phi_j \rangle$. We can diagonalize this matrix on a computer and obtain an n-tuple of eigenvalues $\hat{E}_1 \leq \hat{E}_2 \leq \ldots \leq \hat{E}_n$. The mth of these eigenvalues is characterized by

$$\hat{E}_m = \max_{\varphi_1, \ldots \varphi_{m-1} \in \mathfrak{H}_n} \min_{\substack{\phi \in \mathfrak{H}_n \\ \phi \perp [\varphi_1, \ldots \varphi_{m-1}]}} \frac{\langle \phi, \hat{H}\phi \rangle}{\|\phi\|^2}. \tag{D.35}$$

The maximum is actually achieved if $\varphi_1, \ldots, \varphi_{m-1}$ are chosen as eigenvectors of the $n \times n$ matrix H_{ij}. Hence, it does not matter if we take the maximum over all $\varphi_1, \ldots \varphi_{m-1} \in \mathfrak{H}$:

$$\hat{E}_m = \max_{\varphi_1, \ldots \varphi_{m-1} \in \mathfrak{H}} \min_{\substack{\phi \in \mathfrak{H}_n \\ \phi \perp [\varphi_1, \ldots \varphi_{m-1}]}} \frac{\langle \phi, H\phi \rangle}{\|\phi\|^2}. \tag{D.36}$$

We replaced $\hat{H}$ with H, because $\langle \phi, \hat{H}\phi \rangle = \langle \phi, H\phi \rangle$ for all $\phi \in \mathfrak{H}_n$. But if we now take the minimum over the larger set $\phi \in \mathfrak{D}(H)$, we obtain

$$\hat{E}_m \geq \max_{\varphi_1, \ldots \varphi_{m-1} \in \mathfrak{H}} \min_{\substack{\phi \in \mathfrak{D}(H) \\ \phi \perp [\varphi_1, \ldots \varphi_{m-1}]}} \frac{\langle \phi, H\phi \rangle}{\|\phi\|^2}. \tag{D.37}$$

The right side is equal to E_n by Eq. (D.34). Hence, we find

$$\hat{E}_m \geq E_m, \quad \text{for } m = 1, \ldots, n. \tag{D.38}$$

This observation is the basis of numerical methods for finding the eigenvalues of the Schrödinger equation $H\psi = E\psi$ (*Rayleigh-Ritz technique*).

Appendix E

Adiabatic and Geometric Phases

We consider a Hamiltonian $H(\gamma)$ depending on one or several parameters γ. We assume that the eigenvalue problem has been solved for each value of the parameter γ. What happens if we let γ change slowly with time? What can we say about the solutions of the Schrödinger equation with the time-dependent Hamiltonian $H\gamma(t))$?

E.1. The adiabatic approximation

Assume that $H(\gamma)$ has for all γ under consideration an orthonormal basis of eigenvectors $\psi_j(\gamma)$. For simplicity, we assume that all eigenvalues are nondegenerate. We have

$$H(\gamma)\,\psi_j(\gamma) = E_j(\gamma)\,\psi_j(\gamma). \tag{E.1}$$

This time, we assume that for each value of γ, the eigenvalues and eigenfunctions are already known. The eigenvectors for different values of γ have nothing to do with each other, but we assume that the eigenvectors are chosen in such a way that they depend on γ in a differentiable way. Moreover, we assume

$$\big\langle\, \psi_j(\gamma)\,,\, \frac{d}{d\gamma}\,\psi_j(\gamma)\,\big\rangle = 0. \tag{E.2}$$

This condition replaces (B.29). It states that infinitesimally small changes of ψ_j are orthogonal to ψ_j. The fact that the norm of ψ_j is independent of γ implies that

$$\frac{d}{d\gamma}\langle\psi_j(\gamma), \psi_j(\gamma)\rangle = 2\,\mathrm{Re}\,\big\langle\, \psi_j(\gamma)\,,\, \frac{d}{d\gamma}\,\psi_j(\gamma)\,\big\rangle = 0. \tag{E.3}$$

Now we assume that the parameter γ depends on time in a differentiable way. We look for a solution of

$$\mathrm{i}\,\frac{d}{dt}\,\Psi(t) = H\big(\gamma(t)\big)\,\Psi(t). \tag{E.4}$$

The functions $\psi_j\big(\gamma(t)\big)$ form an orthonormal basis of eigenvectors of $H\big(\gamma(t)\big)$. Clearly, we expand the solution at time t in this basis.

$$\Psi(t) = \sum_j c_j(t)\, \psi_j\big(\gamma(t)\big). \tag{E.5}$$

We insert this into the time-dependent Schrödinger equation.

$$\begin{aligned} \mathrm{i}\,\frac{d}{dt}\,\Psi(t) &= \mathrm{i}\,\sum_j \Big(\frac{d}{dt}\,c_j(t)\Big)\,\psi_j\big(\gamma(t)\big) + \mathrm{i}\,\sum_j c_j(t)\,\frac{d}{dt}\,\psi_j\big(\gamma(t)\big), \\ &= H\big(\gamma(t)\big)\,\Psi(t) = \sum_j c_j(t)\,E_j\big(\gamma(t)\big)\,\psi_j\big(\gamma(t)\big). \end{aligned} \tag{E.6}$$

We note that

$$\frac{d}{dt}\,\psi_j\big(\gamma(t)\big) = \frac{d}{d\gamma}\,\psi_j(\gamma)\Big|_{\gamma=\gamma(t)}\,\frac{d\gamma(t)}{dt}. \tag{E.7}$$

We take the scalar product of (E.6) with the vector $\psi_k\big(\gamma(t)\big)$. Using the orthonormality of the basis vectors and (E.2), we obtain

$$\mathrm{i}\,\frac{d}{dt}\,c_k(t) + \mathrm{i}\,\frac{d\gamma(t)}{dt}\,\sum_{j\neq k} c_j(t)\,\big\langle\,\psi_k(\gamma)\,,\,\frac{d}{d\gamma}\,\psi_j(\gamma)\,\big\rangle\Big|_{\gamma(t)} = c_k(t)\,E_k\big(\gamma(t)\big). \tag{E.8}$$

Solving this coupled system of equations for the coefficients $c_j(t)$ solves the time-dependent Schrödinger equation.

If γ varies only slowly with t, we can assume that the sum proportional to $d\gamma/dt$ can be neglected. This is the *adiabatic approximation*.

$$\mathrm{i}\,\frac{d}{dt}\,c_k(t) = \,c_k(t)\,E_k\big(\gamma(t)\big). \tag{E.9}$$

$$c_k(t) = c_k(0)\,\exp\Big(-\mathrm{i}\int_0^t E_k\big(\gamma(s)\big)\,ds\Big). \tag{E.10}$$

Whenever the state starts out in an eigenstate of $H\big(\gamma(0)\big)$, say

$$\Psi(0) = \psi_k\big(\gamma(0)\big), \quad \text{that is,} \quad c_j(0) = \delta_{jk}, \tag{E.11}$$

then the solution of the Schrödinger equation in the adiabatic approximation will be

$$\Psi(t) = c_k(t)\,\psi_k\big(\gamma(t)\big) = \exp\Big(-\mathrm{i}\int_0^t E_k\big(\gamma(s)\big)\,ds\Big)\,\psi_k\big(\gamma(t)\big). \tag{E.12}$$

If the Hamiltonian depends on time very slowly, then the solution clings to an eigenstate.

The adiabatic approximation is only valid if $d\gamma/dt$ is very small and if the scalar products $\langle\,\psi_k\,,\,\psi_j'\rangle$ for $j \neq k$ do not get large for some value of

γ (the prime denotes differentiation with respect to γ). Differentiate the eigenvalue equation (E.1) (omitting the argument γ)

$$H'\,\psi_j + H\,\psi_j' = E_j'\,\psi_j + E_j\,\psi_j' \tag{E.13}$$

and take the scalar product with ψ_k. Using the self-adjointness of H and the orthonormality of the eigenvectors, we obtain

$$\langle\,\psi_k\,,\,H'\,\psi_j\,\rangle + E_k\,\langle\,\psi_k\,,\,\psi_j'\,\rangle = E_j'\,\delta_{kj} + E_j\,\langle\,\psi_k\,,\,\psi_j'\,\rangle. \tag{E.14}$$

This gives, for $k \neq j$,

$$\langle\,\psi_k(\gamma)\,,\,\frac{d}{d\gamma}\,\psi_j(\gamma)\,\rangle = \frac{\langle\,\psi_k(\gamma)\,,\,H'(\gamma)\,\psi_j(\gamma)\rangle}{E_j(\gamma) - E_k(\gamma)}. \tag{E.15}$$

We see that we have to assume that during the whole change of γ the eigenvalues $E_j(\gamma)$ and $E_k(\gamma)$ stay well separated from each other.

E.2. The Berry phase

Very often, the Hamiltonian does not only depend on one but on several parameters. If a quantum system is adiabatically moved along a curve C in a higher-dimensional parameter space, then a new type of phase factor $\exp(\mathrm{i}\lambda(C))$ appears. Its presence was discovered by M.V. Berry in 1983, and hence $\lambda(C)$ is called the *Berry phase*.

We assume that the parameters $\boldsymbol{\gamma} = (\gamma_1, \dots \gamma_N)$ lie in some subset U of $\mathbb{R}^N$ and consider a Hamiltonian operator $H(\boldsymbol{\gamma})$ depending on these parameters. We also assume that there is an orthonormal basis of non-degenerate eigenvectors for each value of $\boldsymbol{\gamma}$ in U,

$$H(\boldsymbol{\gamma})\,\psi_j(\boldsymbol{\gamma}) = E_j(\boldsymbol{\gamma})\,\psi_j(\boldsymbol{\gamma}). \tag{E.16}$$

Again we assume that eigenvectors and eigenvalues depend on $\boldsymbol{\gamma}$ in a differentiable way. The eigenvectors are all normalized, hence

$$\frac{\partial}{\partial\gamma_k}\,\|\psi_j(\boldsymbol{\gamma})\|^2 = 2\,\mathrm{Re}\,\langle\,\psi_j(\boldsymbol{\gamma})\,,\,\frac{\partial}{\partial\gamma_k}\,\psi_j(\boldsymbol{\gamma})\rangle = 0, \tag{E.17}$$

for all j, all $k = 1, 2, \dots, N$, and all $\boldsymbol{\gamma} \in U$. Can we impose a normalization condition similar to (E.2)? In general,

$$\mathrm{Im}\,\langle\,\psi_j(\boldsymbol{\gamma})\,,\,\frac{\partial}{\partial\gamma_k}\,\psi_j(\boldsymbol{\gamma})\rangle \neq 0. \tag{E.18}$$

Let us multiply the eigenvector $\psi_j(\boldsymbol{\gamma})$ by a phase factor,

$$\phi(\boldsymbol{\gamma}) = \mathrm{e}^{\mathrm{i}\lambda(\boldsymbol{\gamma})}\,\psi_j(\boldsymbol{\gamma}), \tag{E.19}$$

where $\lambda : U \to \mathbb{R}$ is real-valued and differentiable. The normalization of ϕ implies that (E.17) holds with ψ_j replaced by ϕ. Can we adjust the phase λ

such that also the expression for the imaginary part corresponding to (E.18) vanishes? With the abbreviations

$$\partial_k = \frac{\partial}{\partial \gamma_k}, \qquad \langle \psi_j, \partial_k \psi_j \rangle = \big\langle\, \psi_j(\boldsymbol{\gamma})\,,\, \frac{\partial}{\partial \gamma_k}\, \psi_j(\boldsymbol{\gamma}) \big\rangle, \quad \text{etc.}, \tag{E.20}$$

we compute

$$\langle \phi, \partial_k \phi \rangle = \mathrm{e}^{-\mathrm{i}\lambda} \big\langle\, \psi_j\,,\, \mathrm{i}(\partial_k \lambda)\mathrm{e}^{\mathrm{i}\lambda}\psi_j + \mathrm{e}^{\mathrm{i}\lambda}\partial_k \psi_j \,\big\rangle = \mathrm{i}(\partial_k \lambda) + \langle \psi_j, \partial_k \psi_j \rangle. \tag{E.21}$$

If this expression should vanish, then we must choose λ in such a way that

$$\mathrm{i}\langle \psi_j, \partial_k \psi_j \rangle = \partial_k \lambda. \tag{E.22}$$

This equation states that the real-valued vector field $\vec{v} : U \to \mathbb{R}^N$,

$$\vec{v}(\boldsymbol{\gamma}) = (v_1(\boldsymbol{\gamma}), v_2(\boldsymbol{\gamma}), \ldots, v_N(\boldsymbol{\gamma})) \quad \text{with} \quad v_k(\boldsymbol{\gamma}) = \mathrm{i}\,\langle \psi_j, \partial_k \psi_j \rangle \tag{E.23}$$

has to be the gradient (with respect to $\boldsymbol{\gamma}$) of a scalar function $\lambda(\boldsymbol{\gamma})$. This is equivalent with the condition that the exterior derivative of $\vec{v}$ with respect to the parameters γ vanishes. We are going to show that this condition need not be satisfied. In the following section, we are going to demonstrate that this condition need not hold in general. Our example involves a three-dimensional parameter space, where the exterior derivative is just the curl of the vector field $\vec{v}$:

$$\vec{v}(\boldsymbol{\gamma}) = \nabla_\gamma\, \lambda(\boldsymbol{\gamma}) \quad \text{if and only if} \quad \nabla_\gamma \times \vec{v} = 0 \qquad (N = 3). \tag{E.24}$$

We are going to show that, in general, the curl of $\vec{v}$ is nonzero. Hence, there is no scalar function λ such that (E.22) holds.

We note that in the case of a one-dimensional parameter space ($N = 1$), we can always find λ by a simple integration. For $N = 3$, we can choose a differentiable curve $C : [0, T] \to U$ with $C(0) = \boldsymbol{\gamma}_0$ and $C(T) = \boldsymbol{\gamma}_1$ and define

$$\lambda(\boldsymbol{\gamma}_1) = \int_C \vec{v}(\boldsymbol{\gamma}) \cdot d\boldsymbol{\gamma} = \int_0^T \vec{v}(\boldsymbol{\gamma}(t)) \cdot \frac{d\boldsymbol{\gamma}(t)}{dt}\, dt. \tag{E.25}$$

If the vector field $\vec{v}$ is not a gradient, then the value of λ at γ_1 depends on the chosen curve C. In particular, if C is a closed curve such that $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_1$, then, using Stokes' theorem of vector analysis,

$$\lambda(C) = \oint_C \vec{v}(\boldsymbol{\gamma}) \cdot d\boldsymbol{\gamma} = \int\int \nabla_\gamma \times \vec{v}(\boldsymbol{\gamma})\, df \qquad (N = 3). \tag{E.26}$$

In our case,

$$\begin{aligned} \lambda(C) &= \mathrm{i} \oint_C \big\langle\, \psi_j(\boldsymbol{\gamma})\,,\, \nabla_\gamma\, \psi_j(\boldsymbol{\gamma}) \,\big\rangle \cdot d\boldsymbol{\gamma} \\ &= \mathrm{i} \int_0^T \sum_{k=1}^{3} \big\langle\, \psi_j(\boldsymbol{\gamma})\,,\, \frac{\partial}{\partial \gamma_k} \psi_j(\boldsymbol{\gamma}) \,\big\rangle \Big|_{\boldsymbol{\gamma}(t)} \frac{d\boldsymbol{\gamma}(t)}{dt}\, dt. \end{aligned} \tag{E.27}$$

The fact that we cannot define a unique phase $\lambda(\boldsymbol{\gamma})$ on all of parameter space has important physical consequences. In order to see this, we combine the considerations of this section with the results about the adiabatic limit.

Assume that we start at $t = 0$ with one of the eigenvectors $\Psi(0) = \psi_j(\boldsymbol{\gamma}(0))$. We move the parameter along some curve C in the parameter space

$$C : t \longrightarrow \boldsymbol{\gamma}(t), \quad \text{with } \boldsymbol{\gamma}(t) \in U \text{ for all } t. \tag{E.28}$$

If this change is done slowly enough that the state can adjust, the system will evolve adiabatically and finally at time T be in a state proportional to $\psi_j(\boldsymbol{\gamma}(T))$. However, the computation leading to (E.8) in Section E.1 is changed in one respect. The summand with $j = k$ does not vanish. This changes (E.9) to

$$\mathrm{i}\,\frac{d}{dt}\,c_k(t) = c_k(t)\,E_k\big(\boldsymbol{\gamma}(t)\big) - \mathrm{i}c_k(t)\,\frac{d\boldsymbol{\gamma}(t)}{dt} \cdot \big\langle\, \psi_k(\boldsymbol{\gamma})\,,\, \nabla_{\boldsymbol{\gamma}}\, \psi_k(\boldsymbol{\gamma})\,\big\rangle\Big|_{\boldsymbol{\gamma}(t)}. \tag{E.29}$$

By integration along the curve C in parameter space, we may compute the coefficient $c_k(t)$. In (E.29), we recognize the integrand of (E.27). Hence, integration of (E.29) around a closed loop gives

$$c_k(T) = c_k(0)\,\exp\Big(-\mathrm{i}\int_0^T E_k\big(\boldsymbol{\gamma}(s)\big)\,ds + \mathrm{i}\lambda(C)\Big) \tag{E.30}$$

with $\lambda(C)$ as in (E.26) and (E.27). One might argue that the additional term in (E.29) vanishes in the adiabatic limit, as $d\boldsymbol{\gamma}/dt \to 0$. But, this is not true for a closed curve C, because the value $\lambda(C)$ does not depend on how fast we move along the loop, it just depends on the geometry of the curve C. This geometric phase $\lambda(C)$ can be measured by comparing a system that has been moved adiabatically around a closed loop in parameter space with a system that has not been moved.

Exercise E.1. *Consider an eigenvector $\psi(\boldsymbol{\gamma})$ of $H(\boldsymbol{\gamma})$ that depends on the parameters in a differentiable way. Changing the parameters slowly along a curve C from $\boldsymbol{\gamma}_0$ to $\boldsymbol{\gamma}$ changes the initial vector $\psi(\boldsymbol{\gamma}_0)$ into a vector that differs from $\psi(\boldsymbol{\gamma})$ at most by a phase factor (in the adiabatic limit). Determine this phase factor by inserting $\Psi(t) = \exp(\mathrm{i}\varphi(t, C))\,\psi(\boldsymbol{\gamma}(t))$ into the Schrödinger equation*

$$\mathrm{i}\,\frac{d}{dt}\,\Psi(t) = H(\boldsymbol{\gamma}(t))\,\Psi(t). \tag{E.31}$$

E.3. Example: Spin in magnetic field

In this section, we consider a qubit in a magnetic field. The Hamiltonian is of the form

$$H(\boldsymbol{\gamma}) = \boldsymbol{\sigma} \cdot \boldsymbol{\gamma} = \sigma_1\gamma_1 + \sigma_2\gamma_2 + \sigma_3\gamma_3. \tag{E.32}$$

Here, $\boldsymbol{\gamma}$ involves the magneton $\boldsymbol{\mu}$ and the magnetic field strength $\mathbf{B}$. Hence, the parameter space is three-dimensional. The two eigenvalues of $H(\boldsymbol{\gamma})$ are $\pm|\boldsymbol{\gamma}|$. Normalized eigenvectors are the spinors with spin-up and spin-down with respect to the direction of $\boldsymbol{\gamma}$ (compare (4.19))

$$\psi_{\pm}(\boldsymbol{\gamma}) = \frac{1}{\sqrt{2|\boldsymbol{\gamma}|^2 \pm 2\gamma_3|\boldsymbol{\gamma}|}} \begin{pmatrix} \gamma_3 \pm |\boldsymbol{\gamma}| \\ \gamma_1 + \mathrm{i}\gamma_2 \end{pmatrix}. \tag{E.33}$$

Let us now consider the eigenvector with "spin-up" in the direction of $\boldsymbol{\gamma}$. A little calculation shows that

$$\vec{v}(\boldsymbol{\gamma}) = \mathrm{i}\langle\, \psi_+(\boldsymbol{\gamma})\,,\, \nabla_\gamma\, \psi_+(\boldsymbol{\gamma})\,\rangle = \frac{1}{2|\boldsymbol{\gamma}|^2 + 2\gamma_3|\boldsymbol{\gamma}|}\,(\gamma_2, -\gamma_1, 0). \tag{E.34}$$

Finally, we obtain

$$\vec{w}(\boldsymbol{\gamma}) = \nabla_\gamma \times \vec{v}(\boldsymbol{\gamma}) = -\frac{1}{2}\,\frac{\boldsymbol{\gamma}}{|\boldsymbol{\gamma}|^3}. \tag{E.35}$$

Because the curl of $\vec{w}$ is a spherically symmetric vector field, the flux of $\nabla_\gamma \times \vec{v}$ through a surface bounded by C is equal to the flux through the surface of the unit sphere that is bounded by the projection of the curve C in radial direction onto that surface. (We assume that the curve does not go through the origin where the eigenvalue of $H(\boldsymbol{\gamma})$ is degenerate). On the surface of the unit sphere $\vec{w}$ is just $-\mathbf{n}_r/2$, with $\mathbf{n}_r$ being the outward normal unit vector. Hence, the flux is just $-1/2$ times the area of the area on the unit sphere, that is, the solid angle $\Omega(C)$ under which the curve C is seen from the origin in parameter space:

$$\lambda_+(C) = -\frac{1}{2}\,\Omega(C). \tag{E.36}$$

Here, the index '+' refers to the eigenvector ψ_+. For the eigenvector ψ_-, one obtains the geometric phase

$$\lambda_-(C) = \frac{1}{2}\,\Omega(C). \tag{E.37}$$

A simple special case is obtained when we move the vector $\boldsymbol{\gamma}$ along a circle with center at the origin. Then, the curve is seen under the solid angle $\Omega(C) = 2\pi$ (given by half the surface of the unit sphere). We conclude that Berry's phase in this case is just $\lambda(C) = -\pi$. Hence, when the eigenvector ψ_+ of H is adiabatically transported around this circle, it gets rotated through an angle 2π and acquires a geometric phase factor

$$\mathrm{e}^{\mathrm{i}\lambda(C)} = \mathrm{e}^{-\mathrm{i}\Omega(C)/2} = \mathrm{e}^{-\mathrm{i}\pi} = -1. \tag{E.38}$$

Hence, we have obtained again our result in Section 4.4.2, that a rotation of a qubit through an angle 2π turns a state vector into its negative.



CD 4.18 shows the time evolution (oscillating phase) of the spin-up eigenvector in a time-constant magnetic field in the z-direction. This situation is compared with one where the magnetic field slowly changes its direction. The state adjusts to the magnetic field, and the direction of the spin vector adiabatically follows the direction of the magnetic field. Compared to the eigenvector at rest, the eigenvector that has been slowly turned around has acquired the phase factor -1.

Let us now consider a more general situation with a three-dimensional parameter space. We want to compute the curl of the vector field

$$\vec{v}(\boldsymbol{\gamma}) = \mathrm{i}\left\langle\, \psi_j(\boldsymbol{\gamma})\,,\, \nabla_\gamma\, \psi_j(\boldsymbol{\gamma})\,\right\rangle. \tag{E.39}$$

Consider, for example, the first component

$$\begin{aligned}(\nabla_\gamma \times \vec{v})_1 &= \partial_2 v_3 - \partial_3 v_1 = \mathrm{i}\langle\partial_2\psi_j, \partial_3\psi_j\rangle - \mathrm{i}\langle\partial_3\psi_j, \partial_2\psi_j\rangle \\ &= -2\,\mathrm{Im}\,\langle\partial_2\psi_j, \partial_3\psi_j\rangle = -2\,\mathrm{Im}\sum_k \langle\partial_2\psi_j, \psi_k\rangle\langle\psi_k, \partial_3\psi_j\rangle \\ &= -2\,\mathrm{Im}\sum_{k\neq j} \langle\partial_2\psi_j, \psi_k\rangle\langle\psi_k, \partial_3\psi_j\rangle. \end{aligned} \tag{E.40}$$

The summand with $k = j$ is real because of (E.18) and hence it vanishes after taking the imaginary part. A computation similar to the one leading to (E.15) gives

$$\langle\psi_k, \partial_m\psi_j\rangle = \frac{\langle\psi_k, (\partial_m H)\psi_j\rangle}{E_j - E_k} \tag{E.41}$$

and we obtain as the final result

$$(\nabla_\gamma \times \vec{v})_1 = -2\,\mathrm{Im}\sum_{k\neq j} \frac{\langle(\partial_2 H)\psi_j, \psi_k\rangle\langle\psi_k(\partial_3 H)\psi_j\rangle}{(E_k - E_j)^2}. \tag{E.42}$$

Similar formulas are easily obtained for the other components.

Appendix F

Formal Scattering Theory

F.1. Bound states and scattering states

The states of a particle are usually distinguished into *bound states* and *scattering states*, according to their behavior in space and time. We define the *subspace of bound states* $\mathfrak{H}_{\text{bound}}(H)$ as the Hilbert space spanned by the linear combinations of eigenstates of the Hamiltonian H. In general, bound states are not stationary, because linear combinations of eigenstates depend on time in a nontrivial way. Geometrically, one may characterize bound states by the property that they remain essentially localized within a sufficiently large sphere. More precisely, a particle is in a bound state if the probability of finding it outside a sphere of radius R can be made arbitrarily small for all times by choosing R large. This property is described by the formula

$$\lim_{R\to\infty} \sup_{t\in\mathbb{R}} \int_{|\mathbf{x}|>R} \big|\psi(\mathbf{x},t)\big|^2 \, d^3x = 0 \qquad \text{(bound state)}. \tag{F.1}$$

Scattering states, on the other hand, escape toward infinity, as time goes to infinity. The probability of finding a scattered particle inside a sphere of radius R becomes arbitrarily small for large times $|t|$, no matter how large R is. The wave function $\psi(t,\mathbf{x})$ describes a scattering state if for all $R>0$

$$\lim_{|t|\to\infty} \int_{|\mathbf{x}|\le R} \big|\psi(\mathbf{x},t)\big|^2 \, d^3x = 0 \qquad \text{(scattering state)}. \tag{F.2}$$

In this chapter, we are concerned with the scattering states of a particle moving under the influence of an external force. Bound states can only occur if the force attracts the particle to some region of space. Scattering states can occur if the external force is repulsive or if the binding force goes to zero for large distances from the origin. In most cases of physical relevance,[1] the *subspace of scattering states* consists of precisely those states that are

[1]Other types of behavior are possible if the Hamiltonian has a so-called singularly continuous spectrum (which is usually not the case). Here, we assume that the scattering states belong to the absolutely continuous spectrum of H.

orthogonal to all bound states, and we have

$$\mathfrak{H} = \mathfrak{H}_{\text{bound}}(H) \oplus \mathfrak{H}_{\text{scatt}}(H). \tag{F.3}$$

We assume that the Hamiltonian operator H is of the form

$$H = H_0 + V \quad \text{with } H_0 = \text{generator of free time evolution.} \tag{F.4}$$

The operator H_0, the *free-particle Hamiltonian*, describes the kinetic energy of the particle. This could be, for example, the Schrödinger operator $(-1/2\mathrm{m})\Delta$ in the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^n)$, or the Dirac operator $c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta\mathrm{m}c^2$ in the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^3)^4$. For the free-particle Hamiltonian H_0 we have

$$\mathfrak{H}_{\text{bound}}(H_0) = \emptyset, \quad \mathfrak{H}_{\text{scatt}}(H_0) = \mathfrak{H}. \tag{F.5}$$

The operator V, usually an operator of multiplication by a function $V(\mathbf{x})$ of the position, describes the potential energy of the particle. We assume that $V(\mathbf{x})$ goes to zero, as $|x| \to \infty$. Hence, for large times, any scattering state leaves the region in space where it is influenced by the external force, and its time evolution will asymptotically resemble the free time evolution. The asymptotic comparison of the actual motion with a simpler time evolution (the free motion) is one of the main goals of scattering theory.

F.2. Asymptotic completeness

We are now going to formulate the main requirements for scattering theory. These requirements can only be fulfilled if the potential goes to zero sufficiently fast, as $|\mathbf{x}| \to \infty$. First one demands that there are sufficiently many scattering states.

Existence:

For every vector $\phi \in \mathfrak{H}$ there exist scattering states $\psi_\pm \in \mathfrak{H}_{\text{scatt}}(H)$ such that

$$\mathrm{e}^{-\mathrm{i}H_0 t}\phi - \mathrm{e}^{-\mathrm{i}Ht}\psi_\pm \to 0, \quad \text{as } t \to \pm\infty. \tag{F.6}$$

In many cases of physical interest, the following assertion will also be true. It says that the motion of every scattering state is asymptotically (in time) equal to a freely evolving state.

Completeness:

For every scattering state $\psi \in \mathfrak{H}_{\text{scatt}}(H)$ there are vectors ϕ_{in} and ϕ_{out} in $\mathfrak{H}$ such that

$$\mathrm{e}^{-\mathrm{i}Ht}\psi - \mathrm{e}^{-\mathrm{i}H_0 t}\phi_{\substack{\text{out}\\\text{in}}} \to 0, \quad \text{as } t \to \pm\infty. \tag{F.7}$$

The freely evolving vectors ϕ_{in} and ϕ_{out} in (F.7) are called the *incoming* and *outgoing asymptotes* of the scattering state ψ. Hence, the existence property means that every state in the Hilbert space of the system is the asymptote of a scattering state. Completeness means that every scattering state has asymptotes.

Asymptotic completeness:

A scattering system is called *asymptotically complete* if it has both the existence and the completeness property.

F.3. Wave operators

The condition (F.6) is equivalent to

$$\lim_{t\to\pm\infty} \|\mathrm{e}^{-\mathrm{i}H_0 t}\phi - \mathrm{e}^{-\mathrm{i}Ht}\psi_\pm\| = \lim_{t\to\pm\infty} \|\mathrm{e}^{\mathrm{i}Ht}\mathrm{e}^{-\mathrm{i}H_0 t}\phi - \psi_\pm\| = 0. \tag{F.8}$$

This follows from the fact that $\|\mathrm{e}^{\mathrm{i}Ht}\psi\| = \|\psi\|$ for all ψ (that is, from the unitarity of $\mathrm{e}^{\mathrm{i}Ht}$). The existence condition requires that for each $\phi \in \mathfrak{H}$, we can find vectors $\psi_\pm$ such that the above limit exists, that is, existence of

$$\Omega_{\substack{\text{out}\\\text{in}}}\phi = \lim_{t\to\pm\infty} \mathrm{e}^{\mathrm{i}Ht}\,\mathrm{e}^{-\mathrm{i}H_0 t}\,\phi \quad \text{for all } \phi \in \mathfrak{H}. \tag{F.9}$$

Here, we introduced the *Møller wave operators* Ω_{in} and Ω_{out}. These operators are well defined (everywhere on the Hilbert space) if and only if the existence property holds. The Møller operators are so-called *strong limits* of the unitary operators $U(t) = \mathrm{e}^{\mathrm{i}Ht}\mathrm{e}^{-\mathrm{i}H_0 t}$. The following mathematical theorem applies to this situation.

THEOREM F.1. *Let Ω be the (strong) limit of unitary operators, that is,*

$$\Omega\,\psi = \lim_{t\to\infty} U(t)\,\psi \quad \textit{for all } \psi \in \mathfrak{H}. \tag{F.10}$$

Then Ω is isometric, that is,

$$\langle \Omega\psi, \Omega\phi\rangle = \langle \psi, \phi\rangle \quad \textit{for all } \psi \textit{ and } \phi \in \mathfrak{H}. \tag{F.11}$$

The range of Ω is a closed subspace of $\mathfrak{H}$, characterized by an orthogonal projection operator P,

$$\Omega : \mathfrak{H} \longrightarrow \mathrm{Ran}\,\Omega = P\mathfrak{H}. \tag{F.12}$$

We have

$$\Omega^\dagger\Omega = \mathbf{1}, \qquad \Omega\Omega^\dagger = P, \tag{F.13}$$

and

$$\Omega^\dagger\,\psi = \lim_{t\to\infty} U(t)^\dagger\,P\,\psi \quad \textit{for all } \psi \in \mathfrak{H}. \tag{F.14}$$

Equation (F.13) shows that the restriction of $\Omega^\dagger$ to $\mathrm{Ran}\,\Omega$ is the inverse of Ω. This restriction is an isometric operator that maps $\mathrm{Ran}\,\Omega$ onto $\mathfrak{H}$.

The completeness property (F.7) means that for all $\psi \in \mathfrak{H}_{\mathrm{scatt}}(H)$, we can find $\phi_\pm$ such that

$$\psi - \mathrm{e}^{\mathrm{i}Ht}\mathrm{e}^{-\mathrm{i}H_0t}\phi_\pm \to 0, \quad \text{as } t \to \pm\infty. \tag{F.15}$$

This just means that every $\psi \in \mathfrak{H}_{\mathrm{scatt}}(H)$ is in the range of the Møller operators

$$\mathrm{Ran}\,\Omega_{\substack{\mathrm{out}\\ \mathrm{in}}} = \mathfrak{H}_{\mathrm{scatt}}(H). \tag{F.16}$$

Asymptotic completeness holds whenever the Møller operators have the domain $\mathfrak{H}$ and the range $\mathfrak{H}_{\mathrm{scatt}}(H)$:

Asymptotic completeness:

The scattering system characterized by the Hamiltonians H and H_0 is asymptotically complete if and only if

$$\Omega_{\substack{\mathrm{out}\\ \mathrm{in}}} : \mathfrak{D}(\Omega_{\substack{\mathrm{out}\\ \mathrm{in}}}) = \mathfrak{H} \longrightarrow \mathrm{Ran}\,\Omega_{\substack{\mathrm{out}\\ \mathrm{in}}} = \mathfrak{H}_{\mathrm{scatt}}(H). \tag{F.17}$$

Ψ *Proof of Theorem F.1*: The isometry (F.11) follows from the continuity of the scalar product in both factors. $\Omega^\dagger\Omega = \mathbf{1}$ follows from

$$\langle \psi\,, \Omega^\dagger\Omega\,\phi\rangle = \langle \Omega\psi\,, \Omega\phi\rangle = \langle \psi\,, \phi\rangle \quad \text{all } \psi,\, \phi \in \mathfrak{H}. \tag{F.18}$$

Define $P = \Omega\Omega^\dagger$. This operator is defined everywhere, self-adjoint

$$P^\dagger = (\Omega\Omega^\dagger)^\dagger = \Omega^{\dagger\dagger}\Omega^\dagger = \Omega\Omega^\dagger = P, \tag{F.19}$$

and idempotent

$$P^2 = \Omega\Omega^\dagger\Omega\Omega^\dagger = \Omega\mathbf{1}\Omega^\dagger = P, \tag{F.20}$$

and hence an orthogonal projection operator. The definition $P = \Omega\Omega^\dagger$ implies that $\mathrm{Ran}\,P \subset \mathrm{Ran}\,\Omega$. On the other hand,

$$\Omega = \Omega\mathbf{1} = \Omega\Omega^\dagger\Omega = P\Omega \tag{F.21}$$

implies that $\operatorname{Ran}\Omega \subset \operatorname{Ran} P$. Hence, $\operatorname{Ran}\Omega = \operatorname{Ran} P = P\mathfrak{H}$. Unitarity of $U(t)$ implies with (F.10) that

$$0 = \lim_{t\to\infty} \|\Omega\psi - U(t)\psi\| = \lim_{t\to\infty} \|U(t)^\dagger \Omega\psi - \psi\|, \tag{F.22}$$

or

$$\psi = \lim_{t\to\infty} U(t)^\dagger \Omega\psi. \tag{F.23}$$

Let $\psi = \Omega^\dagger\phi$ for some (arbitrary) $\phi \in \mathfrak{H}$. Then

$$\Omega^\dagger\phi = \lim_{t\to\infty} U(t)^\dagger \Omega\Omega^\dagger\phi = \lim_{t\to\infty} U(t)^\dagger P\phi, \tag{F.24}$$

which proves (F.14). □

F.4. The scattering operator

For an asymptotically complete scattering system, the Møller operators are isometric with

$$\Omega_{\text{out}}^\dagger \Omega_{\text{out}} = \mathbf{1}, \qquad \Omega_{\text{out}} \Omega_{\text{out}}^\dagger = P_{\text{scatt}}(H) \tag{F.25}$$

(and similarly for Ω_{in}). Here, $P_{\text{scatt}}(H)$ is the orthogonal projection operator onto the subspace of scattering states.

We define the scattering operator

$$S = \Omega_{\text{out}}^\dagger \, \Omega_{\text{in}}. \tag{F.26}$$

In case of an asymptotically complete scattering system, Ω_{in} maps $\mathfrak{H}$ isometrically onto $\mathfrak{H}_{\text{scatt}}(H)$, and $\Omega_{\text{out}}^\dagger$ maps $\mathfrak{H}_{\text{scatt}}(H)$ isometrically onto $\mathfrak{H}$. Hence, S is an isometry with domain $\mathfrak{H}$ and range $\mathfrak{H}$, that is, S is a unitary operator.

Unitarity of the scattering operator:

For every $\psi \in \mathfrak{H}$, we have

$$S\,\psi = \Omega_{\text{out}}^\dagger \, \Omega_{\text{in}}\, \psi = \lim_{t\to\infty} \mathrm{e}^{\mathrm{i}H_0 t}\, \mathrm{e}^{-2\mathrm{i}Ht}\, \mathrm{e}^{\mathrm{i}H_0 t}\, \psi. \tag{F.27}$$

The scattering operator S is unitary if and only if the scattering system is asymptotically complete.

The scattering operator S maps an incoming asymptote onto the corresponding outgoing asymptote as shown in Figure F.1. A non-unitary scattering operator would mean that there are incoming asymptotes for which there are no outgoing asymptotes.

It is interesting to note that asymptotic completeness in the form stated above does not hold for the Coulomb potential. The Coulomb potential is a so-called long-range potential that keeps influencing the particles even asymptotically. This situation requires a special treatment. Asymptotic

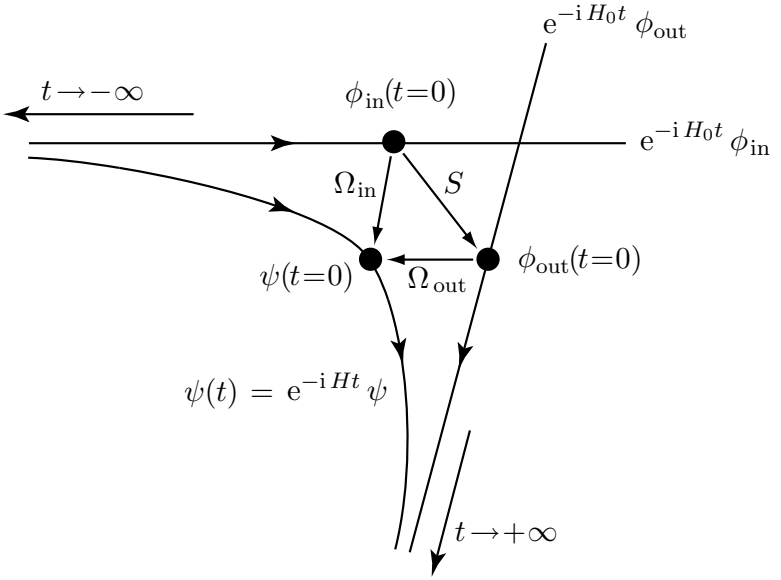


FIGURE F.1. The freely evolving asymptotes of a scattering state are symbolized by straight lines. The Møller operators map the asymptotic states at $t = 0$ onto the scattering state at $t = 0$. The unitary scattering operator maps the incoming asymptote to the outgoing asymptote.

completeness in the sense described above can only hold for short-range potentials, which go to zero faster than the Coulomb potential, as $|\mathbf{x}| \to \infty$. Let us quote the following theorem:

THEOREM F.2. *Assume that $H = H_0 + V(\mathbf{x})$ is self-adjoint on $\mathfrak{D}(H_0)$, where $H_0 = -(1/2\mathrm{m})\,\Delta$ is the free-particle Schrödinger operator. Let $V(\mathbf{x})$ be a short-range potential,*[2] *that is, there is some $R > 0$ such that*

$$|V(\mathbf{x})| \leq \frac{1}{|\mathbf{x}|^{1+\delta}} \quad \textit{for some } \delta > 0 \textit{ and all } |\mathbf{x}| > R. \tag{F.28}$$

Then the scattering system is asymptotically complete.

F.5. Properties of wave and scattering operators

Every self-adjoint operator H generates a one-parameter unitary group e^{-iHt} (see Book One, Sections 6.1–6.3). Using the continuity of the operator e^{-iHt}

[2]We refer to the mathematical physics literature for more general short-range conditions.

and the group property, we find

$$\begin{aligned}
\mathrm{e}^{-\mathrm{i}Ht}\,\Omega_{\substack{\mathrm{out}\\\mathrm{in}}} &= \mathrm{e}^{-\mathrm{i}Ht}\lim_{\tau\to\pm\infty}\mathrm{e}^{\mathrm{i}H\tau}\,\mathrm{e}^{-\mathrm{i}H_0\tau} = \lim_{\tau\to\pm\infty}\mathrm{e}^{-\mathrm{i}Ht}\,\mathrm{e}^{\mathrm{i}H\tau}\,\mathrm{e}^{-\mathrm{i}H_0\tau}\\
&= \lim_{\tau\to\pm\infty}\mathrm{e}^{\mathrm{i}H(\tau-t)}\,\mathrm{e}^{-\mathrm{i}H_0\tau} = \lim_{\tau-t\to\pm\infty}\mathrm{e}^{\mathrm{i}H(\tau-t)}\,\mathrm{e}^{-\mathrm{i}H_0\tau}\\
&= \lim_{s\to\pm\infty}\mathrm{e}^{\mathrm{i}Hs}\,\mathrm{e}^{-\mathrm{i}H_0(s+t)} = \lim_{s\to\pm\infty}\mathrm{e}^{\mathrm{i}Hs}\,\mathrm{e}^{-\mathrm{i}H_0s}\,\mathrm{e}^{-\mathrm{i}H_0t}\\
&= \Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,\mathrm{e}^{-\mathrm{i}H_0t} \qquad\qquad (\mathrm{F}.29)
\end{aligned}$$

This "commutation relation" is called the *intertwining property* of the Møller operators. The intertwining relation carries over to the generators H and H_0. We write the generator in terms of the unitary group:

$$H\,\psi = \mathrm{i}\,\frac{d}{dt}\,\mathrm{e}^{-\mathrm{i}Ht}\,\psi\Big|_{t=0} = \mathrm{i}\lim_{t\to0}\frac{1}{t}(\mathrm{e}^{-\mathrm{i}Ht}-\mathbf{1})\,\psi, \qquad (\mathrm{F}.30)$$

for all $\psi\in\mathfrak{D}(H)$. Hence, we obtain, using (F.29) and the fact that the Møller operators are bounded and hence continuous,

$$\begin{aligned}
\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,H_0\,\psi &= \mathrm{i}\lim_{t\to0}\frac{1}{t}\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,(\mathrm{e}^{-\mathrm{i}H_0t}-\mathbf{1})\,\psi\\
&= \mathrm{i}\lim_{t\to0}\frac{1}{t}(\mathrm{e}^{-\mathrm{i}Ht}-\mathbf{1})\,\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,\psi = H\,\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,\psi. \qquad (\mathrm{F}.31)
\end{aligned}$$

Hence, whenever $\psi\in\mathfrak{D}(H_0)$, we find that $\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,\psi\in\mathfrak{D}(H)$ and

$$\Omega_{\substack{\mathrm{out}\\\mathrm{in}}}\,H_0 = H\,\Omega_{\substack{\mathrm{out}\\\mathrm{in}}} \quad \text{on } \mathfrak{D}(H_0). \qquad (\mathrm{F}.32)$$

An analogous relation holds for the adjoint operators

$$\Omega^\dagger_{\substack{\mathrm{out}\\\mathrm{in}}}\,H = H_0\,\Omega^\dagger_{\substack{\mathrm{out}\\\mathrm{in}}} \quad \text{on } \mathfrak{D}(H). \qquad (\mathrm{F}.33)$$

From this we conclude that the scattering operator $S=\Omega^\dagger_{\mathrm{out}}\Omega_{\mathrm{in}}$ commutes with the free-particle Hamiltonian.

Conservation of the kinetic energy:

In an asymptotically complete scattering system, the S operator commutes with H_0,

$$S\,H_0 = H_0\,S \quad \text{on } \mathfrak{D}(H_0). \qquad (\mathrm{F}.34)$$

The kinetic energy of the incoming particles is the same as the kinetic energy of the outgoing particles.

F.6. Scattering operator in the energy representation

Mathematically, the most important consequence of $[S, H_0] = 0$ is that the scattering operator S is diagonal in the spectral representation of H_0. The *spectral representation* of H_0 is that representation where the operator is just multiplication by E. Here, it is called *energy representation*, because H_0 describes the energy of a free particle (see also Book One, Section 3.8). The transformation that links the Hilbert space $L^2(\mathbb{R}^3)$ with the spectral representation space is called the *spectral transformation*. For $H_0 = -\Delta/(2\mathrm{m})$ the spectral transformation is given by the Fourier transformation, the transition to polar coordinates in momentum space, and the variable substitution $k \to E = k^2/(2\mathrm{m})$. For a given wave packet ψ, we consider its Fourier transformation $\hat{\psi}(\mathbf{k})$ and introduce spherical coordinates in momentum space, $\mathbf{k} \equiv (k, \omega)$, where $\omega = (\vartheta, \varphi)$ are spherical angles. Then we define

$$g(E, \omega) = (2\mathrm{m}^3 E)^{1/4}\, \hat{\psi}(\sqrt{2\mathrm{m}E}, \omega). \tag{F.35}$$

The function g is considered an element of the Hilbert space

$$\mathfrak{K} = L^2([0, \infty), dE) \otimes L^2(S^2, d\omega). \tag{F.36}$$

Here, as usual, S^2 is the unit sphere, and $d\omega = \sin\vartheta\, d\vartheta\, d\varphi$. Note that

$$\begin{aligned} \|\psi\|^2 &= \int |\psi(\mathbf{x})|^2\, d^3x = \int |\hat{\psi}(\mathbf{k})|^2\, d^3k \\ &= \int_0^\infty k^2\, dk \int_{S^2} d\omega\, |\hat{\psi}(k, \omega)|^2 \\ &= \int_0^\infty dE \int_{S^2} d\omega\, (2\mathrm{m}^3 E)^{1/2}\, |\hat{\psi}(\sqrt{2\mathrm{m}E}, \omega)|^2. \end{aligned} \tag{F.37}$$

In the last step, we performed the variable substitution $k^2 = 2\mathrm{m}E$, $dk = \sqrt{\mathrm{m}/2E}\, dE$. Hence, we obtain

$$\|\psi\|^2 = \int |g(E, \omega)|^2\, dE\, d\omega, \tag{F.38}$$

that is, the norm of ψ in $L^2(\mathbb{R}^3, d^3x)$ equals the norm of g in the Hilbert space $\mathfrak{K}$. Hence, there is a unitary relation between $\psi(\mathbf{x})$ and $g(E, \omega)$. We know already that in momentum space, H_0 just becomes multiplication by $k^2/(2\mathrm{m})$. For the functions $g(E, \omega)$, this just amounts to multiplication by the variable E:

Energy representation:

The energy representation g of a wave packet ψ is given by (F.35). In the energy representation, the operator H_0 acts as multiplication by the variable E

$$H_0 : g(E,\cdot) \to E\, g(E,\cdot). \tag{F.39}$$

Hence, if ψ corresponds to g as described above, then $H_0\psi$ corresponds to Eg. Note that for fixed $E \geq 0$, $g(E,\cdot)$ is a square-integrable function on the unit sphere S^2, that is, an element of $L^2(S^2)$. One can prove the following result:

Scattering operator:

The action of any bounded (self-adjoint, unitary) operator S that commutes with H_0 can be described in the spectral representation of H_0 by

$$S : g(E,\cdot) \to S(E)\, g(E,\cdot), \tag{F.40}$$

where $S(E)$ is for every E a bounded (self-adjoint, unitary) operator in $L^2(S^2)$.

In scattering theory, the operator $S(E)$ is called the *on-shell scattering operator*.

F.7. Stationary Scattering Theory

The stationary free-particle Schrödinger equation $H_0\phi = E\phi$ has plane wave solutions $\phi(\mathbf{k},\mathbf{x}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$, with $k^2/(2\mathrm{m}) = E$. For energies E in the continuous range of scattering energies,[3] we expect that the stationary Schrödinger equation

$$H\psi = E\psi \tag{F.41}$$

also has plane-wave like solutions that are not square-integrable. These solutions are asymptotically similar to plane waves for $|\mathbf{x}| \to \infty$.

We can obtain these solutions heuristically as follows. Scattering states are related to solutions of the free-particle Schrödinger equation by the Møller operator Ω_-. Hence, a free plane-wave ϕ should be related to a plane-wave like solution ψ by

$$\phi = \Omega_-^\dagger \psi = \lim_{t\to-\infty} \mathrm{e}^{\mathrm{i}H_0 t}\, \mathrm{e}^{-\mathrm{i}Ht}\, \psi. \tag{F.42}$$

[3]Usually, if the potential goes to zero as $|\mathbf{x}| \to \infty$, these are the positive energies $0 < E < \infty$.

For any self-adjoint operator H we have

$$\frac{d}{dt}\,\mathrm{e}^{-\mathrm{i}Ht} = -\mathrm{i}H\,\mathrm{e}^{-\mathrm{i}Ht} = \mathrm{i}\,\mathrm{e}^{-\mathrm{i}Ht}\,H. \tag{F.43}$$

Hence, using the product rule of differentiation, we may write

$$\frac{d}{ds}\,\mathrm{e}^{\mathrm{i}H_0 s}\,\mathrm{e}^{-\mathrm{i}Hs} = \mathrm{i}\,\mathrm{e}^{\mathrm{i}H_0 s}\,(H_0 - H)\,\mathrm{e}^{-\mathrm{i}Hs} = -\mathrm{i}\,\mathrm{e}^{\mathrm{i}H_0 s}\,V\,\mathrm{e}^{-\mathrm{i}Hs}. \tag{F.44}$$

Now we obtain, by a formal application of the fundamental theorem of calculus,

$$\begin{aligned}\mathrm{e}^{\mathrm{i}H_0 t}\,\mathrm{e}^{-\mathrm{i}Ht} &= \mathbf{1} + \int_0^t \Big(\frac{d}{ds}\,\mathrm{e}^{\mathrm{i}H_0 s}\,\mathrm{e}^{-\mathrm{i}Hs}\Big)\,ds \\ &= \mathbf{1} - \mathrm{i}\int_0^t \mathrm{e}^{\mathrm{i}H_0 s}\,V\,\mathrm{e}^{-\mathrm{i}Hs}\,ds. \end{aligned}\tag{F.45}$$

From $H\psi = E\psi$, we conclude that $\mathrm{e}^{-\mathrm{i}Hs}\psi = \mathrm{e}^{-\mathrm{i}Es}\psi$, and hence

$$\begin{aligned}\phi &= \psi - \mathrm{i}\lim_{t\to-\infty}\int_0^t \mathrm{e}^{\mathrm{i}H_0 s}\,V\,\mathrm{e}^{-\mathrm{i}Hs}\,\psi\,ds \\ &= \psi - \mathrm{i}\lim_{t\to-\infty}\int_0^t \mathrm{e}^{\mathrm{i}(H_0-E)s}\,V\,\psi\,ds \\ &= \psi - \mathrm{i}\int_0^{-\infty} \mathrm{e}^{\mathrm{i}(H_0-E)s}\,V\,\psi\,ds. \end{aligned}\tag{F.46}$$

The problem with this expression is that we cannot expect that $\mathrm{e}^{\mathrm{i}(H_0-E)s}$ goes to zero for $s \to -\infty$. Therefore, the integral above does not exist, unless we introduce some regularization. Hence, we write

$$\phi = \psi - \mathrm{i}\lim_{\epsilon\to+0}\int_0^{-\infty} \mathrm{e}^{\epsilon s}\,\mathrm{e}^{\mathrm{i}(H_0-E)s}\,ds\,V\,\psi. \tag{F.47}$$

The exponential factor $\exp(\epsilon s)$ decays exponentially, as $s \to -\infty$, whenever $\epsilon > 0$. In this case, we expect the integral to converge, and (by an elementary integration)

$$-\mathrm{i}\int_0^{-\infty} \mathrm{e}^{\epsilon s}\,\mathrm{e}^{\mathrm{i}(H_0-E)s}\,ds = \big(H_0 - (E + \mathrm{i}\epsilon)\big)^{-1}. \tag{F.48}$$

Hence, stationary scattering theory has to investigate the limit of the resolvent $(H_0 - z)^{-1}$, as z approaches the positive real axis from above. The equation (F.46) now becomes the *Lippmann-Schwinger equation*

$$\psi = \phi - \lim_{\epsilon\to+0}\big(H_0 - (E + \mathrm{i}\epsilon)\big)^{-1}\,V\,\psi. \tag{F.49}$$

In case of the Schrödinger equation, free plane waves with energy $E => 0$ are given by $\phi(\mathbf{k}, \mathbf{x}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$ with $E = k^2/2\mathrm{m}$. The free resolvent

is a well-known integral operator, and the Lippmann-Schwinger equation becomes

$$\psi(\mathbf{k},\mathbf{x}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} - \frac{\mathrm{m}}{2\pi}\int_{\mathbb{R}^3} \frac{\mathrm{e}^{\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}\, V(\mathbf{y})\, \psi(\mathbf{k},\mathbf{y})\, d^3y. \tag{F.50}$$

Solutions of the Lippmann-Schwinger equation give the plane-wave like solutions of the stationary Schrödinger equation. We may use these stationary plane waves to build wave packets, as described earlier:

$$\psi(\mathbf{x}) = \int_{\mathbb{R}^3} g(\mathbf{k})\, \psi(\mathbf{k},\mathbf{x})\, d^3k. \tag{F.51}$$

will be a wave packet in the subspace of scattering states. Its time evolution will be given by

$$\psi(\mathbf{x},t) = \int_{\mathbb{R}^3} g(\mathbf{k})\, \exp\bigl(-\mathrm{i}\frac{k^2}{2\mathrm{m}}t\bigr)\, \psi(\mathbf{k},\mathbf{x})\, d^3k. \tag{F.52}$$

Ψ The integral kernel of the resolvent,

$$G^+(E,\mathbf{x},\mathbf{y}) = \frac{\mathrm{m}}{2\pi}\,\frac{\mathrm{e}^{\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}, \quad \text{where } k=\sqrt{2\mathrm{m}E} \tag{F.53}$$

is the Green function of the free-particle Schrödinger equation, that is, a solution of the distributional equation

$$\Bigl(-\frac{1}{2\mathrm{m}}\Delta - E\Bigr)\, G(E,\mathbf{x},\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}) \tag{F.54}$$

(where δ is Dirac's delta distribution). This equation has two solutions,

$$G^\pm(E,\mathbf{x},\mathbf{y}) = \frac{\mathrm{m}}{2\pi}\,\frac{\mathrm{e}^{\pm\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}, \quad \text{where } k=\sqrt{2\mathrm{m}E}. \tag{F.55}$$

These solutions are distinguished by their behavior at infinity. G^+ describes an outgoing spherical wave and G^- an incoming wave. Using the inverse Fourier transform, the Green functions can be written as

$$G^\pm(E,\mathbf{x},\mathbf{y}) = \frac{\mathrm{m}}{2\pi}\int_{\mathbb{R}^3} \frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{k^2/2\mathrm{m} - E \pm \mathrm{i}\epsilon}\, d^3k. \tag{F.56}$$

F.8. Scattering amplitude

We want to investigate the asymptotic behavior of the solutions of the Lippmann-Schwinger equation (F.50). In order to avoid technicalities, we assume that the potential $V(\mathbf{x})$ vanishes outside a finite region. We consider $\mathbf{x}$ with $|\mathbf{x}| \gg |\mathbf{y}|$ for all $\mathbf{y}$ in the region contributing to the integral in (F.50). Then, we can approximate

$$|\mathbf{x}-\mathbf{y}| = \sqrt{x^2 - 2\mathbf{x}\cdot\mathbf{y} + y^2} \approx |\mathbf{x}| - \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|} + \mathrm{O}\Bigl(\frac{1}{x^2}\Bigr). \tag{F.57}$$

Keeping only the leading terms,

$$\frac{\mathrm{e}^{\mathrm{i}k\,|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} \approx \frac{\mathrm{e}^{\mathrm{i}k\,|\mathbf{x}|}}{|\mathbf{x}|}\,\exp\Big(-\mathrm{i}k\frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|}\Big). \tag{F.58}$$

Let us introduce the vector

$$\mathbf{k}' = k\,\frac{\mathbf{x}}{|\mathbf{x}|} \quad \text{with } |\mathbf{k}'| = |\mathbf{k}| = k. \tag{F.59}$$

The asymptotic form of the Lippmann-Schwinger equation (F.50) for large $|\mathbf{x}|$ now becomes

$$\psi(\mathbf{k},\mathbf{x}) \approx \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} - \frac{\mathrm{e}^{\mathrm{i}k\,|\mathbf{x}|}}{|\mathbf{x}|}\,\frac{\mathrm{m}}{2\pi}\int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\mathbf{k}'\cdot\mathbf{y}}\,V(\mathbf{y})\,\psi(\mathbf{k},\mathbf{y})\,d^3y. \tag{F.60}$$

By choosing a suitable frame of reference, we may assume that $\mathbf{k}$ points in the positive z-direction, such that $\mathbf{k} = (0,0,k)$. The integral can then be written as a function of the spherical coordinates (k,ϑ,φ) of $\mathbf{k}'$.

Asymptotic behavior for $|\mathbf{x}| \to \infty$:

The asymptotic form of a solution $\psi(\mathbf{k},\mathbf{x})$ of the Lippmann-Schwinger equation is given by

$$\psi(\mathbf{k},\mathbf{x}) \approx \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} + \frac{\mathrm{e}^{\mathrm{i}k\,|\mathbf{x}|}}{|\mathbf{x}|}\,f(k,\vartheta,\varphi), \quad \text{as } |\mathbf{x}| \to \infty. \tag{F.61}$$

Here, (ϑ,φ) are the spherical angles of $\mathbf{x}$ in a coordinate frame where $\mathbf{k} = (0,0,k)$. Equation (F.61) is the superposition of an *incoming plane wave* $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}$ moving in the positive z-direction and an outgoing spherical wave with an angular-dependent amplitude

$$f(k,\vartheta,\varphi) = -\frac{\mathrm{m}}{2\pi}\int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}\mathbf{k}'\cdot\mathbf{y}}\,V(\mathbf{y})\,\psi(\mathbf{k},\mathbf{y})\,d^3y, \quad \mathbf{k}' = k\,\frac{\mathbf{x}}{|\mathbf{x}|}, \tag{F.62}$$

called the *scattering amplitude*.

F.9. Scattering cross section

The prototype of a scattering experiment is Rutherford's classical investigation of the diffraction of alpha particles by a gold foil, which he began in 1909 and which led to the discovery of the atomic nucleus. This type of scattering experiment is depicted schematically in Figure F.2. It shows a source sending a beam of particles toward a target. As indicated, the frame of reference is usually oriented such that the incoming particles move in the z-direction. A detector is located far away from the target in a direction

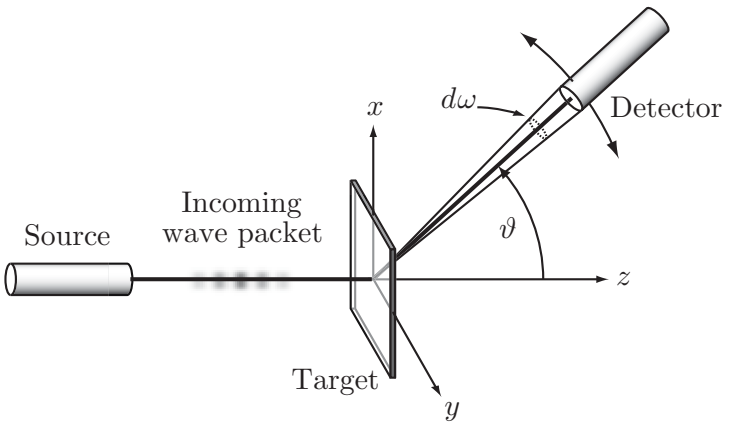


FIGURE F.2. A typical scattering experiment.

given by spherical angles $\omega = (\vartheta, \varphi)$. It counts the number dN of particles that are deflected per unit time into a cone of opening angle $d\omega$ about ω.

The experiment is usually done in such a way that the incoming particles have a well-defined momentum. Hence, we may assume that the energy is sharply concentrated around some average energy E. In a situation where the energy of the incident particle is not changed by the collision, the outgoing particles might have different momenta, but the same energy distribution as the incoming particles.

We assume further that the particles in the incident beam do not interact with one another. This condition can be met if the beam has a low intensity (that is, if it consists of very few particles per unit volume). A measure for this is the incident flux f_{in}, the number of particles crossing per unit time a unit surface perpendicular to the direction of the incident beam.

Here, we assume that the target consists of individual scattering centers (atoms, nuclei) that are sufficiently separated that each particle of the incident beam interacts only with one of the scattering centers in the target. Often, the target is only a thin foil in order to reduce the probability of multiple scattering events.

If all these conditions (low density of particles in the incoming beam and in the target) are met, we can assume that the whole experiment actually consists of a large ensemble of independent scattering processes. Each elementary experiment consists in the scattering of a single particle by the external force produced by a single scattering center. In Rutherford's scattering experiment, the external force is the Coulomb force of an atomic nucleus at some fixed position inside the target.

The experimental setup is designed to determine the probability that an incoming particle after interaction with the target is finally found in a cone defined by the solid angle $d\omega$ about ω. The number $dN(E,\omega)$ of particles scattered per unit time into the cone will depend on the energy of the incoming particles and on the direction ω. Moreover, we may assume that for fixed E and ω, the number $dN(E,\omega)$ is directly proportional to the size of the solid angle $d\omega$ and to the incoming flux f_{in},

$$dN(E,\omega) \propto f_{\text{in}}\, d\omega. \tag{F.63}$$

The constant of proportionality is known as the *differential scattering cross section*

$$\frac{d\sigma(E,\omega)}{d\omega} = \frac{dN(E,\omega)}{f_{\text{in}}\, d\omega}. \tag{F.64}$$

The integral

$$\int_{B\subset S^2} \frac{d\sigma(E,\omega)}{d\omega}\, d\omega \tag{F.65}$$

will be proportional to the probability that an incoming particle with energy E is scattered into a conical region defined by the angles $\omega = (\vartheta,\varphi) \in B \subset S^2$. (Here, S^2 is the unit sphere in three dimensions, that is, the set of all directions $\omega = (\vartheta,\varphi)$ with the volume element $d\omega = \sin\vartheta\, d\vartheta\, d\varphi$).

The *integral* or *total scattering cross section* $\sigma_{\text{tot}}(E)$ is defined by integrating the differential cross section over all directions,

$$\sigma_{\text{tot}}(E) = \int_{S^2} \frac{d\sigma(E,\omega)}{d\omega}\, d\omega. \tag{F.66}$$

It has the physical dimension of a surface.

In Book One, we defined the current density

$$\mathbf{j} = \frac{\mathrm{i}}{2\mathrm{m}} \left((\nabla\overline{\psi})\,\psi - \overline{\psi}\,(\nabla\psi) \right) \tag{F.67}$$

which describes the flow of the density $|\psi|^2$. In stationary scattering theory, $\mathbf{j}\cdot\mathbf{n}$ is assumed to be proportional to the flux of particles in the beam (that is, the number of particles that go per unit time through a unit surface defined by the normal vector $\mathbf{n}$). For the incoming wave $\psi = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})$, the current is

$$\mathbf{j}_{\text{in}} = \frac{\mathbf{k}}{\mathrm{m}}, \tag{F.68}$$

hence, the incoming flux is

$$f_{\text{in}} = c\,\mathbf{j}_{\text{in}}\cdot\mathbf{e}_z = c\frac{k}{\mathrm{m}}. \tag{F.69}$$

For the outgoing wave, we compute the radial current with the help of the radial derivative $\nabla_r = \partial/\partial r$,

$$\mathbf{j} \cdot \mathbf{e}_r = j_{\text{rad}} = \frac{\mathrm{i}}{2\mathrm{m}} \Big(\big(\frac{\partial \overline{\psi}}{\partial r}\big)\, \psi - \overline{\psi}\, \big(\frac{\partial \psi}{\partial r}\big) \Big). \tag{F.70}$$

For the radial wave $(1/r)\, \exp(\mathrm{i}kr)\, f(k, \vartheta, \varphi)$, we easily compute the radial current at a distance r from the origin as

$$j_{\text{rad}} = \frac{k}{\mathrm{m}\, r^2} |f(k, \vartheta, \varphi)|^2. \tag{F.71}$$

The number of particles going per unit time through a spherical surface element $r^2 d\omega$ about some angle $\omega = (\vartheta, \varphi)$ is therefore

$$dN(E, \omega) = c\, j_{\text{rad}}\, r^2\, d\omega = c \frac{k}{\mathrm{m}} |f(k, \vartheta, \varphi)|^2\, d\omega. \tag{F.72}$$

Here, $E = k^2/2\mathrm{m}$ and $\omega = (\vartheta, \varphi)$. Finally, we obtain the scattering cross section as

$$\frac{d\sigma(E, \omega)}{d\omega} = \frac{dN(E, \omega)}{f_{\text{in}}\, d\omega} = |f(k, \vartheta, \varphi)|^2. \tag{F.73}$$

Appendix G

Books

This bibliography lists the books cited in the main text. A more comprehensive list of books dealing with quantum mechanics in general was given in Book One [**11**].

Bibliography

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 10th printing in 1972. ISBN: 0-486-61272-4. See also `http://dlmf.nist.gov/`.

[2] D. Bouwmeester, A.K. Ekert, and A. Zeilinger *The Physics of Quantum Information*: *Quantum Cryptography, Quantum Teleportation, Quantum Computation*. Springer-Verlag, New York, 2000. ISBN: 3-540-66778-4.

[3] J. D. Jackson. *Classical Electrodynamics*, 3rd edition. John Wiley & Sons, 1998. ISBN: 0-471-30932-X.

[4] T. Kato. *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, Heidelberg, 1980 (reprint 2003). ISBN: 3-540-58661-X.

[5] M.A. Nielsen and I.L. Chuang . *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000. ISBN: 0-521-63503-9.

[6] J. Preskill. *Physics 229: Quantum Computation and Information*, California Institute of Technology, Pasadena, 1998.
See `http://www.theory.caltech.edu/people/preskill/ph229/`.

[7] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: IV Analysis of Operators*. Academic Press, New York, 1978. ISBN: 0-125-85004-2.

[8] H. Rauch and S.A. Werner. *Neutron Interferometry: Lessons in Experimental Quantum Mechanics*. Oxford University Press, Oxford, 2000. ISBN: 0-198-50027-0.

[9] R.U. Sexl and H.K. Urbantke. *Relativity, Groups, Particles: Special Relativity and Relativistic Symmetry in Field and Particle Physics*, Revised edition. Springer-Verlag, New York, 2001. ISBN 3-211-83443-5.

[10] B. Thaller. *The Dirac Equation*. Springer-Verlag, Berlin, Heidelberg, New York, 1992. ISBN: 3-540-54883-1.

[11] B. Thaller. *Visual Quantum Mechanics*. Springer-Verlag, New York, Berlin, Heidelberg 2000. ISBN: 0-387-98929-3.

[12] C.P. Williams and S.H. Clearwater. *Explorations in Quantum Computing*. Springer-Verlag, New York, 1998. ISBN: 0-387-94768-X.

Appendix H

Movie Index

0. Introduction

CD 0.1. **Welcome**
(1) Hello, and a quick overview
(2) What's new
(3) Sort of a disclaimer

CD 0.2. **Organization**
(1) The main menu
(2) Chapters and sections
(3) Sections and topics

CD 0.3. **Buttons**
(1) Navigation
(2) Getting additional information
(3) Hyperlinks

CD 0.4. **Movies and controllers**
(1) Animation of a time-dependent function
(2) Interactive image (parameter dependence)

CD 0.5. **Keyboard shortcuts**

CD 0.6. **About Book One**

CD 0.7. **Adjust colors**

CD 0.8. **Visualization of complex numbers**
(1) Mapping phases to colors
(2) Colors for phase and absolute value

CD 0.9. **Mathematica notebooks**

CD 0.10. **Examples from Book One – 1**
(1) Free fall – classical motion and quantum wave packet
(2) Quantum-mechanical effects dominate

CD 0.11. **Examples from Book One – 2**
(1) Particle hitting an impenetrable barrier
(2) Interference effects in front of the barrier

CD 0.12. **Examples from Book One – 3**
(1) Double-slit experiment
(2) Single-slit experiment

1. Spherical Symmetry

CD 1.1. **The rotation group**
(1) Exploring the group manifold
(2) Closed orbits on the surface
(3) Closed orbit inside
(4) Closed orbit that cannot shrink to a point

CD 1.2. **Angular momentum**
(1) Classical motion on a circle
(2) Center of motion not at the origin
(3) Linear motion
(4) Classical Coulomb motion

CD 1.3. **The rotation group**
(1) Rotations do not commute with each other
(2) The commutator of two rotations (small angles)

CD 1.4. **Spherical symmetry**
(1) Force field of the harmonic oscillator
(2) Repulsive Coulomb force

CD 1.5. **Spherical coordinates**
(1) Polar and azimuthal angle
(2) Uniform motion in a straight line
(3) Discontinuity of the azimuthal angle
(4) Motion on a circle (rigid rotator)

CD 1.6. **Spherical coordinate space**
(1) Linear motion
(2) Linear motion (discontinuous φ)
(3) Circular motion (rigid rotator)
(4) Map of the world

CD 1.7. **Spherical harmonics**
(1) Introduction
(2) Spherical plot of the absolute value
(3) Spherical plot of orbitals
(4) Map of complex values
(5) Color plot on the unit sphere

CD 1.8. **Rigid rotator**
(1) Classical orbits with given L
(2) Time-dependence of polar angle
(3) Classical vs. quantum probabilities

CD 1.9. **Quantum rotator**
(1) Superposition $|1,1\rangle + |2,2\rangle$
(2) Superposition $|1,1\rangle + |2,-1\rangle$

CD 1.10. **Spherical harmonics Quantum rotator**
(1) Superposition $|1,1\rangle + \ldots + |4,1\rangle$
(2) Superposition $|1,1\rangle + \ldots + |4,4\rangle$

CD 1.11. **Particle of a sphere**
(1) Oscillating state $|0,0\rangle + |1,0\rangle$
(2) Superposition of eigenstates
(3) Superposition of orbitals

CD 1.12. **Particle of a sphere**
(1) Superposition with random coefficients
(2) "Boltzmann distribution" of energies

CD 1.13. **Gaussian on a sphere**
(1) Particle near the equator
(2) Initial state "at rest"

CD 1.14. **Ψ on a sphere**

CD 1.15. **The Riccati-Bessel functions**

CD 1.16. **Expansion of a plane wave**

2. Coulomb Problem

CD 2.1. **Classical Coulomb problem**
(1) Motion on a Kepler ellipse
(2) Various orbits
(3) Dependence on the coupling constant
(4) Orbits with the same energy

CD 2.2. **Coulomb spectrum**
(1) Degeneracy of eigenvalues
(2) Transitions and spectroscopy
(3) Structure of spectrum in two dimensions

CD 2.3. **Radial Coulomb eigenfunctions**
(1) Introduction
(2) Gallery of radial eigenfunctions
(3) Dependence on the coupling constant

CD 2.4. **Radial Coulomb motion in two and three dimensions**
(1) Superposition with $\ell = 1$
(2) Superposition with $\ell = 3$

CD 2.5. **Coulomb functions in two dimensions**
(1) Introduction

(2) Gallery of complex-valued eigenfunctions
(3) Real-valued eigenfunctions (orbitals)

CD 2.6. **Time-evolution of eigenstates**
(1) Eigenstate $|8, 3\rangle$
(2) Orbital $|8, 3\rangle + |8, -3\rangle$

CD 2.7. **Some flowers**
(1) Eigenstate $|6, 5\rangle + |6, -5\rangle + |6, 4\rangle$
(2) Eigenstate $|6, 1\rangle + |6, -2\rangle + |6, 4\rangle$
(3) Eigenstate $|6, 5\rangle + |6, -2\rangle + |6, 4\rangle$
(4) Eigenstate $|6, 5\rangle + |6, -5\rangle + |6, 2\rangle$

CD 2.8. **Radial oscillations**
(1) $|5, 3\rangle + |6, 3\rangle$
(2) $|5, 4\rangle + \ldots + |9, 4\rangle$
(3) Centered Gaussian

CD 2.9. **Rotating states**
(1) $|5, -1\rangle + |6, 2\rangle$
(2) $|4, -1\rangle + |7, 4\rangle$
(3) $|6, -2\rangle + |7, 3\rangle$
(4) $|6, 3\rangle + |7, -3\rangle$

CD 2.10. **Simple superpositions**
(1) $|7, -3\rangle + |5, -1\rangle + |5, 1\rangle + |7, 3\rangle$
(2) Superposition with radial quantum number 3.
(3) $|m| \leq 4$, Boltzmann distribution

CD 2.11. **Various Gaussians**
(1) Centered, with nonzero momentum
(2) Initially at rest
(3) Orbiting state
(4) Revivals of the initial state

CD 2.12. **Eigenstates in three dimensions**
(1) Variable isosurface of $|7, 4, 2\rangle$
(2) Variable isosurface of $|11, 5, 3\rangle$
(3) The state $|18, 9, 5\rangle$ from all sides

CD 2.13. **Hydrogen** $|11, 5, 3\rangle$
(1) Slice parallel to the yz-plane
(2) Slice parallel to the xy-plane

CD 2.14. **Hydrogen orbitals in three dimensions**
(1) Variable isosurface of $|11, 5, |3|\rangle$ (1)
(2) Variable isosurface of $|11, 5, |3|\rangle$ (2)
(3) Variable isosurface of $|11, 5, 0\rangle$

CD 2.15. **Eigenfunctions in three dimensions**

(1) Introduction
(2) Visualization via isosurfaces
(3) Visualization via slice planes
(4) Orbitals of the first kind
(5) Orbitals of the second kind

CD 2.16. **Stationary superposition**
(1) Superposition of eigenstates $|2,4,2\rangle + |2,4,-1\rangle$
(2) Variable slice plane

CD 2.17. **Oscillations**
(1) The superposition $|3,1,0\rangle + |4,2,0\rangle$
(2) The superposition $|3,1,0\rangle + |4,1,0\rangle$
(3) The superposition $|3,1,1\rangle + |4,2,1\rangle$
(4) The superposition $|3,1,1\rangle + |4,1,1\rangle$

CD 2.18. **Rotating states**
(1) The superposition $|3,1,0\rangle + |4,2,2\rangle$
(2) The superposition $|3,1,1\rangle + |4,2,2\rangle$
(3) The superposition $|3,1,1\rangle + |4,2,-1\rangle$

CD 2.19. **Rotating state**
(1) Superposition of eigenstates $|5,1,-1\rangle + |6,3,2\rangle$
(2) Data slice of the initial state (yz-plane)

CD 2.20. **Motion of a bound state**
(1) Superposition of four eigenstates
(2) Data slice of the initial state (yz-plane)

CD 2.21. **Rydberg states**
(1) Angular part of the wave function
(2) Radial part of the wave function
(3) Radial distribution (large angular momenta)
(4) Angular motion (on the equator)

CD 2.22. **Rydberg states**
(1) Radial and angular motion
(2) Radial and angular motion
(3) Time step = Kepler period

CD 2.23. **Rydberg states**
(1) Very high angular momenta
(2) Revival of the initial state

CD 2.24. **Rydberg states in two dimensions**
(1) Radial and angular motion
(2) Radial and angular motion
(3) Radial and angular motion

CD 2.25. **Coulomb eigenstates**
(1) Separation in parabolic coordinates

3. Spin

CD 3.1. **Inhomogeneous magnetic field**
(1) Neutral particle with a magnetic moment
(2) Acceleration in direction of the field gradient
(3) Motion in a more realistic magnetic field
(4) Approximation of 1 by a realistic field

CD 3.2. **Homogeneous magnetic field**
(1) Classical electron moving on a circle
(2) Charged particle, g-factor < 2
(3) Charged particle, g-factor > 2

CD 3.3. **Homogeneous B-field**
(1) Charged particle with g-factor 4
(2) Classical electron moving on a helix

CD 3.4. **Stern-Gerlach - classical**
(1) Neutral particles with various spins
(2) Trajectories for particles with random spins
(3) Two- and four-wire fields

CD 3.5. **Model of Stern-Gerlach experiment**
(1) Type-1 particles in an inhomogeneous field
(2) Particles of type 2
(3) Superposition of the two types
(4) Another superposition

CD 3.6. **Stern-Gerlach experiment**
(1) Component with spin-up
(2) Component with spin-down

CD 3.7. **Stern-Gerlach filter**
(1) Preparing a spin-up wave packet
(2) Preparing a spin-down wave packet

CD 3.8. **Spinor-field visualization**
(1) Vectors indicating the local spin direction
(2) Method using colored "magnetic needles"
(3) Bipartite enlarged data pixels

CD 3.9. **Homogeneous B-field**
(1) The direction of the spin
(2) Interlaced colored density plots
(3) Comparison with scalar wave packet

CD 3.10. **Color map for vector fields**
(1) The colored sphere
(2) Slice parallel to the xz-plane
(3) Slice parallel to the xy-plane

CD 3.11. **Spinor to colors**
(1) Splitting into parts with spin-up/down
(2) Landau orbit in homogeneous B-field

CD 3.12. **Spin in Stern-Gerlach device**
(1) x-polarized particle through analyzer
(2) Projection onto spin-up state
(3) Projection onto spin-down state

CD 3.13. **Spinor harmonics**
(1) Introduction
(2) Color density plot of the two components
(3) Map of the spin-vector field
(4) Spin-vectors on the unit circle (xz-plane)
(5) Spin-vectors on the unit sphere

CD 3.14. **Hydrogen atom with spin**
(1) Graphics gallery

4. Qubits

CD 4.1. **States of a qubit**

CD 4.2. **Stern-Gerlach analyzer**

CD 4.3. **State preparation**

CD 4.4. **Spin transition**
(1) With observation
(2) Without observation

CD 4.5. **Transition probabilities**

CD 4.6. **Determinative measurement**
(1) Special preparation (restricted set of states)
(2) Complete measurement of a general state

CD 4.7. **Qubit in Bloch sphere**

CD 4.8. **Qubit rotations**
(1) Rotation around the z-axis
(2) Rotation around the y-axis

CD 4.9. **Stern-Gerlach interferometer**

CD 4.10. **Fun with interferometers**
(1) A double-slit experiment
(2) Rotating the qubits in one path

CD 4.11. **Interaction-free measurement (Bomb Quest)**

CD 4.12. **Spin precession**

(1) Vertical magnetic field
(2) Magnetic field in another direction
(3) Slowly moving magnetic field vector

CD 4.13. **Theory of spin resonance**
(1) Transition probability to orthogonal state
(2) The resonance curve
(3) The time needed to return to the initial state

CD 4.14. **Spin resonance – 1**
(1) Below the resonance frequency
(2) At the resonance frequency
(3) Above the resonance frequency
(4) Perturbation with an even higher frequency

CD 4.15. **Spin resonance – 2**
(1) Large perturbation, below the resonance
(2) Still below the resonance
(3) At the resonance frequency
(4) Above the resonance frequency

CD 4.16. **Spin resonance – 3**
(1) Small perturbation, below the resonance
(2) Still below the resonance
(3) At the resonance frequency
(4) Above the resonance frequency

CD 4.17. **Spin dynamics**
(1) B-field with an oscillating x-component
(2) Large amplitude, slow oscillation
(3) Small perturbation, resonance
(4) Small perturbation, rapid oscillation

CD 4.18. **Berry phase**
(1) Rotation through 2π and geometric phase
(2) Adiabatic motion and geometric phase

5. Composite Systems

CD 5.1. **Two-particle systems in one dimension**
(1) Two independent free particles
(2) Two independent oscillating particles

CD 5.2. **Two-particle configurations**
(1) Configurations of two free particles
(2) Configurations of oscillating particles

CD 5.3. **Separable and entangled**
(1) Entangled state of two free particles

(2) Counter example: A separable state

CD 5.4. **Interacting system**
(1) Two particles bound by oscillator force
(2) Oscillator with equal masses

CD 5.5. **Individual position densities**
(1) Harmonic oscillator system
(2) Oscillator consisting of equal masses
(3) Both particles heavy

CD 5.6. **Two identical particles**
(1) Antisymmetric state (fermion system)
(2) Symmetric state (boson system)

CD 5.7. **States of a two-qubit system**
(1) Bases of qubit states
(2) Some product states
(3) Realization by harmonic oscillator

CD 5.8. **Pure and mixed states**
(1) Determinative measurement on an ensemble

CD 5.9. **Bell's singlet state – 1**
(1) The spins are totally anticorrelated

CD 5.10. **Bell's singlet state – 2**
(1) Classical model of anticorrelation

CD 5.11. **Determine a Bell state**
(1) Local measurements with comparison

CD 5.12. **Remote state preparation**
(1) Alice can predict Bob's result
(2) ... but there is no information for Bob
(3) Classical information prepares a state

CD 5.13. **CHSH inequality – 1**
(1) Violation of local realism

CD 5.14. **CHSH inequality – 2**
(1) Local hidden variables

CD 5.15. **Quantum correlations**
(1) Violations of Bell's inequality

CD 5.16. **Classical hidden variables**
(1) Local realism and Bell's inequality

6. Relativistic Systems

CD 6.1. **Introduction**

(1) Spinor-valued wave functions
(2) A solution of the Dirac equation

CD 6.2. **Relativistic plane waves**
(1) Plane waves with positive energy
(2) Plane waves with negative energy
(3) Eigenspinors of the Dirac operator

CD 6.3. **Kinematics of plane waves**
(1) Dependence of phase velocity on k
(2) Phase velocity for negative energy
(3) Superposition with negative energy
(4) Positive plus negative energy, same momenta
(5) Opposite energies and momenta

CD 6.4. **Wave packets "at rest"**
(1) Positive-energy wave packet
(2) Contributions of relativistic momenta
(3) Extreme relativistic case

CD 6.5. **Wave packets in motion**
(1) Spreading of a positive-energy solution
(2) Motion in momentum space
(3) Wave packet with a high average momentum
(4) Wide momentum distribution

CD 6.6. **Solutions with negative energy**
(1) Wave packet "at rest"
(2) Spreading of a moving wave packet
(3) Time-evolution in momentum space

CD 6.7. **Wave-packet anatomy**
(1) Various parts and representations
(2) Time evolution in position space
(3) Time evolution in momentum space

CD 6.8. **Special initial conditions**
(1) Only an upper component at $t = 0$
(2) A wider momentum distribution
(3) Only an upper component at $t = 0$
(4) Projection of 1 onto negative energies

CD 6.9. **Special initial conditions**
(1) Equal momenta for both components
(2) Similar to 1, only faster
(3) Phase shift between spinor components

CD 6.10. **Special initial conditions**
(1) Components with opposite momenta
(2) Similar to 1, only faster

(3) Phase shift between spinor components

CD 6.11. **Special initial conditions**
(1) Upper component with positive momentum
(2) Lower component with positive momentum

CD 6.12. **Strange superpositions**
(1) Positive and negative energies with equal velocities
(2) Interference at relativistic velocities
(3) Positive and negative energies with opposite velocities
(4) Yet another fine example of Zitterbewegung

CD 6.13. **Superluminal motion?**
(1) Shapes moving faster than light
(2) Phase velocity and interference patterns
(3) Classical interference effect: Moiré pattern
(4) Signals have a limiting velocity

CD 6.14. **Lorentz transformations**
(1) Boost of a positive-energy wave packet
(2) Boost of a negative-energy wave packet

CD 6.15. **Lorentz transformations**
(1) Lorentz-boost of a superposition
(2) Positive- and negative-energy parts
(3) Lorentz transformation in momentum space
(4) Another example

CD 6.16. **Solutions in two dimensions**
(1) Only an upper spinor-component at $t = 0$
(2) Gaussian initial function in both components

CD 6.17. **Wave packet moving in two dimensions**
(1) Only an upper spinor-component at $t = 0$
(2) The two components are equal
(3) Components have opposite momenta

CD 6.18. **External field**
(1) Behavior of a positive-energy wave packet
(2) Behavior of a negative-energy wave packet

CD 6.19. **Constant force field**
(1) A wave packet splits into two parts
(2) Initial state with positive kinetic energy
(3) Electronic wave packet
(4) Positronic wave packet

CD 6.20. **Klein's paradox**
(1) Scattering at a small potential step
(2) Potential step of intermediate size

(3) Total reflection at a high potential step
(4) Nonzero transmission for very high steps

CD 6.21. **Spinor in momentum space**
(1) Scattering at a small potential step
(2) Scattering and partial reflection
(3) Negative-energy wave packet

CD 6.22. **Klein's paradox revisited**
(1) Total reflection in momentum space
(2) Klein's paradox in momentum space
(3) Scattering at an even higher step

CD 6.23. **Scalar potential step**
(1) Scattering at a small scalar step
(2) No Klein paradox for scalar fields

CD 6.24. **Relativistic potential well**
(1) Confinement in an electrostatic well
(2) Wave packet escapes from a deep well
(3) No Klein paradox in a scalar well

CD 6.25. **Particle in a double well**

List of Symbols

$\in$	is contained in, 2
$\subset$	set inclusion, 132
$\approx$	approximately equal to, 16
$\equiv$	by definition equal to, 27
$\cong$	isomorphic to, 70
$[\cdot, \cdot]$	commutator, 14
$\langle \cdot, \cdot \rangle$	scalar product, 433
$\| \cdot \|$	norm, 433
$\langle \cdot \,\vert\, \cdot \rangle$	scalar product (Dirac notation), 440
$\langle \cdot \vert, \vert \cdot \rangle$	Dirac's bra and ket symbols, 440
$\dagger$	(superscript) adjoint of a matrix or operator, 435
$\top$	(superscript) transpose of matrix or vector, 8
$\otimes$	tensor product, 216
$\oplus$	orthogonal direct sum, 127
$\mathbf{0}_n$	$n \times n$ zero matrix, 310
$\mathbf{1}$	identity operator, 11
$\mathbf{1}_n$	$n \times n$ identity matrix, 8
$\mathbf{1}$	identity matrix, identity operator, 438
A	linear operator, 6
A, B	parts of a composite system, 216
$\mathbf{A}$	magnetic vector potential, 135
$A^\dagger$	adjoint of A, 435
$A_\ell^\pm$	ladder operators, 71
a_0	Bohr radius, 96
$a_\pm(k)$	auxiliary quantities, 380
B	some region in space, 437
$\mathbf{B}$	magnetic induction, 116
$\mathbb{C}$	set of complex numbers, 433
c	speed of light, 96
$\mathfrak{D}$	domain of an operator, 438
d	parameter of an ellipse, 65
∂	partial derivative, 11
∂	space-time derivative, 368

$d_\pm$	auxiliary quantities, 333
dx	length element, 437
d^3x	volume element, 434
$d\Omega$	area element on unit sphere, 31
E	energy, eigenvalue of the Hamiltonian, 19
E	relativistic energy, 324
$\mathbf{E}$	electric field strength, 148
$E_n^{(0)}$, $E_n^{(1)}$	unperturbed eigenvalue, first-order perturbation, 443
e	Euler's number, 16
e	elementary charge, 96
$\mathbf{e}_r$, $\mathbf{e}_\vartheta$, $\mathbf{e}_\varphi$	coordinate unit vectors, 27
F	force, 62
$\mathbf{F}$	force vector, 18
$\mathcal{F}$	Fourier transform, 330
G	some region in momentum space, 331
g	Landé g-factor, 116
H_0	free Hamiltonian (Schrödinger), 18
H_0	free Hamiltonian (Dirac), 326
H_{sqrt}	square-root Klein-Gordon operator, 324
$H^{(0)}$, $H^{(1)}$	unperturbed Hamiltonian, perturbation, 443
H	Hamiltonian, 435
$\mathfrak{H}$	Hilbert space, 4
$\hat{\mathfrak{H}}$	set of states, 4
$\mathfrak{H}_{\text{pos}}$, $\mathfrak{H}_{\text{neg}}$	subspaces with positive/negative energy, 381
$\mathbf{h}_\ell$	radial Schrödinger operator, 51
$\hbar$	Planck's constant, 96
i	imaginary unit, 435
$J_\nu(z)$	Bessel functions, 45
$\hat{\jmath}_\ell(kr)$	Riccati-Bessel functions, 45
$\mathbf{K}$	Runge-Lenz vector, 64
k	wave number, momentum, 437
$\mathbf{L} = (L_1, L_2, L_3)$	angular-momentum operator, 11
ℓ	orbital angular momentum quantum number, 29
$\hat{\mathbf{L}}$, $\hat{L}^2$, $\hat{L}_3$	angular momentum in spherical coordinates, 28
L_3	third component of angular momentum, 11
L^2	square of orbital angular momentum, 26
L^2	Hilbert space of square-integrable functions, 10
$L^2(\mathbb{R}^3)^4$	Hilbert space for Dirac equation, 381
M	total mass, 222
m	mass of a particle, 34
m	magnetic quantum number, 29

m_e	electron mass, 96
m_p	proton mass, 96
N	generator of Lorentz boosts, 371
$\mathbf{N}$	generator of boosts in three dimensions, 387
$N_\nu(z)$	Neumann functions, 45
n	principal quantum number, 69
$\mathbf{n}$	unit vector, 14
n_r	radial quantum number, 50
$\hat{n}_\ell(kr)$	Riccati-Neumann functions, 45
$\mathcal{O}$	Landau symbol, 447
P	projection operator, 436
P	parity transformation, 389
$\mathcal{P}$	Poincaré transformation, 383
P_{pos}, P_{neg}	projection operators, 381
$P_\ell(z)$	Legendre polynomial, 32
$P_\ell^m(z)$	associated Legendre functions, 32
p	momentum in one dimension, 325
p	transition probability, 169
$\mathbf{p}$	momentum in three dimensions, 12
q	charge, 115
R_{H}	Rydberg constant (hydrogen), 59
$\mathbf{R}(\boldsymbol{\alpha})$	rotation matrix, 8
$\mathbb{R}$	set of real numbers, 434
$\mathbb{R}^3$	three-dimensional Euclidean space, 2
Ran	range of an operator, 478
r	radial coordinate, radius, 26
S	scattering operator, 479
$\mathbf{S}$, S_k	spin operators, 127
S^2	square of the spin, 127
S^2	unit sphere, 31
$SO(3)$	rotation group, 9
T	time period, 65
T	time reversal transformation, 389
$\mathbf{T}$	representation of Lorentz transformations, 385
t, t_0	time parameter, initial time, 437
t_0	atomic time unit, 96
u_{pos}, u_{neg}	plane-wave spinors, 381
U	unitary matrix or operator, 437
$U(\boldsymbol{\alpha})$	unitary rotation, 10
$\mathbf{u}(k)$	matrix diagonalizing Dirac operator, 333
$\mathbf{v}$	velocity, 125

V_{cov}	covariant potential, 384
$\mathrm{V}_{\mathrm{elm}}$	electromagnetic potential matrix, 382
V_{el}	electrostatic potential matrix, 382
$V(\mathbf{x})$	potential energy, 63
x	space-time four vector, 384
$\mathbf{x} = (x_1, x_2, x_3)$	coordinates in position space, 2
$Y_\ell^m(\vartheta, \varphi)$	spherical harmonics, 32
$\overline{z}$	complex conjugate of z, 433
α	fine structure constant, 96
α	Dirac alpha matrix (one space dimension), 325
α, $\boldsymbol{\alpha}$	rotation angle and vector, 8
$\boldsymbol{\alpha}$	vector of Dirac matrices, 378
β	Dirac beta matrix, 325
Γ_k	matrices of Dirac algebra, 389
γ	perturbation parameter, 443
γ	Coulomb coupling constant, 62
γ	gyromagnetic ratio, 116
γ	Dirac gamma matrices, 368
γ_0	coupling constant for hydrogen, 96
$\gamma(v)$	factor in Lorentz transformation, 366
Δ	Laplace operator, 43
Δ_ψ	uncertainty, 436
δ_{ik}	Kronecker delta symbol, 8
∇	gradient operator (nabla), 435
$\hat{\nabla}$	gradient in spherical coordinates, 28
ϵ	eccentricity of an ellipse, 65
ϵ_{ikm}	totally antisymmetric tensor, 8
ϵ_0	permittivity of vacuum, 96
ϑ	polar angle, 26
$\Lambda(\omega)$	matrix of Lorentz boost, 366
$\lambda(k)$	relativistic energy, 333
μ	reduced mass, 62
$\boldsymbol{\mu}$	magnetic dipole moment, 115
μ_{B}	Bohr magneton, 116
μ_{N}	nuclear magneton, 125
π	the number π, 6
$\sigma(H_0)$	spectrum of H_0, 333
$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$	vector of Pauli matrices, 129
σ_k	Pauli matrices, 25
$\boldsymbol{\tau}$	torque, 116
φ	azimuthal angle, 26

ϕ_{sc}	scalar potential function, 391
ϕ_{el}	electrostatic potential function, 381
ϕ_{in}, ϕ_{out}	incoming and outgoing asymptotes, 477
$\chi_B(x)$	characteristic function, 437
ψ	vector in Hilbert space, 4
ψ	wave function, 10
$[\psi]$	one-dimensional subspace (ray), 4
$\hat{\psi}$	Fourier transform of ψ, 330
ψ_0	initial state, 37
$\psi_n^{(0)}$, $\psi_n^{(1)}$	unperturbed eigenvector, first-order perturbation, 443
$\psi_{\pm}$	scattering states, 476
Ω_{in}, Ω_{out}	Møller wave operators, 477
ω	parameter for Lorentz boost, 383
ω_{L}	Larmor frequency, 126

Index

active transformation, 2
adiabatic approximation, 467
adiabatic limit, 471
adiabatic phase, 467
alkali atoms, 451
analytic perturbation theory, 456
analyticity, 446
angular momentum, 12
angular-momentum operator, 13, 69, 70, 77, 84
angular-momentum quantum numbers, 25, 80, 88, 121, 123
angular-momentum subspace, 20, 78, 263, 415
anharmonic oscillator, 456
anomalous magnetic moment, 124
anticorrelation, 283
antisymmetric state vector, 252
antisymmetric subspace, 253
antisymmetric wave function, 257
antisymmetrizer, 253
antiunitary operator, 5
anyon, 252
associated Legendre functions, 32
asymptotes, 477
asymptotic completeness, 477
atomic number, 58

balanced function, 317
Bell basis, 219
Bell inequality, 295
Bell state, 272
Bell's singlet state, 277
Bell's theorem, 297
Bell-state analyzer, 277
Bernstein-Vazirani problem, 320
Berry phase, 469
Biedenharn-Johnson-Lippmann operator, 428
bipartite system, 216
bit, 158
black box, 316
Bloch sphere, 241
Bohr magneton, 116
Bohr radius, 98, 103, 453
Boolean function, 316
boost, 383
bosonic Hilbert space, 253
bosonic system, 252
bosons, 254
bound states, 475
bra, 440

Cauchy's formula, 457
center-of-mass coordinates, 222
centrifugal barrier, 44
centrifugal potential energy, 44
charge conjugation, 394
CHSH inequality, 296
circular Rydberg states, 106
circularly polarized, 181
classical teleportation, 170
classical velocity operator, 351
Clebsch-Gordan coefficients, 265
CNOT, 310
coherent superposition, 239
coincidence probability, 286
collapse, 165, 439
commutation relations – angular momentum, 14

complete, 433
complete basis measurement, 306
complete set of observables, 147
completeness, 477
conditional probability, 290
configuration, 212
constant function, 317
constant of motion, 17, 438
continuous spectrum, 436
controlled-NOT, 310
convex linear combination, 242
convex set, 243
Coulomb coupling constant, 62
Coulomb potential, 18, 445
Coulomb problem, 54, 57, 147, 451
  relativistic, 422
counterfactual, 284
coupling constant, 443
covariant Dirac equation, 384
critical point, 461
cross section, 488
current density, 488

Darwin term, 409
database search, 320
decoherence theory, 439
degenerate eigenvalue, 20
degenerate eigenvalues, 449
degree of degeneracy, 20
dense coding protocol, 301
destructive interference, 193
detector, 486
Deutsch-Jozsa problem, 318
differential cross section, 488
dilation, 99
dimensionless quantities, 98
Dirac equation, 327, 384
  covariant form, 384
Dirac gamma matrices, 384
Dirac matrices, 378
Dirac operator
  electromagnetic field, 382
  free particles, 326, 379
  momentum space, 331, 380
  spectrum, 333
  spherical coordinates, 411
Dirac spinors, 327
Dirac-Coulomb problem, 427
discrete eigenvalue, 462
discrete Lorentz transformation, 383
double-slit experiment, 192
Dyson expansion, 205

eccentricity, 65
eigenspace, 20
electromagnetic potential matrix, 382
electrostatic field, 445
elementary experiments, 167
elementary measurement, 163
energy functional, 461
energy representation, 482
ensemble measurements, 167
entangled, 215, 218, 272
EPR protocol, 289
error correction, 322
essential spectrum, 462
even parity states, 219
exchange operator, 251
exchange phase, 251
existence (of scattering states), 476
expectation value, 436
external force, 475
extremal point, 243

fermionic Hilbert space, 253
fermionic system, 252
fermions, 254
fine structure, 61
first-order perturbation, 447
Foldy-Wouthuysen method, 409
Fourier transform, 321
fractional quantum Hall effect, 252
free-particle Dirac operator, 326
free-particle Hamiltonian, 476
functional, 461

geometric phase, 467
global measurement, 273
global transformation, 308
global unitary transformation, 310
Gram-Schmidt representation, 246

Green function, 485
ground state, 35, 36, 51, 120, 142, 187, 400, 422, 430
ground state (helium), 452, 464
ground state (hydrogen), 58, 72, 80, 102
ground state (upper bound), 463
Grover's algorithm, 320
gyromagnetic ratio, 116

Hadamard transformation, 176, 316, 319
Hamiltonian, 435
Hankel transformation, 49
Heisenberg equation, 438
helium, 452
Hermitian matrix, 160
hidden variables, 201
Hilbert space, 433
hydrogen atom, 58, 445
hyperfine structure, 61

identical, 250
incident flux, 487
incoherent superposition, 239
incoming, 477
indistinguishable, 250
inertial frame, 365
infimum, 463
integral cross section, 488
interaction-free measurement, 194
interference, 191, 285, 318, 356, 358, 360, 374
interferometer, 189
invariance, 7
invariance transformation, 7
isolated eigenvalue, 443
isometric, 477
isotropy of space, 2

joint probability, 285

Kato-Rellich theorem, 456
ket, 440
ket-symbol, 160
kinetic energy, 476
kinetic-energy correction, 408
Klein paradox, 395, 398, 402
Kratzer's potential, 53
Kronecker sum, 225
Kummer's equation, 90

ladder operator, 23, 141, 153, 188, 266
Lamb shift, 61
Landé g-factor, 116
Laplace-Beltrami operator, 43
Larmor frequency, 126
Legendre function, 32
Legendre polynomial, 32
linear potential, 445
linearly polarized, 181
Lippmann-Schwinger equation, 484, 486
local measurement, 225, 273
local observable, 226
local realism, 296
local unitary transformation, 226, 308
logic gates, 307
Lorentz force, 125
Lorentz group, 383
Lorentz transformation, 6, 383

Møller operators, 477
magnetic field, 444, 471
magnetic quantum number, 51, 82
magneton, 115, 472
maximally entangled state, 247, 272
maximally mixed, 247
maximally mixed state, 241, 272
mean value, 436
metric, 382
metric tensor, 43
mild solution, 338
Minkowski space, 382
mixed state, 234
Moiré pattern, 358
multiplicity, 20
muonic atom, 103

needle in a haystack, 320
Neumann series, 457
no-cloning theorem, 170
Noether's theorem, 12, 17
nonlocal interaction, 278

norm, 433
nuclear magneton, 125
nucleus, 18, 58, 61, 62, 106, 115, 122, 149, 221, 431, 451, 452, 486

observable, 435
odd parity states, 219
one-particle density function, 215
one-time pad, 197
operator
  unitary, 5
oracle, 316
orbital angular momentum, 13
orbitals, 80
orthogonal projection, 287
orthogonal projection operator, 88, 194, 211, 226, 234, 242, 253, 286, 342, 436, 478, 479
outgoing, 477

parabolic coordinates, 91
parameter of an ellipse, 65
parity bit, 272
parity transformation, 383
particle interchange operator, 254
passive transformation, 2
Pauli equation, 135
Pauli exclusion principle, 253
Pauli matrices, 25, 129, 160, 173, 325, 379
Pauli operator, 135
Pauli's exclusion principle, 258
period finding, 321
permittivity, 62
perturbation parameter, 443
perturbed eigenvalue, 443
perturbed operator, 443
phase bit, 272
plane wave with negative energy, 335
plane wave with positive energy, 335
Poincaré transformation, 383
potential energy, 476
prime factors, 322
principal quantum number, 76, 86
principle of relativity, 2
projection operator, 434
projective representation, 179
projective space, 5
proper orthochronous Lorentz group, 383
property, 226
pure state, 234
purification, 248

quantum algorithm, 307, 316
quantum Fourier transform, 321
quantum parallelism, 307
quantum register, 305
quantum teleportation, 171
qubit, 158

radial current, 489
radial quantum number, 50
radial Schrödinger equation, 45, 50
radial Schrödinger operator, 44, 51
random experiment, 163
random variable, 163, 285
ray, 4, 434
ray representation, 179
ray transformation, 5
Rayleigh-Ritz technique, 466
Rayleigh-Schrödinger series, 446
real orbitals, 52, 80
reduced mass, 35
register, 305
regular solution, 85
relative boundedness, 455
relative coordinates, 222
Rellich's theorem, 460
representation, 2
resolvent, 456
resolvent set, 456
Riccati-Bessel functions, 45
Riemannian manifold, 43
Riemannian metric, 43
rotation, 7, 383, 472
rotation matrix, 8
rotational invariance, 17
RSA method, 322
Runge-Lenz vector, 428

Rutherford, 58, 486
Rydberg atom, 106
Rydberg constant, 102
Rydberg constant for hydrogen, 59, 102
Rydberg states, 106

sample space, 163
scalar product, 433
scaling transformation, 99
scattering amplitude, 486
scattering cross section, 488
scattering states, 475
Schrödinger equation, 438
  Coulomb problem, 66
  radial, 45, 50
  time dependent Hamiltonian, 204
  two particles, 213
self-adjoint, 435
semibounded, 462
separable, 218
separable state, 215
separation of variables, 18
Shor's algorithm, 322
Simon's algorithm, 321
single measurements, 167
single-particle interference, 191, 193
singlet state, 268, 277
Slater determinant, 258
spectral gap, 381
spectral representation, 482
spectral transformation, 482
spectroscopy, 58
spectrum, 333, 456
spherical harmonics, 32, 452
spherical symmetry, 17, 51, 70, 78, 150, 410, 452
spin-orbit interaction, 149
spin-orbit operator, 151, 411
spin-orbit term, 409
spin-up direction, 173
spinor harmonics, 152, 413
spinor-wave functions, 128
spinors, 128, 158
square-root Klein-Gordon equation, 324
standard acceleration, 351
standard deviation, 436
standard interpretation, 329
standard position operator, 348
  time evolution, 353
standard position probability density, 329
standard representation, 129, 159, 326
standard velocity operator, 350
Stark effect, 91, 445
state, 434
state estimation, 169, 439
state of the subsystem, 233
stationary point, 461
stationary scattering theory, 483
stationary state, 438
statistical mixture, 228
Stern-Gerlach filter, 166
Stokes' theorem, 470
Stone's Theorem, 438
strict solution, 338
subspace of bound states, 475
subspace of scattering states, 475
subsystem observable, 226
superluminal speed, 357
superposition, 435
symmetric state vector, 252
symmetric subspace, 253
symmetric wave function, 257
symmetrizer, 253
symmetry, 7
symmetry transformation, 4, 179, 251, 366, 382, 393

target, 486
tensor product, 216, 217
theorem
  of Aharonov and Casher, 143
  of Bell, 297
  of Kato and Rellich, 456
  of Noether, 17
  of Rellich, 460
  of Stokes, 470
  of Wigner, 6
  virial, 104
Thomas precession, 149, 383

time reversal, 383
total angular momentum, 125, 260
total cross section, 488
transformation function, 459
transition probability, 5, 437
triangular condition, 264
triplet states, 268
two-particle system, 212
two-state system, 158

uncertainty, 436
unentangled, 218
unitary group, 10, 438
unitary operator, 5
unperturbed operator, 443

valence electron, 451
variational collapse, 399
variational method, 461, 465
vector-addition coefficients, 265
vector-coupling coefficients, 265
velocity transformation, 383
Vernam cipher, 197
virial theorem, 103, 464

wave operators, 477
wave packet with negative energy, 340
wave packet with positive energy, 339
which-way information, 162
Wigner coefficients, 265
Wigner's theorem, 6

XOR-gate, 310

Zitterbewegung, 349, 353, 409