

Graduate Texts in Physics

Hans Lüth

# Quantum Physics in the Nanoworld

Schrödinger's Cat and the Dwarfs

*Second Edition*



Springer

# **Graduate Texts in Physics**

## **Series editors**

Sadri Hassani, Normal, USA  
W.J. Munro, Kanagawa, Japan  
Richard Needs, Cambridge, UK  
William T. Rhodes, Boca Raton, USA  
Martin Stutzmann, Garching, Germany  
Andreas Wipf, Jena, Germany

## **Graduate Texts in Physics**

Graduate Texts in Physics publishes core learning/teaching material for graduate- and advanced-level undergraduate courses on topics of current and emerging fields within physics, both pure and applied. These textbooks serve students at the MS- or PhD-level and their instructors as comprehensive sources of principles, definitions, derivations, experiments and applications (as relevant) for their mastery and teaching, respectively. International in scope and relevance, the textbooks correspond to course syllabi sufficiently to serve as required reading. Their didactic style, comprehensiveness and coverage of fundamental material also make them suitable as introductions or references for scientists entering, or requiring timely knowledge of, a research field.

More information about this series at <http://www.springer.com/series/8431>

Hans Lüth

# Quantum Physics in the Nanoworld

Schrödinger's Cat and the Dwarfs

Second Edition



Springer

Hans Lüth  
Peter Grünberg Institut (PGI),  
PGI-9: Semiconductor Nanoelectronics  
Forschungszentrum Jülich GmbH  
Jülich  
Germany

and

Jülich Aachen Research Alliance (JARA)  
Aachen, Jülich  
Germany

Translation from the German language edition: *Quantenphysik in der Nanowelt*  
by Hans Lüth, © 2009 Springer-Verlag Berlin Heidelberg. All rights reserved

ISSN 1868-4513                    ISSN 1868-4521 (electronic)  
Graduate Texts in Physics         ISBN 978-3-319-14668-3         ISBN 978-3-319-14669-0 (eBook)  
ISBN 978-3-319-14668-3         DOI 10.1007/978-3-319-14669-0

Library of Congress Control Number: 2015939994

Springer Cham Heidelberg New York Dordrecht London  
© Springer International Publishing Switzerland 2013, 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media  
([www.springer.com](http://www.springer.com))

*To Roswitha*

# Preface to the Second Edition

This textbook on quantum physics is in some aspects different from most books on this topic. While the essential mathematical formalism—in the simplest possible form—both of non-relativistic single particle quantum mechanics and of quantum field theory are presented, selected experiments play an important role in the foundation of the theory and for making contact with modern applications. Hereby a special focus is on nanostructures and nanoelectronics as the subtitle “Schrödinger’s Cat with the Dwarfs (in Greek: *nanos*)” indicates. The structure of atoms and of the Periodic Table of Elements, for example, is introduced on the basis of the electronic structure of semiconductor quantum dots rather than by considering the hydrogen atom and its extrapolation to multi-electron atoms. “Schrödinger’s Cat” in the subtitle paradigmatically describes the other aim of the book, namely to discuss more in extension than commonly the philosophical background and the counterintuitive aspects of quantum physics.

Why now a second edition of the book after a relatively short time? From discussions with colleagues and students I got the impression that both specific aspects of the book might be deepened somewhat more. For this purpose I have added some more relevant experiments with nanostructures: The quantum point contact in connection with the conductance quantum is introduced and its use as a charge detector in nanoelectronic circuits is explained. As a direct application interference experiments in a nanoscaled Aharonov Bohm ring with an additional probe for “Which Way” information are presented. Furthermore, the realisation and the study of the electronic properties of an artificial quantum dot molecule are presented.

Already in the first edition of the book I had briefly mentioned that non-locality of quantum physics should be better discussed within the frame of quantum field theory. In this new edition I have extended and deepened this idea, that particle-wave duality and non-locality in the Einstein-Podolsky-Rosen (EPR) paradox are much better understood on the basis of quantum field theory than in the frame of single particle Schrödinger quantum mechanics. Correspondingly, an additional new section on the particle picture in quantum field theory and the non-locality of

quantum fields is devoted to this issue. Some counterintuitive aspects of quantum physics, thus, become more acceptable to our understanding.

Apart from these two major additions to the book I have incorporated two interesting new developments having been awarded with the Nobel prize, the realisation of atomic Bose–Einstein condensates and the detection of the Higgs particle. Both topics being relevant to quantum physics are briefly explained in the corresponding context. Also, a quantum interference experiment with giant C<sub>60</sub> buckyball molecules is reported as an example for present research in the direction of elucidating the border line between classical and quantum behaviour. Some minor errors have been removed in the new edition and some new problems have been added.

I want to thank Gregor Mussler for his help in the preparation of most of the new figures. Stefan Fölsch has supplied nice figures of his work on Indium quantum dot molecules and has critically read the related text; also thanks to him. Thanks are also due to Claus Ascheron of Springer Verlag for his encouragement and his effort in editing this new edition.

Aachen and Jülich, Germany  
June 2015

Hans Lüth

# Preface to the First Edition

The original German edition of this book was published in 2009. Because of the positive response I have got from students and colleagues I translated the book into English and furthermore added some new problems, the last chapter “synopsis,” and an additional Appendix about the reduced density matrix.

What was the reason to write this book? There are a large number of excellent textbooks on quantum mechanics on the market. Nearly all of these books have in common that quantum mechanics is presented as one of the most important and successful theories to solve physical problems. This is totally in the sense of most physicists, who applied, until the 1970s of the twentieth century, in a first quantum revolution quantum mechanics with overwhelming success not only to atom and particle physics but also to nearly all other science branches as chemistry, solid state physics, biology, or astrophysics. Because of the success in answering essential questions in these fields, fundamental open problems concerning the theory itself were approached only in rare cases. This situation has changed since the last decade of the twentieth century. Since then there are new sophisticated experimental tools in quantum optics, atom and ion physics, and in nanoelectronics, which can touch inherent quantum physical questions and allow interesting tests of the theory itself. Such questions, as for example, origin and consequences of superposition and entanglement, are of predominant importance for fields as quantum teleportation, quantum computing, and quantum information in general.

From this “second quantum revolution” as this continuing further development of quantum physical thinking is called by Alain Aspect, one of the pioneers in this field, one expects a deeper understanding of quantum physics itself but also applications in engineering. There is already the term “quantum engineering” which describes scientific activities to apply particle wave duality or entanglement for practical purposes, for example, nanomachines, quantum computers, etc.

This background in mind I have written the present book. Particular quantum phenomena are more at the center of interest rather than the mathematical formalism. I prefer a more pictorial and sometimes intuitive description of the phenomena, and recent experimental findings from research on nanoelectronic systems are often presented to support the theory. Also, connections to other science

branches such as elementary particle physics, quantum electronics, or nuclear magnetic resonance in biology and medicine are made.

Concerning the formalism, I generally restrict myself to first approximation steps, which are relevant for experimental physicists and engineers in applying the theory or to estimate the order of magnitude of experimental results or data. On the other hand, the Dirac bra-ket notation is introduced in analogy to three-dimensional vectors and it is used for simplicity reasons in many cases. Similarly, commutator algebra is introduced as essentially adding or subtraction of symbols (operators). The mathematical background necessary to read the book is quite simple. Only the knowledge of simple functions, simple differential equations, and basics of matrix algebra is required.

Rather than axiomatically introducing important quantities and equations I have preferred to make the invention of basic equations or the mathematical tools for field quantization plausible by physically reasonable conclusions and extrapolations.

The book was written on the basis of manuscripts of lectures on quantum physics and nanoelectronics, which I have given to physics and electrical engineering students at the Aachen University of Technology (RWTH). Essential extensions are, of course, due to my own research in quantum electronics. In particular, supervising PhD students in this field and the many discussions with them had great influence on the way of presentation. I want to thank all of them for the interesting discussions which also helped me to a deeper insight into the fascinating field of quantum physics.

Furthermore, I want to thank my former coworkers, meanwhile all in academic teaching and research positions, Arno Förster, Michel Marso, Michael Indlekofer, and Thomas Schäpers for many exciting disputes, which contributed to further elucidation of difficult questions.

During the translation of the original German edition into English Margrit Klöcker sometimes improved and corrected my English grammar; also thanks to her.

I owe very special thanks to my late wife Roswitha. She supported me all the time during which I wrote the original German manuscript and she invented the subtitle “Schrödinger’s Cat and the Dwarfs.” This subtitle accurately expresses the main focus of the book, namely a more thorough diving into the physical and philosophical content of quantum mechanics (paradigm: Schrödinger’s cat), and this in the context of the nanoworld (world of the dwarfs). Roswitha found the right words for this aspect of the book that I lacked.

Aachen and Jülich, Germany  
September 2012

Hans Lüth

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	General and Historical Remarks	1
1.2	Importance for Science and Technology	3
1.3	Philosophical Implications	5
References		8
<b>2</b>	<b>Some Fundamental Experiments</b>	9
2.1	Photoelectric Effect	9
2.2	Compton Effect	12
2.3	Diffraction of Massive Particles	15
2.4	Particle Interference at the Double Slit	21
2.4.1	Double Slit Experiments with Electrons	21
2.4.2	Particle Interference and “Which-Way” Information	24
References		27
<b>3</b>	<b>Particle-Wave Duality</b>	29
3.1	The Wave Function and Its Interpretation	29
3.2	Wave Packet and Particle Velocity	33
3.3	The Uncertainty Principle	37
3.4	An Excursion into Classical Mechanics	39
3.5	Observables, Operators and Schrödinger Equation	42
3.6	Simple Solutions of the Schrödinger Equation	48
3.6.1	“Locked-Up” Electrons: Confined Quantum States	49
3.6.2	Particle Currents	56
3.6.3	Electrons Run Against a Potential Step	58
3.6.4	Electrons Tunnel Through a Barrier	61
3.6.5	Resonant Tunneling	66
3.7	Single Electron Tunneling	75
3.8	The Quantum Point Contact as Charge Detector	82
References		88

<b>4</b>	<b>Quantum States in Hilbert Space</b>	89
4.1	Eigenvectors and Measurement of Observables	89
4.2	Commutation of Operators: Commutators	95
4.3	Representation of Quantum States and Observables	97
4.3.1	Vectors of Probability Amplitudes and Matrices as Operators	97
4.3.2	Rotations of Hilbert Space	103
4.3.3	Quantum States in Dirac Notation	106
4.3.4	Quantum States with a Continuous Eigenvalue Spectrum	110
4.3.5	Time Evolution in Quantum Mechanics	115
4.4	Games with Operators: The Oscillator	118
4.4.1	The Classical Harmonic Oscillator	118
4.4.2	Upstairs-Downstairs: Step Operators and Eigenvalues	119
4.4.3	The Anharmonic Oscillator	126
	References	130
<b>5</b>	<b>Angular Momentum, Spin and Particle Categories</b>	131
5.1	The Classical Circular Motion	131
5.2	Quantum Mechanical Angular Momentum	133
5.3	Rotational Symmetry and Angular Momentum; Eigenstates	140
5.4	Circulating Electrons in a Magnetic Field	147
5.4.1	The Lorentz Force	147
5.4.2	The Hamilton Operator with Magnetic Field	148
5.4.3	Angular Momentum and Magnetic Moment	150
5.4.4	Gauge Invariance and Aharonov–Bohm-Effect	153
5.5	The Spin	160
5.5.1	Stern–Gerlach Experiment	160
5.5.2	The Spin and Its 2D Hilbert Space	164
5.5.3	Spin Precession	168
5.6	Particle Categories: Fermions and Bosons	171
5.6.1	Two and More Particles	171
5.6.2	Spin and Particle Categories: The Pauli Exclusion Principle	175
5.6.3	Two Different Worlds: Fermi and Bose Statistics	180
5.6.4	The Zoo of Elementary Particles	188
5.7	Angular Momentum in Nanostructures and Atoms	199
5.7.1	Artificial Quantum Dot Atoms	199
5.7.2	Atoms and Periodic Table	206
5.7.3	Quantum Rings	211
	References	215

<b>6 Approximate Solutions for Important Model Systems . . . . .</b>	217
6.1 Particles in a Weakly Varying Potential: The WKB Method . . . . .	218
6.1.1 Application: Tunneling Through a Schottky Barrier . . . . .	220
6.2 Clever Guess of a Wave Function: The Variational Method . . . . .	223
6.2.1 Example of the Harmonic Oscillator . . . . .	226
6.2.2 The Ground State of the Hydrogen Atom . . . . .	229
6.2.3 Molecules and Coupled Quantum Dots . . . . .	232
6.2.4 Experimental Realisation of a Quantum Dot Molecule . . . . .	240
6.3 Small Stationary Potential Perturbations: The Time-Independent Perturbation Method . . . . .	244
6.3.1 Perturbation of Degenerate States . . . . .	248
6.3.2 Example: The Stark Effect in a Semiconductor Quantum Well . . . . .	251
6.4 Transitions Between Quantum States: The Time-Dependent Perturbation Method . . . . .	254
6.4.1 Periodic Perturbation: Fermi's Golden Rule . . . . .	256
6.4.2 Electron-Light Interaction: Optical Transitions . . . . .	259
6.4.3 Optical Absorption and Emission in a Quantum Well . . . . .	262
6.4.4 Dipole Selection Rules for Angular Momentum States . . . . .	265
6.5 Electronic Transitions in 2-Level Systems: The Rotating Wave Approximation . . . . .	272
6.5.1 2-Level Systems in Resonance with Electromagnetic Radiation . . . . .	272
6.5.2 Spin Flip . . . . .	277
6.5.3 Nuclear Spin Resonance in Chemistry, Biology and Medicine . . . . .	282
6.6 Scattering of Particles . . . . .	288
6.6.1 Scattered Waves and Differential Scattering Cross Section . . . . .	290
6.6.2 Scattering Amplitude and Born Approximation . . . . .	292
6.6.3 Coulomb Scattering . . . . .	297
6.6.4 Scattering on Crystals, on Surfaces and on Nanostructures . . . . .	301
6.6.5 Inelastic Scattering on a Molecule . . . . .	309
References . . . . .	313
<b>7 Superposition, Entanglement and Other Oddities . . . . .</b>	315
7.1 Superposition of Quantum States . . . . .	316
7.1.1 Scattering of Two Identical Particles: A Special Superposition State . . . . .	318

7.2	Entanglement . . . . .	322
7.2.1	Bell's Inequality and Its Experimental Check . . . . .	327
7.2.2	“Which Way” Information and Entanglement: A Gedanken Experiment . . . . .	334
7.2.3	“Which Way” Probing in an Aharonov-Bohm Interference Experiment . . . . .	338
7.3	Pure and Mixed States: The Density Matrix . . . . .	343
7.3.1	Quantum Mechanical and Classical Probabilities . . . . .	343
7.3.2	The Density Matrix . . . . .	347
7.4	Quantum Environment, Measurement Process and Entanglement . . . . .	351
7.4.1	Subsystem and Environment . . . . .	351
7.4.2	Open Quantum Systems, Decoherence and Measurement Process . . . . .	354
7.4.3	Schrödinger’s Cat . . . . .	359
7.5	Superposition States for Quantum-Bits and Quantum Computing . . . . .	360
7.5.1	Coupled Quantum Dots as Quantum-Bits . . . . .	361
7.5.2	Experimental Realization of a Quantum-Bit by Quantum Dots . . . . .	365
	References . . . . .	370
<b>8</b>	<b>Fields and Quanta . . . . .</b>	<b>373</b>
8.1	Ingredients of a Quantum Field Theory . . . . .	374
8.2	Quantization of the Electromagnetic Field . . . . .	376
8.2.1	What Are Photons? . . . . .	383
8.2.2	2-Level Atom in the Light Field: Spontaneous Emission . . . . .	387
8.2.3	Atom Diffraction by Light Waves . . . . .	393
8.2.4	Once Again: “Which Way” Information and Entanglement . . . . .	399
8.2.5	The Casimir Effect . . . . .	404
8.3	The Quantized Schrödinger Field of Massive Particles . . . . .	407
8.3.1	The Quantized Fermionic Schrödinger Field . . . . .	414
8.3.2	Field Operators and Back to the Single Particle Schrödinger Equation . . . . .	418
8.3.3	The Particle Picture in Quantum Field Theory . . . . .	424
8.3.4	Electrons in Crystals: Back to the Single Particle Approximation . . . . .	426
8.3.5	The Band Model: Metals and Semiconductors . . . . .	430

8.4	Quantized Lattice Waves: Phonons . . . . .	439
8.4.1	Phonon–Phonon Interaction . . . . .	448
8.4.2	Electron–Phonon Interaction. . . . .	453
8.4.3	Absorption and Emission of Phonons . . . . .	457
8.4.4	Field Quanta Mediate Forces Between Particles . . . . .	460
	References . . . . .	465
<b>9</b>	<b>Synopsis</b> . . . . .	467
	<b>Appendix A: Interfaces and Heterostructures</b> . . . . .	471
	<b>Appendix B: Preparation of Semiconductor Nanostructures</b> . . . . .	479
	<b>Appendix C: The Reduced Density Matrix</b> . . . . .	489
	<b>Problems</b> . . . . .	493
	<b>Index</b> . . . . .	505

# Chapter 1

## Introduction

Quantum physics is thought, without doubt, to be one of the greatest intellectual achievements of the 20th century. Its history began at the turn from the 19th to the 20th century. But we are confronted with its profound scientific, technological and philosophical implications today even more than ever. Not only in scientific original papers and text books but also in popular science literature and fiction more and more frequently book titles appear which contain terms as quantum theory, quantum mechanics, quantum physics, quantum world or quantum entrainment etc. Sometimes these titles are abused to supply quite questionable and esoteric treatises with a quasi-scientific background. What, therefore, is it all about with this field of quantum physics, which plays a central role in the education of physicists and, hopefully soon, also of chemists, biologists and engineers.

### 1.1 General and Historical Remarks

Isaac Newton created, more than 300 years ago, classical mechanics by finding the laws of motion for solids and of gravitation between masses. This theory was so successful for the deterministic description of motions, in particular for the planets in our solar system, that Newton was led to the assumption that also light has corpuscular character. On the basis of light particles, which propagate along a straight line in a light beam, he could consistently explain a number of optical phenomena including the reflection and diffraction of light. The diffraction and interference experiments of Christian Huygens living at Newton's time and a little bit later, at the beginning of the 19th century, of Thomas Young and Augustin Fresnel, however, paved the way for the wave theory of light, at that time still waves in a not understood ether.

The triumph of wave theory could not be stopped anymore when the prominent Scottish physicist James Clark Maxwell successfully described the nature of light by a wave-like propagation of electrical and magnetic fields. He, thus, unified the two classical branches of optics and electricity in one and the same theory. By the detec-

tion of radio waves at around 1887, Heinrich Hertz finally established the familiar theoretical system of electrodynamics and electromagnetic waves.

Simultaneously, during the 19th century, an atomistic and molecular view of matter emerged and became more and more important, and this against various philosophical objections. Milestones in the development of an atomistic picture of matter were certainly the statistical kinetic gas theory of Ludwig Boltzmann around the end of the 19th century and the explanation of the Brownian motion in terms of collisions between liquid molecules and pollen particles suspended in the liquid by Einstein in 1905.

At the beginning of the 20th century, then, experimental results accumulated which contributed essentially to the emergence of a new physics, quantum physics. Among these there must be mentioned the detection of cathode rays in vacuum tubes, of X-rays and of radio activity. In particular, the Rutherford model of the atom must be emphasized, which was suggested by Ernest Rutherford in order to explain his scattering experiments of  $\alpha$ -particles on metal foils. Rutherford's atom is already imagined to consist of a massive small nucleus containing almost the entire atomic mass and an extended electronic cloud which determines the spatial extension of the atom.

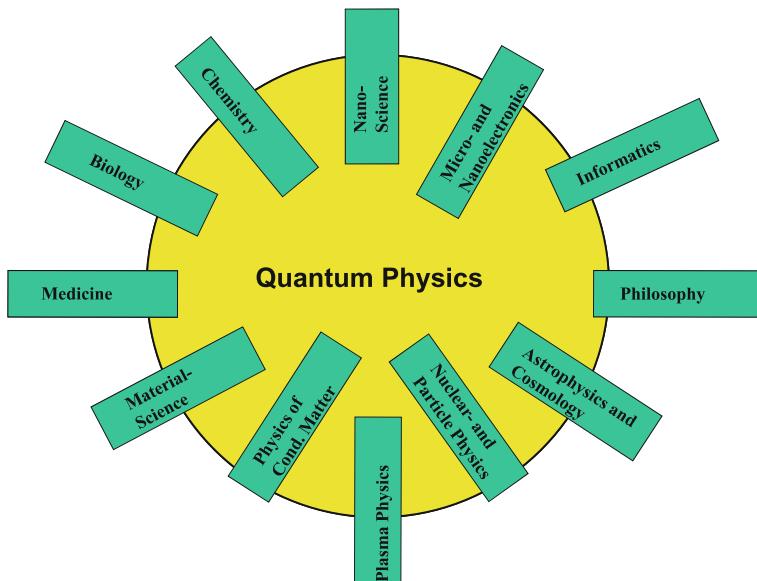
This breakthrough in the understanding of the atom might be thought of as the beginning of the era of quantum physics. In a next step, the emission of sharp spectral lines of excited atoms being in contradiction to the successful theory of electrodynamics by Maxwell was explained. In 1913 Bohr interpreted, or better made plausible, the emitted line spectrum of hydrogen atoms on the basis of heuristic postulates about stable electron orbits around the positive nucleus, the proton.

A little bit earlier, already Max Planck had broken new ground into the direction of quantum physics. Around the end of the 19th century there was the puzzle of black body radiation. A so-called black body emits a continuous spectrum of electromagnetic radiation whose shape strongly depends on the temperature of the emitter. By means of classical electromagnetic theory, the spectrum for the shortest wavelengths always was calculated to diverge into infinity, the so-called ultraviolet (UV) catastrophe. Planck, who was a quite conservative physicist, made the revolutionary assumption that a black body interacts with the electromagnetic field by exchange of energy only in small quanta rather than in a continuous way. The UV catastrophe could thus be removed and the experimental black body emission theoretically be described correctly. In a kind of desperation, he must have drawn this conclusion which was in strict contradiction to Maxwell's electromagnetic field theory of continuous electric and magnetic fields. The assumption, indeed, led back to the rejected corpuscular theory of light by Newton. Planck created the term quantum which gave the whole field its name. In his theoretical assumption, the quanta carry an energy  $E$  which is proportional to the light frequency  $\nu$ . The constant  $h = E/\nu$  has been named Planck's constant in honor of its inventor. A number of illuminating detections followed (Chap. 2) which finally led to the formulation of quantum mechanics in its present form. In particular, the explanation of the photoelectric effect by Einstein (Sect. 2.1) shall be mentioned.

## 1.2 Importance for Science and Technology

While quantum theory was originally intended to explain the world of atoms, molecules and elementary particles, in particular the electron, it became clear meanwhile, that the theory has universal importance for the understanding of the whole surrounding world, up to cosmological questions. This is by no means astonishing since our world consists of atoms, elementary particles and energy fields which closely interact with matter. Thus, the stability of matter can only be understood on the basis of quantum theory (Sect. 5.7.2).

The fundamental principles of quantum theory as particle-wave duality, the uncertainty principle and the random behavior on the atomic level, therefore, have to be taken into account in almost every natural or engineering science. This is true, although, because of historical or practical reasons, models of classical physics, mechanics or chemistry are used in many of these sciences. This is shown in a somewhat qualitative way in Fig. 1.1. Each science field plotted by one of the boxes participates more or less in the general field of quantum physics. The amount by which it reaches into the quantum circle should indicate to what extent theoretical models and experimental tools of quantum physics are used in the field. A partial overlap of a science field with the quantum circle does not mean that only part of the phenomena or systems considered there obey the laws of quantum physics. According to our understanding everything in this world, matter and fields, be it in



**Fig. 1.1** Qualitative representation of the overlap between important science branches and the field of quantum physics. The amount of overlap with the “quantum circle” indicates how far quantum physical methods, theoretical and experimental ones, are used in the particular science disciplines

microelectronics, in medicine, in chemistry or in astrophysics is totally subject to the laws of quantum physics. A partial overlap (Fig. 1.1) only indicates qualitatively to what extent one uses typically quantum physical methods and considerations in this field. Partially, this is dependent on the degree of atomistic thinking in a particular science field.

As an example take chemistry. All what happens in a chemical laboratory or in a chemical plant is related to chemical bonds and reactions and thus obeys the laws of quantum theory. Nevertheless a chemist working in the laboratory must not always think about quantum physical laws. During the long history of chemical sciences typically chemical rules about reactivity between molecules and radicals have been established, which have to be applied in order to produce a certain product. But being confronted with novel problems of chemical bonding or reactivity a theoretical chemist using quantum mechanical calculations has to be asked for an efficient solution.

Similarly in medicine, for the interpretation of images from NMR (nuclear magnetic resonance, Sect. 6.5.3) or PET (positron emission tomography) usually the skills of the special medical education are sufficient. But in difficult cases, at the front of research, one has to dig into the basics of the quantum physical elementary processes of spin precession or decay times etc. in order to reach a certain level of understanding. The same is true for all nuclear medical methods of cancer treatment. The interaction of high energy particle radiation with biomolecules and cells can only be approached by means of quantum physical methods.

Biology presents an extremely broad field of scientific activity reaching from animal observation, evolution biology (theory), cell biology down to molecular biology. This latter branch of biology, which has an ever more growing influence on the explanation of biological phenomena on the atomic and molecular level became possible only on the basis of quantum theory. Decoding of the DNA and its function in genetics was achieved on the basis of quantum theory. The study of folding of proteins and the related biological activity requires the use of supercomputers and algorithms being based on quantum mechanics.

Astrophysics and cosmology reach into the quantum circle only halfway. In these research fields relativity theory certainly plays an equally important role as quantum physics. Similarly, in plasma-physics (nuclear fusion) magneto-hydrodynamics contributes to the understanding of problems as much as quantum physics does.

Nuclear- and elementary particle physics as well as condensed matter physics penetrate the quantum circle almost completely. Both disciplines arose on the basis of quantum physics and can only be understood within the frame of quantum theory. Classical physical models are sometimes used only for analogy reasons.

Material science, micro- and nanoelectronics and nanoscience (treats nanostructured materials) are of particular interest. These disciplines penetrate the quantum circle by a significant amount, since many theoretical models and experimental techniques stem from quantum physics. Examples are the description of the electrical resistance which is due to scattering of charge carriers on crystal defects and lattice vibrations, as well as the scanning electron tunneling microscope which allows imaging of single atoms and atomic orbitals on a solid surface. On the other hand, there

exist many classical, microscopic analysis and preparation techniques in these fields, which work without using explicitly quantum physics. Probes for mechanical hardness and the design of micro- and nanoelectronic circuits shall be mentioned. In the considered disciplines, however, a clear trend to more and more atomistic thinking and to structures on the nanoscale is observed (transistors with 5–10 nm dimensions). In the near future, therefore, quantum physical techniques will be much more important and the corresponding boxes in Fig. 1.1 will move more into the quantum circle.

Informatics characterized by its historical roots, Shannon's entropy (information measure) and the Turing machine (abstract model for computer), managed without using quantum physics. This situation has changed since quantum information (Sect. 7.1) has become an interesting and growing field within information science. Superposition states being characteristic for quantum physics allow extremely parallel data processing which is by no means possible within a classical computer with von Neumann architecture. The realization of quantum computers and correspondingly adapted algorithms is meanwhile an important branch in physical and information research.

Similarly as in science the impact of quantum physics on every day life can not be estimated highly enough. Many industrial products which we use without one single thought would just not exist without quantum physics. The development of lasers, a product of quantum physics, enabled important applications in ophthalmology, material engineering and, of course, the familiar CD (compact disk) player. Our satellite antennas for TV reception contain, in the first amplifier stage, a low noise transistor (HEMT: high electron mobility transistor) which was developed by using principles of quantum physics. For the function of the navigation system (GPS) atomic clocks are essential, also products of quantum physics. This is similarly true for all imaging systems in medicine as NMR, CT, PET etc. The information age is based on integrated semiconductor circuits the development of which was possible after the electronic structure of semiconductors was understood from the laws of quantum mechanics (Sect. 8.3.4). Weather forecast with high predictive quality and climate models require calculations on supercomputers, products of modern semiconductor technology.

Quantum physics is an essential basis of our modern world. There is an estimate that almost a quarter of the gross national product in highly developed countries arises from products being directly or indirectly related to quantum physics.

### 1.3 Philosophical Implications

In Fig. 1.1, even philosophy penetrates into the quantum circle to some extent. No other physics theory excited philosophers, at least those with a view on natural science and epistemology, to such an extent as quantum theory did. No other theory in physics interferes so much with philosophical questions as what is real, what can we recognize, in how far is our knowledge about nature pure imagination.

Let us start with the question, what means quantum theory for the whole edifice of physical science. Its fundamental issues, random behavior on the atomic scale, particle-wave duality (Chap. 3), uncertainty relation (Sect. 3.3), and the principles of field quantization (Chap. 8) form a non-classical frame of thinking which is relevant in all sub-disciplines of physics such as elementary particle physics, physics of condensed matter, astrophysics etc. There are no experimental results in all these fields which are in contradiction to quantum theory so far. Quantum physics, in its non-relativistic Schrödinger formulation for condensed matter physics and the highly sophisticated relativistic field theories of the standard model in elementary particle physics (Sect. 5.6.4) describe nature equally well on all scales, even up to cosmology. Quantum theory must, thus, be considered as a hyper-theory, which has to be matched also by future theories about so far unsolved problems such as quantum-gravity or dark matter and energy.

Theory of relativity and Darwin's theory of biological evolution certainly also belong into this class of hyper-theories. No serious biologist or natural scientist in general would dare to make assumptions which are in contradiction to Darwin's theory, to its central statements, not to minor derivations. Similarly theory of relativity yields the general frame for our understanding of space and time as well as of gravitation. A restriction, however, has to be made. In the theory of relativity, welldefined curves in space and time do exist. The wave-particle dualism and the uncertainty principle do not exist, relativity theory is a classical theory in that sense. We therefore expect that in a future unification of quantum and relativity theory the latter one has to adapt to quantum theory. First approaches to quantum-gravity as loop or string theory point into this direction.

It is worth mentioning that in both hyper-theories, quantum theory and the theory of biological evolution, accident, that is, random behavior, plays a dominant role. Random mutations in biology enable the emergence of something new on the cellular level. ("Le hazard et la nécessité" how it is expressed very accurately by Monod [1] in his famous book). Hereby, the term mutation in biology is intimately related with random behavior as it is defined in quantum physics.

The strongest interference of quantum physics with philosophy is certainly given in the field of the theory of knowledge. Two fundamental issues of quantum physics, in particular, have troubled philosophers, the inherently random, that is, non-deterministic behavior on the atomic level and the interference of the human observer with the physical measurement process, that is, the co-determination of our knowledge about nature by the observing subject. For a long time, the opinion prevailed that the collapse of a wave packet upon a measurement and the transition of the wave function into an eigenstate of the measured observable (Sect. 3.5) demonstrate the dependence of our knowledge on the measurement. Our knowledge should, thus, be determined to an essential part by the measurement and the observer rather than by an externally existing reality. The Copenhagen interpretation of quantum mechanics (Bohr, Heisenberg) sometimes shows features of a subjective and idealistic philosophy, in which a reality beyond our perception horizon is denied. Both a better understanding of the physical measurement process in terms of entanglement

(Sect. 7.4) and philosophical developments as in evolutionary epistemology [2] have caused a return to a critical, realistic interpretation of quantum mechanics.

Particularly, philosophical branches as *Evolutionary Epistemology* [2] in connection with *Hypothetical Realism* [3] are appropriate to quantum mechanics and form a wider frame for quantum mechanical thinking. Popper presents a detailed analysis on realism and subjectivism in physics and concludes [4]:

There is, therefore, no reason whatever to accept either Heisenberg's or Bohr's subjectivist interpretation of quantum mechanics. Quantum mechanics is a statistical theory because the problems it tries to solve—spectral intensities, for example—are statistical problems. There is, therefore, no need here for any philosophical defence of its non-causal character...

To sum up, there is no reason whatsoever to doubt the realistic and objectivistic character of all physics. The role played by the observing subject in modern physics is in no way different from the role he played in Newton's dynamics or in Maxwell's theory of the electric field: the observer is essentially the man who tests the theory.

The statement about the statistical nature of quantum physics must be seen in connection with the fact that quantum physics is non-deterministic on the level of elementary events; but the calculation of probabilities and average measurement results for large ensembles of particles is performed in a deterministic way by means of differential equations with boundary and initial conditions (Sect. 3.5).

The problem of the measurement process in quantum physics has posed many questions and caused much discussion about perception of reality and subjectivism in the past. Meanwhile, these discussions have been eased due to recent fundamental experiments on the participation of the observer in a measurement (Sects. 2.4.2 and 8.2.4) and due to the recognition of the importance of entanglement between the system under study and the measurement apparatus (Sect. 7.2). In this modern context the human experimentalist merely plays the role of an observer rather than an integral part of the system under study. The entanglement (specific quantum correlation) between measurement apparatus and the real object being studied connects both of them and simultaneously separates the cognizing human observer from the reality of the outside world. Consequently, experiments yield an image of the externally existing reality, but we can achieve step by step an ever better image of that reality.

As is worked out in the epistemology of hypothetical realism, all statements about the world have hypothesis character. According to Popper [4], these hypotheses must be falsified to establish new improved hypotheses in a trial and error procedure. By means of ever better hypotheses, reality is described step by step more adequately. The “invention” of Schrödinger's equation or of field quantization (Sect. 3.5, Chap. 8) are good examples for the establishment of hypotheses. These hypotheses in quantum physics could not be falsified in their corresponding validity ranges (non-relativistic range for Schrödinger equation). They must be assumed to be valid for the description of reality so far.

It is essential that modern quantum physics does not deny the existence of a structured reality beyond our senses and our perception. In this context Vollmer remarks [2]:

We assume that a real world does exist, that it has particular structures and that these structures are partially recognizable. We test how far we can come with these hypotheses (translation from the German by the author).

In this context, we always have to remember that philosophical realism can not be proven; it can neither be verified nor falsified [5]. But according to Popper [4] and other philosophical realists, it is certainly the most reasonable hypothesis to get along with the every-day environment as a human being.

In this sense of philosophical realism, the counter-intuitive character of quantum physics, for example, the particle-wave duality, does not cause difficulties. In the evolutionary epistemology, human recognition is essentially determined by limitations of our sensual perception and the structure of our brain. Both are results of the biological evolution of man who had to adapt to a macroscopic rather than to an atomic scale environment. In this sense, Shimony [6] remarks:

Human perceptual powers are as much a result of natural selection as any feature of organisms, with selection generally favoring improved recognition of objective features of the environment in which our pre-human ancestors lived.

## References

1. J. Monod, *Le Hazard et la Necessite* (Editions du Seuil, Paris, 1970)
2. G. Vollmer, *Evolutionäre Erkenntnistheorie*, 3rd edn. (S. Hirzel Verlag, Stuttgart, 1983)
3. D.T. Campbell, *Inquiry* **2**, 152 (1959)
4. K.R. Popper, *Objective Knowledge, An Evolutionary Approach* (Oxford at the Clarendon Press, Oxford, 1979), p. 303
5. B. Russel, *The problems of Philosophy* (1912), and German translation: Probleme der Philosophie (Suhrkamp-TB 1967)
6. A. Shimony, *J. Philosophy* **68**, 571 (1971)

# Chapter 2

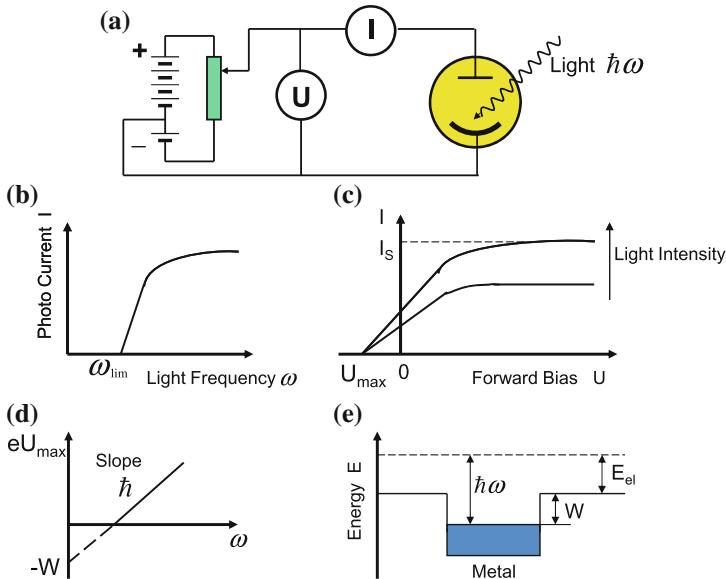
## Some Fundamental Experiments

It is interesting to follow the development of today's quantum physics by considering difficulties in the interpretation of important experimental results. In particular, around the end of the 19th and the beginning of the 20th century empirical facts accumulated which demonstrated the limits of interpretations on the basis of classical physics, Newton's mechanics and Maxwell's theory of electromagnetic fields. Such a historic approach is not intended in the present book. Instead, I want to select some few fundamental experiments, which indicate directly the peculiarities of atomic systems. The experiments are chosen such that they intuitively motivate the basic assumptions of quantum mechanics.

### 2.1 Photoelectric Effect

When a metal surface is irradiated with light of frequency  $\omega$  (ultraviolet or visible for alkali metals), electrons are emitted from the metal. In an appropriate experiment, the electron emitting metal can be the cathode in a vacuum tube and the electrons are sucked up by a positively biased anode (Fig. 2.1). This set-up is the basic element of every secondary electron multiplier in which a series of additional electrodes amplifies the electron beam in a sort of avalanche process before it reaches the last anode and is detected.

Also at negligible acceleration voltage and even under de-acceleration bias (illuminated metal positive) electrons are emitted under illumination. The emitted current vanishes not before a certain maximum de-acceleration voltage  $U_{\max}$  is exceeded (Fig. 2.1c). Thus, the energy of the emitted electrons can be determined from the energy difference  $eU_{\max}$  which can be overcome by the propagating electrons. With  $v$  as electron velocity one has  $eU_{\max} = mv^2/2$ . According to classical electrodynamics the energy flux density in the light beam is given by the Pointing vector  $\mathbf{S} = \mathcal{E} \times \mathbf{H}$ . For low light intensities one would, thus, expect that only after sufficient time enough energy for the emission of electrons has been transferred to the



**Fig. 2.1** a–e Photo-effect: **a** Experimental set-up. By light irradiation (photon energy  $\hbar\omega$ ) electrons are emitted from a photo-cathode; they produce a photo-current  $I$  under the action of a bias voltage  $U$ . **b** Photo-current  $I$  as function of light frequency  $\omega$ . **c** Photo-current  $I$  as function of applied voltage  $U$ . Positive bias defines the illuminated electrode as cathode.  $U_{\text{max}}$  is the maximum negative bias which can be overcome by the emitted electrons due to their kinetic energy. The saturation current height  $I_s$  depends on the irradiated light intensity. **d** Maximum deceleration energy  $eU_{\text{max}}$  as function of light frequency  $\omega$ . From this plot the natural constant  $\hbar$  is obtained as slope; the onset of the curve (straight line) at  $\omega = 0$  yields the work function  $W$  of the cathode material. **e** Explanation of the photo-effect by means of the potential box model of free metal electrons (shaded). The photon energy  $\hbar\omega$  of the irradiated light is sufficient for the electrons to overcome the energy barrier of the work function  $W$ ; on top they carry an additional amount of kinetic energy  $E_{\text{el}}$

metal. Furthermore, the energy  $eU_{\text{max}}$  of the photoelectrons determined from the de-acceleration voltage should increase with growing radiation power. This is not observed in the experiment. The energy of the photoelectrons does not depend on light intensity, that is, radiation power. Instead, a characteristic dependence of the effect on the light frequency  $\omega$  is observed. A lower frequency limit  $\omega_{\text{lim}} = 2\pi\nu_{\text{lim}}$  does exist, below which electrons are not emitted from the metal (Fig. 2.1b). This frequency limit is specific for the material. Furthermore, the emission of electrons starts already at very low light intensities though with very low emission currents, i.e. very small numbers of emitted electrons. A plot of the energy  $E_{\text{el}}$  of the emitted electrons ( $=eU_{\text{max}}$ , determined from de-acceleration voltage) versus light frequency exhibits a linear dependence:

$$E_{\text{el}} = eU_{\text{max}} = \frac{1}{2}mv^2 = \hbar\omega - W, \quad (2.1a)$$

$W$  is the so-called work function of the metal which has to be overcome by the evading electron before it reaches the vacuum. The constant

$$\hbar = h/2\pi = 6.6 \times 10^{-16} \text{ eV} \quad (2.1b)$$

is Planck's constant, which can be measured in the described way by the photoelectric effect.

An explanation of these phenomena is obviously not possible on the basis of classical Maxwell's electrodynamic theory, it became possible by means of Einstein's light quantum hypothesis [1] (1905, Nobel prize 1921). In Einstein's revolutionary new assumption light consists of small particles, the photons, which carry the energy  $\hbar\omega = h\nu$ . Energy can be transferred from the light beam to the metal only in portions of these quanta. Each electron which leaves the metal with an energy  $E_{el}$  (2.1a) has taken over the energy of a photon. The intensity of a light beam with frequency  $\omega$  is proportional to the number of photons with energy  $\hbar\omega$  in the beam. Thus, the emission current is also proportional to the number of photons. These assumptions consistently explain the photoelectric effect (Fig. 2.1).

Further properties of photons can be derived by means of relativity theory, where the speed of light is the absolutely highest possible velocity, and this in all inertial systems moving against each other with certain velocities. Photons as light quanta, thus, move with the speed of light  $c$  in the direction of light propagation described by the light wave vector  $\mathbf{k}$ . From the existence of a maximum constant light velocity relativity theory gives an expression for the energy of a mass  $m$  moving with a momentum  $p$ :

$$E = \sqrt{p^2 c^2 + m^2 c^4}. \quad (2.2)$$

Light, that is, also its constituting particles, the photons, have no mass. Together with the light dispersion relation  $\omega = ck$  one obtains from (2.2)

$$E = \hbar\omega = \hbar ck = pc. \quad (2.3)$$

We, thus, must attribute a momentum  $p = \hbar k$  to the mass-less photons. We conclude that the electromagnetic field being continuous on the macroscopic scale is built up by small particles, the photons, to which we attribute the specific photon energy

$$E = \hbar\omega = h\nu \quad (2.4a)$$

and a momentum

$$\mathbf{p} = \hbar\mathbf{k}. \quad (2.4b)$$

The continuous field of classical Maxwell theory obviously has a granular character in reality which is not seen in phenomena on macroscopic scale.

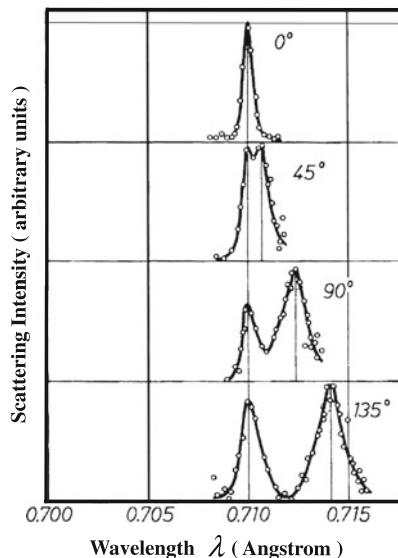
## 2.2 Compton Effect

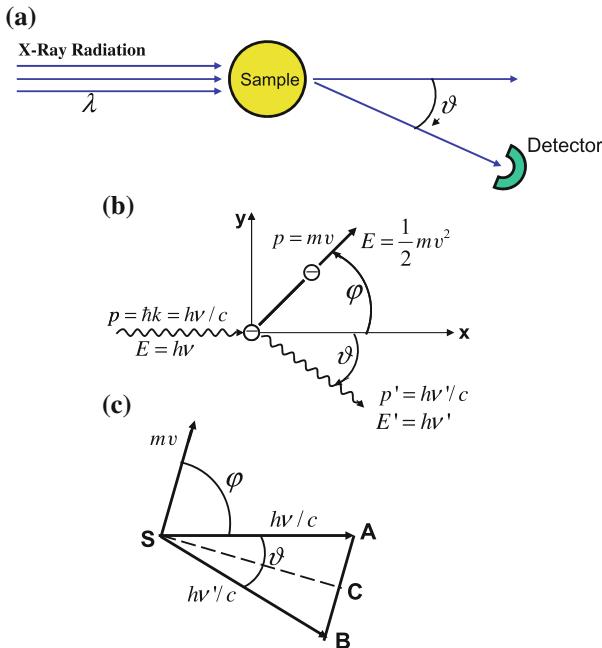
The particle character of electromagnetic radiation is also very clearly seen in the Compton effect, which was detected by Compton and Simon [2] in 1925. When X-rays with photon energies between  $10^3$  and  $10^6$  eV are scattered on free or weakly bound electrons, beside elastically scattered Rayleigh radiation (equal wavelength  $\lambda$  as incident radiation) there appears a second contribution of scattered radiation which is shifted in wavelength by  $\Delta\lambda$ , independent on the material of the scattering target (Fig. 2.2). In the elastic Rayleigh scattering process the oscillating electric field of the incoming X-rays excites electron oscillations (e.g. in the field of the positive nuclei) with the X-ray frequency. These electrons then again emit secondary radiation of the same frequency, the Rayleigh radiation. The additionally emitted radiation whose wavelength is shifted against the Rayleigh scattered one exhibits a characteristic dependence of the wavelength shift  $\Delta\lambda$  on the scattering angle  $\vartheta$  (Fig. 2.2). This phenomenon can be explained quantitatively only under the assumption of an elastic collision with energy and momentum conservation between the electron and the light particle, the photon. We try this approach and write down the following ansatz for momentum conservation in  $x$ - and  $y$ -direction (Fig. 2.3b).

$$\frac{h\nu}{c} = \frac{h\nu'}{c} \cos \vartheta + mv \cos \varphi, \quad (2.5a)$$

$$o = \frac{h\nu'}{c} \sin \vartheta - mv \sin \varphi. \quad (2.5b)$$

**Fig. 2.2** Original measurement data of Compton effect [2]. A graphite sample is irradiated by  $K_\alpha$  radiation from Mo under different angles ( $0^\circ$ – $135^\circ$ ) with regard to the direction of incidence. The radiation is scattered elastically ( $\lambda = 0.71$  Å) without  $\lambda$  shift, partially inelastically with increased wavelength  $\lambda$





**Fig. 2.3 a–c** Scheme of Compton effect. **a** Experimental set-up. **b** Explanation of scattering parameters and particle parameters of X-rays ( $h\nu$  photon energy,  $h\nu/c$  photon momentum) as well as of scattered electron ( $mv^2/2$  energy,  $mv$  momentum). **c** Momentum conservation in a Compton scattering experiment

Hereby,  $m$  is the mass of the propagating electron which is related to its rest-mass  $m_0$  according to relativity theory by

$$m = m_0 (1 - v^2/c^2)^{-1/2}. \quad (2.6)$$

As in the interpretation of the photoelectric effect (2.4b), the photon carries the momentum  $p = \hbar k = h/\lambda = h\nu/c$ . The observed frequency shift  $\Delta\nu = \nu - \nu'$  of the X-rays after scattering can therefore be related to a momentum change of the X-ray photons during the collision with an electron. Apart from momentum conservation (2.5a), (2.5b), also the relativistic energy conservation must hold for the particles, that is, with (2.4a) the energy of the photons must obey the relation

$$h\nu + m_0 c^2 = h\nu' + mc^2. \quad (2.7)$$

Hereby, the electron was assumed to be at rest (rest mass  $m_0$ ) before the collision. By squaring (2.7) and by using (2.6) one obtains the following expression for the frequency change  $\Delta\nu$ :

$$h^2(\Delta\nu)^2 + 2m_0c^2h\Delta\nu = m_0^2c^4 \frac{v^2}{c^2 - v^2}. \quad (2.8)$$

In (2.5a), (2.5b),  $\sin\varphi$  and  $\cos\varphi$  can be eliminated by using the relation  $\sin^2\varphi + \cos^2\varphi = 1$ . After some calculation, one obtains

$$h^2[(\Delta\nu)^2 + 2v(v + \Delta\nu)(1 - \cos\vartheta)] = m_0^2c^4 \frac{v^2}{c^2 - v^2}. \quad (2.9)$$

A comparison of (2.8) and (2.9) shows the equality also of the left sides of the equations which yields

$$m_0c^2h\Delta\nu = h^2v(v + \Delta\nu)(1 - \cos\vartheta), \quad (2.10)$$

with

$$\Delta\lambda = \frac{c}{v} - \frac{c}{v + \Delta\nu} = \frac{c\Delta\nu}{v(v + \Delta\nu)}. \quad (2.11a)$$

Equation (2.10) yields

$$\Delta\lambda = \frac{h}{m_0c}(1 - \cos\vartheta) = \lambda_C(1 - \cos\vartheta). \quad (2.11b)$$

The constant  $\lambda_C$  is called Compton wavelength of the electron, it amounts to:

$$\lambda_C = \frac{h}{m_0c} = 2.4 \times 10^{-10} \text{ cm}. \quad (2.12)$$

The Compton wavelength depends only on natural constants and is, therefore, of general interest. The quantum energy of radiation with a wavelength  $\lambda_C$  corresponds just to the rest mass  $m_0$  of the electron:

$$\frac{hc}{\lambda_C} = hv = m_0c^2 = 511 \text{ keV}. \quad (2.13)$$

Equations (2.11a), (2.11b) describe quantitatively the frequency or wavelength shift  $\Delta\nu$  or  $\Delta\lambda$  as a function of scattering angle  $\vartheta$  as it is observed in the Compton effect.

Someone who feels stressed by the relativistic calculation (2.5a)–(2.12) can obtain the result for the limit of small frequency changes by a non-relativistic treatment (Fig. 2.3c) where the electron mass is approximated by its rest mass ( $m \approx m_0$ ). Inspection of Fig. 2.3c easily shows that for the limit  $v \approx v'$  the momentum vectors of the incident and the scattered light are almost equal ( $hv/c \approx hv'/c$ ). By considering the two rectangular triangles SCB and SCA momentum conservation ( $mv = AB$ ) yields the following relation:

$$\frac{1}{2}mv = \frac{h\nu}{c} \sin \frac{\vartheta}{2}. \quad (2.14)$$

By means of (2.14), the change of the kinetic energy of the electron (initially at rest) which scatters the photon can be expressed by a change of photon energy ( $\hbar\nu - \hbar\nu'$ ):

$$\frac{1}{2}mv^2 = \frac{1}{2} \frac{(mv)^2}{m} = \frac{4h^2\nu^2 \sin^2 \vartheta/2}{2mc^2} = h\nu - h\nu'. \quad (2.15)$$

With  $\nu \approx \nu'$  and by dividing nominator and denominator by  $h\nu^2$ , one obtains:

$$\frac{2h}{mc^2} \sin^2 \frac{\vartheta}{2} = \frac{\nu - \nu'}{\nu^2} \approx \frac{1}{\nu'} - \frac{1}{\nu}. \quad (2.16)$$

Written as a wavelength change,  $\Delta\lambda$  it follows

$$\Delta\lambda = \lambda' - \lambda = \frac{2h}{mc} \sin^2 \frac{\vartheta}{2} = \frac{h}{mc} (1 - \cos \vartheta). \quad (2.17)$$

In the non-relativistic limit ( $m \approx m_0$ ), this equation is identical with the more general relativistic relation (2.11a), (2.11b).

## 2.3 Diffraction of Massive Particles

While photoelectric and Compton effect can only be interpreted on the basis of the particle character of electromagnetic radiation, there are meanwhile numerous diffraction experiments (typical for waves) with all kinds of massive particle beams as electrons, neutrons, atoms, molecules etc. which doubtlessly demonstrate the wave-like propagation of these particles.

Already in 1919, Davisson and Germer detected intensity modulations in the reflection of low energy electrons from crystalline surfaces as a function of the observation angle [3]. The explanation of these observations became possible by De Broglie's hypothesis that the propagation of electrons obeys the laws of waves [4]. In analogy to the photon, the mass-less light particle, De Broglie assumed the validity of the fundamental relation  $p = h/\lambda$  (2.4b) between momentum and wavelength also for massive particles as electrons. Relating the momentum  $p = mv$  to the kinetic energy  $E_{\text{kin}} = mv^2/2$  of a moving particle, one calculates the wavelength of a propagating electron as

$$\lambda = h(2mE_{\text{kin}})^{-\frac{1}{2}}, \quad (2.18a)$$

that is, electrons which have been accelerated by a voltage  $U$  possess a wavelength

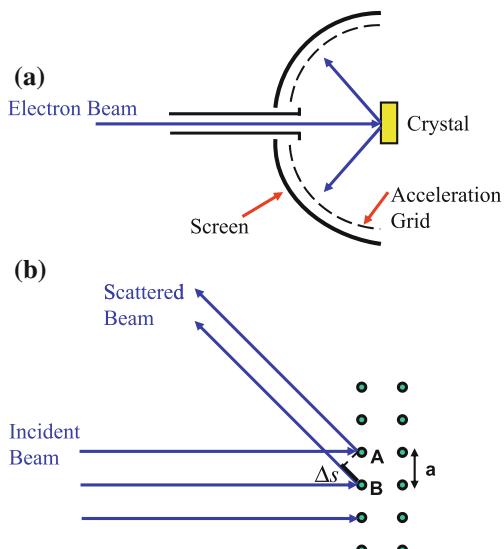
$$\lambda = 12.3 \text{ \AA}/\sqrt{U}. \quad (2.18b)$$

The experiments of Davisson and Germer have initiated the development of a standard characterization method for the atomic structure of solid surfaces, the LEED technique (low energy electron diffraction). The experimental set-up for LEED studies is meanwhile found in every surface science laboratory around the world. The schematic representation of such an experiment is shown in Fig. 2.4. The solid surface under study is arranged in front of a curved fluorescent screen in a vacuum vessel, usually an ultrahigh vacuum (UHV) chamber with background pressure below  $10^{-10}$  Torr. Through an opening in the screen, an electron beam with well defined kinetic particle energy  $E_{\text{kin}} = eU$  obtained by acceleration in a bias between 30 V and 200 V is irradiated on the crystal surface. The electrons backscattered from the sample surface have to pass an acceleration grid in front of the fluorescent screen and an acceleration voltage of some 1000 V in order to have enough energy to become visible on the fluorescent screen. When the sample surface under study is crystalline one always observes more or less bright intensity peaks on the screen, the so-called LEED reflexes. In Fig. 2.5, the LEED reflexes observed on a clean ZnO surface prepared in UHV are shown. The interpretation of this reflex (LEED) pattern is only possible by attributing the propagating electrons in the primary beam a wave. When this electron wave hits the surface atoms of the sample, each atom in the lattice emits a spherical wave. All these spherical waves superimpose and interfere constructively in certain directions and destructively in others. Since electrons with low energies in the order of 100 eV are scattered preferentially on the uppermost atomic layer, the scattering target is 2-dimensional to first approximation. According to Fig. 2.4b, the path difference between two partial waves originating from atoms A and B is  $\Delta s = a \sin \vartheta$  with  $a$  as the interatomic distance within the surface. For constructive interference,  $\Delta s$  must equal a multiple of the electron wavelength  $\lambda$ , which yields the condition

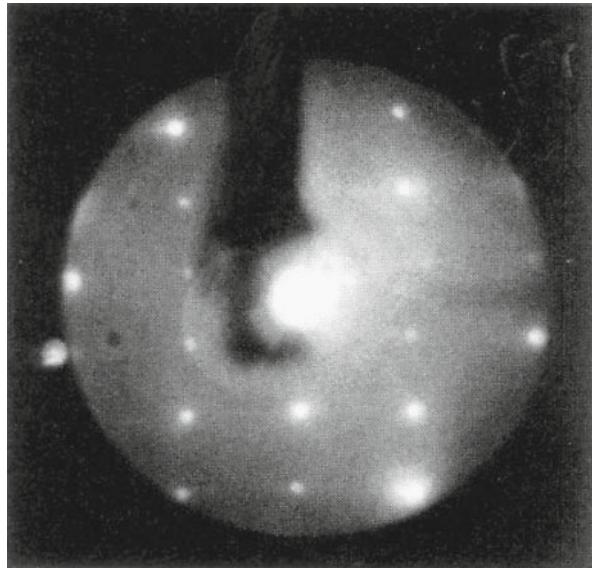
$$a \sin \vartheta = n\lambda. \quad (2.19)$$

**Fig. 2.4 a, b** Scheme of a LEED diffraction experiment (LEED is Low Energy Electron Diffraction) with slow electrons.

**a** Experimental set-up in ultra-high vacuum (UHV). **b** Schematic representation of the diffraction of an incident electron on the upper most atomic layer of a crystal. The atoms A and B are the origin of scattered spherical waves which interfere constructively (Bragg reflection peak) or destructively depending on path difference  $\Delta s$ .  $a$  is distance between atoms



**Fig. 2.5** LEED diffraction pattern of electrons of a kinetic energy  $eU = 140$  eV on a ZnO(1010) surface. The electrons are incident normal to the crystal surface; *bright spots* are Bragg reflection spots due to constructive superposition of waves. The *dark shadow* in the diffraction pattern is due to the crystal holder

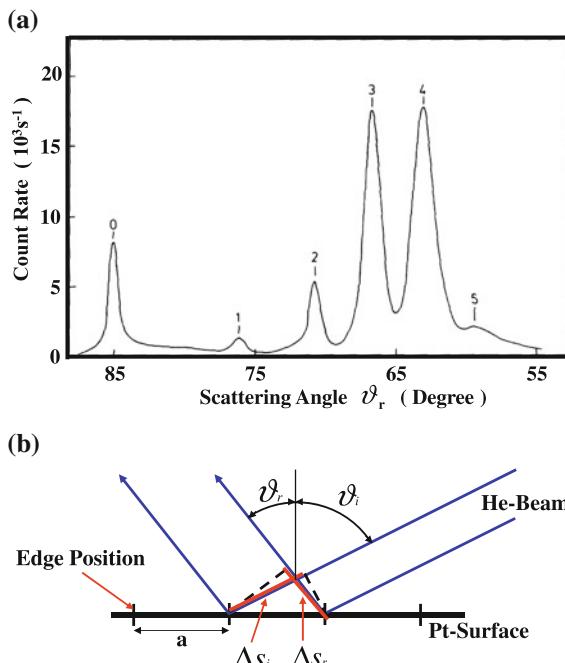


Diffraction intensity is thus expected on a cone with opening angle  $(\pi/2 - \vartheta)$  around the atom row along  $A$  and  $B$ . Since the arrangement of scattering atoms is 2-dimensional a second condition for constructive interference, analogously to (2.19), must be fulfilled in a direction normal to  $AB$  in the surface. The two conditions together limit the spatial range for constructive interference to only one direction, that is, the direction of a particular LEED reflex (bright spot in Fig. 2.5). The different diffraction spots in Fig. 2.5 belong to higher diffraction orders, that is, to different numbers  $n$  in (2.19) and the corresponding second equation. For the interpretation of a LEED pattern as in Fig. 2.5, one calculates the electron wavelength from the kinetic energy of the primary electrons, or respectively from the acceleration voltage according to (2.18a), (2.18b). By means of (2.19), the observation angle for a particular LEED reflex yields information about the interatomic distance, more accurately the periodicity interval, within the sample surface. LEED is meanwhile a standard analysis technique in surface science. Each LEED experiment, many times performed around the globe, demonstrates the wave character of propagating electrons.

Not only moving electrons but also other particles obey the laws of wave propagation. Already in 1930 Estermann and Stern demonstrated that He and H<sub>2</sub> beams undergo diffraction phenomena on solid surfaces [5]. A clear example from recent time are diffraction experiments with He beams on clean, UHV prepared Pt surfaces [6]. The Pt surfaces exhibit a series of regularly spaced monoatomic steps (distance  $a = 2$  nm) which are produced by cutting the crystal at the appropriate angle and annealing in vacuum. The atomic He beam used in the experiment is produced by a supersonic expansion of the gas from a nozzle. The interaction between the atoms in the expanding gas produces a velocity distribution that is significantly

sharper than the Maxwell distribution present before the expansion. The energetically sharp He beam is irradiated on the Pt surface under UHV conditions (background pressure below  $10^{-10}$  Torr). In Fig. 2.6a, the diffracted intensity of He atoms is shown as a function of the scattering (reflection) angle  $\vartheta_r$ , with a fixed angle of incidence  $\vartheta_i = 85^\circ$  against the surface normal. The intensity maxima correspond to the diffraction orders of the periodic lattice of terraces, that is, steps on the Pt surface rather than from the lattice of individual atoms. The steps act as scattering centers, they form a 1-dimensional array. Thus, for the interpretation of the scattering distribution (Fig. 2.6a) relation (2.19) can directly be applied. Only the path difference between two neighboring scattered beams contains the amounts  $\Delta s_i$  and  $\Delta s_r$  of the incident and the reflected (scattered) wave. The position of the diffraction maxima is thus given by

$$a(\sin \vartheta_i - \sin \vartheta_r) = n\lambda. \quad (2.20)$$



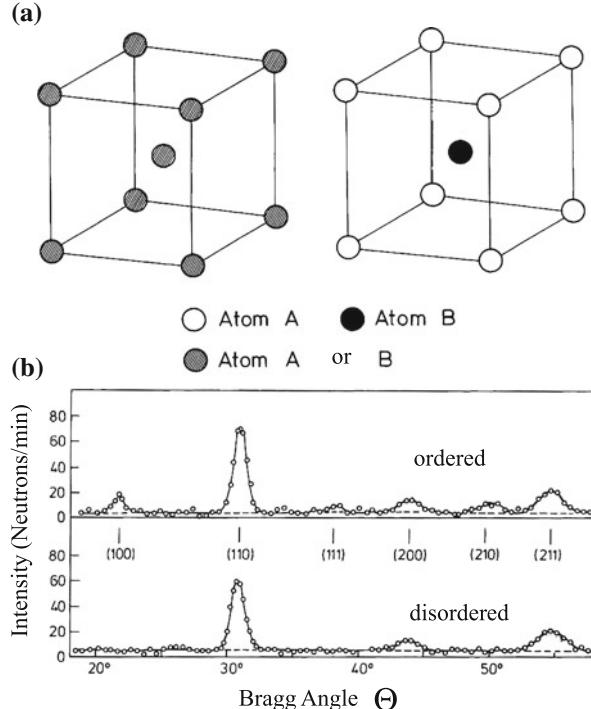
**Fig. 2.6** **a, b** Diffraction of a He atom beam on a Pt surface with a regular step array, step distance  $a = 2 \text{ nm}$  [6]. Like for an Echelette grating in light optics maximum diffraction intensity is obtained in diffraction orders which appear under specular direction with regard to the interaction potential. **a** Diffracted intensity as function of scattering angle  $\vartheta_r$ ; angle of incidence  $\vartheta_i = 85^\circ$  with regard to the Pt surface normal. The reflection angles indicated by 0, 1, 2, ..., 5 are calculated for a step distance  $a = 2 \text{ nm}$ . **b** Scheme of the diffraction geometry. The path differences  $\Delta s_i$  and  $\Delta s_r$  determine the reflection angle, under which the diffraction peak appears

From (2.18a) the wavelength of the He atoms in the beam is obtained as  $0.56 \text{ \AA}$ . With the step distance  $a = 2 \text{ nm}$  the intensity maxima numerated by  $n = 0, 1, 2, 3, \dots$  in Fig. 2.6a are calculated. The agreement between theory and experiment is excellent. As in the case of an optical echelon grating, the direction corresponding to specular (mirror) reflection from the terraces (maxima 3 and 4) is favored in the intensity distribution.

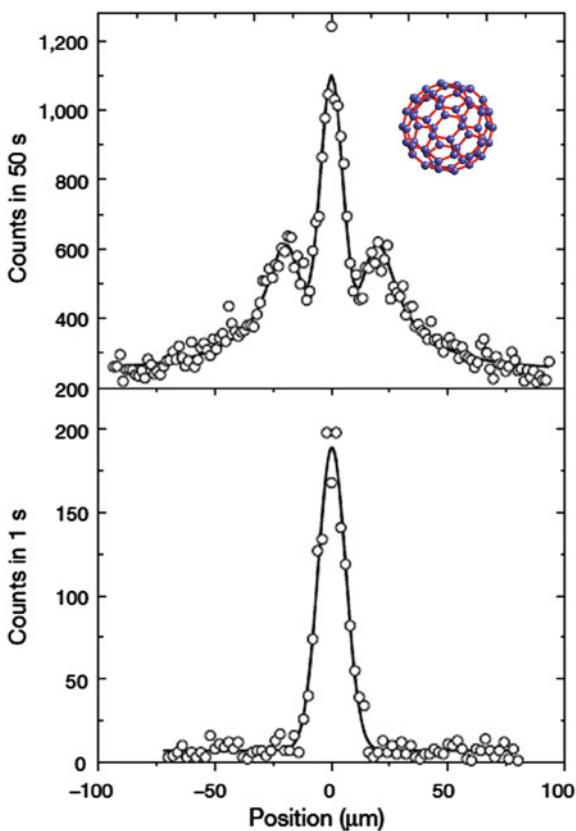
Neutrons interact only extremely weakly with matter because of their missing charge. They penetrate relatively thick solid samples without a significant loss of beam intensity. But also in this case, neutrons which are irradiated on a solid crystalline sample, produce, beside the directly transmitted beam, well-defined sharp beams of neutrons which are diffracted into certain angles with respect to the primary beam direction (Fig. 2.7). The interpretation of the experimental results is based, similarly as in the case of electrons or He atoms, on the assumption of the propagation of neutron waves and their diffraction on the regularly arranged atomic nuclei in the crystal [7].

Interference patterns have meanwhile been observed even with gigantic molecules as  $C_{60}$  [8]. This *fullerene* molecule, sometimes called *buckyball* (named in honour of the British architect Richard Buckminster Fuller, who constructed similar cupolas) consists of 60 carbon atoms bonded in a quasi-planar  $sp^2$  configuration within

**Fig. 2.7** a, b Neutron diffraction on a FeCo alloy [7]. a disordered (left) and ordered (right) phase of FeCo. b Neutron diffractogram of the ordered and disordered phase of FeCo. Because of low counting rates in neutron diffraction long measurement times are needed



**Fig. 2.8** **a** Interference pattern produced by C<sub>60</sub> fullerene molecules diffracted on a lithographically prepared grid [8]. The zeroth and first-order maxima can be seen. The solid curve is calculated from experimental data by means of grid diffraction theory. **b** Control experiment: The molecular beam profile without the grating in the path of the molecules



hexagons and pentagons and forming a football shaped sphere (Fig. 2.8, inset). The molecule has a diameter of about 1 nm (van der Waals diameter) and a weight of  $1.2 \times 10^{-21} \text{ g}$ , i.e. 720 times the weight of a proton.

In the diffraction experiment these molecules have been evaporated from an oven at temperatures around 1000 K. A molecular beam with an average molecule velocity of about 220 m/s is formed by apertures and focussed on a lithographically (Appendix B) produced grid (SiN<sub>x</sub>, 50 nm slits at a distance of 100 nm). Typical distances between the apertures and between grid and detector are in the meter range. For detection the C<sub>60</sub> molecules are ionized by laser light and collected by an electron optics in front of a chaneltron arrangement.

In Fig. 2.8a the observed interference pattern consisting of a central peak and two 1st order side peaks is shown. In the control spectrum measured without grid in the molecular beam (Fig. 2.8b) the interference peaks are missing. The solid curve in Fig. 2.9a calculated by means of grid diffraction theory is based on the (De Broglie) wavelength  $\lambda = h/vm_{C60} = 2.5 \times 10^{-10} \text{ cm} = 2.5 \text{ pm}$ , which is attributed to the moving molecules (v velocity, m<sub>C60</sub> mass) according to (2.4b).

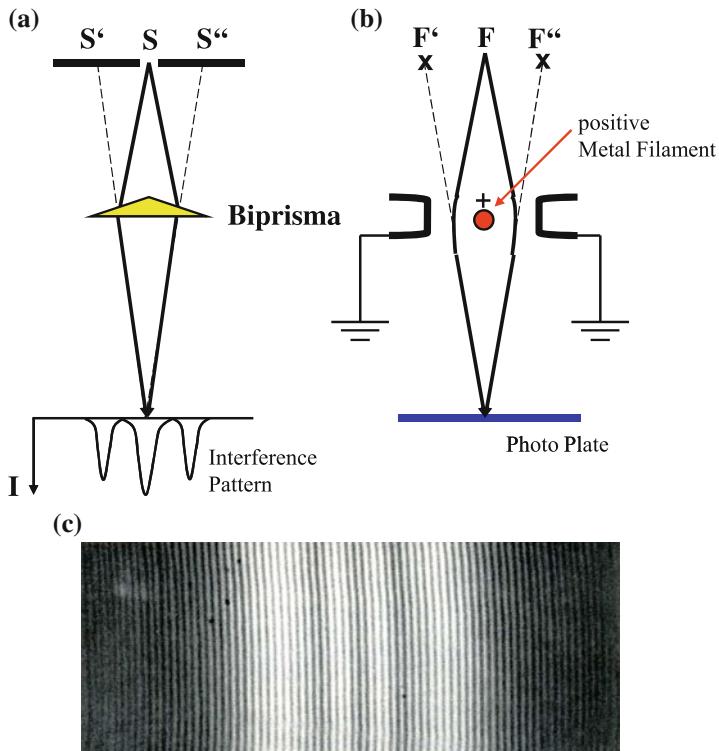
All these experiments with particle beams demonstrate clearly and doubtlessly, that the propagation of massive particles as electrons, neutrons, molecules etc. must be described in terms of wave expansion. Otherwise, we could not understand the occurrence of diffraction and interference phenomena observed with these particles and which are used meanwhile worldwide in standard characterization and analysis techniques in solid state and surface physics. Present cutting edge research in this field aims at the physical limits for the observation of particle interference with bigger and bigger particles. The interesting question is, at what particle size is the quantum character lost and the particle starts to behave classically.

## 2.4 Particle Interference at the Double Slit

Interference experiments with a double slit, that is, the appearance of diffraction fringes on a screen after a light beam has passed the double slit arrangement, lead Th. Young already in 1802 to the interpretation of light as a wave. Instead of a double slit A.J. Fresnel used a bi-prism (Fig. 2.9a) for the demonstration of double slit interferences. In this particular set-up a monochromatic light beam originating from a single slit  $S$  illuminates a double prism with small prism angles. This bi-prism splits the primary beam into two partial beams which are superimposed on a remote screen. As is seen from Fig. 2.9a, the two partial beams seem to originate from two virtual slits  $S'$  and  $S''$ . The interference pattern observed on the screen, thus, is identical with one produced by a double slit arrangement as in Young's experiment. The intensity  $I$  of the interference pattern reaches a maximum when the path difference between the two partial waves from  $S'$  and  $S''$  equals a multiple of the light wavelength  $\lambda$ . Destructive interference, that is, intensity minima appear on the screen for path differences of odd multiples of  $\lambda/2$ . These types of double slit interferences can only be explained in terms of wave propagation, a non-local phenomenon. An interpretation on the basis of a particle picture is excluded.

### 2.4.1 Double Slit Experiments with Electrons

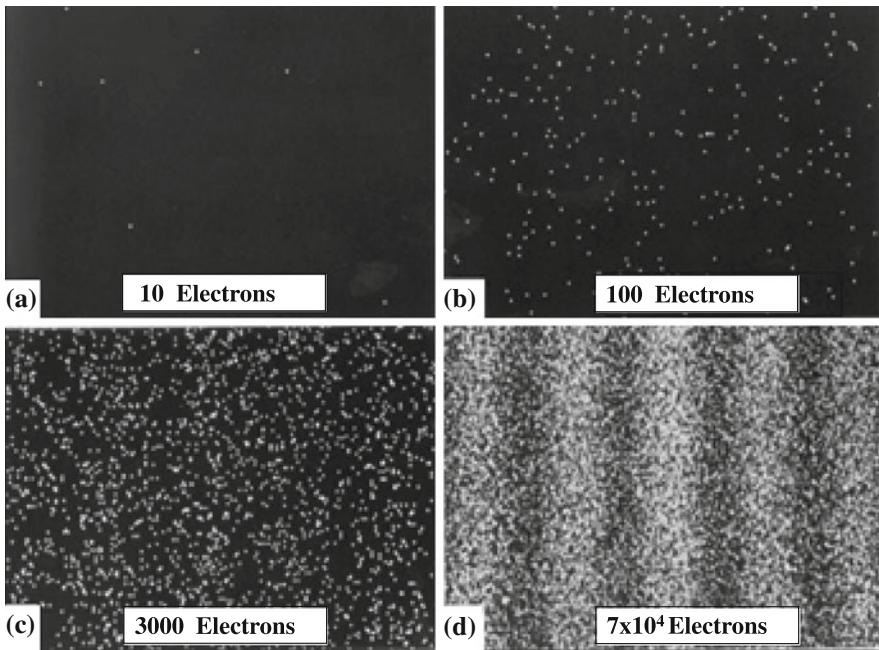
Already in 1956, Möllenstedt and Düker performed a double slit experiment with electrons by means of a bi-prism [9]. The bi-prism for electrons in this experiment consisted of a positively charged metallic filament arranged between two planar electrodes on ground potential (Fig. 2.9b). This set-up is incorporated into an electron microscope column, where an electron beam is focused in a focal point  $F$  (Fig. 2.9b). The double prism arrangement splits the electron beam into two partial beams, similarly as in the optical analogon, and deflects the two beams to the center again. The electric field of the positive filament is proportional to  $r^{-1}$  ( $r$  distance from filament). An electron passing the wire in close vicinity is strongly deflected horizontally, but



**Fig. 2.9** a–c Double slit diffraction of light and of electrons. **a** Set-up for the observation of optical double beam interference with monochromatic light. The two light beams are produced by an optical biprism. **b** Analogue equipment for the observation of electron double beam interference. The biprism is realized by a positively charged metal filament in an electron microscope column. **c** Electron double slit interference pattern produced by the experimental set-up in (b) [9, 10]

only for a short time. An electron passing further away experiences a smaller force, but this for a longer time. The total deflection angle of the electrons in the field of the wire surprisingly depends only on the electron energy and not on the distance from the wire. Thus the two partial electron beams are focused and superimposed on a photosensitive screen behind. An interference pattern with bright and dark fringes is observed (Fig. 2.9c). Electrons with a fixed energy thus behave as light waves passing Fresnel's bi-prism or Young's double slit, a further demonstration of the wave character of electrons.

The experiment of Möllenstedt and Düker was repeated by Tonomura et al. [11] in 1989 with more sophisticated experimental tools. A particular advancement was the use of extremely sensitive, space resolving (imaging) semiconductor detectors. A whole field of highly sensitive pixel detectors enables the detection of one single electron at one pixel and, thus, the computer aided construction of an image of the spatial distribution of the electrons having passed the double slit. The results of the



**Fig. 2.10 a–d** Successive formation of a two-beam (*double slit*) electron interference pattern. The diffraction experiment has been performed by means of a biprism set-up as depicted in Fig. 2.9b [11]. The electron density in the beam is such low that only one single electron passes the electron microscope column at a time. Only single distinct electrons are detected, one after each other, on the 2-dimensional spatially resolving pixel detector screen. The diffraction patterns (a–d) are recorded after increasing electron numbers have passed the apparatus

experiment (Fig. 2.10) clearly show the unexpected and weird behavior of electrons propagating in space.

Electrons expand in space according to the laws of waves, they produce interference patterns, just as light does. But the interference fringes become visible only after the observation of a sufficiently large number of electrons. The observation of only 10 electrons which have passed the bi-prism (Fig. 2.10a) yields a random flash of one pixel somewhere on the screen. An interference pattern can not be recognized. Collecting 100, 3000, or 70,000 events of electrons which have passed the bi-prism builds up step by step the double slit interference pattern (Fig. 2.10d). Only for an ensemble with huge numbers of electrons the laws of wave propagation are valid. One single electron behaves randomly; totally unexpected and statistically the response of a pixel on the screen is caused by an impinging electron which transfers its kinetic energy to the point-like pixel detector.

It must be emphasized at this point that an electron-electron interaction can be excluded while the electrons pass the double prism arrangement to form the interference fringes. Two subsequent electrons do not “see” each other in space and time. The intensity of the electron beam current is so low that only after the detection of

one electron in a pixel detector the next electron leaves the cathode of the microscope column.

Single electrons have the choice to take one or the other path—through this or the other slit—they are detected as point-like particles in a pixel detector, but randomly distributed over the screen. We do not know their individual history, but as an ensemble they build up the interference pattern without having information about each other. This particle-wave duality, which is absolutely counter-intuitive, weird in our imagination, is at the heart of quantum mechanics. Feynman [12] describes this behavior being apparent in the double slit experiment as “impossible, absolutely impossible to explain in any classical way, and has in it the heart of quantum mechanics”. We have to get familiar with the idea, that nature behaves completely different from our everyday experience on an atomic scale or below. For human beings, the natural length scale is that of centimeters and meters corresponding to the perception horizon in our macroscopic surrounding. It would be astonishing, on the other hand, if our sense organs and our brain, which have adapted during more than 100 million years of biological evolution to a macroscopic environment, could perceive the reality of the whole cosmos, the smallest and largest on subatomic and cosmological length scales. In these periods of adaptation it was much more important for human survival to correctly estimate the width of a creek or the distance between two branches of an arbor than the path of an electron. We should, therefore, not be surprised that the atomic and sub-atomic world as it appears in quantum physics is not accessible to our limited senses and imagination. We should, however, be surprised that mathematics opens the way to create an abstract picture of the atomic behavior which allows even quantitative predictions of experimental results. The most straightforward explanation is certainly that a structured reality does exist beyond human perception and imagination which obeys the laws of logic. Mathematics and logic obviously go beyond the reality accessible to our senses and enable the invention of theoretical systems as quantum theory which can correctly describe wide fields of reality extending much further than our meter and centimeter environment.

#### ***2.4.2 Particle Interference and “Which-Way” Information***

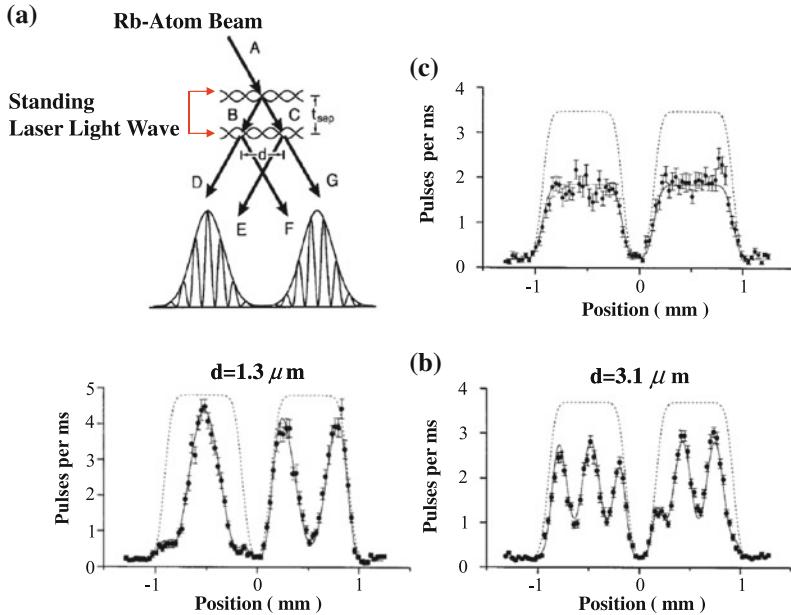
The behavior of atomic and sub-atomic particles becomes even more strange when we ask the question through which particular slit has the particle moved in the double slit experiment (Sect. 2.4.1). Is this question for the detailed way of the particle compatible with the observation of the double slit interference pattern? Already in the early days of quantum mechanics, around 1920, this question was discussed extensively in gedanken (thought) experiments by Heisenberg, Einstein and others and later by Feynman [12]. The essential conclusion of all these discussions always was that the interference pattern can only be observed without additional experiments to elucidate the detailed path (“which-way” information) of the particles. Every measurement of the detailed way, e.g. by scattering of a photon (see Compton effect, Sect. 2.2) in front of one of the two slits transfers so much momentum  $p = \hbar k$

to the electron that interference of the electron waves is not possible anymore, the fringe pattern is washed out due to phase shifts. According to the arguments of Heisenberg and Feynman the photon energy of the probing light can be decreased to such an extent that its effect on the electron is negligible. But simultaneously one has to increase the wavelength of the light, because of  $p = \hbar k = h/\lambda$ , to an amount which does not resolve the spatial distance between the two slits anymore. Microscopic imaging of a structural dimension  $d$ , namely, requires  $\lambda < d$ . In the gedanken experiment, the measurement of the detailed particle path requires a light wavelength  $\lambda < \text{slit distance}$ , which simultaneously is accompanied by a momentum transfer to the electron high enough to destroy the interference pattern.

In recent time, now, experiments became possible, where the “which-way” information can be obtained without significant momentum transfer to the diffracted particle in a double slit experiment. But look, the interference pattern disappears without momentum transfer. The interference fringes can only be seen, when the detection apparatus for the “which-way” information is switched off. Dürr et al. [13, 14] have performed an experiment with a beam of Rb atoms which are diffracted on a standing laser light wave. As we will see later in Chap. 8, high intensity standing light waves with their spatially fixed intensity maxima and knots (intensity = 0) act as a diffraction grating for atoms, with a grating period of half the light wavelength, similarly as the periodic array of atoms in a crystal (Sect. 2.3). According to Fig. 2.11a, diffraction of the Rb atoms on a first standing wave produces, beside the transmitted beam  $C$  (0th order) a beam  $B$  diffracted in 1st order. These two atom beams hit a second standing light wave where they are diffracted into the beams  $D, E$  and  $F, G$ , which pair-wise interfere with each other. Thus, two interference patterns phase shifted against each other are produced in a space resolving imaging detector behind. Figure 2.11b shows the experimentally observed interference patterns for two different laser light wavelengths with knot distances (periodicity period)  $d = 1.3$  and  $3.1 \mu\text{m}$ .

A special property of this experiment is due to the fact that the diffracted Rb atoms are characterized, beside their spatial information, that is, the probability of being somewhere, also by internal degrees of freedom as spin excitations etc. We will be able to understand details of the described experiment only much later in this book (Sect. 8.2.4) after we have learnt a lot more about quantum theory. Nevertheless, it should be anticipated at this point, that irradiation of microwave radiation with a frequency of 3 GHz excites the Rb atoms into an excited state before entering the first diffraction grating (1st standing wave). A second microwave pulse irradiated after the splitting into the two partial beams  $B$  and  $C$  allows the distinction between the two possibilities if the interference pattern (beams  $D$  and  $E$  respectively,  $F$  and  $G$ ) originates from an atom of the partial beam  $B$  or  $C$ .

In this experiment, the two beams of the double slit experiment are realized by the partial beams  $B$  and  $C$ . By means of microwave pulses before and after passing the first diffraction grating (1st standing light wave) one can distinguish between the ways  $B$  and  $C$  which could have been taken by the atom. It is easily estimated (Sect. 8.2.4) that a photon of 3 GHz microwave radiation can not transfer enough momentum to the relatively heavy Rb atom such that the interference pattern is washed out. Nev-



**Fig. 2.11** a–c Two-beam interference of two Rb atom beams. An internal degree of freedom (spin orientation of the outer Rb shell electron) can yield information about the path of a single electron (“Which Way Information”) [13, 14]. **a** Scheme of the atom interferometer: By Bragg reflection on an intense standing laser light wave the incident atomic beam  $A$  is split into two partial beams  $B$  and  $C$ . A second standing laser light wave splits these two beams into the partial beams  $D$  and  $E$  (negative spatial coordinate), respectively  $F$  and  $G$  (positive spatial coordinate). These beams pairwise interfere with each other. Irradiation of microwaves before entering the first diffraction grating (1st standing laser light wave) can excite the Rb atoms in an excited internal state. A 2nd microwave pulse irradiated between the two diffraction gratings (1st and 2nd standing laser light waves) allows the read-out of the “Which Way Information”, i.e., the detailed path of the two interfering atomic beams (see also Sect. 8.2.4 and Fig. 8.5). **b** Measured atomic beam interference pattern originating from the superposition of the partial beams  $D$  and  $E$ , respectively,  $F$  and  $G$ ; for these measurements the “Which-Way Information” was not recorded (no microwave pulses); results with two different grating periods (node distance of standing laser waves)  $d = 1.3 \mu\text{m}$  and  $d = 3.1 \mu\text{m}$ . The solid lines are calculated results. **c** Measured beam intensities upon superposition of partial beams  $D$  and  $E$ , respectively,  $F$  and  $G$  and recording the “Which-Way Information” using microwave pulses

ertheless switching on the microwave radiation as the measurement probe destroys the interference (Fig. 2.11c). Only a monotonous intensity background corresponding to the average Rb atom density in the beams  $D$  and  $E$  respectively,  $F$  and  $G$  is detected. This experimental result is found independently on the observation by a human experimentalist; only the read-out of the which-way information by the corresponding hard-ware probe is essential for the appearance or disappearance of the interference pattern.

What do we learn from this experiment? First, we see that not the human observer has an effect on the outcome of the interference experiment, only the switched on

measurement probe for the which-way information is responsible for the destruction of the interference pattern.

Real world does not worry if it is observed by a human being (Realism instead of Idealism!). Furthermore, there must exist a correlation between the observed particle and the measurement probe, which can not be reduced to energy or momentum transfer between particle and measurement set-up. This phenomenon which is inherently of quantum mechanical character is typical for atomic and sub-atomic systems and beyond our macroscopic perception. It is called “entanglement” (Verschränkung in German, as Schrödinger called it), we will better understand what it means after having learnt more about quantum physics (Chap. 7).

## References

1. A. Einstein, Ann. Phys. **17**, 132 (1905)
2. A.H. Compton, A. Simon, Phys. Rev. **25**, 306 (1925)
3. C.J. Davisson, L.H. Germer, Phys. Rev. **30**, 705 (1927)
4. L. de Broglie, Comptes Rendus Acad. Sci. Paris **177**, 507 (1923)
5. I. Estermann, O. Stern, Z. Phys. **61**, 95 (1930)
6. G. Comsa, G. Mechtersheimer, B. Poelsema, S. Tomoda, Surf. Sci. **89**, 123 (1979)
7. C.G. Shull, S. Siegel, Phys. Rev. **75**, 1008 (1949)
8. M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, A. Zeilinger, Nature **401**, 680 (1999)
9. G. Möllenstedt, H. Dücker, Z. Phys. **145**, 366 (1956)
10. C. Jónsson, Z. Phys. **161**, 454 (1961)
11. A. Tonomura, J. Endo, T. Matsuda, T. Kaeasaki, E. Ezawa, Am. J. Phys. **57**, 157 (1989)
12. R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics—Quantum Mechanics* (Addison-Wesley, Reading, 1965)
13. S. Dürr, T. Nonn, G. Rempe, Nature **395**, 33 (1998)
14. S. Dürr, G. Rempe, Adv. At. Mol. Opt. Phys. **42**, 29 (2000)

# Chapter 3

## Particle-Wave Duality

### 3.1 The Wave Function and Its Interpretation

The experiments described in Chap. 2 doubtlessly show that both light waves propagating in space as well as atomic and subatomic particles as electrons moving from one to another spot have one thing in common: Their propagation obeys the laws of wave expansion. On the other hand, also the particle character, which shows up in scattering experiments as the Compton effect and in the detection process, can not be denied. In a simplifying fashion one can say: everything, matter and energy fields, are simultaneously wave and particle.

We, thus, arrive at the correspondence between particle and wave picture which follows from the experiments in Chap. 2 and which is expressed by the following relations between particle energy  $E$ , particle momentum  $\mathbf{p}$  and frequency  $\omega$  respectively, wavelength  $\lambda$ :

$$E = \frac{1}{2}mv^2 = \hbar\omega, \quad (3.1)$$

$$\mathbf{p} = m\mathbf{v} = \hbar\mathbf{k} = \hbar \frac{2\pi}{\lambda} \frac{\mathbf{k}}{|k|}. \quad (3.2)$$

Mass  $m$  and velocity  $v$  are quantities characteristic for the particle picture while frequency  $\omega$ , wavelength  $\lambda$  and wave vector  $\mathbf{k}$  make contact to the wave description. Equations (3.1) and (3.2) might be used in a first approach to quantitatively describe experiments as in Chap. 2, but there is need for a stringent and coherent theory for the dynamics of particles, which combines the concepts of wave and particle propagation being at first glance contradictory. This highly challenging goal was first achieved with Schrödinger's wave mechanics.

The propagation of a particle, e.g. of an electron is described by a wave function  $\psi$ . In the simplest case of motion along a straight line a plane wave

$$\psi(\mathbf{r}, t) = ce^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad (3.3)$$

describes the propagation of the particle, where wave vector  $\mathbf{k}$  and frequency  $\omega$  are connected to the particle picture by (3.1) and (3.2). The wave function  $\psi$  is a quantity which is analogous to the wave amplitude of a light field. Its absolute square is identified with an observed intensity after collecting a huge number of electrons on a screen (Sect. 2.4.1). In particular, the interference pattern in a double slit experiment with electrons is obtained by superimposing two waves  $\psi_1$  and  $\psi_2$  originating from two slits 1 and 2 at the positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$  on a remote screen at  $\mathbf{r}$  (Fig. 3.1).

At a long distance from the source both spherical and cylinder waves (circular holes or slits at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ ) can be approximated by plane waves. At the observation point  $\mathbf{r}$  on the remote screen, the superposition of the two wave functions thus yields

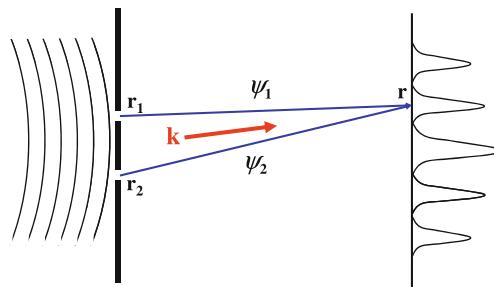
$$\psi = \psi_1 + \psi_2 \quad \text{with } \psi_i = ce^{i[\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_i) - \omega t]}. \quad (3.4)$$

In analogy to light waves the absolute square of the wave function (wave amplitude), the intensity  $I$ , describes the observed intensity contrast on the screen, the sequence of bright and dark fringes as seen in Fig. 2.9:

$$I = |\psi(\mathbf{r}, t)|^2 = |\psi_1|^2 + |\psi_2|^2 + 2c^2 \cos \mathbf{k} \cdot (\mathbf{r}_2 - \mathbf{r}_1). \quad (3.5)$$

In (3.5), the wave vectors of the two partial beams coming from  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are approximated by one and the same  $\mathbf{k}$  (Fig. 3.1). The cosine term in (3.5), in which the wavelength  $\lambda = 2\pi/k$  of the particle and the slit distance  $(\mathbf{r}_1 - \mathbf{r}_2)$  determine the periodicity of the interference fringes, describes very accurately the observed interference pattern.

$\mathbf{k}(\mathbf{r}_2 - \mathbf{r}_1)$  is the path difference between the two partial waves  $\psi_1$  and  $\psi_2$  in units of the particle wave length, it determines the appearance of interference maxima and minima, while  $|\psi_1|^2$  and  $|\psi_2|^2$  describe a homogeneous intensity background. This background intensity is exactly the signal (Fig. 2.10c) which is observed in the double slit experiment when “which-way” information is read out by an additional



**Fig. 3.1** Scheme of double slit interference of two particle waves  $\psi_1$  and  $\psi_2$ . An approximately plane wave (in reality a spherical wave) generates two new waves at the slits (holes) at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . These two partial waves interfere at the position  $\mathbf{r}$ . For large enough distances between the detection screen ( $\mathbf{r}$ ) and the double slit ( $\mathbf{r}_1, \mathbf{r}_2$ ), the wave vectors of the two particle waves  $\psi_1$  and  $\psi_2$  can be assumed to be identical as  $\mathbf{k}$

measurement (Sect. 2.4.2). Information about the detailed path of the particle, through slit 1 or slit 2, therefore, washes out the cosine term in (3.5).

From the double slit experiment with a varying number of electrons (Fig. 2.9), we have learnt that the interference pattern typical for wave expansion is only observed for a sufficiently large number of electrons having passed the double slit. Each single electron is detected fully randomly on the observation screen. Details about the path of a single particle are not reasonable questions in a quantum physical description, the single particle behaves in a statistical manner, fully randomly. Only the ensemble of many particles obeys the laws of waves.

From this observation, we conclude that the wave function  $\psi(\mathbf{r}, t)$  is a statistical quantity, which describes only ensemble properties. The more electrons are observed in the experiment of Fig. 2.9 the better (3.5) describes the interference pattern. This leads directly to the conclusion that  $|\psi(\mathbf{r}, t)|^2$  in (3.5) is proportional to the probability to find an electron at the position  $\mathbf{r}$  at time  $t$ . The probability  $dP$  to find an electron in a volume element  $d^3r$  is proportional to the volume and, of course, to the probability  $|\psi(\mathbf{r}, t)|^2$ , that is,

$$dP \propto |\psi(\mathbf{r}, t)|^2 d^3r, \quad (3.6)$$

$|\psi(\mathbf{r}, t)|^2$  is, thus, the probability density to find an electron at  $\mathbf{r}$  and time  $t$ . Correspondingly the wave function  $\psi(\mathbf{r}, t)$  is called a probability amplitude.

When an electron, or more generally a particle, must be present in a certain volume  $V$  of an experimental set-up, but we do not know where, we can only say the particle is present in the volume  $V$  with certainty. It might be somewhere at  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_n$  within the volume  $V$ . The total probability to find the particle whether at  $\mathbf{r}_1$  or  $\mathbf{r}_2, \dots$ , or  $\mathbf{r}_n$ , that is, somewhere in  $V$ , must be one (certainty). Probabilities of independent events—whether or—add up to the total probability that one or the other event occurs. Therefore, the total probability to find the particle somewhere in the volume  $V$  must be written as

$$P(\text{particle in } V) = \int_V d^3r |\psi(\mathbf{r})|^2 = 1. \quad (3.7)$$

The wave function as a probability density must be normalized, in the sense of (3.7), over the volume of the whole system considered. Depending on the particular problem the considered volume  $V$  might be the whole universe.

As is commonly done for the description of electromagnetic waves and alternating currents, we have adopted a complex valued wave function in (3.4) for the representation of particle propagation. In electromagnetism and electricity theory this is done simply for convenience reasons, one can more easily calculate with exponential than with sine or cosine functions. All real, measured physical quantities as currents or electric and magnetic fields are obtained as real or imaginary parts of the complex wave amplitudes. This is fundamentally different for wave functions of particle waves in quantum mechanics. This becomes clear from the following consideration. Imagine a spatially extended homogeneous electron beam, in

which a propagating electron is described by a plane wave (3.3). The probability density  $|\psi|^2 = \psi^* \psi$ , then, is a constant all over the beam, just what is required to find an electron with equal probability at each location. If we would allow only real valued functions for the electronic wave functions as sine or cosine, then, the probabilities for finding an electron would have  $\sin^2$  or respectively  $\cos^2$  character, that is, spatially and timely restricted regions where electrons are found or not found. This is certainly not an appropriate description for a spatially homogeneous electron beam. The wave description of particle propagation, thus, requires complex valued wave functions  $\psi(\mathbf{r}, t)$ . We will see that in some particular cases for special boundary conditions, of course, also real valued wave functions are obtained as solutions for a problem. But we want to stress that particle waves  $\psi(\mathbf{r}, t)$  in general must have an amplitude and a phase or a real and imaginary part, respectively.

We summarize: Atomic and subatomic particles are described by a complex valued wave function  $\psi(\mathbf{r}, t)$  which is normalized to one according to (3.7). The free propagation of a particle along a straight line in space is represented by a plane wave (3.3), where the wave vector  $\mathbf{k}$  indicates the propagation direction. A wave function does not give any information about one single particle, its fate is inherently undetermined and stochastic. A wave function is a statistical quantity, its absolute square  $|\psi(\mathbf{r}, t)|^2$  describes the probability to find a particle at  $\mathbf{r}$  at a time  $t$ . For a large ensemble of particles, the description in terms of  $|\psi(\mathbf{r}, t)|^2$  is correct as is seen from the double slit experiment in Fig. 2.9. Since the fate of a single particle is principally undetermined, the statistical interpretation of the wave function in quantum physics is more fundamental and rigorous than in classical statistical mechanics. The dynamical equations in classical mechanics are deterministic. From known initial conditions (position and velocity), classical mechanics allows the prediction of the future behavior of a particle. Of course, quasi-un-deterministic behavior might arise from tiny fluctuations in the initial conditions (butterfly problem). Nevertheless, the statistical description of a large particle ensemble in classical statistical theory (kinetic gas theory, Brownian motion etc.) is a construct by which definite quantities of a large ensemble can be calculated without knowing details about the motion of single particles. In quantum physics, however, the un-deterministic stochastic behavior of a particle is an inherent property. According to our present knowledge so far—and our knowledge is quite profound and based on a wealth of experimental data—there are no “hidden variables” which govern the behavior of a single atomic or subatomic particle below its description level in terms of a wave function  $\psi$ . The wave function is ascribed to a certain particle, but only a large ensemble of these particles shows the behavior in an experiment which is described by  $|\psi|^2$ .

## 3.2 Wave Packet and Particle Velocity

The energy-frequency relation (3.1), the connection between particle momentum and wave number  $k$  (3.2) as well as the description of particle propagation by a wave function and its statistical interpretation (3.3), (3.4) and (3.6) are the starting point for the formal description of the particle-wave duality. There is a severe difficulty with our formal description so far: For a spatially extended wave—in the extreme limit, over the whole space—the velocity of a particle can not be described. The term velocity contains inherently the movement of a particle, an entity, which is more or less limited in its spatial extension. How can we make our picture of an extended wave compatible with that of a propagating particle. The key to the solution of the problem is the mathematical formalism of the Fourier-transform. Every “non-pathological” function (mathematical details shall not be considered) can be represented by a superposition of plane waves (3.3), that is, by a sum or an integral over infinitely many waves with densely distributed wave numbers  $k$ .

A particle with a spatial extension  $\Delta x$  in one dimension might be described in simple approximation by a wave function  $\psi$  having Gaussian shape.  $\psi^* \psi$ , then, is the probability to find the particle in the limited spatial region defined by the Gaussian bell-shaped curve, essentially the spread of the Gauss curve. The Gauss function  $\psi$  is evaluated in a Fourier series

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a(k) e^{ikx} dk, \quad (3.8a)$$

where  $a(k)$  denotes the distribution of wave numbers  $k$  which are required to build up the Gaussian function on the  $x$ -axis in real space. For simplicity reasons, we first consider the wave function at the time  $t = 0$ . The Fourier transform (3.8a) can be inverted to yield the  $k$ -distribution:

$$a(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) e^{-ikx} dx. \quad (3.8b)$$

We assume a Gauss function for  $\psi$  in (3.8a) and denote its spread (width) on the  $x$ -scale by  $(\Delta x)^2$  which describes the average spatial extension of the particle:

$$\psi(x) = [2\pi(\Delta x)^2]^{-\frac{1}{4}} \exp\left(-\frac{x^2}{4(\Delta x)^2}\right). \quad (3.9)$$

The choice of the pre-factor is due to the normalization condition for  $|\psi|^2$  (3.7). We, thus, obtain, according to (3.8b), the following expression for the  $a(k)$  distribution which builds up the packet of waves representing the  $\psi$  function in real space.

$$\begin{aligned} a(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [2\pi(\Delta x)^2]^{-\frac{1}{4}} \exp\left(-\frac{x^2}{4(\Delta x)^2}\right) \exp(-ikx) dx \\ &= \frac{1}{\sqrt[4]{(2\pi)^3(\Delta x)^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{4(\Delta x)^2}\right) \exp(-ikx) dx. \end{aligned} \quad (3.10)$$

With

$$\begin{aligned} -\frac{x^2}{4(\Delta x)^2} - ikx &= \frac{(-1)}{4(\Delta x)^2} (x^2 + i4(\Delta x)^2 kx - 4(\Delta x)^4 k^2 + 4(\Delta x)^4 k^2) \\ &= \frac{(-1)}{4(\Delta x)^2} [(x + 2i(\Delta x)^2 k)^2 + 4(\Delta x)^4 k^2] \end{aligned} \quad (3.11a)$$

and

$$\gamma = \frac{(x + 2i(\Delta x)^2 k)}{2(\Delta x)} \quad \text{or} \quad \frac{d\gamma}{dx} = \frac{1}{2(\Delta x)} \quad (3.11b)$$

follows

$$a(k) = \frac{2(\Delta x)}{\sqrt[4]{(2\pi)^3(\Delta x)^2}} \exp[-(\Delta x)^2 k^2] \int_{-\infty}^{\infty} \exp(-\gamma^2) d\gamma. \quad (3.12a)$$

The last integral equals  $\sqrt{\pi}$  and we obtain finally:

$$a(k) = \left(\frac{2}{\pi}\right)^{\frac{1}{4}} (\Delta x)^{\frac{1}{2}} \exp[-(\Delta x)^2 k^2] = \left[\frac{4(\Delta x)^2}{2\pi}\right]^{\frac{1}{4}} \exp[-(\Delta x)^2 k^2]. \quad (3.12b)$$

The distribution of  $k$  vectors  $a(k)$  building up the Gaussian wave packet is, thus, again a Gauss function. The wave vectors of waves which form a wave function with Gaussian shape in real space are therefore Gauss distributed (3.12b). If we compare (3.12b) with the common representation of a Gauss distribution as function of  $k$  with width  $\Delta k$

$$a(k) = \frac{1}{2\pi(\Delta k)^2} \exp\left[-\frac{k^2}{4(\Delta k)^2}\right], \quad (3.12c)$$

we obtain the following relation between spatial width  $\Delta x$  of the wave packet and the spread or width of the corresponding wave vector distribution  $\Delta k$ :

$$\Delta k \Delta x = \frac{1}{2}. \quad (3.13)$$

We can summarize: Particle and wave picture are unified by attributing a superposition of infinitely many plane waves  $\exp(ikx)$  with a Gaussian  $k$  vector distribution (3.10), (3.12a)–(3.12c) to the spatially restricted particle wave function  $\psi$ . The propagation of a particle which is a spatially limited entity can be described by the expansion of infinitely many plane waves in space. For a bell-shaped Gaussian wave

function with spatial width  $\Delta x$  a Gaussian distribution of  $k$  vectors  $a(k)$  is required, with width  $\Delta k = 1/2(\Delta x)^{-1}$ .

So far we have considered the wave packet  $\psi(x)$  and the related  $k$  distribution  $a(k)$  at time  $t = 0$ . The time evolution of the wave packet is easily derived from that of the participating plane waves by means of their frequencies  $\omega(k) = (1/\hbar)E(k)$ . The propagating wave packet (particle) is then represented as a superposition of expanding plane waves moving in time:

$$\psi(x, t) = \int dk a(k) e^{i(kx - \omega t)} \propto \int dk e^{-(\frac{k-k_0}{2\Delta k})^2} e^{i(kx - \omega(k)t)}. \quad (3.14)$$

Hereby, the participating plane waves are assumed to be distributed around a central wave number  $k_0$  with a spread (width)  $\Delta k$ . All these waves propagate with different frequencies  $\omega(k)$ . For a sufficiently narrow Gaussian distribution  $a(k)$ , we can evaluate  $\omega(k)$  around the central wave number  $k_0$ :

$$\omega(k) = \omega(k_0) + \left. \frac{\partial \omega}{\partial k} \right|_{k_0} (k - k_0). \quad (3.15)$$

Using (3.15) and by multiplying (3.14) with  $\exp(ik_0x) \exp(-ik_0x) = 1$ , we obtain

$$\psi(x, t) \propto e^{i(k_0x - \omega(k_0)t)} \int dk e^{-(\frac{k-k_0}{2\Delta k})^2} e^{i(x - \frac{\partial \omega}{\partial k}t)(k - k_0)}. \quad (3.16)$$

This wave function is easily interpreted as a Gaussian wave packet whose maximum is located at  $x - (\partial \omega / \partial k)k_0 t$  rather than  $x$ . The wave packet is multiplied by the plane wave  $\exp[i(k_0x - \omega(k_0)t)]$  with the central wave number  $k_0$ . The maximum of this wave packet thus propagates along the  $x$ -axis with the velocity

$$v = \left. \frac{\partial \omega}{\partial k} \right|_{k_0}. \quad (3.17)$$

Equation (3.17) is called *group velocity* of the wave packet, in contrast to the phase velocity  $v_{\text{phase}} = \omega/k$  of a single plane wave which contributes to the formation of the whole wave packet. In the particle picture, the group velocity (3.17) is identified with the velocity of the moving particle. In three-dimensional (3D) space the same formalism holds, that is, the integrals (3.8a)–(3.10) have to be extended over 3D vectors  $\mathbf{r}$  and  $\mathbf{k}$  by means of volume elements  $d^3r$  and  $d^3k$ . The group velocity in 3D space, then, is the gradient of the frequency or the energy of the particle, respectively:

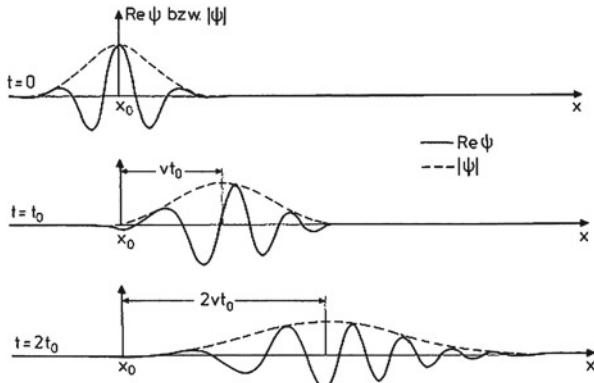
$$v = \nabla_k \omega(k) = \frac{1}{\hbar} \nabla_k E(k). \quad (3.18)$$

For light waves in vacuum with  $\omega = ck$  group velocity  $\partial \omega / \partial k$  and phase velocity  $\omega/k$  are identical with light velocity  $c$ . All light waves which build up a wave

packet propagate with the same velocity  $c$ . The wave packet keeps its shape during propagation in space.

For massive particle waves, however, the energy-wave vector dispersion relation is  $E = \hbar\omega = \hbar^2 k^2 / 2m$ , that is, the group velocity of a particle  $\partial\omega/\partial k = \hbar k/m = p/m$  is not identical with the phase velocity  $\omega/k = \hbar k/2m = p/2m$  of the single waves which build up the wave packet. While the center of mass of the wave packet moves with the particle velocity  $p/m$ , the single constituting waves  $(\omega, k)$  propagate with different speed. Short wave length waves (larger  $k$ ) are faster than those with longer wave lengths, which are passed over. This effect causes a broadening of the wave packet during propagation of the particle (Fig. 3.2). The phenomenon is called *dispersion*. Massive particle waves exhibit dispersion, while light waves in vacuum propagate without dispersion because of the proportionality  $\omega \propto k$ .

There must be consistence between the formal introduction of the group velocity  $\partial\omega/\partial k$  of a wave packet and the energy-frequency (3.1) and momentum-wave vector (3.2) relations. Because of the classical particle energy  $E = (1/2)mv^2$  the velocity of a particle in the particle picture  $v = \partial E/\partial p$  (with  $p = mv$ ) can consistently be related to the velocity in the wave picture  $v = \partial\omega/\partial k$  only, if both  $E$  is proportional to  $\omega$  and  $p$  proportional to the wave number  $k$  via one and the same constant. This constant  $\hbar = h/2\pi$  (2.1b) was introduced in Sect. 2.1 in connection with the interpretation of the photoelectric effect. We will see later that the sum of wave vectors is conserved upon scattering of waves on each other as does the sum of momenta ( $\hbar k$ ) in a particle scattering process, that is, also in this context  $\hbar$  and  $h$ , respectively, must be a universal constant with general importance for all kinds of waves and particles.



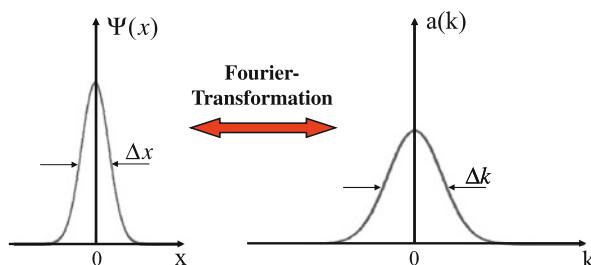
**Fig. 3.2** Schematic representation of a wave packet ( $\text{Re } \psi$ : solid line,  $|\psi|$ : broken line), which describes the propagation of a spatially localized free electron. The wave packet is shown for different times  $t = 0, t_0, 2t_0$ . The center of the wave packet, i.e. in the particle picture the electron itself, propagates with the group velocity  $v = \partial\omega/\partial k$ . The spectral width of the wave packet  $|\psi|$  broadens with time. During the broadening of the packet the wavelength of the  $\text{Re } \psi$  oscillations decreases at the front side while it increases at the back side

### 3.3 The Uncertainty Principle

From the representation of a particle by means of a wave packet, we conclude directly that the width  $\Delta k$  of the distribution of wave vectors  $a(k)$  which constitute the wave packet is inversely proportional to the spread, that is, the spatial extension of the wave packet (3.13). For a Gaussian wave packet we quantitatively obtain the relation (3.13). The Gaussian packet, of course is a special case of a wave function which describes a spatially limited probability amplitude to find a particle localized within a certain volume at a particular position. Many other mathematical forms of wave functions can be imagined which describe a localized particle by means of a spatially limited function defined in a particular range  $\Delta x$ . Examples may be a rectangular box extended along a width  $\Delta x$  or the function  $\sin^2 x/x^2$  which is defined along the whole  $x$ -axis, but with non-negligible values only in a limited  $x$  range. Depending on the particular problem all these wave packets can be used to describe the propagation of a particle. An evaluation of these wave packets being restricted in some way to a limited space region  $\Delta x$  in a Fourier series always yields, as Fourier transform, a distribution  $a(k)$  of  $k$  vectors which are spread over a limited range  $\Delta k$ , only (Fig. 3.3). Because of the general rules of Fourier transformation a relation similar to (3.13) is always valid:

$$\Delta x \cdot \Delta k \sim 1. \quad (3.19)$$

This relation between the spatial width  $\Delta x$  of a wave packet and the spread  $\Delta k$  of its Fourier transform leads to an important, typically quantum physical phenomenon. According to the probability interpretation of the wave function we have to interpret the width  $\Delta x$  of the wave packet as that spatial range where we find the particle in a position measurement. The exact detailed position of its detection within  $\Delta x$  is stochastic and not determined in quantum physics. Analogously also the wave vector  $k$  of the particle is determined only within an uncertainty range  $\Delta k$ . With  $p = \hbar k$  as the particle momentum both position and momentum of the particle are inherently undetermined as exact values, they can only be obtained within certain margins  $\Delta x$



**Fig. 3.3** Gaussian wave packet  $\psi(x)$  with a spatial extension (full width at half maximum)  $\Delta x$  and its Fourier representation  $a(k)$  in the wave number space  $k$ .  $a(k)$  are amplitudes of harmonic waves which build up  $\psi(x)$  when they are integrated (added up) over  $k$ . The distribution  $a(k)$  has again Gaussian shape

and  $\Delta p$  which obey the relation

$$\Delta x \cdot \Delta p \sim \hbar. \quad (3.20)$$

The measurement of a “sharp” position coordinate ( $\Delta x = 0$ ) would require an infinitely washed out information about the particle momentum ( $\Delta p \rightarrow \infty$ ), that is, no knowledge about its momentum anymore and vice versa. This important relation (3.20) was detected by Heisenberg [1] and is called the *uncertainty principle or relation*. If the Planck constant  $h$  or  $\hbar$  would be negligibly small as is the case in classical physics, then position and momentum of a particle could be measured simultaneously with infinite precision, as is assumed in Newtonian mechanics. The uncertainty principle is, thus, fundamental and inherently connected with the particle-wave duality in quantum mechanics and the un-deterministic behavior of atomic and subatomic particles. We will see that apart from position and momentum there are many other physical quantities which obey an uncertainty relation similar to (3.20) and can therefore not be measured simultaneously with infinite precision. Such quantities are called *complementary* or *incommensurable* and the principle behind it *principle of complementarity*. Quantities which can be measured simultaneously with any accuracy are called *commensurable*.

At this point it is worth mentioning that already in classical mechanics exactly those complementary quantities of quantum mechanics were marked as special canonical variables. We will learn a bit more about this topic in the next section.

The uncertainty principle, of course, must be valid in general, i.e. also in the macroscopic world. Consider, e.g., a bullet, which moves with supersonic velocity of  $v = 10^5$  cm/s and an uncertainty in speed of  $\Delta v = 10^{-2}$  cm/s ( $\Delta p = m \cdot 10^{-2}$  cm/s). Then, the uncertainty in the determination of its position amounts to  $\Delta x \approx (1/m) \cdot 10^{-25}$  g cm. For a bullet mass of only  $10^{-3}$  g a spatial uncertainty  $\Delta x$  of about  $10^{-22}$  cm, that is, about  $10^{-14}$  atomic radii results. This is far below any detection possibility. Even for small macroscopic bodies the uncertainty principle is without any relevancy, it becomes important only on the atomic scale.

From the position-momentum uncertainty (3.20), we can derive a further “secondary” uncertainty relation, that for energy and time. A particle with a position uncertainty  $\Delta x$  crosses during its movement a position  $x$  within a certain time interval  $\Delta t$ . Then,

$$\Delta t = \frac{\Delta x}{v} = \frac{m \Delta x}{p} \quad (3.21)$$

is the time, during which a wave packet with linear expansion  $\Delta x$  crosses the position  $x$ . Because of  $E = p^2/2m$  and  $\Delta E = p \Delta p/m$  (3.20) together with (3.21) yield

$$\Delta E \cdot \Delta t = \Delta x \Delta p \sim \hbar. \quad (3.22)$$

Thus, also for energy and time an uncertainty relation holds, which, however, has a different character than the position-momentum uncertainty principle. Position and momentum are observable quantities, while time is not “observable”, it plays the role

of a parameter in non-relativistic physics, along which we observe certain outcomes of experiments. Nevertheless, we want to emphasize that an energy measurement with an accuracy  $\Delta E$  requires a time interval of at least  $\Delta t \approx \hbar/\Delta E$ .

The energy-time uncertainty relation (3.22) is also of importance for decay processes where an excited state of an atom, a radioactive nucleus or an unstable elementary particle is de-excited by emission of a new particle. If the excited, unstable state has an average lifetime  $\tau$ , the emitted particle is in contact or interaction with the unstable object during this time  $\tau$ . The time uncertainty for the emission of the particle, thus, is  $\tau$ , too. The particle is emitted with an energy uncertainty

$$\Delta E \approx \hbar/\tau. \quad (3.23)$$

The energy width of the spectral emission line on the energy scale of the emitted particle, thus, yields information about the lifetime of the excited state from which the particle was emitted.

### 3.4 An Excursion into Classical Mechanics

The uncertainty relation (3.20) is an essential ingredient for quantum mechanics; for macroscopic bodies, even small ones, it loses importance because of the huge number of atoms being involved. In the limit of large numbers of atomic particles building up a macroscopic system the laws of quantum mechanics approach those of classical, deterministic Newtonian mechanics. Our familiar macroscopic environment is described correctly by classical mechanics. Engineers calculate bridges, cars and air planes according to its laws and deliver systems in which we survive safely.

Classical mechanics can, thus, be considered as an extrapolation of quantum mechanics into macroscopic dimensions. Accordingly we expect a very close similarity between quantum mechanical principles and the rules of classical mechanics. Otherwise, we could not understand the correspondence between both theories. This correspondence helps us to guess or understand the laws of quantum mechanics: we look at classical relationships and try to extend them, with reasonable additional assumptions, into quantum physics. This principle of similarity between classical mechanics and quantum mechanics is called *correspondence principle*.

The correspondence principle is already found in the uncertainty principle, where two variables, position and momentum, are incommensurable, that is, can not be measured simultaneously with infinite accuracy. Exactly these two variables appear as canonically conjugate variables in classical mechanics, in its Hamiltonian formulation. The so-called Hamilton formalism was invented by Hamilton (1805–1865) as a fully equivalent version of Newtonian mechanics, but easier to handle for more complex problems.

According to Newton the basic dynamic equation for the acceleration  $\ddot{x}$  of a mass point (mass  $m$ ) is

$$m\ddot{x} = K(x) = -\frac{dV}{dx}, \quad (3.24)$$

where the acting force  $K(x)$  results as the gradient of the potential  $V(x)$ . Hereby  $p = m\dot{x}$  is the momentum, which is conserved in collisions and, thus, plays an important role in solving mechanical problems. Hamilton found out that for complex systems the dynamic equation (3.24) can be solved much easier if one starts with the so-called *Hamilton function*, the *Hamiltonian*  $H$ , which is nothing else than the total energy of the mechanical system. For a simple mass point the kinetic energy is  $T = p^2/2m$  and its potential energy  $V(x)$ , thus the Hamiltonian follows as

$$H = T + V = \frac{p^2}{2m} + V(x). \quad (3.25)$$

Within the framework of this description, the solution of mechanical problems is performed by using the so-called *Hamilton equations*, which are written down in a wonderful symmetrical form as

$$\dot{x} = \frac{\partial H}{\partial p}, \quad (3.26a)$$

$$\dot{p} = -\frac{\partial H}{\partial x}. \quad (3.26b)$$

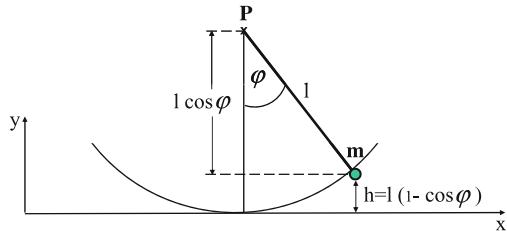
When we apply (3.26a) to (3.25) we obtain  $p = m\dot{x}$ , the relation between momentum and velocity. Equation (3.26b) applied to (3.25) yields

$$\dot{p} = m\ddot{x} = -\frac{\partial H}{\partial x} = -\frac{\partial V}{\partial x} = K(x), \quad (3.27)$$

that is, the dynamical equation of Newton. The variables  $x$  and  $p$  are assumed to be independent on each other. The Hamilton formalism expressed by (3.25), (3.26a) and (3.26b), thus, completely represents Newtonian mechanics.

The Hamilton formalism, however, can be extended to dynamical variables other than  $x$  and  $p$ . This might lead to mathematical simplifications if a moving body must obey scleronomic (rheonomic) constraints, that is, if certain forces built-in into the mechanic system exclude some degrees of freedom for the motion. Let us consider, as an example, the simple pendulum. The pendulum is an oscillating system composed of a weight (mass  $m$ ) and a string which is attached at the top end to a pivot  $P$  (Fig. 3.4). The string's length  $l$  is a constant which restricts the degrees of freedom for motion of the weight within the gravitation field ( $mg$  = gravity force) to oscillations on a circle with radius  $l$  around  $P$  (scleronomic constraint). Rather than describing the two-dimensional movement of the weight by means of the coordinates  $x$  and  $y$  it is advantageous to take the scleronomic constraint into account and to introduce the angle  $\varphi$  (Fig. 3.4) as a generalized position coordinate.  $s = l\varphi$  is the way of the mass  $m$  on the circle around  $P$ , that is, we can write down the pendulum's Hamilton function as

**Fig. 3.4** Ideal pendulum:  
The mass point  $m$  oscillates  
on a circle at distance  $l$  from  
the pivot  $P$



$$\begin{aligned} H &= T + V = \frac{1}{2}m\dot{s}^2 + mgl(1 - \cos \varphi) \\ &= \frac{1}{2}ml^2\dot{\varphi}^2 + mgl(1 - \cos \varphi). \end{aligned} \quad (3.28)$$

In the Hamilton formalism (3.25), (3.26a) and (3.26b) we, thus, substitute  $x$  by the generalized coordinate  $q = \varphi$ , that is,  $\dot{q} = \dot{\varphi}$  and obtain from (3.26b) and (3.28)

$$\dot{p} = -\frac{\partial H}{\partial \varphi} = -mgl \sin \varphi. \quad (3.29a)$$

In analogy to the conjugate coordinate couple  $(x, p = m\dot{x})$ , we conclude  $p \propto \dot{\varphi}$  and derive from (3.26a) and (3.28)

$$\dot{q} = \dot{\varphi} = \frac{\partial H}{\partial p} \propto \frac{\partial H}{\partial \dot{\varphi}} = ml^2\dot{\varphi}. \quad (3.29b)$$

Equation (3.29b) can be written as

$$\dot{q} = \frac{\partial H}{\partial (ml^2\dot{\varphi})} = \frac{\partial H}{\partial p}, \quad (3.29c)$$

that is,

$$p = ml^2\dot{\varphi}.$$

From (3.29a)–(3.29c), we obtain the dynamic equation for the pendulum as

$$\dot{p} = ml^2\ddot{\varphi} = -mgl \sin \varphi,$$

respectively,

$$\ddot{\varphi} + \frac{g}{l} \sin \varphi = 0. \quad (3.30)$$

For small elongations ( $\varphi \ll \pi$ ) the familiar pendulum equation

$$\ddot{\varphi} + \omega^2\varphi = 0, \quad (3.31)$$

follows with  $\omega^2 = g/l$  as oscillation frequency. Equation (3.30) describes sine-like angle oscillations  $\varphi(t)$  of the weight around its zero position  $\varphi = 0$ .

The simple example of the pendulum clearly demonstrates the importance of two independent canonically conjugated variables,  $q = \varphi$  and  $p = ml^2\dot{\varphi}$  for the pendulum. The derivative of the Hamilton function with respect to one of these yields the time derivative of the respective other conjugated variable (3.26a), (3.26b). Exactly these canonically conjugated variables of Hamilton mechanics show up as incommensurable variables in quantum mechanics, they obey the uncertainty principle. As is required in the uncertainty relation (3.20), the product of these variables always has units of an *action* (=energy  $\times$  time) as Planck's constant  $h = 2\pi\hbar$ , the quantum of action has. Check that the unit of  $ml^2\varphi\dot{\varphi}$  is just Joule-Second.

A further advantage of the Hamilton formalism of classical mechanics derives from the fact that it is easily extended to many-body problems by using a many-particle Hamiltonian  $H$  and the Hamilton equations:

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad \text{and} \quad \dot{q}_i = \frac{\partial H}{\partial p_i}, \quad (3.32)$$

$p_i$  and  $q_i$  are the generalized momentum and position coordinates of the particles numerated by the index  $i$ . For a system with  $n$  degrees of freedom (3.32) yields  $2n$  equations of 1st grade.

This correspondence between incommensurable quantities as position and momentum in quantum mechanics and the canonically conjugate variables in Hamilton mechanics will serve as a guiding line in the following to find further incommensurable quantities in quantum physics. Furthermore, we will use the Hamilton function, the Hamiltonian (3.25), to "invent" the fundamental dynamic equation of quantum physics, the Schrödinger equation.

### 3.5 Observables, Operators and Schrödinger Equation

Wave functions  $\psi(\mathbf{r}, t)$  describe as  $\psi^*\psi d^3r$  probabilities to find a particle at time  $t$  at a position  $\mathbf{r}$  within a volume element  $d^3r$ . The outcome of a position measurement is fully random in detail, only a certain probability for the position of the particle can theoretically be given in quantum physics. Thus, the result of such a position measurement on a particle can only be given in terms of an average position  $\langle r \rangle$  which is obtained for an ensemble of a huge number of particles being observed under exactly the same experimental conditions (see double slit experiment in Sect. 2.4.1).

For a formal description of the issue we, therefore, want to remind classical statistics. Here, the mean value  $\langle Y \rangle$  of a randomly distributed discrete quantity  $Y_i$  which appears with probability  $w_i$  is

$$\langle Y \rangle = \frac{\sum_i w_i Y_i}{\sum_i w_i}. \quad (3.33)$$

Mostly the probabilities  $w_i$  are normalized with  $\sum_i w_i = 1$ . In analogy, the mean value, descriptively called expectation value, of a position measurement, that is, the average outcome of many measurements on an ensemble is written as

$$\langle \mathbf{r} \rangle = \int d^3r |\psi|^2 \mathbf{r} = \int d^3r \psi^*(\mathbf{r}, t) \mathbf{r} \psi(\mathbf{r}, t). \quad (3.34)$$

We have used the normalization of the wave function in the sense of (3.7) (Sect. 3.1).

For the wave function of a spatially limited wave packet as for example, a Gaussian packet (Sect. 3.2), the average coordinate  $\langle x \rangle$  indicates the maximum of the spatial  $\psi$  distribution.

Its time derivative  $\partial \langle x \rangle / \partial t$  is the group velocity  $v = \partial \omega / \partial k$  of the wave packet which is identified with the particle velocity.

Because of the uncertainty relation (Sect. 3.3) also the momentum  $\mathbf{p}$  of the wave packet can only be defined as an expectation value  $\langle \mathbf{p} \rangle$  which is calculated as a mean value from measurement results on an ensemble of many particles. Each single momentum measurement on a particular particle yields statistically varying numbers for its momentum. Thus, in analogy with (3.34) the average momentum of all these measurement results on the ensemble is

$$\langle \mathbf{p} \rangle = \int d^3r \psi^*(\mathbf{r}, t) \mathbf{p} \psi(\mathbf{r}, t). \quad (3.35)$$

Particle waves exhibit dispersion, that is, the particle energy is related to its momentum via  $E(k) = \hbar\omega(k) = \hbar^2 k^2 / 2m = p^2 / 2m$ . In case of a statistical momentum distribution, therefore, the particle energy must also be defined as an average value over the ensemble:

$$\langle E \rangle = \langle \hbar\omega \rangle = \int d^3r \psi^*(\mathbf{r}, t) \hbar\omega \psi(\mathbf{r}, t). \quad (3.36)$$

For plane waves  $\exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$  and all kinds of wave functions as for example, wave packets, which are built up by plane waves by means of a Fourier series, the two last terms in the integrals (3.35) and (3.36) can also be obtained by differentiating the wave function:

$$\mathbf{p}\psi(\mathbf{r}, t) = \frac{\hbar}{i} \nabla \psi(\mathbf{r}, t), \quad (3.37)$$

$$E\psi = \hbar\omega\psi(\mathbf{r}, t) = -\frac{\hbar}{i} \frac{\partial}{\partial t} \psi(\mathbf{r}, t). \quad (3.38)$$

The average (expectation) values (3.35) and (3.36) might, therefore, also be written as

$$\langle \mathbf{p} \rangle = \int d^3r \psi^* \left( \frac{\hbar}{i} \nabla \right) \psi, \quad (3.39)$$

$$\langle E \rangle = \int d^3r \psi^* \left( i\hbar \frac{\partial}{\partial t} \right) \psi. \quad (3.40)$$

In these expressions, of course, the sequence  $\psi^*$ —Differentiation— $\psi$  is relevant. A changed order of wave functions and differentiation operation would yield a wrong expression for the expectation values.

Since in quantum physics a measurement of a certain quantity—mostly called *observable*—can only yield randomly distributed numbers (for a large ensemble of particles a well defined average value is obtained), we better depart from the familiar classical description of a physical quantity, an observable, in terms of sharp, well defined numbers for the outcome of an experiment. This is fully in accordance with the uncertainty principle (Sect. 3.3), which prohibits sharp, well defined values simultaneously for momentum and space coordinate and other incommensurable observables. Instead, we attribute, by looking at (3.37) and (3.38), so-called *operators*  $\hat{\mathbf{p}}$  and  $\hat{H}$  to the observables momentum and energy, respectively. In this particular case of momentum and energy the corresponding operators are defined by a differentiation of the wave function as

$$\hat{\mathbf{p}} = \frac{\hbar}{i} \nabla, \quad (3.41)$$

$$\hat{H} = i\hbar \frac{\partial}{\partial t}. \quad (3.42)$$

Other observables might be described by different types of operators, different operations on the wave function as e.g. integration or squaring etc. Defining the momentum by means of an operator rather than by a simple number also avoids a problem which we have stealthily suppressed in the expression (3.35) for the momentum expectation value  $\langle \mathbf{p} \rangle$ . Since  $\mathbf{p}$  and  $\mathbf{r}$  can not be measured simultaneously with any accuracy because of the uncertainty principle, the expectation values (3.34)–(3.36) can not be calculated, because measured numbers for the observables  $\mathbf{p}$  and  $\mathbf{r}$  do not exist simultaneously.

Operators, that is, rules for operating on a wave function, always have to be positioned in front of the wave function. As in (3.41) and (3.42) we always denote operators in this book by means of a roof symbol  $\hat{\phantom{x}}$  above the letter, in order to distinguish them from simple numbers or functions. The energy operator  $\hat{H}$  is commonly denoted as *Hamilton operator* or *Hamiltonian*, in analogy to the Hamilton function (Sect. 3.4). The corresponding average or expectation values for momentum and energy, thus, are written as

$$\langle \mathbf{p} \rangle = \int d^3r \psi^* \hat{\mathbf{p}} \psi, \quad (3.43)$$

$$\langle E \rangle = \int d^3r \psi^* \hat{H} \psi. \quad (3.44)$$

Energy or momentum measurements on a single particle, however, rather than on an ensemble of many particles give single numbers for  $E$  and  $\mathbf{p}$  as result, but randomly varying from measurement to measurement. All such numbers taken together (ensemble), of course are distributed according to (3.43) and (3.44). The numbers  $\mathbf{p}$  or  $E$  resulting from measurements on a single particle obey (3.37) and (3.38) or expressed by the corresponding operators

$$\hat{\mathbf{p}}\psi(\mathbf{r}, t) = \mathbf{p}\psi(\mathbf{r}, t), \quad (3.45)$$

$$\hat{H}\psi(\mathbf{r}, t) = E\psi(\mathbf{r}, t). \quad (3.46)$$

In analogy, we can ascribe a position operator  $\hat{\mathbf{x}}$  (or  $\hat{\mathbf{r}}$ ) to the position observable. This operator does nothing else but multiplying the wave function with the space coordinate  $x$  or  $\mathbf{r}$ :

$$\hat{\mathbf{x}}\psi(\mathbf{r}, t) = \mathbf{r}\psi(\mathbf{r}, t). \quad (3.47)$$

All functions of  $\mathbf{r}$  or  $x$ , for example, the potential  $V(\mathbf{r})$ , in which a mass is moving, are transformed into operators, which also mean just multiplication of that function with the wave function. The potential operator  $\hat{V}$  does nothing else but multiplying the wave function  $\psi$  with the potential  $V(\mathbf{r})$ .

The relations (3.45)–(3.47) are called eigenvalue equations for the operators  $\hat{\mathbf{p}}$ ,  $\hat{H}$  and  $\hat{\mathbf{x}}$ . From linear algebra and the mathematics of matrices eigenvalue equations are well known. In matrix algebra they have the familiar form (Sect. 4.3.1)

$$\underline{A} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (3.48)$$

A matrix  $\underline{A}$ , two-dimensional (2D) in (3.48), multiplied with a vector (the matrix acts on the vector, comparable to an operator acting on a function) yields the same vector scaled by a number, in some cases a complex number. Equation (3.48) can be fulfilled only under certain conditions. The number  $\lambda$  in (3.48) is called eigenvalue of the matrix, it represents the action of the whole matrix on the vector. For this particular case of a 2D matrix two eigenvalues  $\lambda_1$  and  $\lambda_2$  exist. For higher dimensional problems, as many eigenvalues exist as the number of dimensions is. In analogy, we say that the operators  $\hat{\mathbf{p}}$ ,  $\hat{H}$  and  $\hat{\mathbf{x}}$  have momentum, energy and position eigenvalues which are the measured numbers resulting from the corresponding experiments on a single particle. A measurement on a single particle always yields one of the whole variety of possible eigenvalues of the respective observable, randomly varying from one to the other measurement. Many measurements on a whole ensemble of particles yield the average or expectation values  $\langle \mathbf{p} \rangle$ ,  $\langle E \rangle$ ,  $\langle \mathbf{r} \rangle$ .

As a general rule, we want to keep in mind: Observables  $\Omega$  are described in quantum mechanics by operators  $\hat{\Omega}$  which act on a wave function  $\varphi$ . The operator's eigenvalues  $\omega$  determined from the eigenvalue equation

$$\hat{\Omega}\varphi = \omega\varphi, \quad (3.49)$$

are the possible measurement results obtained as real numbers from the measurement of the observable  $\Omega$ . The outcome of a single measurement is random; only measurements on a large ensemble determine a mean value  $\langle \omega \rangle$ . We will see that not only the operator itself determines the eigenvalues of a problem but also boundary conditions. From the whole variety of possible eigenvalues for a free particle, for example, the boundary condition, particle freely moving or confined to a box, selects the eigenvalues appropriate to the problem. One might imagine the action of an operator in the eigenvalue equation like a frequency or wavelength filter in electronics. The application of the operator  $(\hbar/i)\partial/\partial x = \hat{p}$ , for example, on the wave function of a wave packet filters out a particular plane wave with a special  $k$  vector from the whole variety of vectors constituting the packet.

It is the fundamentally statistical behavior of atomic and subatomic particles which requests that we leave the picture of well defined particle trajectories in classical mechanics and introduce operators for the description of dynamical observables in quantum mechanics.

The straightforward next step, of course, is the replacement of the classical basic dynamical equation, Newton's or Hamilton's equations by a quantum mechanical analogon. Having in mind the correspondence principle (Sect. 3.4) we start, just as Schrödinger did, with the classical Hamiltonian (3.25) and replace the expressions for the kinetic and potential energy by the corresponding quantum mechanical operators. Thus, in  $p^2/2m$  the squared momentum is replaced by the operator  $\hat{p}^2$  and the potential  $V(x)$  by its operator (only multiplication with  $V$ ). Using (3.41), we thus obtain the Hamilton operator or Hamiltonian as

$$\hat{H} = \frac{\hat{p}^2}{2m} + \hat{V}(\mathbf{r}) = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}). \quad (3.50)$$

On the other hand, already in (3.42) we have found an expression for the Hamilton operator. Both forms operating on a general wave function must yield the same result. From the identity of (3.50) and (3.42) we, thus, obtain the fundamental dynamical equation for one single particle in quantum physics. This differential equation named after its inventor *Schrödinger equation* [2] allows the calculation of the wave function  $\psi(\mathbf{r}, t)$  for a particular problem, if also boundary conditions are given. In three dimensions, with  $\nabla^2 = \Delta$  (squared nabla operator = delta operator), it is written as:

$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, t) = \left( -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right) \psi(\mathbf{r}, t).$

(3.51)

In one dimension, the Schrödinger equation is

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \psi(x, t). \quad (3.52)$$

As in all fundamental physical theories, the basic equations can not be derived deductively, they are guessed or invented, but clever, with a great amount of previous knowledge. All facts known from experiment have to be considered and should be described correctly by the equation. In successful cases, these fundamental equations allow the prediction of a wealth of other new phenomena not known so far. We have tried to make this way of guessing the Schrödinger equation a little bit plausible, maybe not the way Schrödinger did it. It should be anticipated, that the Schrödinger equation describes all phenomena for atomic and subatomic particles in the non-relativistic limit extremely well. Relativistic extensions (Klein–Gordon and Dirac equations) not treated in this book do the same for relativistic particles (Sects. 5.6.4 and 8.3).

Having such a differential equation as (3.51) at hand a wave function  $\psi(\mathbf{r}, t)$  for a particle can be calculated under certain boundary conditions imposed by the particular problem. By means of (3.34), (3.39) and (3.40) expectation values for interesting variables are calculated which might be compared with the corresponding experimental results obtained on an ensemble of many particles. So far experimental observations within the non-relativistic validity range always confirmed the Schrödinger equation.

We analyze (3.51) a little bit further by considering that the Hamiltonian (3.50) does not explicitly depend on time. In this case, the Schrödinger equation (3.51)

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, t) = \hat{H} \psi(\mathbf{r}, t) \quad (3.53)$$

can be separated with respect to space and time. For the wave function, we make the separation ansatz

$$\psi(\mathbf{r}, t) = f(t)\varphi(\mathbf{r}). \quad (3.54)$$

Equation (3.54) inserted in (3.53) yields

$$\frac{1}{f(t)} i\hbar \frac{\partial}{\partial t} f(t) = \frac{1}{\varphi(\mathbf{r})} \hat{H} \varphi(\mathbf{r}) = E. \quad (3.55)$$

In (3.55), the left side depends only on the time  $t$  while the right side only contains the position coordinate  $\mathbf{r}$ . Both sides of the equation, thus, must have an identical constant value which we call  $E$ . From the units of the right side we infer that  $E$  must be the total energy of the system.  $E$  is a constant of motion in this case.

We conclude the following time dependence of the wave function

$$i\hbar \frac{\partial}{\partial t} f(t) = Ef(t), \quad (3.56a)$$

that is,

$$f(t) = e^{-iEt/\hbar}. \quad (3.56b)$$

The spatial dependence of the wave function, then, must obey the equation

$$\hat{H}\varphi(\mathbf{r}) = E\varphi(\mathbf{r}). \quad (3.57)$$

This is the so-called *time-independent Schrödinger equation*, a type of eigenvalue equation as in (3.45)–(3.48). For a physical problem with particular boundary conditions, the solution of (3.57) yields the energy eigenvalues, that is, the possible numerical results of an energy measurement on the system. These energy values are constants of motion as in classical mechanics, when the Hamiltonian does not explicitly depend on time. In this case, the wave function  $\psi(\mathbf{r}, t)$  as a solution of (3.53) is represented, because of (3.54) and (3.56b), as

$$\psi(\mathbf{r}, t) = \varphi(\mathbf{r})e^{-iEt/\hbar}. \quad (3.58)$$

Even though the possible energies  $E$  of the system are constant in time, the wave function  $\psi$  exhibits a time dependence with  $iEt/\hbar$  as the phase of the wave function. Nevertheless, the probability density  $|\psi|^2$  and the expectation values as (3.43) etc., that is, observable quantities, do not depend on time. A wave function of the type (3.58) always describes a stationary physical system, the Hamilton operator of which does not depend on time.

## 3.6 Simple Solutions of the Schrödinger Equation

The Schrödinger equation as the fundamental dynamic equation of quantum mechanics is a linear differential equation, of 1st order in time and of 2nd order in the position coordinate. The solution of this equation requires boundary conditions, which fix the wave function  $\psi(\mathbf{r}, t)$  on a spatial border line given by the particular physical problem. This boundary or border line might be located in infinity. In that case,  $\psi(\mathbf{r}, t)$  must be infinitely small in infinity, otherwise the normalization condition (3.7) could not be fulfilled. Only those solutions of the Schrödinger equation are relevant in physics, which can be normalized according to (3.7). This is required by the probability interpretation of the wave function (Sect. 3.1).

Furthermore, for stationary problems with a time independent Hamiltonian  $\hat{H}$  (3.58) we only have to solve the time independent eigenvalue equation (3.57) rather than the complete Schrödinger equation (3.51). The complete wave function with time dependence is simply obtained by multiplying the time independent eigensolu-

tion of (3.57) with the factor  $\exp(-iEt/\hbar)$ , where  $E$  is the energy eigenvalue obtained from the solution of (3.57).

In the following section we want to consider some simple examples of solutions of the Schrödinger equation, which are nevertheless of considerable importance for application.

### 3.6.1 “Locked-Up” Electrons: Confined Quantum States

In a metal free electrons are “locked-up”, they are confined in their motion to the volume of the solid. They can only leave the metal by overcoming the potential barrier of the work function [Sect. 2.1, (2.1a)]. In modern nanostructures, electrons are confined to volume dimensions in the range of several ten nanometers (nm). Also the valence electron in the hydrogen atom must be considered as “locked-up” by the Coulomb potential  $-e^2/4\pi\epsilon_0 r$  of the positive nucleus, the proton. Here, the electron has a radius for free motion of about 0.1 nm. The most simple model for the description of electron confinement is the so-called potential box with square well potential. A cubic box with side length  $L$  is assumed as confinement volume. The potential  $V(\mathbf{r})$  in the Schrödinger equation is constant within the box. Because of freedom in the choice of the zero level of the energy scale we define the potential in the box as  $V(\mathbf{r}) = V_0 = 0$ . On the surfaces of the box, an infinitely high potential barrier is assumed which prohibits the electrons inside from leaving the box (ideal confinement), that is,

$$V(\mathbf{r}) = \begin{cases} V_0 = 0 & \text{for } 0 < x, y, z < L \\ \rightarrow \infty & \text{at } x, y, z = 0; x, y, z = L. \end{cases} \quad (3.59)$$

We will see that electron confinement in two-dimensional (2D), one-dimensional (1D) and even zero-dimensional (0D) potentials is of considerable interest for real nanostructures. Depending on their dimensions they are called quantum films (2D), quantum wires (1D) and quantum dots or boxes (0D). Since the potential does not depend on time, we solve the tim-independent Schrödinger equation for the cubic box:

$$-\frac{\hbar^2}{2m} \Delta \varphi(\mathbf{r}) = E\varphi(\mathbf{r}). \quad (3.60)$$

Since an electron can not leave the box, the wave function  $\varphi$  must vanish on the surfaces of the box, that is, the boundary conditions are

$$\varphi(x = 0, y = 0, z = 0) = 0, \quad (3.61a)$$

$$\varphi(x = L, y = L, z = L) = 0. \quad (3.61b)$$

The differential equation (3.60) is solved by the plane wave  $\varphi(\mathbf{r}) = A \exp(i\mathbf{k} \cdot \mathbf{r}) = A(\cos \mathbf{k} \cdot \mathbf{r} + i \sin \mathbf{k} \cdot \mathbf{r})$ , that is, also by the cosine and sine parts separately. Since

the boundary condition (3.61a) can not be fulfilled by  $\cos \mathbf{k} \cdot \mathbf{r}$ , the solution of the problem is

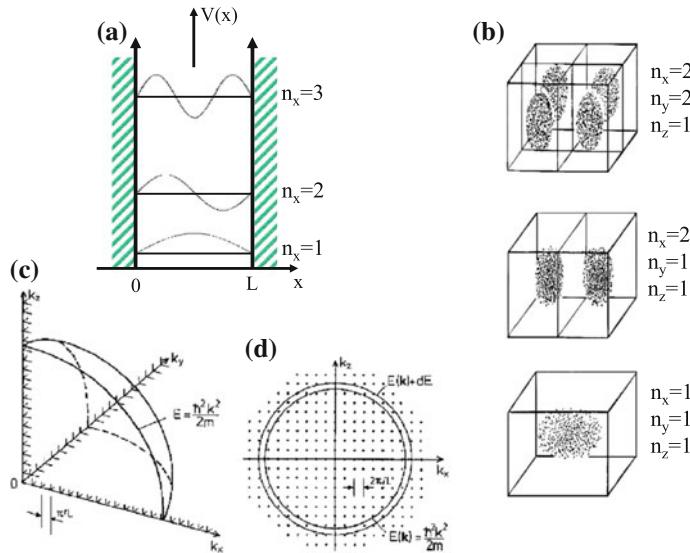
$$\varphi(\mathbf{r}) = C(\sin k_x x)(\sin k_y y)(\sin k_z z). \quad (3.62)$$

The second boundary condition (3.61b) requires quantized wave numbers  $\mathbf{k}$ , that is, only discrete values for  $k_x, k_y, k_z$  because of

$$\begin{aligned} k_x &= n_x \pi / L, \quad n_x = 1, 2, 3 \dots, \\ k_y &= n_y \pi / L, \quad n_y = 1, 2, 3 \dots, \\ k_z &= n_z \pi / L, \quad n_z = 1, 2, 3 \dots. \end{aligned} \quad (3.63)$$

At the zero point  $(n_x, n_y, n_z) = \mathbf{0}$  the wave function  $\varphi$  vanishes and can therefore not be normalized. This solution has no physical meaning and must be skipped. Furthermore, negative numbers  $n_x, n_y$  and  $n_z$  only change the sign of the wave function (3.62) and do not produce new, linearly independent solutions. Thus, the variety of meaningful solutions is simply represented as a point lattice within the positive octant of reciprocal  $\mathbf{k}$  space (Fig. 3.5c). The points representing an electronic state  $\varphi(\mathbf{r})$  (3.62) have a linear separation of  $\pi/L$ . In  $\mathbf{k}$  space a state occupies a volume of  $(\pi/L)^3$ . In Fig. 3.5a, the three lowest wave functions for a 1D quantum wire are plotted. These solutions of the Schrödinger equation are standing waves with wavelengths matching the length of the wire: Multiples of half the wavelength must equal the wire length for each eigensolution. For a 3D quantum box these properties are required for the three dimensions (3 axis in space) and we can qualitatively represent the three lowest eigenstates by a “cloud” picture as in Fig. 3.5b. Here, the point density is a qualitative measure for the absolute value of the probability density  $|\varphi(\mathbf{r})|^2$  at a certain position in space. Similar pictures are obtained for the 2D eigenstates in a quantum film. In this case the 2D cloud picture of  $|\varphi(\mathbf{r})|^2$  resembles Chladni figures of a resonating plate. On a plate or membrane covered with powder 2D oscillations are excited by a violin bow and the powder film orders into a characteristic node pattern (Chladni figure) which is an image of the amplitude of the 2D standing wave. The wave functions of the 1D quantum wire in Fig. 3.5a can similarly compared with standing waves on a vibrating violin string tied at both ends. Waves of confined electrons behave as classical standing waves which might be described as the superposition of forth and back propagating waves. As in the case of the oscillating violin string, only discrete vibration modes are possible (Fig. 3.5a, b). Correspondingly the energy eigenvalues as solutions of (3.60) are discrete, that is, quantized. They follow from (3.60) as  $E = \hbar^2 k^2 / 2m$ . With the  $\mathbf{k}$  quantization (3.63) resulting from the boundary conditions, one obtains

$$\begin{aligned} E_{n_x, n_y, n_z} &= \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2}{2m} \left( \frac{\pi}{L} \right)^2 (n_x^2 + n_y^2 + n_z^2), \\ n_x, n_y, n_z &= 1, 2, 3 \dots \end{aligned} \quad (3.64)$$



**Fig. 3.5** a–d Free electrons in a potential box. **a** Ground state ( $n_x = 1$ ) and lowest excited states ( $n_x = 2, 3$ ) of an electron in a 1D square well potential with length  $L$  and infinitely high potential walls. **b** Spatial cloud-like plot of the probability densities  $|\psi(x, y, z)|^2$  of the three low index 3D wave functions of an electron in a cubic potential box. *High dot density* indicates high probability density. **c** Representation of the single electron states by means of a point lattice in the reciprocal space of electron wave vectors  $\mathbf{k}$  (2D cut along the  $k_x$ ,  $k_z$  plane). Each point belongs to a quantum state. For fixed boundary conditions the possible  $\mathbf{k}$  values are confined to the positive octant of  $\mathbf{k}$  space and the points have a linear distance  $\pi/L$ . Because of the two spin orientations each point represents two electronic states with opposite spin. **d** For periodic boundary conditions the points cover the whole reciprocal space, but their linear separation is  $2\pi/L$ . In **c** and **d** spheres of constant energy  $E(\mathbf{k})$  and  $E(\mathbf{k}) + dE$  are plotted

The possible energy eigenvalues of a confined (locked-up) electron always form a discrete spectrum of energy levels. This is not only true for the simple confining potential of a rectangular box but rather for all kinds of potential wells including the (screened) Coulomb potentials of atomic nuclei. Confined quantum states have a discrete spectrum of energy eigenvalues. This inevitably explains the sharp discrete energy levels of electrons in an atom and thus the resulting sharp spectral lines upon light emission.

Let us consider the energetic distance  $\Delta E$  between the energy levels of an electron confined in a box. Because of  $E = \hbar^2 k^2 / 2m$ , we conclude

$$\Delta E \sim \frac{\hbar^2}{m} k \Delta k. \quad (3.65)$$

With increasing wave number  $k$  the interval  $\Delta E$  grows proportional to  $k$  (Fig. 3.5a) and  $\Delta k$  changes in steps of  $\pi/L$  (3.63), (3.64).

To estimate the energetic distance between the two lowest energy levels from (3.64), we assume  $n_y = n_z = 0$  and  $n_x = 1$  and 2, respectively. This yields

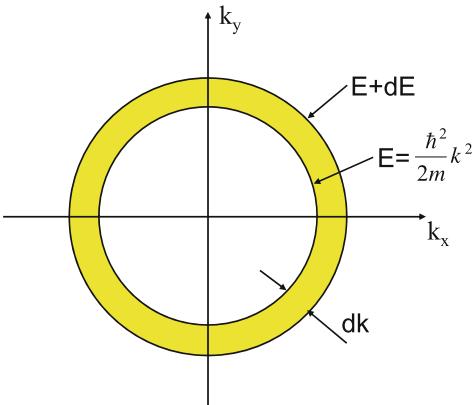
$$\Delta E = E_{2,0,0} - E_{1,0,0} = \frac{3\hbar^2}{2m} \left(\frac{\pi}{L}\right)^2. \quad (3.66)$$

The spatial extension  $L$  of the box, that is, of the confining potential, thus determines the energetic distance of the quantized energy levels.

- For a macroscopic box, for example, a metal cube with 1 cm edge length  $L$ ,  $\Delta E$  (3.66) is in the order of  $10^{-18}$  eV. This is far below any detection limit by which an energy measurement can resolve the quantized character of the energy. For macroscopic bodies classical behavior is found, the energy levels are distributed quasi-continuously. In this case, we could also solve the Schrödinger equation (3.60) for infinite space [or in (3.63)  $L \rightarrow \infty$ ] and obtain a continuous spectrum of  $\mathbf{k}$  vectors rather than (3.63).
- For mesoscopic structures, that is, potential wells with spatial extension in the range of 10 nm the energy levels are spaced by amounts of the order of 0.25 meV. Such mesoscopic structures can be fabricated by modern structuring techniques as for example, electron beam lithography or prepared by self organized epitaxial growth (Appendix B). At low temperatures spectroscopic measurements can resolve the quantum character of the level distribution. At a temperature of 1 K one might assume a thermal “smearing” of the energy levels by about  $kT_{1\text{ K}} \approx 0.13$  meV (see Sect. 5.6.3).
- In nanoscopic structures, for example, big molecules, we have dimensions with characteristic lengths of 1 nm. Then, the energetic distance between the levels is in the order of some electron volt (eV). Even at room temperature ( $kT_{300\text{ K}} \approx 0.04$  eV) spectroscopic measurements can resolve the quantization of the energy. Spectra of molecules exhibit sharp optical absorption lines.
- In atomic binding potentials with spatial extensions in the 0.1 nm range the energy levels of an electron are typically spread by energies  $\Delta E$  (3.66) of the order of 100 eV. These are typical binding energies of electrons in atoms which are known from X-ray spectroscopy.

For macroscopic bodies with more than  $10^{22}$  atoms/cm<sup>3</sup> surface effects are usually not relevant, since a surface typically contains only  $10^{15}$  atoms/cm<sup>2</sup>. The so-called *fixed boundary conditions* with vanishing wave function on the surface of the body (3.61a), (3.61b) can, therefore be relaxed in order to gain a more flexible mathematical description. For fixed boundary conditions the possible wave numbers of electronic states are limited to the positive octant of reciprocal  $\mathbf{k}$  space, an uncomfortable restriction for the mathematical treatment of problems. Because of symmetry reasons mathematics becomes simpler if all  $\mathbf{k}$  vectors of the whole reciprocal space would be allowed for the possible states of an electron in the potential box. This can be achieved by neglecting surface effects for macroscopic bodies. So-called *periodic boundary conditions* are introduced, where it is only required

**Fig. 3.6** Surfaces of constant energy (2D projection of spheres *circles*) for an electron in the potential box. The number of quantum state points (see Fig. 3.5) between the two spherical shells  $E$  and  $E + dE$  determines the state density in  $\mathbf{k}$  space



that  $\varphi(x = 0, y = 0, z = 0) = \varphi(x = L, y = L, z = L)$  rather than a vanishing wave function  $\varphi$  on the boundaries. These boundary conditions seem reasonable for a macroscopic solid which might be imagined as a piece of matter within a large closed ring of that material. One revolution along this ring leads directly to periodic boundary conditions. For periodic boundary conditions, instead of (3.61a), (3.61b), the wave function  $\varphi \propto \exp(i\mathbf{k} \cdot \mathbf{r})$  must obey less stringent requirements:

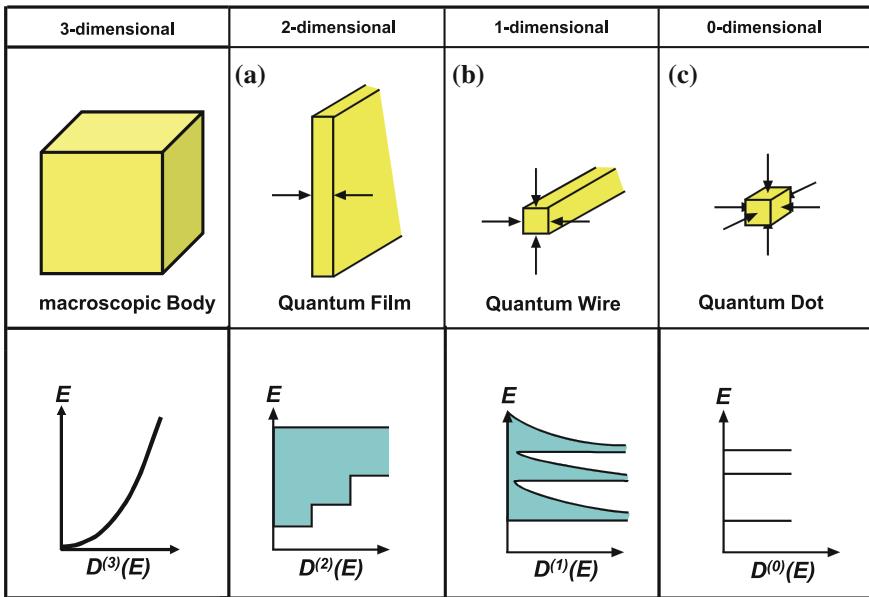
$$e^{ik_x L} e^{ik_y L} e^{ik_z L} = e^0 = 1. \quad (3.67)$$

From (3.67), the following quantization of wave numbers  $\mathbf{k}$  results:

$$\begin{aligned} k_x &= n_x \left( \frac{2\pi}{L} \right), \quad n_x = 0, \pm 1, \pm 2, \dots, \\ k_y &= n_y \left( \frac{2\pi}{L} \right), \quad n_y = 0, \pm 1, \pm 2, \dots, \\ k_z &= n_z \left( \frac{2\pi}{L} \right), \quad n_z = 0, \pm 1, \pm 2, \dots. \end{aligned} \quad (3.68)$$

In contrast to fixed boundary conditions, here,  $n_x = n_y = n_z = 0$  yields  $e^0 = 1$  and, thus a physically meaningful wave function which can be normalized. Furthermore, negative  $k$  values produce new wavefunctions linearly independent on those with positive  $k$ . For periodic boundary conditions (3.68) the points attributed to electronic states, therefore, fill the whole reciprocal  $\mathbf{k}$  space (Fig. 3.5d). Their linear distance, however, has increased to  $2\pi/L$ , double the value as that for fixed boundary conditions. The volume of an electronic state in  $k$  space is  $(2\pi/L)^3$ , eight times that for fixed boundary conditions. The points lie less densely in  $\mathbf{k}$  space as compared with fixed conditions (3.63).

For macroscopic solids, where surface effects might be neglected in good approximation, both types of boundary conditions yield essentially identical results for



**Fig. 3.7** Electronic state densities  $D^{(0)}(E), \dots, D^{(3)}(E)$  for electrons in quantum boxes of different dimensions (0–3)

macroscopic quantities. This is easily seen from a consideration of the so-called *density of states* per volume  $D^{(3)}(E)$ . This quantity indicates how many electronic states (energy levels) are found at an energy  $E$  within an energy interval  $dE$  around  $E$ ; it enters most macroscopic quantities as we will see soon. For electrons in a potential box with  $E = \hbar^2 k^2 / 2m$  the iso-energetic surfaces in reciprocal space are spheres (Fig. 3.6). The energy element  $dE$  is the volume between the two spheres of energy  $E$  and  $E + dE$ , it amounts to  $4\pi k^2 dk$  (Fig. 3.6). For periodic boundary conditions, each state occupies the volume  $(2\pi/L)^3$ , that is, the density of states per volume in real space  $L^3$  follows as:

$$D^{(3)}(E) dE = \frac{1}{(2\pi)^3} 4\pi k^2 dk. \quad (3.69)$$

Because of  $\hbar k = \sqrt{2mE}$  one can also write

$$D^{(3)}(E) dE = \frac{1}{2\pi^2} \frac{m}{\hbar^3} \sqrt{2mE} dE. \quad (3.70)$$

The same expression is obtained also for fixed boundary conditions (3.63). In that case only one eighth part of energy shell  $(E, E + dE)$  contributes, but each state occupies only one eighth of the volume as for periodic boundary conditions.

While the electronic density of states  $D^{(3)}(E)$  of a 3D potential box increases with the square root of the electron energy  $E$  [(3.69) and Fig. 3.7a], the corresponding problem in lower dimensions yields differing functional dependencies (Fig. 3.7). For a 2D electron gas in a quantum film as realized in very thin metal layers or some advanced modern semiconductor devices (heterostructure FETs, HEMTs, Appendix A) also the reciprocal  $\mathbf{k}$  space is two-dimensional. The  $\mathbf{k}$  volume of states between the energies  $E$  and  $E + dE$ , then, is a circular ring with an area  $2\pi k dk$  (Fig. 3.6). For periodic boundary conditions an electronic state occupies an area of  $(2\pi/L)^2$ . Because of  $dE = \hbar^2 k dk/m$  we obtain the following expression for the 2D density of states per volume  $L^2$  in real space:

$$D^{(2)}(E) dE = \frac{2\pi k dk}{(2\pi)^2} = \frac{m}{2\pi\hbar^2} dE. \quad (3.71)$$

The 2D density does not depend on the electron energy, it is a constant  $m/2\pi\hbar$ . For a 1D quantum well, a quantum wire, we ask, how many states with a linear “volume” (extension)  $2\pi/L$  fit into a reciprocal length element  $dk$ . This leads to the 1D density of states

$$D^{(1)}(E) dE = \frac{1}{2\pi} dk = \frac{m}{2\pi\hbar} \frac{1}{\sqrt{2mE}} dE, \quad (3.72)$$

$D^{(1)}(E)$  has a singular pole, the electronic states are not spread over a broad energy range but rather concentrated in a sharp spectral structure (Fig. 3.7). This concentration is even more pronounced in a 0D quantum box or dot, where the density of states consists of single sharp energy levels (Fig. 3.7).

Also in 2D or 1D quantum wells, quantum films or quantum wires, electronic states are described by wave functions  $\psi(\mathbf{r})$  which depend on all three coordinates  $x, y, z$ . In a 2D quantum film the electron is confined only along one direction  $z$  within a mesoscopic or nanoscopic length  $l_z$  (5–200 nm). This dimension is not macroscopic and fixed boundary conditions (3.63) must be applied along the  $z$  direction. We, thus, obtain sine-like wave functions (Fig. 3.5a) within the interval  $l_z$ . In  $x$  and  $y$  direction, however, a quantum film is extended over macroscopic dimensions. Edge effects can be neglected and periodic boundary conditions (3.68) are applied in the  $xy$  plane. Electrons propagate freely within the  $xy$  plane which is described by plane waves appropriate to periodic boundary conditions. For a 2D quantum film the electron wave function, thus, follows as

$$\psi(\mathbf{r}) = C e^{ik_x x} e^{ik_y y} \sin k_z z, \quad (3.73)$$

$k_x$  and  $k_y$  obey the conditions (3.68) with macroscopically large  $L$  values, that is, quasi-continuous wave numbers  $k_x, k_y$ . For  $k_z$  the quantization (3.63) is required, that is,  $k_z = n_z \pi / l_z$ . Because of (3.64), the energy levels of an electron in the quantum film are obtained as

$$E_{k_{||}, n_z} = \frac{\hbar^2 \pi^2}{2ml_z^2} n_z^2 + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m}. \quad (3.74)$$

The first term describes the discrete quantized levels arising from confinement in  $z$  direction. The series of levels is numbered by the numbers  $n_z = 1, 2, 3, \dots$  (Fig. 3.5a). The second term is the kinetic energy of the free motion within the film plane. Because of macroscopic dimensions of the film the wave numbers  $k_x$  and  $k_y$  are quasi-continuous and we have the characteristic parabolic dependence of energy on wave number for free motion. The energy levels (3.74) of an electron in a quantum film, thus, form a series of parabolae on the energy scale, which are numbered by  $n_z$ , the quantum number of  $z$ -quantization. The different parabolae are also called *sub-bands*. Each sub-band characterized by its index  $n_z$  corresponds to free motion of the electron in the film but with a different spatial structure of the wave function in  $z$  direction (different number of nodes, Fig. 3.5a). Each 2D sub-band belongs to modes of the 2D quantum well and, thus, exhibits a density of states  $D^{(2)}(E)$  which is a constant as function of energy (3.71). The density of states of an electron in a quantum film is the superposition of these constant contributions of the different sub-bands, it has the shape of a staircase (Fig. 3.7).

For a 1D quantum wire, nanoscopic dimensions are given for the  $x$  and  $y$  directions, while in one direction ( $z$ ) the wire is macroscopic. In this  $z$  direction the electron can freely propagate. Its kinetic energy is parabolic in  $k_z$ . In  $x$  and  $y$  direction, the electron is confined and fixed boundary conditions have to be applied. In analogy to the 2D case (3.74) we, thus, obtain a series of parabolic (in  $k_z$ ) sub-bands which are numbered by quantum numbers  $n_x$  and  $n_y$  corresponding to 2D confinement. Each 1D sub-band has a peaked density of states according to (3.72). The density of states of an electron in a quantum wire, thus, consists of a series of peaks on the energy scale (Fig. 3.7).

### 3.6.2 Particle Currents

The Schrödinger equation with a time independent potential describes phenomena which do not depend on time, that is, the stationary behavior of a system. Also stationary particle currents including an electrical current which does not change in time belong into this category. For particle currents, there is a *continuity relation* in classical physics. In the stationary case, conservation of particle number requires that the change of the number of particles within a certain volume is reflected by the total number of particles which flow into and out of that volume. This so-called continuity relation is written as

$$\frac{\partial \rho}{\partial t} + \text{div } \mathbf{j} = 0. \quad (3.75)$$

Hereby  $\rho(\mathbf{r}, t)$  is the density of particles at  $\mathbf{r}$  and time  $t$ . It changes in time due to the current density  $\mathbf{j}$  of in and out flowing particles. In quantum mechanics, the particle density is given by the number  $N$  of particles multiplied with the probability density  $\psi^* \psi$  to find a particle at the particular position.  $N$  describes a large ensemble of

particles being all in the same quantum state. Measured quantities are obtained by averaging the measurement results over the large number  $N$  or the results measured on one particle over a long observation time (Sect. 2.4.1). We, therefore, expect a relation similar to (3.75) also for the probability density  $|\psi|^2 = \psi^* \psi$ . For the derivation we differentiate  $|\psi|^2$  with respect to time, that is,

$$\frac{\partial}{\partial t} \psi^* \psi = \psi^* \dot{\psi} + \dot{\psi}^* \psi \quad (3.76)$$

and use the Schrödinger equation to obtain

$$i\hbar \psi^* \dot{\psi} = -\frac{\hbar^2}{2m} \psi^* \Delta \psi + V \psi^* \psi, \quad (3.77a)$$

$$i\hbar \psi \dot{\psi}^* = \frac{\hbar^2}{2m} \psi \Delta \psi^* - V \psi^* \psi. \quad (3.77b)$$

Adding both equations together yields

$$i\hbar (\psi^* \dot{\psi} + \dot{\psi}^* \psi) = -\frac{\hbar^2}{2m} (\psi^* \Delta \psi - \psi \Delta \psi^*). \quad (3.78)$$

Using the simple relation  $\Delta = \nabla^2 = \text{div}(\text{grad})$  we get from (3.76), (3.77a) and (3.77b):

$$\frac{\partial}{\partial t} (\psi^* \psi) = -\text{div} \left[ \frac{\hbar}{2im} (\psi^* \nabla \psi - \psi \nabla \psi^*) \right]. \quad (3.79)$$

This equation is directly identified as the quantum mechanical analogon to the classical continuity relation (3.75). The quantum mechanical particle current probability density is easily recognized as

$$\mathbf{j} = \frac{\hbar}{2mi} (\psi^* \nabla \psi - \psi \nabla \psi^*) = \frac{1}{2m} (\psi^* \hat{\mathbf{p}} \psi + \psi \hat{\mathbf{p}}^* \psi^*). \quad (3.80)$$

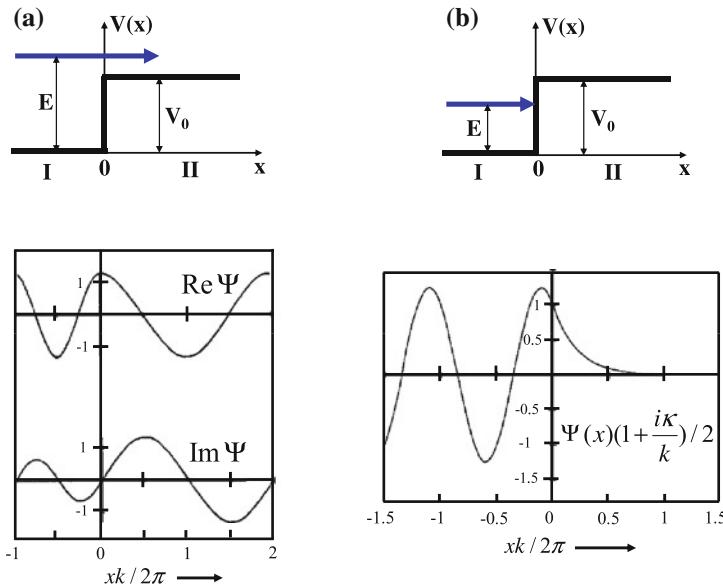
It describes the current probability density of particles in the state  $\psi(\mathbf{r}, t)$ . Note the correspondence between the classical velocity  $\mathbf{v} = \mathbf{p}/m$  of a particle and the operator  $\hat{\mathbf{p}}/m$  in (3.80).  $\mathbf{j}$  in (3.80) multiplied by the large number  $N$  of particles in an ensemble yields the real current density  $\mathbf{j}' = N\mathbf{v}$  which is measured in a particle stream as number of particles per area.  $N\mathbf{j} dA$  is the number of particles which are detected by a detector with area  $dA$  during time unit.  $\mathbf{j}'$  multiplied by the electronic charge is the electric current density in quantum physics.

### 3.6.3 Electrons Run Against a Potential Step

A classical particle which hits a potential barrier is reflected back or continues to propagate with diminished kinetic energy, if its kinetic energy exceeds the potential step height. Because of its wave character, electrons behave differently.

We solve the Schrödinger equation (3.51) for an electron which approaches a potential step of height  $V_0$  (region II in Fig. 3.8) from the left side (region I). Both cases, that of the electron energy  $E$  exceeding the potential height  $V_0$  (Fig. 3.8a) and that with  $E < V_0$  (Fig. 3.8b) are considered. In the latter case, the particle could not overcome the barrier in classical physics. The particle flow is stationary because of a time independent potential, that is, we can apply the time independent Schrödinger equation (3.57). The total time dependent solution  $\psi(\mathbf{r}, t)$  is obtained as usually according to (3.58) if the energy eigenvalue  $E$  has been determined from (3.57). The step potential (Fig. 3.8) is assumed with  $V_0 > 0$  as

$$V(x) = V_0 \Theta(x) \quad \text{with } \Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases} \quad (3.81)$$



**Fig. 3.8** **a, b** Electrons with a kinetic energy  $E$  propagate from a region I with negligible potential against a potential step ( $x = 0$ ) into the region II with a potential  $V_0$ . **a** For  $E > V_0$  the electrons continue to propagate with longer wave length in region II, as is evident from the plot  $\text{Re } \psi$  and  $\text{Im } \psi$  versus  $xk/2\pi$ . **b** For  $E < V_0$  the amplitude of the electron wave incident from region I decays exponentially into region II, as is seen from the plot of the wave function  $\psi(x)$  [reduced by a factor  $2/(1 + ik/k)$ ]

We solve the Schrödinger equation (3.57) separately for the spatial regions I ( $x < 0$ ) and II ( $x > 0$ ), that is,

$$\frac{d^2}{dx^2} \psi = -\frac{2mE}{\hbar^2} \psi \quad \text{in I}, \quad (3.82a)$$

$$\frac{d^2}{dx^2} \psi = -\frac{2m(E - V_0)}{\hbar^2} \psi \quad \text{in II}. \quad (3.82b)$$

For **particle energies exceeding the potential step** ( $E > V_0$ ) different wave numbers  $k_1$  and  $k_2$  are introduced and one obtains for the two regions I and II

$$\text{I: } \frac{d^2}{dx^2} \psi = -k_1^2 \psi; \quad k_1 = \sqrt{2mE}/\hbar, \quad (3.83a)$$

$$\text{II: } \frac{d^2}{dx^2} \psi = -k_2^2 \psi; \quad k_2 = \sqrt{2m(E - V_0)}/\hbar. \quad (3.83b)$$

These are vibration differential equations which are solved by  $\exp(\pm ikx)$ . Since particles do not get lost during their path across the potential step, the particle current must be continuous at the step ( $x = 0$ ). Because of (3.80) this continuity is required both for  $\psi$  and  $d\psi/dx$ , that is,

$$\psi_I(x = 0) = \psi_{II}(x = 0) \quad \text{and} \quad (3.84a)$$

$$\psi'_I(x = 0) = \psi'_{II}(x = 0). \quad (3.84b)$$

In order to fulfill these conditions, we match three plane waves at the potential step: in region I (left side) an incoming wave with wave vector  $k_1$  and a wave (same wave vector) reflected at the step with *reflection amplitude*  $r$ ; furthermore a transmitted wave with wave vector  $k_2$  in region II. The *transmission amplitude* for this wave is  $t$ , that is,

$$\psi_I(x) = e^{ik_1 x} + r e^{-ik_1 x}, \quad (3.85a)$$

$$\psi_{II}(x) = t e^{ik_2 x}. \quad (3.85b)$$

Using the continuity conditions (3.84a), (3.84b) at the step location  $x = 0$ , one immediately obtains

$$1 + r = t \quad \text{and} \quad ik_1(1 - r) = ik_2 t \quad (3.86a)$$

respectively,

$$r = \frac{k_1 - k_2}{k_1 + k_2}, \quad t = \frac{2k_1}{k_1 + k_2}. \quad (3.86b)$$

For a deeper insight into the meaning of the reflection and transmission amplitudes  $r$  and  $t$ , we calculate the current probability density  $j(x)$  within the two spatial regions

I and II according to (3.79):

$$\begin{aligned} j_I(x) &= \frac{\hbar}{2mi} [(e^{-ik_1x} + r^* e^{ik_1x}) ik_1 (e^{ik_1x} - r e^{-ik_1x}) - \text{c.c.}] \\ &= \frac{\hbar}{2mi} [ik_1(1 - |r|^2 - r e^{-2ik_1x} + r^* e^{2ik_1x}) - \text{c.c.}], \\ j_I(x) &= \frac{\hbar k_1}{m} (1 - |r|^2) \equiv j_{\text{in}} - j_{\text{refl}}, \end{aligned} \quad (3.87a)$$

$$j_{\text{II}}(x) = \frac{\hbar k_2}{m} |t|^2 \equiv j_{\text{trans.}} \quad (3.87b)$$

Since  $\hbar k/m = p/m$  is the velocity of a particle,  $j_I$  and  $j_{\text{II}}$  describe current densities as in the classical case, but related to one particle ( $j = v$ ).  $j_I$  is the difference between the incoming and the reflected current density, while  $j_{\text{II}}$  describes the current of particles propagating (transmitted) over the potential step. In addition to the amplitudes  $r$  and  $t$  one might introduce *transmission* and *reflection coefficients* or *probabilities*  $T$  and  $R$ , respectively, by

$$R = \frac{j_{\text{refl}}}{j_{\text{in}}} = |r|^2, \quad T = \frac{j_{\text{trans}}}{j_{\text{in}}} = \frac{k_2}{k_1} |t|^2. \quad (3.88)$$

We summarize: In quantum physics, particles are reflected at a potential step, even when their kinetic energy is sufficiently high ( $E > V_0$ ) to overcome the step. In classical physics, all particles would continue to propagate forwards, though with reduced energy. The reflection process is a typically wave mechanical phenomenon which we know well from optics. Light (photons) is always reflected at the interface between media with different refraction index.

For **particle energies below the potential step** ( $E < V_0$ ) the Schrödinger equation (3.82a) in region I remains unchanged (Fig. 3.8b). As for  $E > V_0$  it is solved by (3.85a). In the spatial range of the potential step (region II), however, we have to assume

$$\frac{d^2}{dx^2} \psi = \kappa^2 \psi; \quad \kappa = \sqrt{2m(V_0 - E)/\hbar}. \quad (3.89a)$$

In contrast to (3.83b),  $\kappa$  is a real number because of  $V_0 > E$ . Equation (3.89a) thus has solutions which exponentially increase or decrease with  $x$ . The solutions follow immediately with (3.83a), (3.83b) by assuming  $k_2 = i\kappa$  as a purely imaginary number [see (3.89a)]. If we transfer the results from (3.83a)–(3.85b), the solution for region II with finite probability density follows as

$$\psi_{\text{II}}(x) = t e^{-\kappa x}. \quad (3.89b)$$

In analogy to (3.86b), we obtain the reflection and transmission amplitudes as

$$r = \frac{k_1 - i\kappa}{k_1 + i\kappa}, \quad t = \frac{2k_1}{k_1 + i\kappa}. \quad (3.90)$$

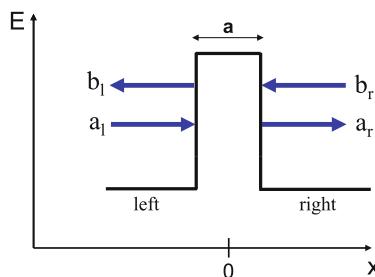
From (3.90), we derive  $|r|^2 = 1$ , which means that all particles are reflected for  $E < V_0$ . Nevertheless, because of (3.89a), (3.89b), particles enter to a certain extent, namely up to the average depth  $\kappa^{-1}$ , into the potential step. This is also a phenomenon, which we know from optics, where light penetrates slightly into a strongly absorbing medium.

Furthermore, these results suggest, that electrons confined in a quantum well with finite potential walls behave differently from those confined between infinitely high walls as in Sect. 3.6.1. The wave function does not vanish on the surface of the potential box but rather penetrates through the walls. Depending on the height of the potential barrier  $V_0$  and on the energy  $E$  of the quantized confined state the electronic wave function (3.62) decays into the exterior of the potential box over a distance  $\kappa^{-1} = (\sqrt{2m(V_0 - E)/\hbar})^{-1}$ .

### 3.6.4 Electrons Tunnel Through a Barrier

We have seen that electrons, because of their wave character, can enter a potential step similarly as light waves do in an absorbing medium. We, thus, expect that electrons can penetrate a sufficiently thin energetic barrier, even when their energy is lower than the potential barrier. This is, of course, impossible for classical particles if they do not have enough kinetic energy to overcome the energetic barrier. But again, the analogy to light waves is given, which can penetrate thin metal films (semi-transparent metal coatings on windows).

For the description of the phenomenon called *electron tunneling*, we assume a rectangular potential barrier with spatial extension  $a$  (width) on the  $x$  axis and an energetic height  $V_B$  on the energy scale (Fig. 3.9). The tunneling problem can be described in terms of a stationary flux of electrons left and right of the barrier and the potential in the time independent Schrödinger equation (3.57) is assumed piece-wise as constant:



**Fig. 3.9** Schematic plot of an electron wave which moves with an amplitude  $a_l$  from the left side towards an energy barrier with thickness  $a$ . For the solution of the problem of an electron tunneling through the barrier a superposition of several waves is required: a wave reflected to the left with amplitude  $b_l$  and two waves with amplitudes  $a_r$  and  $b_r$  in the region on the right side of the barrier

$$V(x) = \begin{cases} 0 & \text{for } x \leq -\frac{a}{2} \text{ (left),} \\ V_B & \text{for } -\frac{a}{2} < x < \frac{a}{2}, \\ 0 & \text{for } x \geq \frac{a}{2} \text{ (right).} \end{cases} \quad (3.91a)$$

For a constant potential (3.57) is solved by plane waves, since the electron energy  $E$  exceeds the potential maximum. Within the barrier region, however, where  $E = \hbar^2 k^2 / 2m < V_B$ , we must assume, in analogy to (3.89a), (3.89b), exponentially decaying solutions. The solutions left and right of the barrier must be matched continuously to those within the barrier. Because of particle (flux) conservation (3.80), both  $\psi$  and  $d\psi/dx$  must be continuous at the barrier borders at  $x = -a/2$  and  $x = a/2$  (Fig. 3.9). The most general ansatz for the wave functions in the three spatial regions is

$$\psi(x) = \begin{cases} a_l e^{ikx} + b_l e^{-ikx}; & x < -a/2, \\ c e^{-\kappa x} + d e^{\kappa x}; & x < |a/2|, \\ a_r e^{ikx} + b_r e^{-ikx}; & x > a/2. \end{cases} \quad (3.91b)$$

Given the complete expression for a plane wave with its time dependence  $\psi \propto \exp(ikx - i\omega t)$  we must interpret  $a_l$  and  $a_r$  as the amplitudes of waves propagating to the right direction in the left ( $x < a/2$ ) and right ( $x > a/2$ ) spatial region (Fig. 3.9). In analogy, the waves with amplitudes  $b_l$  and  $b_r$  propagate to the left within the regions left and right of the barrier (Fig. 3.9). We must make an ansatz with four propagating plane waves which are matched to the solution inside the barrier, in order to fulfill the continuity conditions at the barrier borders. Inserting (3.91a), (3.91b) into the Schrödinger equation (3.57) we obtain, in analogy to (3.82a), (3.82b), the following energy-wave number relations:

$$E = \frac{\hbar^2 k^2}{2m} \quad \text{for } x < -a/2 \text{ and } x > a/2, \quad (3.92a)$$

$$V_B - E = \frac{\hbar^2 \kappa^2}{2m} \quad \text{for } x < |a/2|. \quad (3.92b)$$

Together with (3.92a), (3.92b), the continuity of  $\psi$  and  $d\psi/dx$  at the left barrier side at  $x = -a/2$  requires

$$a_l e^{-ika/2} + b_l e^{ika/2} = c e^{\kappa a/2} + d e^{-\kappa a/2}, \quad (3.93a)$$

$$ik(a_l e^{-ika/2} - b_l e^{ika/2}) = -\kappa(c e^{\kappa a/2} - d e^{-\kappa a/2}). \quad (3.93b)$$

These relations can be written as matrices

$$\begin{pmatrix} e^{-ika/2} & e^{ika/2} \\ e^{-ika/2} & -e^{ika/2} \end{pmatrix} \begin{pmatrix} a_l \\ b_l \end{pmatrix} = \begin{pmatrix} e^{\kappa a/2} & e^{-\kappa a/2} \\ \frac{i\kappa}{k} e^{\kappa a/2} & -\frac{i\kappa}{k} e^{-\kappa a/2} \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix}. \quad (3.93c)$$

We now apply the rules of matrix algebra, which are known already or are explained in a more general context in Sect. 4.3.1, and solve for the vector  $(a_l, b_l)$ , that is, we rewrite (3.93c) into

$$\begin{pmatrix} a_l \\ b_l \end{pmatrix} = \underline{\underline{M}}(a/2) \begin{pmatrix} c \\ d \end{pmatrix}. \quad (3.94a)$$

Hereby the matrix  $\underline{\underline{M}}(a/2)$  has the form

$$\underline{\underline{M}}(a/2) = \frac{1}{2} \begin{pmatrix} (1 + \frac{ik}{\kappa})e^{\kappa a/2 + ika/2} & (1 - \frac{ik}{\kappa})e^{-\kappa a/2 + ika/2} \\ (1 - \frac{ik}{\kappa})e^{\kappa a/2 - ika/2} & (1 + \frac{ik}{\kappa})e^{-\kappa a/2 - ika/2} \end{pmatrix}. \quad (3.94b)$$

Our problem of matching  $\psi$  and  $d\psi/dx$  together at the positions  $-a/2$  and  $a/2$  has mirror symmetry about  $x = 0$  (Fig. 3.9). For the barrier side at  $a/2$  we, thus, obtain the transformation matrix by substituting  $a/2$  by  $-a/2$  in (3.94b):

$$\begin{pmatrix} a_r \\ b_r \end{pmatrix} = \underline{\underline{M}}(-a/2) \begin{pmatrix} c \\ d \end{pmatrix}. \quad (3.94c)$$

By combining (3.94a) with (3.94c), the relation between the wave amplitudes left (source) and the amplitudes in the right region (drain) is obtained as

$$\begin{pmatrix} a_l \\ b_l \end{pmatrix} = \underline{\underline{M}}\left(\frac{a}{2}\right) \left[ \underline{\underline{M}}\left(-\frac{a}{2}\right) \right]^{-1} \begin{pmatrix} a_r \\ b_r \end{pmatrix}. \quad (3.95a)$$

The matrix

$$\underline{\underline{M}}\left(\frac{a}{2}\right) \left[ \underline{\underline{M}}\left(-\frac{a}{2}\right) \right]^{-1} = \underline{\underline{S}}^{-1} \quad (3.95b)$$

is called the inverse of the *transfer-matrix* (transfer from left to right):

$$\underline{\underline{S}} = \underline{\underline{M}}\left(-\frac{a}{2}\right) \left[ \underline{\underline{M}}\left(\frac{a}{2}\right) \right]^{-1}. \quad (3.95c)$$

With the inverse matrix

$$\left[ \underline{\underline{M}}\left(-\frac{a}{2}\right) \right]^{-1} = \frac{1}{2} \begin{pmatrix} (1 - \frac{ik}{\kappa})e^{\kappa a/2 + ika/2} & (1 + \frac{ik}{\kappa})e^{\kappa a/2 - ika/2} \\ (1 + \frac{ik}{\kappa})e^{-\kappa a/2 + ika/2} & (1 - \frac{ik}{\kappa})e^{-\kappa a/2 - ika/2} \end{pmatrix} \quad (3.96)$$

we calculate from (3.95a)–(3.95c) the relation between left and right wave amplitudes as

$$\begin{aligned} \begin{pmatrix} a_l \\ b_l \end{pmatrix} &= \underline{\underline{S}}^{-1} \begin{pmatrix} a_r \\ b_r \end{pmatrix} \\ &= \begin{pmatrix} (\cosh \kappa a + \frac{i\varepsilon}{2} \sinh \kappa a)e^{ika} & \frac{i\eta}{2} \sinh \kappa a \\ -\frac{i\eta}{2} \sinh \kappa a & (\cosh \kappa a - \frac{i\varepsilon}{2} \sinh \kappa a)e^{-ika} \end{pmatrix} \begin{pmatrix} a_r \\ b_r \end{pmatrix}. \end{aligned} \quad (3.97)$$

Hereby the following definitions were introduced

$$\varepsilon = \frac{\kappa}{k} - \frac{k}{\kappa}, \quad (3.98a)$$

$$\eta = \frac{\kappa}{k} + \frac{k}{\kappa} \quad (3.98b)$$

with  $\kappa^{-1}$  as the exponential decay length of the wave function into the barrier (3.92a), (3.92b), (3.93a)–(3.93c).

We now consider the frequent special case of transmission of particles which approach the barrier from the left side. Then  $b_r$  must be assumed to be zero ( $b_r = 0$ ) and we obtain from (3.97)

$$a_l = a_r \left( \cosh \kappa a + \frac{i\varepsilon}{2} \sinh \kappa a \right) e^{ika}, \quad (3.99a)$$

$$b_l = a_r \left( -\frac{i\eta}{2} \right) \sinh \kappa a. \quad (3.99b)$$

The *transmission amplitude*  $\vec{t}$  from left to right, then, describes the tunneling of the electron through the barrier in terms of an attenuation of the incoming wave amplitude upon transmission through the barrier:

$$\vec{t} = \frac{a_r}{a_l} = \frac{e^{-ika}}{\cosh \kappa a + i(\varepsilon/2) \sinh \kappa a}, \quad (3.100)$$

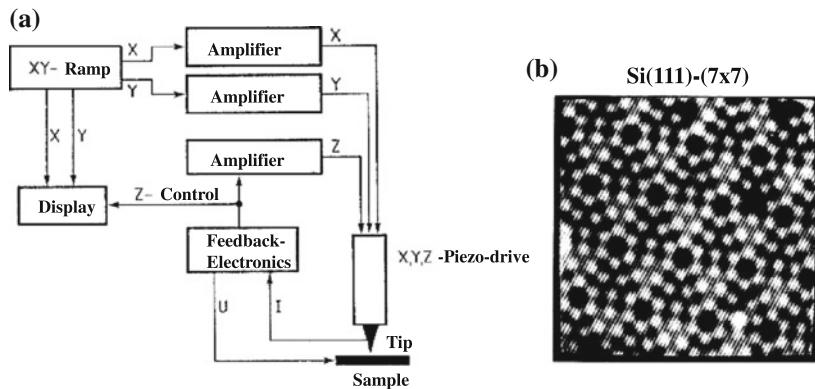
$\vec{t}$  as an amplitude is complex valued, it contains the phase change of the wave function upon tunneling. Its real absolute square  $|\vec{t}|^2$  is the transmission probability from left to right:

$$\vec{T} = |\vec{t}|^2 = \frac{1}{1 + (1 + \frac{\varepsilon^2}{4}) \sinh^2 \kappa a} = \frac{1}{1 + \frac{1}{4}\eta^2 \sinh^2 \kappa a}. \quad (3.101)$$

For high and wide barriers with low transmission probability one has  $\kappa a \gg 1$ , that is,  $\sinh(\kappa a) \approx 1/2$  and  $\exp(\kappa a) \gg 1$ . In this limit of weak tunneling, the transmission or tunneling probability decays exponentially with the barrier width  $a$ :

$$\vec{T} = |\vec{t}|^2 \approx \frac{16E}{V_B} \exp(-2\sqrt{2m(V_B - E)}a/\hbar). \quad (3.102)$$

The tunnel effect described by (3.102) has played an essential role during the development of quantum mechanics. In 1928, Gamow could explain the alpha decay of atomic nuclei by means of tunneling of  $\alpha$  particles (He nuclei) out of the nuclear potential [3]. The nuclear potential forms a wall around the nucleus which confines the protons and neutrons therein. Classically energy had to be supplied to the nucleus



**Fig. 3.10** **a, b** Scanning electron tunneling microscopy. **a** Electrical components and set-up of a Scanning Tunneling Microscope (STM). Tip, piezo-drive and sample are located in an ultra-high vacuum (UHV) vessel. **b** STM image of a Si(111) surface with  $(7 \times 7)$  surface reconstruction prepared in UHV. The bright dots are produced by tunneling electrons originating from occupied Si atomic orbitals [5]

for the  $\alpha$  particle to overcome the barrier and to be emitted. But due to tunneling of the  $\alpha$  particle through the nuclear potential wall the nucleus decays without energy supply from outside.

Also in nuclear fusion of hydrogen to helium a tunnel barrier has to be penetrated. The quantitative description of the effect is based on tunneling of H atoms through the nuclear potential.

Meanwhile the most important application of electron tunneling (3.102) in solid state physics is related to the *scanning electron tunneling microscope* (Scanning Tunneling Microscopy = STM) [4]. As is shown schematically in Fig. 3.10a, the tunneling current of electrons between a metallic tip and a solid surface being the sample under study is measured. Tip and surface are separated by a vacuum gap (about 0.1 to 1 nm wide) which represents the tunnel barrier. The tip is scanned over the surface by piezoelectric drives with a spatial precision in the range of 0.1 nm and the tunnel current is measured and registered pixel-wise as a function of the position on the surface. Because of the exponential dependence of the tunnel current on the gap width between tip and surface (3.102) the tunnel current reacts extremely sensitively to any unevenness of the surface under study. The scanning images of the tunnel current, thus, show steps, dislocations, and at highest resolution, even the position of surface atoms in detail (Fig. 3.10b) [5]. Note that the tunneling electrons originate from the spatially extended atomic orbitals of the surface atoms or of those of the tip (at inverted tip bias). Position and shape of the measured intensity contrast in STM images, thus, is determined by the shape of the atomic orbitals. STM images have been a breakthrough in recognizing the wave nature of matter. For the first time real pictures of the wave-like electronic orbitals at atoms could be made visible experimentally. The solutions of the Schrödinger equation, so far existent

only in mathematics, appeared on an experimental screen. For a further experimental example see Sect. 6.2.4.

A further important application of the tunnel effect has emerged in modern semiconductor device physics. By means of advanced layer deposition (epitaxy) techniques (MBE = molecular beam epitaxy, MOVPE = metal-organic vapor phase epitaxy, Appendix B) semiconductor layer stacks with layer thicknesses down to several atomic layers can be grown perfectly in their crystallographic structure on each other (semiconductor heterostructures). By these techniques thin crystalline AlAs layers, for example, can be built-in into GaAs having a lower electronic band gap than AlAs. In AlAs, quasi-free electrons in the conduction band have an energy higher by about 1 eV than in GaAs. A thin AlAs layer imbedded in a GaAs matrix, thus, acts as a potential barrier for free electrons in the GaAs matrix. Conduction electrons in the GaAs matrix can tunnel through the AlAs barrier, if it is sufficiently thin. Furthermore, the AlAs barrier exhibits a higher electrical resistance (electrons collect in the energetically more favorable GaAs) than the surrounding GaAs region; an electrical bias can be applied between the two GaAs regions on both sides of the AlAs barrier. Thus, by means of tunnel barriers built-in into semiconductor heterostructures novel semiconductor devices with new functionalities can be realized. This will be more extensively considered in the next section.

### 3.6.5 Resonant Tunneling

Two potential barriers arranged in sequence, one behind the other, yield interesting transport properties for electrons which are able to cross the whole structure. Such double barrier arrangements can easily be realized in semiconductor heterostructure technology (Appendix B). Between the two barriers a potential well is formed which gives rise to discrete confined electronic states similarly as in the potential box (Sect. 3.6.1). Because of the finite energetic height of the barriers these confined states do not have vanishing amplitudes at the inner barrier boundaries; they decay exponentially into the barriers (quasi-standing waves). For sufficiently thin and low barriers the exponential tails of these quasi-standing waves couple to the tails of plane waves, the states of free electrons left and right of the double barrier, which also leak into the barriers from left and right (Sect. 3.6.3). As a consequence, one single coherent wave function extends over the whole space, the double barrier region and the regions of free electronic states left and right of the barriers.

The situation resembles that of a Fabry–Perrot interferometer in optics. There, two parallel semi-transparent mirrors (finite potential barriers for light photons) cause the built-up of quasi-standing waves in between when a parallel light wave passes the double mirror arrangement. The transparency of the double mirror set-up reaches a maximum when multiples of half the light wave length match the mirror distance. We will see that a double barrier arrangement in a semiconductor heterostructure is nothing else but a Fabry–Perrot interferometer for electron waves.

For an elegant mathematical description of electron transmission through a double barrier we introduce, beside the transfer-matrix  $\underline{\underline{S}}$  (3.95a)–(3.95c), the so-called *transmission-matrix*. This transmission-matrix couples the amplitudes of waves leaving a potential barrier ( $a_r, b_l$  in Fig. 3.9) with those of waves approaching the barrier ( $a_l, b_r$  in Fig. 3.9). For its calculation one has to reorder the continuity conditions (3.93a)–(3.93c) such that transmitted and reflected wave amplitudes, each propagating to the right and the left, respectively, are described by appropriate transmission ( $t$ ) and reflection-amplitudes ( $r$ ) (Fig. 3.9):

$$a_r = \vec{t} a_l + r_r b_r, \quad (3.103a)$$

$$b_l = r_l a_l + \overleftarrow{t} b_r. \quad (3.103b)$$

The transmission-matrix defined by the coefficients  $t$  and  $r$  is thus obtained as:

$$\begin{pmatrix} a_r \\ b_l \end{pmatrix} = \begin{pmatrix} \vec{t} & r_r \\ r_l & \overleftarrow{t} \end{pmatrix} \begin{pmatrix} a_l \\ b_r \end{pmatrix}. \quad (3.103c)$$

For ideal reflection and transmission through the barrier and because of symmetry reasons, we can assume:

$$t = \vec{t} = \overleftarrow{t}, \quad (3.104a)$$

$$|r|^2 = |r_l|^2 = |r_r|^2, \quad (3.104b)$$

$$|t|^2 + |r|^2 = 1. \quad (3.104c)$$

Taking into account the different signs in the exponent of waves propagating to left and to the right, respectively, we conclude furthermore:

$$\vec{t} r_r^* = -r_l \vec{t}^*, \quad (3.105a)$$

$$\overleftarrow{t} r_l^* = -r_r^* \overleftarrow{t}^*. \quad (3.105b)$$

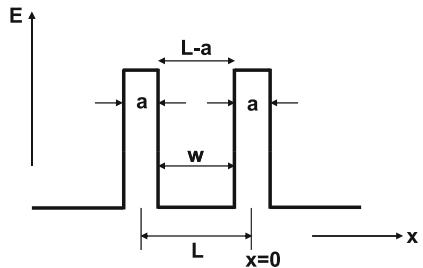
By comparison of (3.103c) with (3.95a)–(3.95c) and reordering of the wave amplitudes  $a_l, b_l, a_r, b_r$  the relation between the transfer-matrix  $\underline{\underline{S}}$  and the elements  $r$  and  $t$  of the transmission-matrix is obtained:

$$\underline{\underline{S}} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\vec{t}^*} & \frac{r_r}{\vec{t}} \\ \frac{r_l^*}{\overleftarrow{t}} & \frac{1}{\overleftarrow{t}} \end{pmatrix}. \quad (3.106)$$

For the determinant of  $\underline{\underline{S}}$  follows

$$\det \underline{\underline{S}} = \frac{\vec{t}}{\overleftarrow{t}} = \frac{\vec{t}^*}{\overleftarrow{t}^*}. \quad (3.107)$$

**Fig. 3.11** Explanation of symbols being used for the description of resonant tunneling of an electron through a double barrier structure



We express the complex transmission amplitude  $t$  in terms of its modulus and its phase  $\phi_t$  as

$$t = |t| \exp(i\phi_t), \quad (3.108a)$$

and obtain by comparison of (3.97) with (3.106)

$$\phi_t = \varphi - ka. \quad (3.108b)$$

Hereby, the angle  $\varphi$  is defined by

$$\tan \varphi = \frac{1}{2} \left( \frac{k}{\kappa} - \frac{\kappa}{k} \right) \tanh \kappa a = -\frac{1}{2} \varepsilon \tanh \kappa a. \quad (3.108c)$$

The angle  $\phi_t$  in (3.108b) describes the phase shift of the electron wave upon transmission through the barrier. It depends sensitively on the wave vector  $k$  and the decay length  $1/\kappa$  of the wave amplitude in the barrier.

In order to derive a formula for the transmission of an electron wave through a double barrier we use the results about tunneling through one barrier at  $x = 0$ , shift this barrier mathematically to  $x = -L$  (Fig. 3.11) and calculate the effect of this shift on the tunneling. Then the effects of both barriers have to be “switched” in series mathematically.

The shift of the barrier to  $x = -L$  is identical with a shift of the coordinate system by  $L$  in positive direction. This means for the wave functions:

$$\psi(x) = a_l e^{ikx} + b_l e^{-ikx} \rightarrow \psi'(x) = a_l e^{ikL} e^{ikx} + b_l e^{-ikL} e^{-ikx}. \quad (3.109)$$

The amplitudes  $a_l$  and  $b_l$  change into  $a'_l = a_l \exp(ikL)$  and  $b'_l = b_l \exp(-ikL)$ , that is, we obtain a transformation formula

$$\begin{pmatrix} a'_l \\ b'_l \end{pmatrix} = \begin{pmatrix} e^{ikL} & 0 \\ 0 & e^{-ikL} \end{pmatrix} \begin{pmatrix} a_l \\ b_l \end{pmatrix} = \underline{T} \begin{pmatrix} a_l \\ b_l \end{pmatrix} \quad (3.110a)$$

and the inverse of  $\underline{T}$

$$\underline{\underline{T}}^{-1} = \begin{pmatrix} e^{-ikL} & 0 \\ 0 & e^{ikL} \end{pmatrix}. \quad (3.110b)$$

As the transfer matrix  $\underline{\underline{S}}$  relates the wave amplitude left with that on the right side

$$\begin{pmatrix} a_r \\ b_r \end{pmatrix} = \underline{\underline{S}} \begin{pmatrix} a_l \\ b_l \end{pmatrix}, \quad (3.111)$$

we still have to calculate the transfer matrix  $\underline{\underline{S}}_L$  for the barrier which has been shifted to left. For this purpose, we transform, analogously to (3.110a), the amplitudes into the shifted coordinate system (upper dash) in (3.111):

$$\begin{pmatrix} a_r \\ b_r \end{pmatrix} = \underline{\underline{T}}^{-1} \begin{pmatrix} a'_r \\ b'_r \end{pmatrix} = \underline{\underline{S}} \underline{\underline{T}}^{-1} \begin{pmatrix} a'_l \\ b'_l \end{pmatrix}. \quad (3.112a)$$

We obtain for the shifted barrier the relation

$$\begin{pmatrix} a'_r \\ b'_r \end{pmatrix} = \underline{\underline{T}} \underline{\underline{S}} \underline{\underline{T}}^{-1} \begin{pmatrix} a'_l \\ b'_l \end{pmatrix} \quad (3.112b)$$

and finally the transformation matrix for the barrier at  $x = -L$  as

$$\underline{\underline{S}}_L = \underline{\underline{T}} \underline{\underline{S}} \underline{\underline{T}}^{-1}. \quad (3.113)$$

$\underline{\underline{S}}$  and  $\underline{\underline{S}}_L$  connect wave amplitudes right of the respective barrier with those on the left side. The total effect of the double barrier arrangement on the wave amplitudes is therefore obtained by multiplication of both transfer matrices:

$$\underline{\underline{S}}_{\text{tot}} = \underline{\underline{S}}_L \underline{\underline{S}}. \quad (3.114)$$

Together with (3.106), (3.110a), (3.110b) and (3.113), we finally obtain the total transfer matrix as

$$\underline{\underline{S}}_{\text{tot}} = \begin{pmatrix} \frac{1}{t^{*2}} + e^{-2ikL} |\frac{r}{t}|^2 & \frac{r}{t} (\frac{e^{-2ikL}}{t} + \frac{1}{t^*}) \\ \frac{r^*}{t^*} (\frac{e^{2ikL}}{t^*} + \frac{1}{t}) & \frac{1}{t^2} + e^{2ikL} |\frac{r}{t}|^2 \end{pmatrix}. \quad (3.115)$$

According to (3.106) the total transmission amplitude  $t_{\text{tot}}$  through the double barrier structure is, then, obtained from the matrix element  $S_{22}^{\text{tot}}$  as

$$t_{\text{tot}} = \frac{1}{S_{22}^{\text{tot}}} = \frac{t^2}{1 + \frac{t^2}{|t|^2} |r|^2 e^{2ikL}} = \frac{t^2}{1 + |r|^2 e^{2i(kL + \phi_t)}} \quad (3.116)$$

and the transmission probability as

$$T_{\text{tot}} = |t_{\text{tot}}|^2 = \frac{T^2}{|1 + |r|^2 e^{2i(kL + \phi_t)}|^2} = \frac{(1 - |r|^2)^2}{|1 + |r|^2 e^{2i(kL + \phi_t)}|^2}. \quad (3.117)$$

In the limit of low reflection  $|r|^2 \ll 1$  the transmission probability approaches the value one. It can, then, be written as a product of transmission probabilities of the two single barriers. Interference effects between the barriers (standing waves in the quantum well) can be neglected.

The transmission probability through the double barrier is exactly one, if  $T^2/|1 - |r|^2|^2 = T^2/|t|^2 = 1$ . This condition is fulfilled, if

$$\exp[2i(kL + \phi_t)] = -1. \quad (3.118a)$$

For this singular case it is required, that

$$2(kL + \phi_t) = (2n + 1)\pi, \quad n = 0, 1, 2, \dots. \quad (3.118b)$$

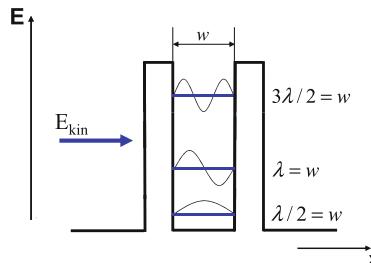
For high barriers with strong reflection and low transmission (limit:  $V_B \rightarrow \infty$ ) (3.92b) yields  $\kappa \rightarrow \infty$ , and with (3.108c), we obtain  $\tan \varphi \rightarrow -\infty$  and respectively  $\varphi \rightarrow -\pi/2$ . Because of  $\phi_t = \varphi - ka$  (3.108b), then (3.118b) yields the condition

$$2\left(kL - \frac{\pi}{2} - ka\right) = (2n + 1)\pi, \quad n = 0, 1, 2, \dots, \quad (3.118c)$$

and because of  $w = L - a$  (Fig. 3.11)

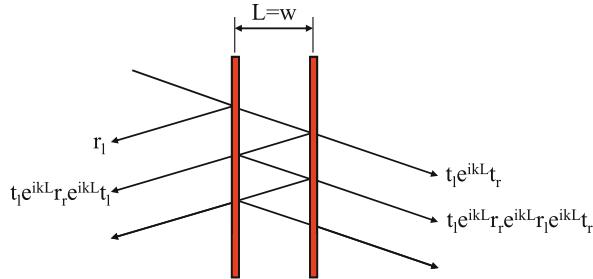
$$w = (n + 1)\lambda/2. \quad (3.119)$$

In this limit of high barriers with high reflection the double barrier arrangement exhibits a transmission probability of one (ideal transmission), if multiples of half the electron wave length match the width  $w$  of the quantum well between the barriers. The related kinetic electron energy is identical with the energies of the discrete confined electronic states between the barriers (standing waves) (Fig. 3.12). In this



**Fig. 3.12** Schematic plot of the three lowest energy bound states in a double barrier potential well structure of width  $w$ : wave functions with  $\lambda = 2w$ ,  $\lambda = w$ ,  $\lambda = 2w/3$ . Depending on their kinetic energy  $E_{\text{kin}}$  electrons can resonantly tunnel through these bound states

**Fig. 3.13** Another representation of resonant electron tunneling through a double barrier: Partial electron waves being many times transmitted and reflected with transmission and reflection amplitudes  $t_l$ ,  $r_l$  and  $t_r$ ,  $r_r$  left ( $l$ ) and right ( $r$ ) are superimposed



case, one single coherent electronic state is formed by the incoming (from the left) and the transmitted (to the right) plane wave as well as the standing wave between the two barriers (Fig. 3.14). This phenomenon is called *resonant tunneling* (through a confined state).

At the beginning of the section, we have already emphasized the analogy between resonant electron tunneling and light waves passing two parallel semitransparent mirrors in a Fabry–Perrot interferometer. Correspondingly, we can calculate the transmission amplitude  $t_{\text{tot}}$  for an electron tunneling through the double barrier by summing up the contributions of partial waves which have passed the barriers after multiple reflection and transmission (Fig. 3.13). For simplicity, we assume thin barriers with negligible thickness at a distance  $L$  (approximately width  $w$  of the potential well, Fig. 3.13). In that case, the phase change of the electron waves upon tunneling through a single barrier can be neglected, that is,  $t^* \approx t \approx |t|$ . For sake of clearness, we denote the transmission and reflection amplitudes, respectively, of the left barrier by  $t_l$ ,  $r_l$  and those of the right barrier by  $t_r$ ,  $r_r$ . According to Fig. 3.13 each partial wave, which contributes to the total transmitted signal after two additional reflections left and right in the potential well, contains an additional amplitude factor  $\exp(ikL)$ . The total transmission amplitude is, then, obtained by superposition of all partial waves:

$$t_{\text{tot}} = t_l e^{ikL} t_r + t_l e^{ikL} r_r e^{ikL} r_l e^{ikL} t_r + t_l e^{ikL} r_r e^{ikL} r_l e^{ikL} r_r e^{ikL} r_l e^{ikL} t_r + \dots \quad (3.120)$$

We now assume both barriers, left and right, as equal, that is,  $r_l = r_r$  and  $t_l = t_r$ , and obtain

$$\begin{aligned} t_{\text{tot}} &= t^2 e^{ikL} + t^2 r^2 e^{3ikL} + t^2 r^4 e^{5ikL} + \dots \\ &= t^2 e^{ikL} [1 + r^2 e^{2ikL} + (r^2 e^{2ikL})^2 + \dots]. \end{aligned} \quad (3.121)$$

Because of  $|r^2 \exp(2ikL)| < 1$  we can sum up the geometrical series (3.121) and obtain

$$t_{\text{tot}} = \frac{t^2 e^{ikL}}{1 - r^2 e^{i2kL}}. \quad (3.122)$$

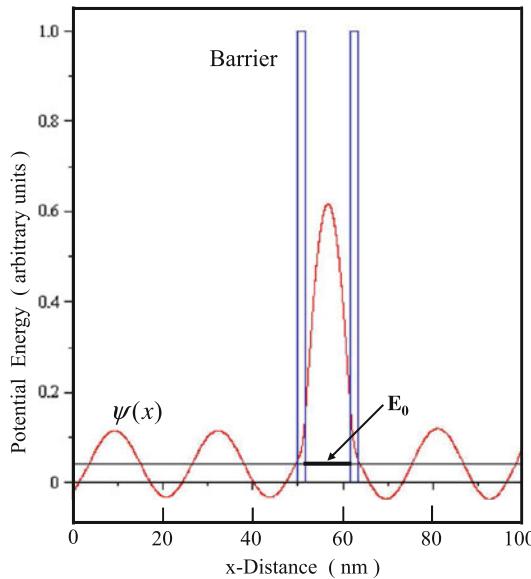
The similarity with the more rigorous expression (3.116) is striking. Differences in comparison with (3.116) are due to the simplifications assumed for the barriers in the latter derivation (neglecting phase shift in barriers).

Again we consider the ideal situation of maximum transmission  $|t_{\text{tot}}|^2 = 1$ . From (3.122), the requirement is  $t^2 = 1 - r^2$ , that is,  $\exp(i2kL) = 1$ . This condition is fulfilled by  $2kL = 2\pi \cdot n$  with  $n = 0, 1, 2, 3, \dots$ . The resulting condition for maximum transmission is analogous to (3.119):

$$w \approx L = n \frac{\lambda}{2}, \quad n = 1, 2, 3, \dots \quad (3.123)$$

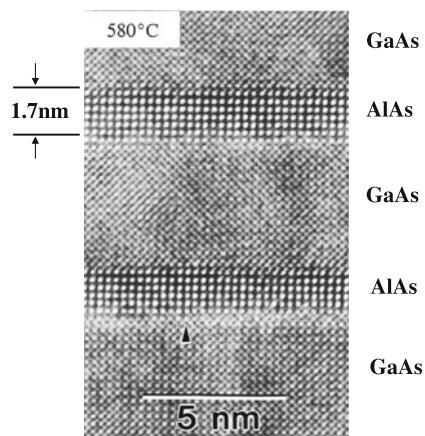
The same physics determines maximum transmission  $|T| = 1$  through the double barrier (Fig. 3.13), namely multiples of half the electron wave length  $\lambda$  must fit into the potential well between the two barriers (Fig. 3.12).

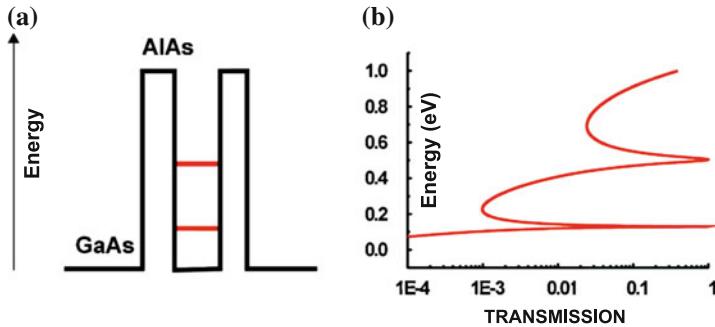
As was emphasized before, upon resonance (3.123) one single coherent wave function extends over the whole structure, from source (left) over the double barrier to drain (right). It carries the electron transport by tunneling. Such a coherent wave function was calculated by the described mathematical formalism for the case that the electron wave coming from the left has a kinetic energy matching the energy  $E_0$  of the lowest confined electronic state within the quantum well (Fig. 3.14). In the left source region, the plane wave character of  $\psi(x)$  is clearly seen. It passes over continuously into the standing wave (approximate lateral extension  $\lambda/2$ ), which forms in the quantum well, and finally into the plane wave leaving the barrier structure within the right drain region. The wave amplitude in the quantum well is strongly enhanced as is expected for many back and forth reflections at the inner walls of the quantum well (Fig. 3.13). The calculation was performed for realistic conditions of two AlAs barriers imbedded into a GaAs layer. Free electrons in AlAs have a minimum potential energy which is higher by about 1 eV than in GaAs (Sect. 8.3.4, Appendix A). For free electrons in the GaAs matrix, therefore, the two AlAs layers form a double barrier structure (Figs. 3.14 and 3.15) as discussed above. Such semiconductor resonance tunneling structures can easily be prepared as epitaxial layer stacks with atomic interface precision by modern deposition techniques (Appendix B) as is seen from the transmission electron micrograph in Fig. 3.15 [6–8]. For the described semiconductor double barrier structure two confined states between the AlAs barriers are calculated, which are shown in a quantitative plot in Fig. 3.16a. The transmission probability calculated on this basis as a function of the kinetic energy of an electron approaching from the left is presented in Fig. 3.16b. Corresponding to the two discrete confined states in the well (Fig. 3.16a) two sharp transmission bands ( $T = 1$ ) appear at energies of about 0.2 and 0.5 eV [8]. The electrons tunnel without any resistance through the quasi-bound states of the quantum well. Because of the finite barrier height the energetic positions of the quasi-bound states obey only approximately the simple  $1/w^2$  dependence of confined states in an ideal quantum well (3.64). Because of the higher electrical resistance of the AlAs barriers as compared with the GaAs matrix an external voltage applied to a layer stack as in Fig. 3.15 drops essentially across the double barrier region. By a variation of the applied voltage the electron energy on one side



**Fig. 3.14** Coherent wave function  $\psi(x)$  (real) of an electron, which approaches a double barrier structure (energy scale on left ordinate) from the *left side*, tunnels resonantly through the double barrier and continues to propagate as a free particle again on the *right side*. The electron propagates with a kinetic energy  $E_0$  which is identical with the energy of the first quasi-bound state of the quantum well between the barriers. Between the barriers the wave amplitude is significantly enhanced due to the quasi-bound state. The calculation has been performed for two AlAs barriers embedded in a GaAs matrix [6, 7]

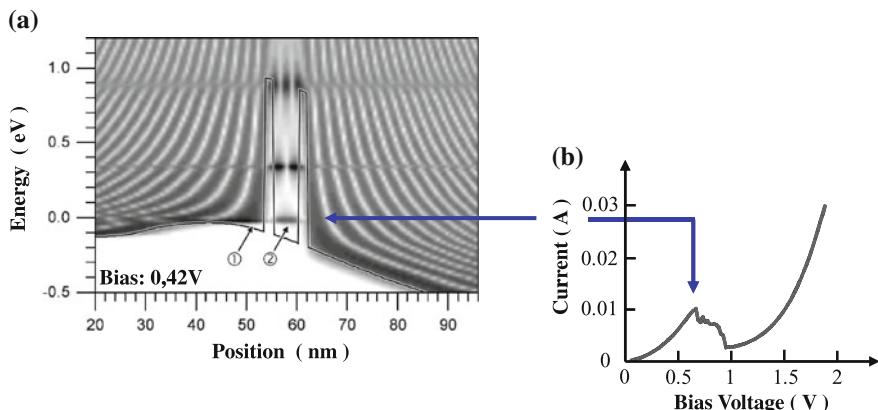
**Fig. 3.15** Transmission electron microscopic (TEM) image of a double barrier structure epitaxially grown by molecular beam epitaxy (MBE) (Appendix B). The two potential barriers are realized by thin AlAs layers imbedded in GaAs. The point pattern of the micrograph shows the atomic resolution of the TEM image. Single points represent atomic rows [8]





**Fig. 3.16** **a, b** Transmission  $T$  of a double barrier structure for resonance tunneling of an electron. **a** Double barrier with energy levels of quasi-bound electronic states. **b** Calculated transmission  $T$  as function of the kinetic energy of an incident electron. The energetic position of the transition maxima corresponds to the energies of quasi-bound states of the double barrier

of the double barrier (source) can be enhanced with respect to the energy on the other side (drain). The energy of electrons in the source region might, thus, be tuned to the lowest confined state in the well. Because of maximum transmission of electrons in this resonance case the electric current measured between two contacts on both sides of the double barrier layer stack reaches a pronounced maximum (Fig. 3.17). A layer stack as in Fig. 3.15 structured into a column laterally and supplied with two electric-



**Fig. 3.17** **a, b** Realistic quantum transport calculation of the measured current ( $I$ )–voltage ( $V$ ) characteristic of a resonance tunneling diode prepared from an AlAs/GaAs multiple-heterostructure [9]. **a** Calculated local electronic density of states (*half tone contures*) plotted together with the lower conduction band edge of the heterostructure (*solid line* showing the two barriers). The AlAs barriers confine three quasi-bound states. The strength of blackening of the half tone contures is a measure of the probability for finding the tunneling electron at that position. The diode is biased with 0.42 V. **b** Experimentally determined  $I$ – $V$  characteristic [8]. The current maximum at about 0.7 V arises from resonant tunneling of electrons through the lowest quasi-bound state in (a)

cal contacts on top and bottom (source and drain) is called *resonance tunneling diode* (RTD). The current–voltage ( $I$ – $V$ ) characteristics measured on such an RTD device is shown in Fig. 3.17b. After the current through the RTD has reached its maximum originating from electron tunneling through the lowest confined well state, it drops again with increasing applied voltage, because the confined state gets out of resonance with the kinetic energy of the incoming electrons (Fig. 3.17). For even higher voltages ( $U > 1$  V in Fig. 3.17b), the current increases again and eventually reaches a second maximum due to resonance with the second confined state in the well.

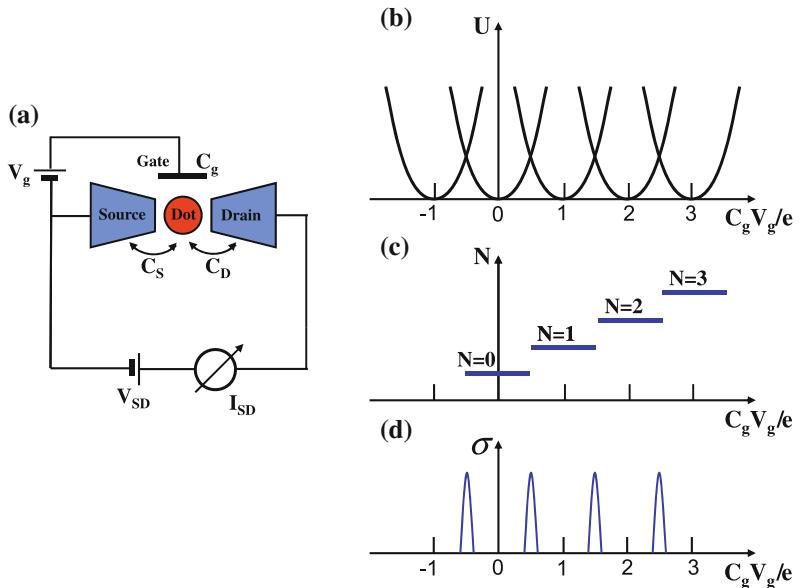
The described RTDs are quantum-electronic devices which exhibit their characteristic quantum behavior even at room temperature. Their complex non-linear  $I$ – $V$  characteristics (Fig. 3.17b) with *negative differential resistance* (NDR) enables interesting new functionalities in electronic circuits: RTDs are used in resonance circuits to generate RF radiation with frequencies up into the THz range. Furthermore, they offer the possibility for realization of novel digital logic circuits.

## 3.7 Single Electron Tunneling

At this point we leave, for a moment, the rigorous description of quantum mechanical systems on the basis of Schrödinger equation and wave function, in order to get familiar with a phenomenon which is of paramount importance for the experimental research in fields as fundamentals of quantum mechanics, quantum electronics and nano-physics. This phenomenon, called *single electron tunneling*, is related to tunneling of an electron through two sequential potential barriers, which form a type of quantum dot or box with linear dimensions of some ten nanometers in between (Fig. 3.18). Across the two barriers the quantum dot is electrically (by tunneling) connected to a left electrode, the *source* ( $S$ ) of electrons, and a right *drain* ( $D$ ) electrode, through which the electrons leave the arrangement (Fig. 3.18a). A third electrode, the *gate* ( $G$ ), is capacitively coupled through an insulating layer (barrier) to the quantum dot. A voltage ( $V_g$ ) applied between gate and source contact allows the variation of the dot potential with respect to that of the source.

There are several ways to realize such a structure experimentally. A resonance tunneling layer structure consisting of GaAs and two built-in AlAs barriers (Fig. 3.15) might be structured lithographically (Appendix B) into a column with lateral dimensions perpendicular to the layer sequence in the 100 nm range. Because of carrier depletion near the surface (Appendix A) electrons are confined laterally within a range of a couple of nanometers and perpendicularly by the two AlAs barriers along a similar spatial extension. Thus, electrons within this spatial region are confined in a quantum dot. As gate electrode an additional metal stripe might be prepared lithographically on the side surface of the column (see also Sect. 5.7.1, Fig. 5.18a).

A further common method to realize a source/quantum dot/drain system with additional gate electrode is based on a two-dimensional electron gas (2DEG) which can be prepared in a semiconductor heterostructure (Appendix A). Metallic electrodes deposited on the surface of the heterostructure, about 30 nm above the interface



**Fig. 3.18** **a–d** Scheme of single electron tunneling through a quantum dot. **a** Schematic measurement circuit for the study of single electron tunneling. By means of a gate electrode, the potential of the quantum dot can be shifted with respect to the source electrode. **b** Ground state energy  $U$  of the quantum dot as function of gate voltage  $V_g$  (plot normalized by  $C_g V_g / e$ ). Different parabolae correspond to different occupations of the dot with increasing number of electrons  $N = 0, 1, 2, \dots$ . **c** Occupation of the dot with  $N = 0, 1, 2, \dots$  electrons as function of gate voltage  $V_g$ . **d** Tunneling conductivity  $\sigma$  between source and drain contact as function of gate voltage  $V_g$ . These so-called Coulomb blockade oscillations exhibit maxima each time when the electronic occupation of the quantum dot changes

(2DEG), deplete the 2DEG from electrons, in particular, if biased with a negative voltage. The negative voltage pushes the electrons beneath away and creates insulating areas below the metal patterns (Appendix B). Appropriately shaped metal electrodes (patterns), thus, separate a conducting quantum dot area from source and drain as well as from gate areas in the 2DEG.

In each case there is a quantum dot (box), in which electrons are spatially confined and discrete energy levels are formed (Sect. 3.6.1). Depending on the size and on the nature of the material (semiconductor or metal) these discrete levels lie quite close to each other, quasi-continuously, or form a discrete level spectrum like in an atom. In contrast to natural atoms, however, the potential of the quantum dot can be varied with respect to its surrounding by changing the electrical bias between dot and source, respectively drain contact. By an appropriate bias, we can even force electrons to tunnel from the source contact into the quantum dot and to occupy one of the confined states there. Depending on barrier thickness and height as well as on bias conditions between the electrodes the electron can leave the quantum dot by tunneling into the

drain contact. This process is called *single electron tunneling*. We will encounter an interesting new phenomenon in this process, the so-called *Coulomb blockade*.

Imagine, some few electrons already occupy the energetically lowest confined states in the quantum dot. We will learn in Sect. 5.6.3 that a quantum state can be occupied by one single electron only (Pauli exclusion principle). The next electron which “wants” to tunnel into the quantum dot from the source contact, therefore, has to occupy the next higher unoccupied state in the dot. Furthermore, this new electron experiences the Coulomb repulsion from all other electrons already present in the dot. Coulomb repulsion is a many body effect, an interaction between two or more electrons, which was never considered so far in our single particle picture of quantum mechanics. All quantum states for spatially constricted volumes (potential well, quantum wire, quantum box, Sect. 3.6.1) and for propagating electrons (Sects. 3.6.2–3.6.5) were single electron states. Interactions with other electrons were neglected. Many body interactions as electron Coulomb repulsion require a much more involved theoretical formalism than it is possible at this stage of understanding of quantum physical phenomena. The reader should wait for Chap. 8 of this book to find a more rigorous treatment of those kind of problems.

We will therefore use a quasi-classical description for the treatment of Coulomb repulsion in single electron tunneling. In classical electrostatics the effect of an additional electron being added to a small body carrying some charge already is described in terms of charging. In the following treatment of the single electron tunneling effect, therefore, the addition of an electron to the quantum dot by tunneling through the source/dot barrier is referred to as charging the dot by one elementary charge. We only keep in mind that the quantum mechanical tunnel effect is responsible for the electron to switch over from the source electrode to the quantum dot (Fig. 3.18a).

For the classical description of the phenomena, we start with an electrically charged metal particle (dot) which is imbedded in a dielectric medium. With respect to infinity the particle has a capacity  $C = q/V$  with  $q$  as the charge on the particle and  $V$  a voltage referred to an infinitely distant electrode. When an additional charge  $dq$  is brought on the particle, its energy increases by

$$dE = V dq = q dq/C. \quad (3.124)$$

An increase of the total charge up to the value  $Q$  causes an energy increase to

$$E = \int_0^Q \frac{q dq}{C} = \frac{1}{2} \frac{Q^2}{C}. \quad (3.125)$$

Therefore, the electrostatic energy of the charged particle amounts to  $E = Q^2/2C$ .

In order to estimate the order of magnitude of the expected effects, we assume the quantum dot to be a little circular conducting disk with the radius  $r = 250$  nm at a distance  $a = 70$  nm from an extended planar metallic gate electrode. This quantum dot shall be imbedded in GaAs with a dielectric constant  $\varepsilon_r \approx 13$ . The capacity of the dot with respect to the planar gate electrode is

$$C = \varepsilon_r \varepsilon_0 \pi r^2 / a, \quad (3.126)$$

and the numerical value for the considered example follows as  $C \approx 10^{-16}$  F. According to (3.125) the addition of one electron to the dot gives rise to a charging energy  $E_C = e^2/2C$  which, in our example, amounts to about 1 meV. In order to resolve such effects spectroscopically one has to perform the experiment at low temperature  $T < 1$  K an observation of charging effects at room temperature requires much lower capacities (3.126), that is, the dimensions of the quantum dot must be in the order of nanometers.

In our classical description, the so-called *orthodox model* of single electron tunneling, the charge transported via source and drain through the quantum dot is assumed to be quantized in units of the electronic elementary charge. On the other hand, the action of the gate bias  $V_g$  is described in terms of a continuous charge  $C_g V_g$  which is induced on the dot through the gate capacity  $C_g$ . From (3.125), we then obtain the electrostatic energy (ground state energy) of the quantum dot as

$$U(N) = [|e|(N - N_0) + C_g V_g]^2 / 2C + \sum_n^N E_n. \quad (3.127a)$$

Here  $N$  is the number of electrons on the dot and  $N_0$  the electron number at vanishing gate voltage  $V_g = 0$ .  $C$  is the total capacity of the dot, namely that against gate ( $C_g$ ), against source ( $C_S$ ) and against drain ( $C_D$ ):

$$C = C_g + C_S + C_D. \quad (3.127b)$$

$E_n$  are the single electron energies of the confined states in the dot (solutions of Schrödinger equation, Sect. 3.6.1), which are filled with electrons up to the number  $N$ . The last term in (3.127a), thus, contains the sum of the single particle energies, while the first term describes the many-body interaction between the electrons in terms of a classical charging energy.

If  $N - 1$  electrons are present on a quantum dot, the addition of the  $N$ th electron requires an energy, the *electrochemical potential*  $\mu_{\text{QD}}(N)$  of the dot with the occupation number  $N$ . The thermodynamical potential  $\mu_{\text{QD}} = \partial U / \partial N$ , of course, contains per definition the many-body electron-electron interaction. In the present case we can write:

$$\begin{aligned} \mu_{\text{QP}}(N) &= U(N) - U(N - 1) \\ &= \frac{1}{2C} \left[ \{|e|(N - N_0) + C_g V_g\}^2 \right. \\ &\quad \left. - \{|e|(N - 1 - N_0) + C_g V_g\}^2 + E_N \right], \end{aligned} \quad (3.128a)$$

$E_N$  is the energy of the highest occupied electronic state in the dot. Summing up the brackets in (3.128a) yields for the electrochemical potential

$$\mu_{\text{QP}}(N) = \left( N - N_0 - \frac{1}{2} \right) \frac{e^2}{C} + |e| \frac{C_g}{C} V_g + E_N. \quad (3.128b)$$

The *addition energy* for adding one further electron to the dot, thus, follows as

$$\begin{aligned} \Delta\mu(N) &= \mu_{\text{QP}}(N+1) - \mu_{\text{QP}}(N) \\ &= U(N+1) - 2U(N) + U(N+1) \\ &= \frac{e^2}{C} + E_{N+1} - E_N = \frac{e^2}{C} + \Delta E. \end{aligned} \quad (3.129)$$

In atomic physics one would call  $A = U(N) - U(N+1)$  electron affinity and  $I = U(N-1) - U(N)$  ionization energy. The addition energy (3.129), then, equals the difference of  $I$  and  $A$ :  $\Delta\mu = I - A$ .

The addition energy  $\Delta\mu(N)$  contains the highest single electron energies, that of the highest occupied level  $E_N$  and that of the next higher empty state  $E_{N+1}$  (into which the additional electron goes) as well as the charging energy  $e^2/C$  arising from the Coulomb repulsion between the electrons already present on the dot and the new one being added (many body interaction).

Later we will see that for metallic quantum dots of sufficient size  $\Delta E$  is negligibly small as compared with the charging energy. For semiconductor quantum dots, however, the sum of single electron energies  $\sum_n^N E_n$  and, therefore, also  $\Delta E$  cannot be neglected and have to be taken into account for a quantitative description of the phenomena.

In the following we consider a metallic quantum dot with the simplifying assumption  $N_0 = 0$ , that is, the quantum dot does not carry any charge at vanishing gate voltage  $V_g = 0$ . In this case the ground state energy is given by

$$U(N, V_g) = \frac{1}{2C} (N|e| + C_g V_g)^2. \quad (3.130)$$

This energy plotted as a function of the gate voltage  $V_g$  (Fig. 3.18b) yields a series of parabolas, each one belonging to an occupation of the quantum dot with  $N = 1, N = 2, N = 3, \dots$  electrons. At the point  $C_g V_g = e/2$  the two parabolas belonging to  $N = 0$  and  $N = 1$  cross each other. For gate voltages exceeding this value, it is energetically more favorable that one additional elementary charge is added to the dot than if the dot would stay uncharged ( $N = 0$ , parabola centered around 0). Further increase of the gate voltage correspondingly increases the occupation of the quantum dot (Fig. 3.18c). The dot changes its occupation with electrons at each crossing point of the parabolas, while in the range in between the occupation is stable with  $N = 0, N = 1, N = 2, \dots$  electrons on the dot. In the stable regime between the crossing points no current can flow, electron transport is blocked. Here the condition

$$U(N \pm 1, V_g) - U(N, V_g) \leq 0, \quad (3.131a)$$

holds, that is, according to Fig. 3.18c

$$\left(N - \frac{1}{2}\right) < C_g V_g / e < \left(N + \frac{1}{2}\right). \quad (3.131b)$$

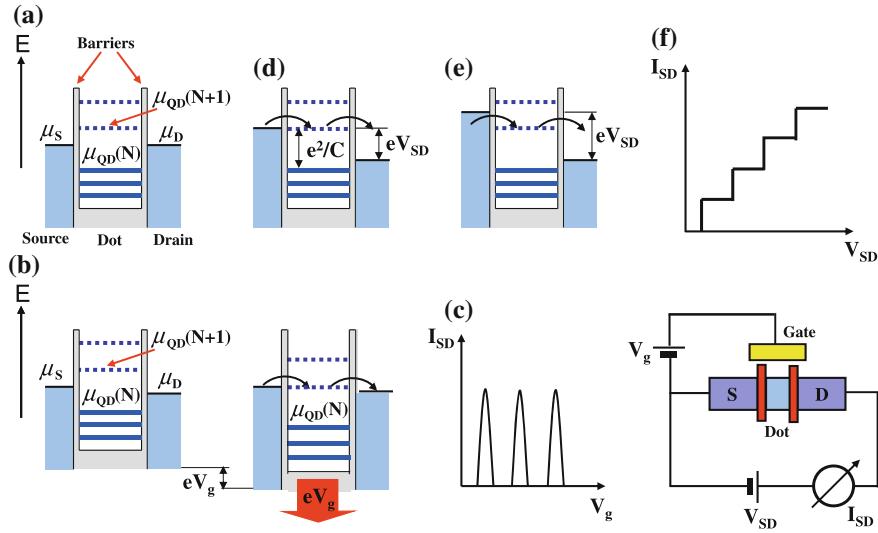
This phenomenon of a blocked current flow through the source/quantum dot/drain arrangement because of a stable occupation of the dot is called *Coulomb blockade*.

At the crossing points of the parabolas, we have

$$U(N \pm 1) = U(N). \quad (3.132)$$

At this particular gate voltage the system does not show any preference for an occupation of the quantum dot with  $N$  or  $N + 1$  electrons. The number of electrons on the dot can fluctuate, electrons can tunnel between source contact, quantum dot and drain contact and a current flows through the arrangement. Its electrical conductance  $\sigma$  as a function of the gate voltage or of  $C_g V_g / e$  exhibits sharp peaks at the crossing points of the parabolas (3.127a), so-called *Coulomb blockade oscillations* (Fig. 3.18d).

The single electron tunneling phenomenon might also be explained in an energy versus spatial position plot, in analogy to the one used for resonant tunneling (Figs. 3.16 and 3.17). In Fig. 3.19, characteristic energies of the system are plotted on an energy scale in the spatial regions source ( $S$ ), quantum dot (QD) and drain ( $D$ ). The three regions are separated by insulating potential barriers, which do not have any electronic states in the energy range considered. The barriers are sufficiently thin that the wave functions in  $S$ ,  $D$  and QD have a significant overlap, which allows electron tunneling between the three regions. In source and drain metallic character is assumed, which is described by the model of the potential box (Sect. 3.6.1): The electronic states lie quasi-continuously dense on the energy scale and are filled, each state by one electron (according to Pauli principle, Sect. 5.6), up to maximum energies  $\mu_S$  and  $\mu_D$ , the chemical potentials of source and drain (blue areas). In the quantum dot, we assume discrete energy levels because of the confinement of the electrons (Fig. 3.19). We take into account the many-body effect of Coulomb repulsion (charging effect) by assuming that addition of one further electron to the dot requires the addition energy  $e^2/C$ .  $\mu_{\text{QD}}(N + 1)$ , here, denotes the many-body energy (chemical potential) of the quantum dot at which the increase of the electron occupation from  $N$  to  $N + 1$  becomes possible. When the chemical potential  $\mu_S$  of the source contact lies below  $\mu_{\text{QD}}(N + 1)$ , no additional electron, exceeding the  $N$  electrons already on the dot, can tunnel into the dot. This situation of the Coulomb blockade is shown in Fig. 3.19a. By application of bias voltages between the S, QD and D contacts the energies, that is, also the chemical potentials in source, quantum dot and drain can be shifted against each other. In Fig. 3.19b an appropriate gate voltage  $V_g$  shifts the dot potential with respect to  $\mu_S$  and  $\mu_D$  such that  $\mu_{\text{QD}}(N + 1)$  equals the chemical potential in the source contact (only slightly higher than in drain). The quantum dot can change its electron occupation by tunneling from source into drain. A tiny source-drain voltage  $V_{SD}$ , i.e. a minimum difference  $\mu_S - \mu_D > 0$ , causes a tunneling current through the quantum dot. The Coulomb blockade is lifted.



**Fig. 3.19 a–f** Explanation of single electron tunneling measured by means of a circuit consisting of source contact ( $S$ ), quantum dot, drain contact ( $D$ ) and gate (right below). **a** Potential well scheme (potential energy versus position coordinate) for source, drain and dot at negligibly small bias between source and drain.  $\mu_S$  and  $\mu_D$  are the chemical potentials of source and drain;  $\mu_{QD}(N)$  and  $\mu_{QD}(N+1)$  are the chemical potentials of the quantum dot being occupied by  $N$ , respectively ( $N+1$ ) electrons (Coulomb blockade). **b** Potential well scheme under a gate bias  $V_g$  which lowers the dot potential  $\mu_{QD}(N+1)$  to allow tunneling of an electron through the dot (removal of Coulomb blockade). **c** Source-drain current  $I_{SD}$  as function of gate voltage  $V_g$ . The sharp bands of the Coulomb blockade oscillations indicate tunneling processes as shown in (b). **d, e** Possible tunneling current transport by applying a convenient source-drain voltage  $V_{SD}$ . **f** Coulomb blockade staircase characteristics  $I_{SD}(V_{SD})$  as it is observed for tunneling processes as in (d) and (e)

This situation occurs periodically upon increasing the gate voltage (Fig. 3.18b). This explains the observation of Coulomb blockade oscillations (Fig. 3.18c).

The Coulomb blockade can also be lifted with constant potential at the quantum dot (fixed gate bias) by increasing the source-drain voltage  $V_{SD}$  (Fig. 3.19d). Because of the voltage drop across the two barriers the chemical potentials  $\mu_{QD}(N)$  and  $\mu_{QD}(N+1)$  are shifted on the energy scale with respect to the potentials  $\mu_S$  and  $\mu_D$ , such that electron tunneling from S via QD into D becomes possible. A tunneling current sets in, when  $\mu_{QD}(N+1)$  reaches the energetic level of  $\mu_S$  (Fig. 3.19d). Further lowering of  $\mu_{QD}(N+1)$  by an increased voltage  $V_{SD}$  keeps the tunneling current flowing, since occupied states in the source region still lie on the level of  $\mu_{QD}(N+1)$  (Fig. 3.19e). When the voltage  $V_{SD}$  is even further enhanced, the quantum dot potential drops further and eventually the chemical potential  $\mu_{QD}(N+2)$  for adding a second electron to the dot reaches the source potential  $\mu_S$ . A further electron tunnels into the dot and the tunneling current increases by a corresponding step. The source-drain current/voltage characteristics  $I_{SD}(V_{SD})$ , thus becomes a stair case function, called *Coulomb blockade staircase* (Fig. 3.19f).

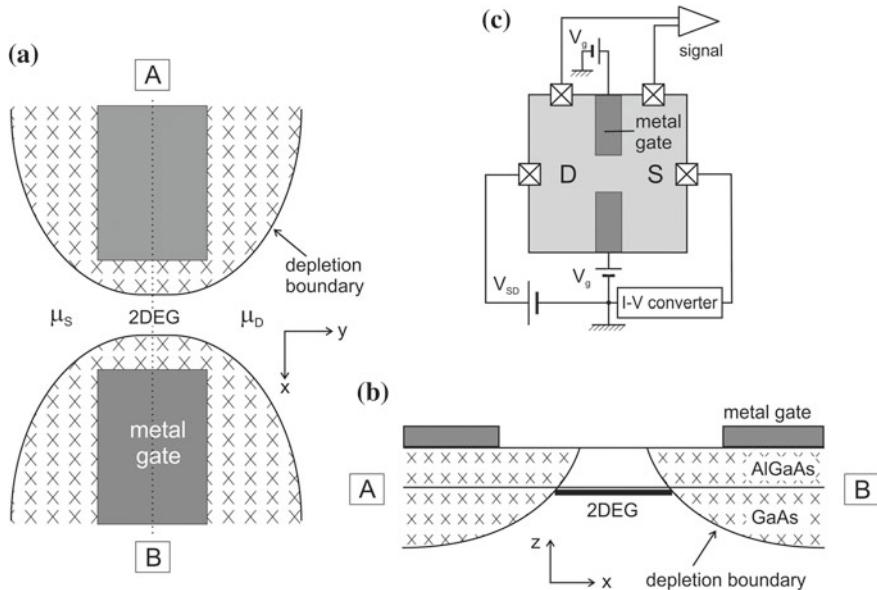
Finally we want to emphasize again that the energy levels plotted in Fig. 3.19 as  $\mu_{\text{QD}}(N)$ ,  $\mu_{\text{QD}}(N + 1)$  etc. are many-body energies in the sense of the chemical potential rather than single electron energies as discussed in connection with resonant tunneling (Sect. 3.6.5).

### 3.8 The Quantum Point Contact as Charge Detector

Another nanostructure device, the so-called *quantum point contact* (QPC), has gained significant importance in fundamental experiments, where the presence or movement of single electronic charges is registered in nanoscaled electronic circuits. The device essentially consists of a short highly conductive channel in which electrons are confined in two dimensions perpendicularly to their propagation between source (S) and drain (D) contacts having chemical potentials  $\mu_S$  and  $\mu_D$ . External gates along the channel allow the induction of charge nearby and thereby widening or tightening of the channel cross section depending on the sign of the induced charge. Depending on the channel width more or less half electron wavelengths can match the channel diameter. The number of electron wave modes, which can pass the channel, can, thus, be controlled by the external gate bias.

The standard and frequently used experimental realisation of the QPC device is based on the two-dimensional electron gas (2DEG) at the AlGaAs/GaAs heterostructure interface (Appendix A and Fig. 3.20). Two metal gate fingers evaporated on top of the heterostructure produce depletion zones (Schottky contacts) beneath, where the free electrons of the underlying 2DEG are removed except for a narrow stripe between the metal fingers (Figs. 3.20a, b). This remaining area of high conductivity, where the electrons have not been removed from the 2DEG, forms the channel in which the electron waves are confined in x direction normal to their propagation direction y. An electron current from the S to the D contact through the channel is induced by the potential difference ( $\mu_S - \mu_D$ ) being established by a convenient bias voltage between S and D (Fig. 3.20c). A varying gate voltage  $V_g$  applied between the metal gate fingers and the ground contact modifies, by charge induction, the extension of the depletion space charge zones below the split metal gate contacts and thus the width of the channel. Because of the geometrical arrangement of the two metal contacts the described device is often called *split gate contact* (SGC). The name quantum point contact (QPC) is used because of historical reasons. The described effect of quantum transport through quantized electronic modes was originally also observed in point-like contacts between the metallic tip of a scanning tunnelling microscope (Sect. 3.6.4) and a conducting surface.

Because of the channel length in the 100 nm range and the high electronic mobility in the 2DEG (Appendix A) the electronic transport through the channel is ballistic at low temperature. Even elastic scattering on crystal defects can largely be excluded on these distances. For the calculation of the electron current through the channel we thus need to know only the electronic states, respectively their density, and the corresponding group velocity of electrons in these states. The origin of the current is

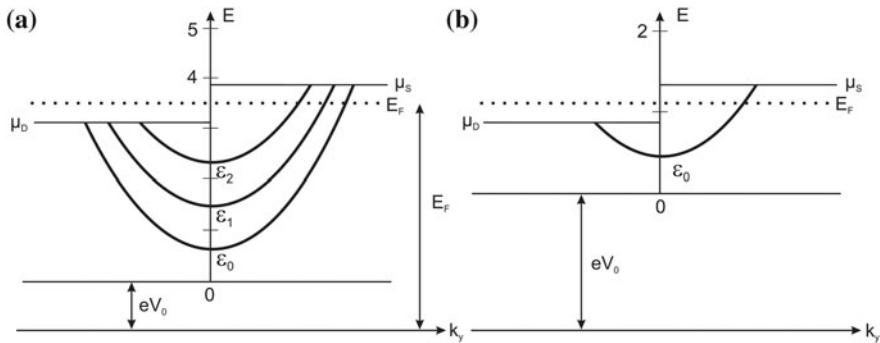


**Fig. 3.20** Quantum point contact (QPC) in split gate technology. **a** Schematic top view on a semiconductor heterostructure (e.g. GaAs/AlGaAs) containing a high mobility 2D electron gas (2DEG) at the interface (Appendix A). Two (split gate) metal electrodes (dark grey) produce depletion space charge layers (Schottky barriers) *below* the surface where the free electrons are removed from the 2DEG. Between the two depletion boundaries a narrow conductive channel between source (S) and drain (D) remains in the 2DEG.  $\mu_S$  and  $\mu_D$  are the chemical potentials in the S and D region. **b** Cross section view of the QPC along the line A–B as indicated in (a). The 2DEG quasi-1D channel is indicated between the depletion zone boundaries. **c** Electric circuit for measuring the conductance of the QPC

due to an imbalance in the occupation of states between the S and the D side (lower occupation) of the channel (Fig. 3.21). Along the channel in  $y$  direction electrons can freely move with a kinetic energy  $\hbar^2 k_y^2 / 2m$ . Perpendicularly, they are confined by the channel width and by the  $z$  extension of the 2DEG. The electrons, thus assume quantized discrete energy levels  $\varepsilon_i$  (Sect. 3.6.1). The total energy, therefore, is obtained as

$$E = \frac{\hbar^2 k_y^2}{2m} + \varepsilon_i + eV_0 \quad . \quad (3.133)$$

$eV_0$  is the electrostatic energy in the channel determined by the external bias at the gate. In good approximation the potential valley between the two opposite depletion space charge layers below the metal gate fingers can be described by a parabola  $\Phi \propto (x - x_0)^2$  with  $x_0$  as the centre of the channel (see Sect. 6.1.1). As will be shown for the parabolic oscillator potential in Sect. 4.4.2 the energy eigenvalues  $\varepsilon_i$  form a ladder of equidistant levels with a groundstate  $\varepsilon_0$  (Fig. 3.21), the mutual distance of which ( $\varepsilon_{i+1} - \varepsilon_i$ ) decreases with increasing width  $w$  of the channel (Sect. 3.6.1). For



**Fig. 3.21** Qualitative plot of the electronic band structure in the QPC of Fig. 3.20. The parabolic sub-bands belonging to the lateral quantisation energies  $\varepsilon_i$  (due to  $x$  and  $z$  confinement) are filled with electrons (two on each single electron level because of Pauli exclusion principle, Sect. 5.6.2) up to different energies  $\mu_S$  at the source and  $\mu_D$  at the drain region (because of SD bias). Electrons can freely move along  $y$  with a momentum  $k_y$ . In thermal equilibrium (zero external bias:  $\mu_S = \mu_D$ ) the sub-bands are filled up to the energy  $E_F$  (Fermi energy, Sect. 5.6.3) **a** Situation for a wide channel. **b** Narrow channel with stronger  $x$  confinement

the simplest assumption of a rectangular confining potential between the split gate we conclude a dependence  $(\varepsilon_{i+1} - \varepsilon_i) \propto 1/w^2$  of the energy eigenvalues on the channel width  $w$ , i.e. also a decrease of the energetic level distance with increasing width  $w$  (Sect. 3.6.1). Plotted along  $k_y$ , the transport wave vector, the electron energies (3.133) form a sequence of parabolas, whose mutual energetic distance decreases with increasing channel width (Fig. 3.21). Details of the  $(\varepsilon_{i+1} - \varepsilon_i)$  dependence, the mutual distance of the parabolas on the channel width, depend on the shape of the confining potential below the split gate. Each parabola of (3.133) in Fig. 3.21 is called a *subband* for electron transport. The subbands are numerated by the index  $i$ , the number of the particular discrete level  $\varepsilon_i$ .

We can calculate the current  $I_i$  through the channel for each subband separately by summing up the product of electronic charge determined by the density of states  $D_i(E)$  within this subband and the electron velocity  $v_i(E)$ . Only those states of the subband contribute which lie energetically between the source and drain potentials  $\mu_S$  and  $\mu_D$  (Fig. 3.21). For all other states current contributions from left to right and vice versa cancel each other because of equal charge values having opposite velocities. For one subband the current, then, follows as:

$$I_i = e \int_{\mu_D}^{\mu_S} D_i(E) v_i(E) dE . \quad (3.134)$$

The conducting channel in the present case is a quantum wire because of confinement in two directions along  $x$  and  $z$ . According to (3.72) the density of states of a quantum wire is  $(dk/dE)/2\pi$ . Because of spin degeneracy each state can be occupied by two electrons (Sect. 5.6) and we obtain for the  $i$ th subband:

$$D_i(E) = \frac{1}{\pi} \frac{dk_y}{dE} = \frac{1}{\pi} \left( \frac{dE}{dk_y} \right)^{-1}. \quad (3.135)$$

According to Sect. 3.2 (3.17) the velocity of an electron in the  $i$ th subband is

$$v_i(E) = \frac{d\omega}{dk_y} = \frac{1}{\hbar} \frac{dE}{dk_y}. \quad (3.136)$$

The product  $D_i(E)v_i(E)$  in the integral (3.134) is a constant and we obtain the current carried by the  $i$ th subband as

$$I_i = \frac{2e}{h} (\mu_S - \mu_D) = \frac{2e^2}{h} V_{SD}, \quad (3.137a)$$

where  $V_{SD}$  is the source drain voltage. Usually a finite number of subbands is occupied and contributes to the total current:

$$I = \sum_{i=1}^{i_{\max}} \frac{2e}{h} (\mu_S - \mu_D). \quad (3.137b)$$

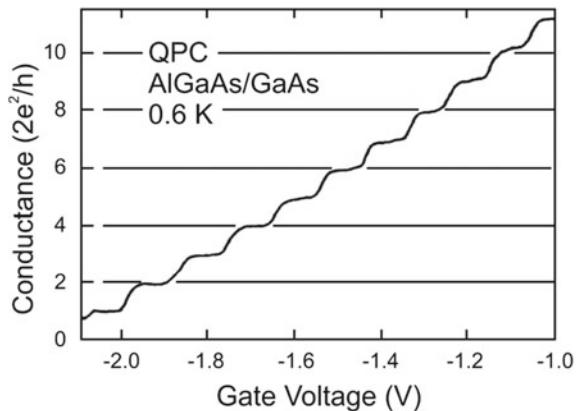
$i_{\max}$  denotes the highest occupied subband (Fig. 3.21). From (3.137) we obtain for the conductance of the QPC:

$$G_{QPC} = \frac{I}{V_{SD}} = \sum_{i=1}^{i_{\max}} \frac{2e^2}{h}. \quad (3.138)$$

(3.138) describes a stepwise increase of the conductance of the QPC with decreasing negative gate voltage. Each time, when a new subband becomes occupied with further widening of the channel, the conductance jumps by an amount of  $(2e^2/h)$ . This is the so-called *conductance quantum*. It is noteworthy, that in spite of ballistic transport without scattering of the carriers within the QPC channel the device has a finite conductance. This phenomenon is due to the quantum nature of the transport, which is reflected by the conductance quantum containing only the elementary charge and Planck's constant  $h$ . Sometimes also half the value  $e^2/h$  is called conductance quantum, depending whether spin degeneracy is taken into account or not.

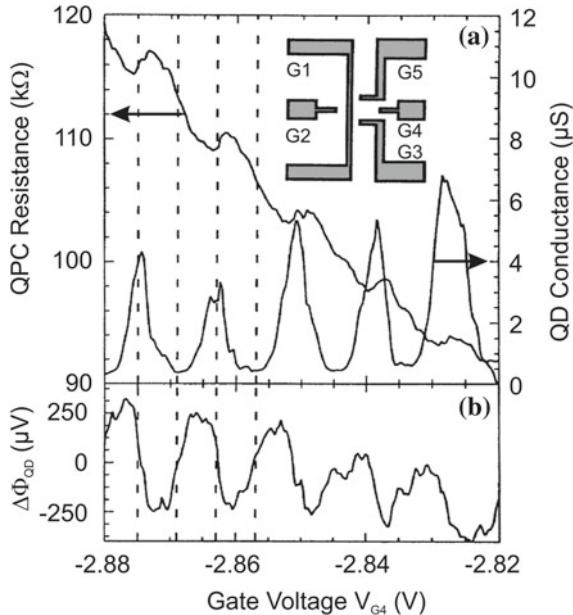
The stepwise increase of the QPC conductance by jumps of  $(2e^2/h)$  has been observed for the first time in a well defined manner on a split gate contact (QPC) prepared on an AlGaAs/GaAs heterostructure [10] just as it has been explained in Fig. 3.20. In Fig. 3.22 the original data measured at a temperature of 0.6 K are shown. The conductance increases with decreasing negative gate voltage, which corresponds to less positive induced charge in the gate depletion space charge layer. Hereby the conducting channel is widened and more and more subbands contribute to the quantum transport (Fig. 3.21).

**Fig. 3.22** Electrical conductance of quantum point contact in units of the conductance quantum  $2e^2/h$  measured at 0.6 K as a function of the gate voltage  $V_g$  (see Fig. 3.20c) [10]



Apart from the interesting fundamental aspects of quantum transport through the QPC the device is also useful for measuring the presence or transport of single charges in nanoelectronics. Single charges in close vicinity to the QPC channel modify the potential locally; hereby the channel width is varied and the QPC changes its conductance. The effect is particularly strong, when the QPC gate voltage is tuned into a region of steep conductance slope, i.e. into the jump region between two plateaus. As an example Fig. 3.23 shows data of a measurement of charging and discharging of a quantum dot (QD) [11]. The inset in Fig. 3.23 qualitatively depicts the arrangement of a QD in close vicinity to a QPC, which is the measurement probe for detecting charge changes on the QD. In the inset metal gate electrodes on top of a 2DEG within an AlGaAs/GaAs heterostructure are plotted. These gate electrodes produce depletion zones beneath in the 2DEG and define energy barriers for electrons, which can freely move in the areas of the 2DEG between the electrodes. The extension of the barriers can be modified by external voltages applied to the gates. In the described way the gates G1 and G2 define the probe QPC, which is separated from the QD on the right side by the bar opposite G2. The QD is defined by this bar and the gate electrodes G5, G3 and G4. The gate G4 defining the right-hand edge of the QD acts as a “plunger”. A voltage applied to G4 varies the potential of the dot. When the voltage on G4 is swept, Coulomb blockade oscillations (Sect. 3.7) are seen in the conductance of the QD. For the measurement of charge changes on the QD the conductance of the QPC channel between G1 and G2 is set at a value where it reacts most sensitively to the surrounding potential, i.e. at the jump region between two conductance plateaus (Fig. 3.22).

In Fig. 3.23a the conductance through the QD (right scale) and the resistance of the split gate QPC detector (left scale, curve with small dips on a rising background) are shown as the plunger voltage at G4 is swept. The little dips on the rising detector resistance ( $R_{QPC}$ ) curve directly correlate with the Coulomb blockade oscillations measured as conductance changes  $G_{QD}$  on the QD. The QPC obviously reacts sensitively to charge changes on the QD. While the Coulomb blockade peaks in Fig. 3.23a



**Fig. 3.23** Quantum point contact (QPC) as detector for charge on a quantum dot (QD) realized in split gate technology on a GaAs/AlGaAs heterostructure [11]. **a** Coulomb blockade oscillations versus gate voltage ( $V_{G4}$ ) through the QD (scale on right side) in comparison with the resistance of the QPC (scale on left side). Inset Metal gate arrangement for defining the QPC (gates G1, G2) and the QD (gates G3, G4, G5) within the 2D electron gas (2DEG) below the surface at the GaAs/AlGaAs interface. **b** Change in QD potential  $\Delta\Phi_{QD}$  versus gate voltage ( $V_{G4}$ ) as calculated from the QPC detector resistance

(right scale) indicate transmission of the QD, i.e. those plunger voltages, where the QD changes its charge (Sect. 3.7), the dips in the QPC resistance  $R_{QPC}$  shifted somewhat on the voltage scale are related to the presence of the particular charge on the QD. This explains the somewhat different plunger gate voltages, at which the peak structures appear. The detector QPC can be calibrated by removing the plunger bias (G4) and applying a voltage to the 2DEG region between the tunnel barriers of the quantum dot region. The measured detector channel resistance  $R_{QPC}$  can thus be transformed into the QD potential  $\Delta\Phi_{QD}$  shown in Fig. 3.23b. This electrostatic potential of the dot is oscillatory with an average amplitude of  $500 \pm 100 \mu\text{V}$ , which is the charging energy of the dot (Sect. 3.7). As expected the dot potential  $\Delta\Phi_{QD}$  has its maxima and minima at plunger voltages between those, where the Coulomb blockade peaks appear, i.e. where the charge on the QD jumps.

The described method for sensitively measuring single charges in nanoelectronic circuits is based on capacitive coupling of the measurement probe (QPC) to the sample under study (QD) rather than on a direct intervention into the charge transport itself. The technique is less disturbing to the measurement process.

## References

1. W. Heisenberg, Z. Phys. **43**, 172 (1927)
2. E. Schrödinger, Ann. Phys. **79**, 361 (1926)
3. G. Gamow, Z. Phys. **51**, 204 (1928)
4. G. Binnig, H. Rohrer, Ch. Gerber, W. Weibel, Appl. Phys. Lett. **40**, 178 (1982)
5. R. Butz, (Research Center Jülich): Private Commun. (1997)
6. A. Förster, J. Lange, D. Gerthsen, C. Dieker, H. Lüth, J. Phys. D.: Appl. Phys. **27**, 127 (1994)
7. A. Förster, Private Commun. (1994)
8. K. Maezawa, A. Förster: Quantum transport devices based on resonant tunneling, in *Nanoelectronics and Information Technology*, ed. by R. Waser (Wiley-VCH, Weinheim 2003), p. 407
9. M. Indlekofer, J. Malindretos, Free download of WinGreen simulation package. <http://www.fz-juelich.de/isi/mbe/wingreen.html>
10. B.J. van Wees, H. Van Houten, C.W.J. Beenacker, J.W. Williamson, L.P. Kouwenhoven, D. Van der Marel, C.T. Foxon, Phys. Rev. Lett. **60**, 848 (1988)
11. M. Field, C.G. Smith, M. Pepper, D.A. Ritchie, J.E.F. Frost, G.A.C. Jones, D.G. Hasko, Phys. Rev. Lett. **70**, 1311 (1993)

# Chapter 4

## Quantum States in Hilbert Space

In the preceding chapters, we have learnt that on the atomic and nanoscopic scale the world is ruled by laws which are different from those which we are familiar with from our every day life experience. The propagation of a particle can no longer be described in terms of trajectories with well defined position and velocity. The uncertainty principle does exist for position and momentum observables. Both quantities can be measured simultaneously only as statistical expectation values. On the atomic and subatomic level physical phenomena are fundamentally random and non-deterministic. By means of wave function and Schrödinger equation, however, we were able to describe the statistical behavior of nature, characterized by the particle-wave duality, in a well defined and rigorous mathematical formalism. The Schrödinger equation, together with boundary conditions, determines unequivocally the wave function, the basis for the statistical description. Hereby, sometimes the term of a quantum state has sneaked into the discussion in a blurred way. Because of the lack of fixed and definite numbers for position and momentum, we have used this term of a quantum state to describe the behavior of an electron, or more precisely, of an ensemble of electrons in terms of a wave function and/or its energy (eigenvalue). In this chapter, we will formulate the term *quantum state* of a system more precisely. For this purpose, generalized vectors in a so-called *Hilbert space* (David Hilbert 1862–1943, famous German mathematician) are used. At first sight, this approach might look quite complex and abstract. But the handling of the mathematical formalism will appear as quite easy and straightforward, if we make use of the analogy with the 3D Euclidean space of our familiar imagination.

### 4.1 Eigenvectors and Measurement of Observables

We start with the solutions of the Schrödinger equation for an electron in a 1D quantum well (Sect. 3.6.1). The eigenvalue problem (3.60) solved with fixed boundary conditions yields

$$\varphi_n(x) = C \sin k_x x. \quad (4.1)$$

Because of the boundary conditions (3.63),  $k_x$  is quantized with values  $k_x = n\pi/L$  where the quantum number  $n$  has integer values  $n = 1, 2, 3, \dots$ . The constant  $C$  is determined by the normalization condition (3.7) required for the wave function over the considered volume, in the present case, the length  $L$ :

$$\int_0^L \varphi_n^*(x) \varphi_n(x) dx = C^2 \int_0^L \sin^2\left(n \frac{\pi}{L} x\right) dx = 1 \quad (4.2)$$

that is, the normalization constant follows as  $C = (2/L)^{-1/2}$  and the normalized eigenfunction (4.1) is

$$\varphi_n(x) = \sqrt{\frac{2}{L}} \sin n \frac{\pi}{L} x. \quad (4.3)$$

If we consider an integral similar to the normalization condition (4.2), but now with two different eigenfunctions (wave functions)  $\varphi_m$  and  $\varphi_n$  ( $m \neq n$ ), we obtain

$$\begin{aligned} \int_0^L \varphi_m^*(x) \varphi_n(x) dx &= \frac{2}{L} \int_0^L \sin\left(\frac{m\pi}{L} x\right) \sin\left(\frac{n\pi}{L} x\right) dx \\ &= \frac{2}{L} \left[ \frac{\sin(m-n)\pi}{2(m-n)\pi/L} - \frac{\sin(m+n)\pi}{2(m+n)\pi/L} \right] = 0. \end{aligned} \quad (4.4)$$

Eigenfunctions  $\varphi_n(x)$  being solutions to the Schrödinger equation (3.60) for an electron in the potential box, therefore, obey the relation

$$\int_0^L \varphi_m^*(x) \varphi_n(x) dx = \delta_{mn} = \begin{cases} 1 & \text{for } m = n, \\ 0 & \text{for } m \neq n, \end{cases} \quad (4.5)$$

with  $\delta_{mn}$  as the so-called *Kronecker* symbol.

Equal relations are derived for the eigensolutions of the Schrödinger equation (3.60) for an electron in the quantum box with periodic boundary conditions (3.68). Using the general solution  $\varphi_k = C \exp(i k_x x)$  with the  $k_x$  quantization  $k_x = n2\pi/L$  we obtain the following relation

$$\int_0^L \varphi_m^*(x) \varphi_n(x) dx = C^2 \int_0^L dx e^{i(n-m)\frac{2\pi}{L}x}. \quad (4.6a)$$

For  $n = m$  the normalization constant follows as  $C = 1/\sqrt{L}$ . For  $n \neq m$  (4.6a) yields

$$\begin{aligned} \int_0^L \varphi_m^*(x) \varphi_n(x) dx &= \frac{1}{L} \int_0^L dx e^{i(n-m)\frac{2\pi}{L}x} \\ &= \frac{1}{L} \frac{L}{2\pi(n-m)} [e^{i(n-m)\frac{2\pi}{L}L} - 1] = 0. \end{aligned} \quad (4.6b)$$

Thus, also in the case of periodic boundary conditions the general relation (4.5) is valid for the set of eigensolutions  $\varphi_n(x)$  of the Schrödinger equation.

Equation (4.5) is the starting point for an important generalization of the interpretation of the system of eigensolutions of the Schrödinger eigenvalue equation (3.60) and the Hamilton operator  $\hat{H}$ , respectively.

We remember the properties of vectors in the familiar three-dimensional (3D) space. There, the scalar product of two vectors  $\mathbf{a}$  and  $\mathbf{b}$  is defined as

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^3 a_i b_i. \quad (4.7a)$$

If  $\mathbf{a}$  and  $\mathbf{b}$  are mutually normal, the expression (4.7a) vanishes. For  $\mathbf{a}$  and  $\mathbf{b}$  being unity vectors (length one) the expression (4.7a) becomes unity, if the two vectors are parallel to each other. For unity vectors  $|\mathbf{a}| = |\mathbf{b}| = 1$ , we thus have

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^3 a_i b_i = \begin{cases} 1 & \text{if } \mathbf{a} \parallel \mathbf{b}, \\ 0 & \text{if } \mathbf{a} \perp \mathbf{b}. \end{cases} \quad (4.7b)$$

Extending this relation to infinitely many dimensions  $x$  (instead of  $i = 1, 2, 3$ ) with  $x$  being a continuous variable (densely lying  $x$  values) the analogy between the sum in (4.7a), (4.7b) and the integral in (4.5) is obvious. The continuous  $x$  values in (4.5) correspond to the discrete vector components  $i = 1, 2, 3$  in (4.7a), (4.7b). In an abstract sense, thus, the integral (4.5) can be considered as a generalized scalar product of two vectors  $\varphi_m^*(x)$  and  $\varphi_n(x)$  being defined in a space with infinitely many, continuously lying dimensions. It is important to emphasize that this continuous vector space is a so-called *dual vector space*, since the vectors  $\varphi_n(x)$  have complex values, that is, a real and an imaginary part, in general. The scalar product, then, has to be calculated from  $\varphi_m^* \varphi_n$ , similarly as the product of two complex numbers  $a^* a$  which yields the modulus  $|a| = (a^* a)^{1/2}$  of the complex number.

In analogy to the definition in 3D space (4.7a), (4.7b), we can consider the solutions  $\varphi_n(x)$ ,  $\varphi_m(x)$  of the Schrödinger eigenvalue equation (3.60) as a system of generalized normalized basis vectors of infinite dimension in an infinite continuous space. This space is called *Hilbert space*. All eigenfunctions of an operator (quantum mechanical observable) span an infinite vector space with densely lying coordinates in which every physically reasonable wave function (abstract vector) can be represented. This is completely analogous to the representation of a general 3D vector in our familiar 3D Euclidean space. We will see later, that there exist also

Hilbert spaces with a finite number of dimensions, even two only, in which specific quantum states of a system are represented.

The analogy between 3D Euclidean and Hilbert space leads to further statements: A general 3D vector  $\mathbf{r}$  can be represented by three orthogonal unit vectors  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  as

$$\mathbf{r} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3 = \sum_{i=1}^3 \alpha_i \mathbf{a}_i. \quad (4.8a)$$

The vector component  $\alpha_j$  in  $j$  direction is obtained, hereby, as

$$\mathbf{a}_j \cdot \mathbf{r} = \sum_{i=1}^3 \alpha_i (\mathbf{a}_j \cdot \mathbf{a}_i) = \sum_{i=1}^3 \alpha_i \delta_{ij} = \alpha_j. \quad (4.8b)$$

Correspondingly a general state vector in the Hilbert space, a general wave function  $\psi(\mathbf{r})$ , can be represented as a linear superposition of orthonormal eigensolutions  $\varphi_n(\mathbf{r})$ , that is, it can be expanded in terms of a series of eigenfunctions:

$$\psi(\mathbf{r}) = \sum_n b_n \varphi_n(\mathbf{r}). \quad (4.9)$$

Hereby, the functions  $\varphi_n(\mathbf{r})$  are the analogue to the three basis vectors in  $x, y, z$  direction of the 3D Euclidean space and  $b_n$  the analogue to the vector components in these directions. In analogy to (4.8b), the vector components in Hilbert space follow as:

$$\begin{aligned} b_m &= \int \varphi_m^*(\mathbf{r}) \psi(\mathbf{r}) d^3r = \sum_n b_n \int \varphi_m^*(\mathbf{r}) \varphi_n(\mathbf{r}) d^3r \\ &= \sum_n b_n \delta_{mn} = b_m. \end{aligned} \quad (4.10)$$

In (4.10), we have used the orthonormality relation (4.5) for the system of eigenfunctions  $\varphi_n(\mathbf{r})$ .

What we have learnt so far for the special case of solutions of the Schrödinger equation for an electron in the potential box (eigenvalue equation:  $\hat{H}\varphi_n = E_n \varphi_n$ ) can be generalized (as mathematicians prove) to the case of a general, physically proper (reasonable) operator  $\hat{\Omega}$  attributed to an observable  $\Omega$ .

A physically proper operator  $\hat{\Omega}$ , thus, must have a system of orthogonal and normalized (orthonormal) eigenfunctions  $\varphi_n(\mathbf{r})$  which solve the eigenvalue equation

$$\hat{\Omega} \varphi_n(\mathbf{r}) = \omega_n \varphi_n(\mathbf{r}). \quad (4.11)$$

According to Sect. 3.5, the eigenvalues  $\omega_n$  are, as measurement values, the possible numerical results of a measurement of the observable  $\Omega$ . Apart from having an

orthonormal eigenfunction system, a physically reasonable operator, thus, must have real eigenvalues  $\omega_n$ . Only real numbers can be the results of a measurement.

In order to guarantee both requirements for a physically reasonable operator, this operator must fulfill the condition

$$\int d^3r \varphi^*(\mathbf{r}) \hat{\Omega} \psi(\mathbf{r}) = \int d^3r (\hat{\Omega} \varphi)^* \psi = \int d^3r \varphi \hat{\Omega}^* \psi^*. \quad (4.12a)$$

Hereby  $\hat{\Omega}^+$  is called the adjoint operator to  $\hat{\Omega}$  if the following relation holds:

$$\int d^3r (\hat{\Omega}^+ \varphi)^* \psi = \int d^3r \varphi^* \hat{\Omega} \psi. \quad (4.12b)$$

The adjoint is to an operator what the complex conjugate is to numbers. An operator which fulfills (4.12a) is called *Hermitian operator* (or *self-adjoint operator*). If we consider for such an operator beside

$$\int d^3r \varphi_n^* \hat{\Omega} \varphi_n = \omega_n \int \varphi_n^* \varphi_n d^3r \quad (4.13a)$$

the complex conjugate relation

$$\int d^3r \varphi_n \hat{\Omega}^* \varphi_n^* = \omega_n^* \int \varphi_n^* \varphi_n d^3r, \quad (4.13b)$$

subtraction of both yields, because of (4.12a), (4.12b), the simple relation

$$\omega_n - \omega_n^* = 0. \quad (4.14)$$

The Hermitian property of the operator  $\hat{\Omega}$  (4.12a) guarantees real eigenvalues, that is, an essential requirement for the operator to be physically reasonable.

Hermiticity of an operator also guarantees a system of orthogonal eigenfunctions, as is easily shown. For two different eigensolutions  $\varphi_m$  and  $\varphi_n$  of a Hermitian operator  $\hat{\Omega}$ , we can write

$$\hat{\Omega} \varphi_m = \omega_m \varphi_m, \quad \hat{\Omega} \varphi_n = \omega_n \varphi_n, \quad (4.15)$$

$$\int d^3r \varphi_m^* \hat{\Omega} \varphi_n = \omega_n \int d^3r \varphi_m^* \varphi_n \quad \text{and} \quad (4.16a)$$

$$\int d^3r \varphi_m^* \hat{\Omega} \varphi_n = \int d^3r (\hat{\Omega}^* \varphi_m^*) \varphi_n = \omega_m^* \int d^3r \varphi_m^* \varphi_n. \quad (4.16b)$$

Because of (4.14) subtraction of (4.16a) and (4.16b) yields

$$0 = (\omega_n - \omega_m) \int d^3r \varphi_m^* \varphi_n. \quad (4.17)$$

For two different real eigenvalues  $\omega_m \neq \omega_n$ , thus, the corresponding eigenfunctions  $\varphi_m$  and  $\varphi_n$  are orthogonal in the sense of (4.5). If several eigenfunctions possess one and the same eigenvalue, they are called *degenerate* (the phenomenon: *degeneracy*). It might easily be shown that degenerate eigenfunctions can be chosen to be orthogonal to each other.

We summarize: Physically proper (reasonable) operators are Hermitian (4.12a), (4.12b); they possess real eigenvalues and a system of orthogonal eigenfunctions. These orthogonal eigenfunctions  $\varphi_n(\mathbf{r})$  can be normalized and they span a generalized vector space, the Hilbert space. A general quantum state represented by a specific vector of the Hilbert space, that is, a wave function  $\psi(\mathbf{r})$ , can be expanded in terms of a series of basis vectors  $\varphi_n(\mathbf{r})$  as shown in (4.9). In analogy to the 3D Euclidean space, the weight  $b_n$  (expansion coefficients), by which the eigenvectors  $\varphi_n(\mathbf{r})$  contribute to  $\psi(\mathbf{r})$ , is calculated according to (4.10) as the generalized scalar product of  $\varphi_n(\mathbf{r})$  and  $\psi(\mathbf{r})$ .

These general fundamental rules of quantum mechanics can only be applied to a particular problem, if the corresponding system of generalized basis vectors, that is, eigenfunctions  $\varphi_n(\mathbf{r})$  is *complete*. “Complete” means, all necessary basis vectors (eigenfunctions) for the description (linear superposition) of a general vector  $\psi(\mathbf{r})$  in that particular Hilbert space are contained in the variety of eigenfunctions (coordinates) of the operator. Imagine, you want to represent a general 3D vector in Euclidean space by its coordinates as in (4.8a). If in the mathematical formalism one particular direction (unity vector  $\mathbf{a}_i$ ) has been forgotten, the representation of a general vector is not possible, because its component in  $\mathbf{a}_i$  direction does not exist. The system of basis vectors is not complete. Such a situation must not occur for a successful description of quantum mechanical phenomena. One, therefore, has to require a *complete, orthonormal system* of eigenfunctions for a physically reasonable operator. The mathematical expression for this requirement of completeness of an eigenfunction system will be given later in Sect. 4.2.

Instead, we want to elucidate the physical meaning of the expansion coefficients  $b_n$  in the superposition (series expansion) representation of the wave function  $\psi(\mathbf{r})$  (4.9). A system is assumed to be in a quantum state described by the wave function  $\psi(\mathbf{r})$  and a measurement of the observable  $\mathcal{Q}$  is performed. As outcome of such an experiment we expect one particular eigenvalue  $\omega_n$  of the operator  $\hat{\mathcal{Q}}$ , which one of all possible values can not be predicted with certainty. The specific result is random. If we perform the measurement on an ensemble or if we repeat the same measurement on one and the same system many times, we obtain, by averaging all measured  $\omega_n$  values, the expectation value  $\langle \mathcal{Q} \rangle$ . A representation of  $\langle \mathcal{Q} \rangle$  by means of an expansion in terms of eigenfunctions (basis vectors)  $\varphi_n(\mathbf{r})$  yields

$$\begin{aligned}
\langle \hat{\Omega} \rangle &= \int d^3r \psi^* \hat{\Omega} \psi = \int d^3r \sum_{mn} b_m^* \varphi_m^* \hat{\Omega} b_n \varphi_n \\
&= \sum_{mn} b_m^* b_n \omega_n \int d^3r \varphi_m^* \varphi_n = \sum_{mn} b_m^* b_n \omega_n \delta_{mn} \\
&= \sum_n |b_n|^2 \omega_n.
\end{aligned} \tag{4.18}$$

This is exactly the expression (3.33) for a statistical average value. We conclude, that  $|b_n|^2$  is the normalized probability for finding the numerical measurement result  $\omega_n$  within the variety of all possible eigenvalues  $\omega_m$ .

The measurement of an observable  $\hat{\Omega}$  on a system being described by the wave function  $\psi(\mathbf{r}, t)$  (its quantum state) forces the system into a new state, one of the possible eigenvectors (states)  $\varphi_n(\mathbf{r})$  of the operator  $\hat{\Omega}$ . The probability, by which a particular eigenstate  $\varphi_n(\mathbf{r})$  is found after the measurement, is given by the term  $|b_n|^2$ , where  $b_n$  is determined through the series representation of the initial wave function  $\psi$  (before the measurement):

$$\psi(\mathbf{r}, t) = \sum_n b_n(t) \varphi_n(\mathbf{r}), \tag{4.19a}$$

$\varphi_n(\mathbf{r})$  are, of course, the time-independent eigensolutions of the operator  $\hat{\Omega}$ . The time dependence of  $\psi$  in (4.19a) expressed by  $b_n(t)$  follows, in the case of an energy measurement, as that of a stationary solution of the Schrödinger equation  $\exp(-i\omega_n t)$ .  $b_n(t)$  in (4.19a) can then be written as

$$b_n(t) = b_n \exp(-i\omega_n t). \tag{4.19b}$$

It is worth to emphasize: The  $\hat{\Omega}$  measurement changes the original wave function  $\psi(\mathbf{r}, t)$  into one of an eigenstate  $\varphi_n(\mathbf{r})$  of the operator  $\hat{\Omega}$  and the measurement result is the corresponding eigenvalue  $\omega_n$  of  $\hat{\Omega}$ . This (previously somewhat mysterious) phenomenon is called *Reduction* or *Collapse of the wave function*. The wave function  $\psi(\mathbf{r}, t)$  might be assumed as the result of a preceding measurement, where  $\psi$  is an eigenfunction of the corresponding operator.  $\psi$  is said to be prepared by this preceding measurement.

## 4.2 Commutation of Operators: Commutators

The immediate question arises, if every measurement of an observable causes the collapse of a given state vector (wave function). The answer is simple: A quantum system is assumed to be in the state  $\varphi_n(\mathbf{r})$  which has been prepared as an eigenstate of the operator  $\hat{A}$  by measurement of the observable  $A$ . Subsequently, an observable  $B$  is measured, the operator  $\hat{B}$  of which has the same system of eigenfunctions  $\varphi_m(\mathbf{r})$

as  $\hat{A}$ . The  $B$  measurement, then, can yield only one of the eigenstates  $\varphi_m$  of  $\hat{B}$ . The eigenstate  $\varphi_n(\mathbf{r})$ , however, has already been prepared by the  $A$  measurement; the  $B$  measurement can do nothing else but yielding the same, already present eigenstate. The state  $\varphi_n(\mathbf{r})$  persists, it does not collapse due to the  $B$  measurement.

We conclude: If two operators  $\hat{A}$  and  $\hat{B}$  (observables  $A$  and  $B$ ) possess the same eigenfunction system  $\{\varphi_m(\mathbf{r})\}$  the sequence of the two types of measurements can be exchanged without collapse of the once obtained eigenstate (wave function). The sequence of two measurements,  $B$  after  $A$ , is formally expressed by the sequential application of the operators, i.e. the operator product  $\hat{A}\hat{B}$ . If two operators have the same system of eigenfunctions, thus, their order of application can be exchanged. What does that mean in detail? Lets assume  $\hat{A}$  and  $\hat{B}$  have the same eigenfunctions  $\varphi_n$ , but with different eigenvalues  $a_n$  and  $b_n$ . This means

$$\hat{A}\varphi_n = a_n\varphi_n, \quad (4.20a)$$

$$\hat{B}\varphi_n = b_n\varphi_n. \quad (4.20b)$$

From (4.20a), (4.20b), we immediately obtain

$$(\hat{A}\hat{B} - \hat{B}\hat{A})\varphi_n = \hat{A}b_n\varphi_n - \hat{B}a_n\varphi_n = a_nb_n\varphi_n - a_nb_n\varphi_n = 0. \quad (4.21)$$

Thus, a new operator can be defined by

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}. \quad (4.22)$$

It allows the distinction between operators which can or cannot be exchanged with respect to their order of application. One also says, two operators commute with each other or they do not commute. Accordingly the operator (4.22) is called *commutator*.

In (4.21), the commutator acts on a wave function and the result is zero. If this is true for any wave function, we can say the commutator vanishes:

$$[\hat{A}, \hat{B}] = 0. \quad (4.23)$$

Equation (4.23) means that the sequence of the  $A$  and  $B$  measurements can be inverted without any influence on the given quantum state. The operators  $\hat{A}$  and  $\hat{B}$  commute, that is, they have the same system of eigenfunctions. A vanishing commutator (4.23) also implies that the two observables  $A$  and  $B$  can be measured simultaneously with infinite precision. The observables are, therefore, also called *commensurable*. They do not obey an uncertainty relation (Sect. 3.3) as for example, position and momentum (3.20). Position and momentum are incommensurable observables and are subject to the uncertainty relation  $\Delta p \Delta x \approx \hbar$ . We prove the link between uncertainty relation and incommensurability by calculating the commutator of the operators  $\hat{x} = x$  and  $\hat{p} = (\hbar/i)(\partial/\partial x)$  and applying it to a general wave function  $\psi(x)$ :

$$[\hat{x}, \hat{p}]\psi = \hat{x}\hat{p}\psi - \hat{p}\hat{x}\psi = \frac{\hbar}{i}x\psi' - \frac{\hbar}{i}x\psi' - \frac{\hbar}{i}\psi = i\hbar\psi. \quad (4.24)$$

Since this relation holds for any wave function  $\psi$ , we can write the operator equation

$$[\hat{x}, \hat{p}] = i\hbar. \quad (4.25)$$

As is shown here for the special case of position and momentum observables, we can generally state: Incommensurable observables are described by non-vanishing commutators as in (4.25). Such observables are subject to an uncertainty relation (Sect. 3.3). The fundamentally random behavior of nature on the atomic and sub-atomic level showing up in the uncertainty relation, thus, is reflected in non-commuting operators in quantum physics. This is in contrast to classical physics, where normal numbers as results of a measurement commute. The commutator relations (4.22)–(4.25) are, therefore, of paramount importance in quantum mechanics. We will use them frequently in the following, when we try to derive, or better to guess, quantum mechanical laws from classical relations.

## 4.3 Representation of Quantum States and Observables

In Sect. 4.1, we have successfully used the analogy between the abstract Hilbert space and the familiar 3D Euclidean space to describe quantum states in terms of wave functions respectively vectors in the Hilbert space. In the following, we will extend this analogy to learn some more details about quantum states and observables.

### 4.3.1 Vectors of Probability Amplitudes and Matrices as Operators

So far, a quantum state has been described by a wave function  $\psi$  as solution to the Schrödinger equation. An equivalent description of a quantum state would be in terms of probability amplitudes  $a_n$  which are the expansion coefficients in the series representation of the same wave function [ $b_n$  in (4.8b) and (4.19a)]. These amplitudes  $a_n$  are, of course, only meaningful in connection with the system of eigenfunctions (vectors)  $\varphi_n$  which are obtained, together with the eigenvalues  $A_n$ , as outcome of an  $A$  (observable) measurement on the state  $\psi(\mathbf{r})$ . This situation is described by the equations

$$\hat{A}\varphi_n = A_n\varphi_n, \quad (4.26)$$

respectively

$$\psi = \sum_n a_n \varphi_n \quad \text{with } a_n = \int d^3r \varphi_n^* \psi. \quad (4.27)$$

The probability to find the particular eigenvalue  $A_n$  as the result of the  $A$  measurement is, of course, given by the probability  $a_n^* a_n = |a_n|^2$ . The set of probability amplitudes  $(a_1, a_2, a_3, a_4, \dots)$  describes the quantum state equally well as the wave function  $\psi$ . Every information obtainable from  $\psi$  can also be derived from the set of amplitudes  $a_n$ . The description of a particular quantum state requires the whole set of amplitudes, as only a probabilistic prediction about the outcome of a measurement is possible. Each eigenvector  $\varphi_n$  with its amplitude  $a_n$  and the eigenvalue  $A_n$  might be the result. The set of amplitudes  $(a_1, a_2, a_3, \dots)$  represents a vector, in this case with an infinite number of discrete components (also a finite number is generally possible) in an infinite vector space, but in contrast to the  $\psi$  vector space with continuous coordinates, here, with discrete coordinates.

We now want to find out, how a general operator  $\hat{\Omega}$  is represented in the discrete vector space of the amplitudes  $(a_1, a_2, a_3, \dots)$ . For this purpose, we calculate the expectation value  $\langle \Omega \rangle$  in the state (wave function)  $\psi$  given as an expansion in the eigenfunctions  $\varphi_n$ :

$$\begin{aligned} \langle \Omega \rangle &= \int d^3r \psi^*(\mathbf{r}) \hat{\Omega} \psi(\mathbf{r}) \\ &= \sum_{mn} a_m^* a_n \int d^3r \varphi_m^* \hat{\Omega} \varphi_n = \sum_{mn} a_m^* \Omega_{mn} a_n, \end{aligned} \quad (4.28)$$

$\Omega_{mn}$  are, hereby, the elements of a quadratic arrangement of complex numbers, which is called a *matrix*  $\underline{\Omega} = \{\Omega_{mn}\}$ , that is,

$$\underline{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \dots \\ \Omega_{21} & \Omega_{22} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad \text{with } \Omega_{mn} = \int d^3r \varphi_m^* \hat{\Omega} \varphi_n. \quad (4.29)$$

In this representation, the expectation value  $\langle \Omega \rangle$  can be written as a product of the matrix  $\{\Omega_{mn}\}$  with the vector  $(a_1, a_2, a_3, \dots)$  and subsequently multiplied from left with the vector  $(a_1^*, a_2^*, a_3^*, \dots)$ , that is,

$$\langle \Omega \rangle = (a_1^*, a_2^*, a_3^*, \dots) \begin{pmatrix} \Omega_{11} & \Omega_{12} & \dots \\ \Omega_{21} & \Omega_{22} & \dots \\ \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \end{pmatrix}. \quad (4.30)$$

In this context, one should remember the formalism of multiplying a 3D quadratic matrix with a 3D vector and the multiplication of two matrices with each other:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}. \quad (4.31a)$$

The resulting  $\gamma$  vector components are defined as  $\gamma_2 = \alpha_{21}\beta_1 + \alpha_{22}\beta_2 + \alpha_{23}\beta_3$  and analogously  $\gamma_1$  and  $\gamma_3$ .

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \quad (4.31b)$$

with  $\gamma_{22} = \alpha_{21}\beta_{12} + \alpha_{22}\beta_{22} + \alpha_{23}\beta_{32}$  and in analogy the other  $\gamma_{ij}$ .

Again, we remind, that the order of the matrices is essential in the calculation of their product. Matrices, generally, do not commute. This is similar to operators when applied to a wave function (Sect. 4.2). We recognize the analogy between the formal description by means of operators and wave functions and that using matrices and vectors of probability amplitudes. The first formalism called *Wave Mechanics* was invented by Schrödinger, while the approach to quantum physics by means of discrete matrices and vectors dates back to Heisenberg and was sometimes called *Matrix Mechanics*. Here, we have shown, that both approaches are equally well suited for the description of single particle quantum dynamics.

Next, we want to examine how two mutually adjoint operators are represented as matrices in the discrete space (also a particular Hilbert space) of probability amplitude vectors. The operator  $\hat{\Omega}^+$  is adjoint to  $\hat{\Omega}$  if

$$\int d^3r (\hat{\Omega}^+ \varphi)^* \psi = \int d^3r \varphi^* \hat{\Omega} \psi. \quad (4.32)$$

By expanding  $\varphi$  and  $\psi$  in terms of a series of orthonormal eigenfunctions  $\varphi_n$ , that is,

$$\varphi = \sum_n a_n \varphi_n, \quad \psi = \sum_m b_m \varphi_m, \quad (4.33)$$

we obtain for the right side of (4.32)

$$\sum_{nm} a_n^* \left( \int d^3r \varphi_n^* \hat{\Omega} \varphi_m \right) b_m = \sum_{nm} a_n^* \Omega_{nm} b_m, \quad (4.34a)$$

and for the left side, respectively

$$\sum_{nm} a_n^* \left[ \int d^3r \varphi_m (\hat{\Omega}^+)^* \varphi_n^* \right] b_m = \sum_{nm} a_n^* \Omega_{mn}^* b_m. \quad (4.34b)$$

By comparison of (4.34a) with (4.34b), we realize that two mutually adjoint operators are represented in the vector space of probability amplitudes by two matrices, in which rows and columns are exchanged (*transposed matrices*) and the elements are the complex conjugate of each other. The matrix representing  $\hat{\Omega}^+$  is the transpose conjugate of the matrix representing  $\hat{\Omega}$ :

$$\Omega_{nm} \rightarrow \Omega_{mn}^*. \quad (4.35)$$

Physically proper operators are self-adjoint or Hermitian, that is, they obey the relation  $\hat{\Omega}^+ = \hat{\Omega}$  (4.12a). Correspondingly, their so-called Hermitian matrices obey

$$\Omega_{mn}^* = \Omega_{nm}. \quad (4.36)$$

Also the fundamental dynamic equation of quantum mechanics, the Schrödinger equation  $i\hbar\dot{\psi} = \hat{H}\psi$  (3.50), (3.51), is transformed into a matrix equation by inserting the expansion of  $\psi$  in terms of an orthonormal eigenfunction system  $\{\varphi_n\}$ :

$$i\hbar \sum_n \dot{a}_n(t) \varphi_n(\mathbf{r}) = \sum_n a_n(t) \hat{H} \varphi_n(\mathbf{r}). \quad (4.37)$$

After multiplication of (4.37) with  $\varphi_m^*$  from the left and integration over the volume of the system, we obtain:

$$i\hbar \sum_n \dot{a}_n(t) \int d^3r \varphi_m^* \varphi_n = i\hbar \sum_n \dot{a}_n \delta_{mn} = \sum_n a_n(t) \int d^3r \varphi_m^* \hat{H} \varphi_n, \quad (4.38a)$$

that is,

$$i\hbar \dot{a}_m(t) = \sum_n H_{mn} a_n(t).$$

(4.38b)

Equation (4.38b) is a system of coupled differential equations for the probability amplitudes  $a_n(t)$  which are obtained as results of a measurement of the observable  $A$  (4.26), (4.27), with

$$H_{mn} = \int d^3r \varphi_m^*(\mathbf{r}) \hat{H} \varphi_n(\mathbf{r}) \quad (4.39)$$

as the matrix elements of the Hamilton operator  $\hat{H}$  in the eigenfunction system  $\{\varphi_n\}$  of the operator  $\hat{A}$ .

For time-independent problems, only the time-independent Schrödinger eigenvalue equation (3.57) must be solved. The representation of the Schrödinger equation  $\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r})$  in terms of the matrix formalism is obtained by inserting the series expansion of  $\psi$  in eigenfunctions  $\varphi_n$ , multiplication with  $\varphi_m^*$  from the left and integrating over the volume of the system:

$$\sum_n a_n \int d^3r \varphi_m^* \hat{H} \varphi_n = E \sum_n a_n \int \varphi_m^* \varphi_n d^3r = E a_m, \quad (4.40a)$$

$$\sum_n H_{mn} a_n = E a_m. \quad (4.40b)$$

In (4.40b), the matrix  $\underline{\underline{H}} = \{H_{mn}\}$  is multiplied with the amplitude vector  $(a_1, a_2, a_3, \dots)$  and the resulting vector is represented as the same vector, only multiplied with the energy eigenvalue  $E$ . As described in (3.48) this is a typical eigenvalue problem of linear algebra. The solution aims at finding a particular set of eigenvectors  $(a_1, a_2, a_3, \dots)$  with corresponding eigenvalues  $E$ . The analogy to the eigenvalue problem in the operator, eigenfunction formalism (4.26) is obvious.

We will see in the following that the matrix formalism of quantum mechanical relations might simplify the model description of certain problems considerably by intuitive assumptions about the matrix elements  $H_{mn}$  of the Hamilton operator. This is particularly true for so-called 2-level systems (Sect. 6.5), where simplifying assumptions about the underlying physics reduce the dimension of  $H_{mn}$  down to two.

The matrix eigenvalue problem for the energy (Hamilton) operator (4.40b) is only one special example of the general class of eigenvalue problems which are obtained from the representation of the general operator equation  $\hat{\Omega}\psi = \omega\psi$  (3.49) in terms of the matrix formalism. By expanding the wave function  $\psi$  in terms of the eigenfunction system  $\{\varphi_n\}$  and performing a calculation analogously to (4.40a), (4.40b), we obtain the matrix equation

$$\sum_n \underline{\Omega}_{mn} a_n = \omega a_m \quad \text{with } \underline{\Omega}_{mn} = \int d^3r \varphi_m^* \hat{\Omega} \varphi_n. \quad (4.41)$$

With  $\mathbf{a} = (a_1, a_2, a_3, \dots)$  as the amplitude vector, (4.41) can be written in a compact form as

$$\underline{\underline{\Omega}} \mathbf{a} = \omega \mathbf{a}. \quad (4.42)$$

With

$$\underline{\underline{1}} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots & \dots \\ 0 & 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad \text{and} \quad \underline{\underline{0}} = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad (4.43)$$

as unity and zero matrix we obtain from (4.42)

$$(\underline{\underline{\Omega}} - \omega \underline{\underline{1}}) \mathbf{a} = \mathbf{0} \quad (4.44a)$$

and respectively,

$$\mathbf{a} = (\underline{\underline{\Omega}} - \omega \underline{\underline{1}})^{-1} \mathbf{0}. \quad (4.44b)$$

At this point we want to remind the solution of this eigenvalue problem of linear algebra a little bit more in detail: The vector/matrix equation (4.44b) can only have non-trivial solutions for the eigenvalues  $\omega$  if the inverse matrix  $(\underline{\underline{\Omega}} - \omega \underline{\underline{1}})^{-1}$  approaches infinity (exactly: their single elements). What does that mean in detail? Let us consider the calculation of an inverse matrix.

The inverse matrix  $\underline{\underline{M}}^{-1}$  of a matrix  $\underline{\underline{M}}$  is defined by  $\underline{\underline{M}}\underline{\underline{M}}^{-1} = \underline{\underline{1}}$ . It is calculated from  $\underline{\underline{M}}$  by the rule

$$\underline{\underline{M}}^{-1} = \frac{1}{\det \underline{\underline{M}}} \underline{\underline{\tilde{M}}}. \quad (4.45)$$

Hereby,  $\underline{\underline{\tilde{M}}}$  is the so-called cofactor matrix. Each matrix element  $\tilde{M}_{ij}$  is calculated by the following multistep procedure:

- Remove the  $i$ th row and  $j$ th column of  $\underline{\underline{M}}$ .
- For each  $i, j$  index doublet the remaining matrix elements form a so-called submatrix with one dimension lower than  $\underline{\underline{M}}$ .
- Calculate the so-called minor determinants  $A_{ij}$  from these submatrices, supply them with a sign according to  $(-1)^{i+j} A_{ij}$  and form a new matrix with these elements (checkered change of signs).
- Transpose this matrix of minor determinants by exchanging rows and columns and obtain the element  $\tilde{M}_{ij}$  of the cofactor matrix.

The following two-dimensional example shall clarify the procedure. From the matrix

$$\underline{\underline{M}} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad (4.46a)$$

we calculate the matrix of minor determinants with alternating (checkered) signs

$$\underline{\underline{M}}^1 = \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix}. \quad (4.46b)$$

The one-dimensional minor determinant, in this case the simple number  $\delta$  in (4.46b), follows from (4.46a) as the remaining element after removing the row  $\alpha, \beta$  and the column  $\alpha, \gamma$ .  $\underline{\underline{M}}^1$  in (4.46b) is transposed and we obtain the cofactor matrix

$$\underline{\underline{\tilde{M}}} = \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}. \quad (4.46c)$$

With  $\det \underline{\underline{M}} = \alpha\delta - \beta\gamma$ , we obtain the inverse matrix

$$\underline{\underline{M}}^{-1} = \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix}. \quad (4.46d)$$

The proof simply follows as

$$\begin{aligned} \underline{\underline{M}}\underline{\underline{M}}^{-1} &= \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix} \\ &= \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \alpha\delta - \beta\gamma & -\alpha\beta + \alpha\beta \\ \gamma\delta - \delta\gamma & \alpha\delta - \beta\gamma \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \underline{\underline{1}}. \end{aligned} \quad (4.47)$$

Because of (4.45) equation (4.44b) has non-trivial solutions for the eigenvalue  $\omega$  only if

$$\det(\underline{\underline{\Omega}} - \omega \underline{1}) = 0. \quad (4.48)$$

For a 3 ( $n$ ) dimensional matrix  $\underline{\underline{\Omega}}$  the determinant in (4.48) is a polynomial of 3rd ( $n$ th) order. The so-called *characteristic equation* (4.48) to the eigenvalue problem, thus, has 3 ( $n$ ) solutions for  $\omega$ , the eigenvalues  $\omega_1, \omega_2, \omega_3, \dots$ . The eigenvalue equation (4.42) fixes the eigenvector only up to an overall scale factor. Multiples of an eigenvector are not treated as distinct eigenvectors. The eigenvectors  $\mathbf{a}$  are calculated from (4.42) by inserting the different eigenvalues  $\omega_i$ . Subsequently, the eigenvectors are normalized in order to fulfill the basic quantum mechanical normalization condition.

### 4.3.2 Rotations of Hilbert Space

In the Euclidean 3D space, one and the same vector  $\mathbf{a}$  can be represented, of course with different Cartesian coordinates, in infinitely many, mutually rotated coordinate systems. A rotation of the coordinate system keeps orientation and length ( $= |\mathbf{a}|^{1/2}$ ) of the vector unchanged. The coordinates in the different mutually rotated coordinate systems are, therefore, connected with each other through a transformation matrix (*rotation matrix*) which is attributed to the corresponding rotation. For a rotation of the  $(x, y)$  coordinate system into the  $(x', y')$  system by the angle  $\varphi$  about the  $z$  axis, the situation is shown in Fig. 4.1. The relation between corresponding vector components is expressed by a rotation matrix containing  $\sin \varphi$  and  $\cos \varphi$  elements:

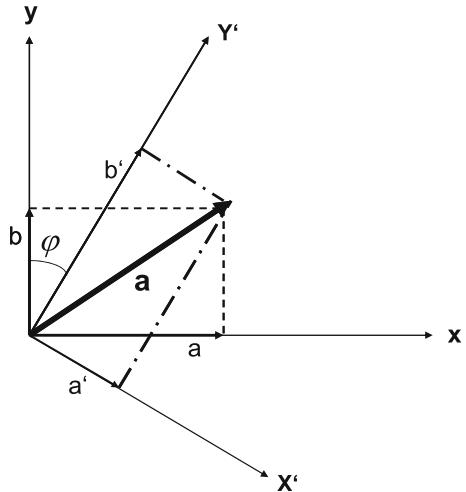
$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \cos \varphi + b \sin \varphi \\ -a \sin \varphi + b \cos \varphi \end{pmatrix}. \quad (4.49)$$

Because of  $\sin^2 \varphi + \cos^2 \varphi = 1$  the squared modulus, that is, the length of the vector is conserved upon rotation:

$$\begin{aligned} a'^2 + b'^2 &= (a \cos \varphi + b \sin \varphi)^2 + (-a \sin \varphi + b \cos \varphi)^2 \\ &= a^2 + b^2. \end{aligned} \quad (4.50)$$

Equation (4.50) results from the orthogonality of the rotation matrix: their rows and columns assumed as vectors are mutually orthogonal. Furthermore, if we mirror the rotation matrix at its diagonal, we obtain the inverse matrix (see Sect. 4.3.1).

**Fig. 4.1** Representation of a vector  $\mathbf{a}$  in the Cartesian  $x, y$ -coordinate system and in the tilted  $x', y'$ -system (rotation angle  $\varphi$ )



Analogous properties can be derived for the abstract vectors  $(a_1, a_2, a_3, \dots)$  of a quantum state in the Hilbert space. A general state, given in terms of a wave function  $\psi$ , can be expanded in two different eigenfunction systems, for example,  $\{\varphi_n\}$  and  $\{\psi_m\}$  being eigensolutions to the operators  $\hat{A}$  and  $\hat{B}$ , respectively. Apart from (4.26) we, thus, have in addition

$$\hat{B}\psi_n = B_n\psi_n. \quad (4.51)$$

$\psi$  can be expanded in both eigenfunction systems as follows:

$$\psi = \sum_n a_n \varphi_n \quad \text{with } a_n = \int d^3r \varphi_n^* \psi, \quad (4.52a)$$

$$\psi = \sum_m b_m \psi_m \quad \text{with } b_m = \int d^3r \psi_m^* \psi. \quad (4.52b)$$

We can construct the connection between both representations of  $\psi$  by expanding the eigenfunctions  $\psi_m$  in terms of the functions  $\varphi_n$ :

$$\psi_m = \sum_n T_{mn} \varphi_n \quad \text{with } T_{mn} = \int d^3r \varphi_m^* \psi_n. \quad (4.53)$$

Hereby, the expansion coefficients  $T_{mn}$  form a transformation matrix, the properties of which we want to elucidate a little bit more in detail. We can represent the relation (4.53) also in terms of the vectors  $\{a_n\}$  and  $\{b_n\}$  of probability amplitudes (4.52a), (4.52b). For this purpose, we multiply the complex conjugate relation of (4.53) with  $\psi(\mathbf{r})$  and integrate over the volume of the system:

$$b_m = \int d^3r \psi_m^* \psi = \sum_n T_{mn}^* \int d^3r \varphi_n^* \psi = \sum_n T_{mn}^* a_n. \quad (4.54a)$$

The transformation recipe between the two representations of  $\psi$  in probability amplitudes is, thus, given by

$$b_m = \sum_n T_{mn}^* a_n. \quad (4.54b)$$

In analogy to 3D space (4.50), the state vector  $\psi(\mathbf{r})$  in Hilbert space must have the same length (modulus) in both coordinate (eigenfunction) systems  $\{\varphi_n\}$  and  $\{\psi_m\}$ . This requires:

$$\int d^3r \psi^* \psi = \sum_{nm} a_n^* a_m \int d^3r \varphi_n^* \varphi_m = \sum_{nm} b_n^* b_m \int d^3r \psi_n^* \psi_m. \quad (4.55a)$$

We have used the expansions of  $\psi$  in terms of  $\varphi_n$  and  $\psi_n$  (4.52a), (4.52b). Because of the orthogonality of both eigenfunction systems, we directly obtain

$$\sum_n a_n^* a_n = \sum_n b_n^* b_n. \quad (4.55b)$$

Equation (4.54b) inserted into (4.55b) yields:

$$\sum_n a_n^* a_n = \sum_{nmm'} T_{nm} T_{nm'}^* a_m^* a_{m'}. \quad (4.56)$$

The condition of equal length of the vector in the two representations (4.55b) requires for the transformation matrix  $\underline{\underline{T}}$ :

$$\sum_n T_{nm}^* T_{nm} = \delta_{m'm}. \quad (4.57a)$$

By comparison with (4.31a), (4.31b), we realize that the order of indices  $n$  and  $m$  in (4.57a) is not correct for the calculation of a matrix product. Summation has to be performed over the inner indices as in (4.31b). In order to write (4.57a), (4.57b) as a matrix product, the indices in  $T_{nm}^*$  have to be exchanged, the matrix must be transposed. Then, (4.57a) means that the matrix  $\underline{\underline{T}}$  multiplied from left with its complex conjugate and transposed yields the unity matrix. We have encountered such matrices already earlier. They belong to mutually adjoint operators (4.32)–(4.36). Usually they are denoted by the symbol  $\underline{\underline{T}}^+$ , such that (4.57a) is written as

$$\underline{\underline{T}}^\dagger \underline{\underline{T}} = \underline{\underline{1}}. \quad (4.57b)$$

Matrices (or transformations)  $\underline{\underline{T}}$  which connect different eigenvector systems for the representation of a general state vector with each other obey the relation (4.57a), (4.57b). They are called *Unitary Matrices* or *Transformations*. Due to (4.57b), unitary matrices fulfill the following relations:

$$\underline{\underline{T}}^\dagger \underline{\underline{T}} \underline{\underline{T}}^\dagger = \underline{\underline{T}}^\dagger, \quad (4.58a)$$

$$\underline{\underline{T}}^\dagger \underline{\underline{T}} = \underline{\underline{T}} \underline{\underline{T}}^\dagger = \underline{\underline{1}} \quad (4.58b)$$

respectively,

$$\underline{\underline{T}}^{-1} = \underline{\underline{T}}^\dagger. \quad (4.59)$$

Unitary transformations or matrices mutually transform different representations (4.52a), (4.52b) of one and the same state vector  $\psi(\mathbf{r})$  in different Hilbert spaces. They describe how measurement results of an observable  $A$  transform into results of a measurement of another observable  $B$ . The different Hilbert spaces are spanned by different eigenfunction systems  $\{\varphi_n\}$  and  $\{\psi_n\}$  which are the possible new quantum states after an  $A$  and a  $B$  measurement, respectively. In analogy to the situation in the 3D Euclidean space (4.49), Hilbert spaces belonging to different measurements (observables) are said to be rotated against each other.

### 4.3.3 Quantum States in Dirac Notation

In the preceding section, we have seen that a state of a quantum system described by a wave function  $\psi(\mathbf{r})$  might be represented in different eigenfunction systems (4.52a), (4.52b):

$$\psi(\mathbf{r}) = \sum_n a_n \varphi_n = \sum_n b_n \psi_n, \quad (4.60)$$

Hereby, the Hilbert spaces spanned by  $\{\varphi_n\}$  and  $\{\psi_n\}$  are rotated against each other. As for rotations the modulus of the state vector  $\psi(\mathbf{r})$  is conserved upon rotation (4.55a), (4.55b):

$$\int d^3r \psi^*(\mathbf{r}) \psi(\mathbf{r}) = \sum_n a_n^* a_n = \sum_n b_n^* b_n. \quad (4.61)$$

It is always one and the same quantum state, which is concerned, independent of its particular representation, be it in the continuous coordinates  $\mathbf{r}$  (integral over  $\psi^* \psi$  is analogous to a sum, Sect. 4.1) or in the discrete vector space of the probability amplitudes  $a_n$  and  $b_n$ . The different representations are analogous to the situation in Fig. 4.1, where one and the same 2D vector  $\mathbf{a}$  can be described by its coordinates  $a, b$  or by  $a', b'$  in the rotated coordinate system. The notation  $\mathbf{a}$ , hereby, is a description of the vector which is not dependent on a special coordinate system;

it is very well suited for a mathematical formalism of great generality. In analogy, Dirac (1902–1984), one of the founders of quantum theory, has invented an elegant general description of quantum states, which is independent of the coordinate (eigenfunction) system [1]. Rather than expressing a quantum state as in (4.60) in terms of its wave function or different components  $a_n$  or  $b_n$  in Hilbert spaces of probability amplitudes Dirac denotes a quantum state by the general and abstract symbol  $|\psi\rangle$ . This symbol describes the quantum state in an abstract way as does the solid symbol  $\mathbf{a}$  for an abstract vector in 3D space. In both cases the notation is independent on a particular coordinate (eigenfunction) system. As our Hilbert spaces are dual spaces with complex numbers and functions, the following correlations are defined:

$$\psi \rightarrow \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} \rightarrow |\psi\rangle, \quad (4.62a)$$

$$\psi^* \rightarrow (a_1^*, a_2^*, a_3^*, \dots) \rightarrow \langle\psi|. \quad (4.62b)$$

From (4.62a), (4.62b), the following expression for the scalar product, the modulus, of  $\psi$  (4.61) is obtained:

$$\langle\psi|\psi\rangle = \int d^3r \psi^*(\mathbf{r})\psi(\mathbf{r}) = \sum_n a_n^* a_n. \quad (4.63)$$

The expression  $\langle\psi|\psi\rangle$  looks like a bracket. Dirac, therefore, called the left part  $\langle\psi|$  a bra vector and the right part  $|\psi\rangle$  a ket vector. Bras and kets are mutually complex conjugate and transposed vectors within a general Hilbert space, without making notice to any special representation in a particular space. The bra-ket notation of quantum states is, thus, independent of any special representation, it describes a physical state of a system in a most general and abstract way. But we will see, that calculations based on *bras* and *kets* are quite easy to handle, if we apply some simple algebraic rules.

A scalar product of two wave functions  $\psi$  and  $\varphi$  is formed in the bra  $\psi$  and ket  $\varphi$  notation as

$$\langle\psi|\varphi\rangle = \int d^3r \psi^* \varphi. \quad (4.64)$$

Accordingly the expansion of a wave function  $\psi$  in terms of eigenfunctions  $\varphi_n$  is written in Dirac notation as

$$|\psi\rangle = \sum_n a_n |\varphi_n\rangle = \sum_n a_n |n\rangle \quad \text{with } a_n = \langle n|\psi\rangle. \quad (4.65)$$

In the last term of (4.65), a common way of denoting a quantum state in the bra-ket formalism is used: Rather than using the expression  $|\varphi_n\rangle$  only the quantum number  $n$  of the state is written into the *ket* symbol. This is sufficient, since the bra-ket formalism is independent on any coordinate system. Of course, the context, in which the quantum numbers  $n$  are defined, must be known. We will see in Sect. 7.4.3 that even large macroscopic systems as a living animal, a cat, can be inserted into the *ket* symbol ( $|\text{cat}\rangle$ ), as long as they are considered as complex multidimensional many particle quantum systems.

With respect to (4.65), it should be emphasized again, that the probability amplitude for finding the eigenstate  $|n\rangle$  in an  $A$  measurement on the  $\psi$  *ket* is the projection of  $|\psi\rangle$  into the  $|n\rangle$  direction, namely  $\langle n|\psi\rangle$ . Using solely the quantum number  $n$  in the *ket* symbol the eigenvalue equation for the operator  $\hat{A}$  is written as

$$\hat{A}|n\rangle = A_n|n\rangle. \quad (4.66)$$

The orthonormality of the eigenfunctions  $\varphi_n$  is expressed in Dirac notation as

$$\langle m|n\rangle = \delta_{mn}. \quad (4.67)$$

Also the condition for completeness (Sect. 4.1) of an orthonormal eigenfunction system  $\{\varphi_n\}$  can be written down in a very elegant way. In a complete system, all eigenvectors  $|n\rangle$  necessary to expand any vector  $|\psi\rangle$  in the Hilbert space do exist. By means of (4.65) we, thus can expand the  $\psi$  *ket* as

$$|\psi\rangle = \sum_n a_n|n\rangle = \sum_n \langle n|\psi\rangle|n\rangle = \sum_n |n\rangle\langle n|\psi\rangle. \quad (4.68)$$

For a complete series representation the right side of (4.68) must equal  $|\psi\rangle$ , that is, the completeness condition for the system of eigenkets  $|n\rangle$  is obtained as:

$$\sum_n |n\rangle\langle n| = \hat{1}, \quad (4.69a)$$

where  $\hat{1}$  is the unity operator or the unity matrix in a discrete Hilbert space.

The operator  $|n\rangle\langle n|$  in (4.69a) looking like a butterfly is called *projection operator*  $\hat{P}_n$  for the  $|n\rangle$  *ket*. It projects the vector  $|\psi\rangle$  onto the eigenket  $|n\rangle$ , that is, it filters out from  $|\psi\rangle$  (4.65) specifically the  $|n\rangle$  direction with its component  $a_n$  (probability amplitude):

$$\begin{aligned} \hat{P}_n|\psi\rangle &= \hat{P}_n \sum_{n'} a_{n'}|n'\rangle = \sum_{n'} a_{n'}|n\rangle\langle n|n'\rangle \\ &= \sum_{n'} a_{n'}|n\rangle\delta_{nn'} = a_n|n\rangle. \end{aligned} \quad (4.69b)$$

In Dirac notation matrix elements  $\Omega_{mn}$  (complex numbers) in the eigenfunction system  $\{\varphi_n\}$  of an operator  $\hat{\Omega}$  are expressed as:

$$\langle m | \hat{\Omega} | n \rangle = \langle \varphi_m | \hat{\Omega} | \varphi_n \rangle = \int d^3r \varphi_m^* \hat{\Omega} \varphi_n = \Omega_{mn}. \quad (4.70)$$

From (4.32), we conclude that for the adjoint operator  $\hat{\Omega}^+$  to  $\hat{\Omega}$  the matrix elements in Dirac notation must obey the relation

$$\langle m | \hat{\Omega} | n \rangle^* = \langle n | \hat{\Omega}^\dagger | m \rangle. \quad (4.71)$$

By twofold application of the completeness relation (4.69a), any operator  $\hat{\Omega}$  can be represented by its matrix elements (4.70):

$$\begin{aligned} \hat{\Omega} &= \sum_{nm} |n\rangle \langle n| \hat{\Omega} |m\rangle \langle m| = \sum_{nm} \langle n | \hat{\Omega} | m \rangle |n\rangle \langle m| \\ &= \sum_{nm} \Omega_{nm} |n\rangle \langle m|. \end{aligned} \quad (4.72)$$

Using (4.71) and (4.69a), (4.69b), a relation between ket  $|\psi'\rangle$  and bra  $\langle\psi'|$  and the action of an operator  $\hat{\Omega}$  can be derived. For this purpose, we assume

$$|\psi'\rangle = \hat{\Omega} |\psi\rangle = \sum_n |n\rangle \langle n| \hat{\Omega} |\psi\rangle \quad (4.73a)$$

and obtain for the corresponding *bra* vector:

$$\begin{aligned} \langle\psi'| &= \sum_n \langle n | \langle n | \hat{\Omega} | \psi \rangle^* = \sum_n \langle n | \langle \psi | \hat{\Omega}^\dagger | n \rangle \\ &= \sum_n \langle \psi | \hat{\Omega}^\dagger | n \rangle \langle n | = \langle \psi | \hat{\Omega}^\dagger. \end{aligned} \quad (4.73b)$$

The action of an operator  $\hat{\Omega}$  from left on a *ket* is equivalent to the action of the adjoint operator  $\hat{\Omega}^+$  from right on the *bra*.

The Dirac bra-ket notation of quantum states in a very general way, without notice to any special Hilbert space, allows a very elegant representation of quantum mechanical formulas. Quantum theoretical calculations often seem like a game with *bra* and *ket* symbols. One has to keep in mind the formulas (4.64), (4.65), (4.67) and particularly (4.69a), (4.69b) in order to play this game.

In the following, we will use all kind of representations of quantum states and operators, depending on what the most comfortable way for the solution of the problem they offer. The Dirac notation, however, often is the most compact and comfortable formalism.

### 4.3.4 Quantum States with a Continuous Eigenvalue Spectrum

For confined states of an electron in a potential box (Sect. 3.6.1), the spectrum of eigenstates and eigenvalues is discrete, that is, they are numerable by discrete quantum numbers  $n$ . For electrons propagating in free space, however, as they tunnel through barriers or surmount potential steps, the quantum numbers, their wave vectors  $\mathbf{k}$  lie densely, they form a continuous spectrum (Sects. 3.6.3, 3.6.4). The reason is the infinite extension of the physical system, the free space. Also an electron localized in space by a propagating wave packet (Sect. 3.2) is described by a dense, continuous variety of  $\mathbf{k}$  vectors or wave numbers  $k$  [distribution  $a(k)$ ] which build up the wave packet.

In all these cases, we can formally evade the problem of a continuous eigenvalue spectrum by assuming a large macroscopic cubic box with length  $L$  as physical system rather than considering the infinite space as the definition volume. By using periodic boundary conditions (3.67) we, then, obtain a discrete quantization of  $k$  values (3.63), but with a quasi-continuous spectrum. In this case integration over the  $\mathbf{k}$  space always requires the following substitution:

$$\frac{L^3}{(2\pi)^3} \int d^3k \rightarrow \sum_k. \quad (4.74)$$

For discrete spectra of states, the formalism considered so far is sufficient. Mathematical problems arise when we try to describe series extensions, orthogonality of eigenvectors etc. for systems with continuous eigenvalue spectra.

Let us consider the momentum operator  $\hat{p} = (\hbar/i)d/dx$  as an example. In infinite space, its spectrum of eigenvalues is continuous. Its eigenvalue equation

$$\frac{\hbar}{i} \frac{d}{dx} \psi_p(x) = p \psi_p(x) \quad \text{or} \quad \hat{p}|\psi_p\rangle = p|\psi_p\rangle \quad (4.75)$$

is solved by familiar plane waves  $\psi_p(x)$ :

$$\psi_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}. \quad (4.76)$$

Because of their property of being eigensolutions of (4.75), the  $\psi_p(x)$  functions form an orthonormal complete functional basis with a continuous eigenvalue ( $p$ ) spectrum. In the orthogonality (4.67) and completeness (4.69a), (4.69b) relations, therefore, the sum over discrete quantum numbers must be replaced by an integral over  $x$  and  $p$ , respectively, that is:

$$\langle \psi_p | \psi_{p'} \rangle = \langle p | p' \rangle = \int dx \psi_p^*(x) \psi_{p'}(x) = \frac{1}{2\pi\hbar} \int dx e^{i(p'-p)x/\hbar}, \quad (4.77a)$$

$$\int dp |p\rangle\langle p| = \int dp \psi_p^*(x')\psi_p(x) = \frac{1}{2\pi\hbar} \int dp e^{ip(x-x')/\hbar}. \quad (4.77b)$$

According to (4.67), (4.69a) and (4.69b), respectively, both relations (4.77a), (4.77b) must yield an expression which is in some way similar to “one”, in the sense of the Kronecker symbol  $\delta_{mn}$  (4.5). More precisely,  $\langle p|p'\rangle$  (4.77a) must equal a normalizable expression for  $p = p'$ , otherwise it should be negligibly small. Analogous behavior must be given for (4.77b) when  $x = x'$ .

The solution to the problem becomes obvious when we consider the Fourier expansion of a function  $f(x)$  as it was used for the description of a wave packet (Sect. 3.2).

With  $g(k)$  as the densely lying, continuous expansion coefficients one has:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk g(k) e^{ikx}. \quad (4.78a)$$

Hereby the Fourier transform  $g(k)$  is written as

$$g(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx' f(x') e^{-ikx'}. \quad (4.78b)$$

If we insert  $g(k)$  from (4.78b) again into (4.78a), exchange of the integrals yields:

$$\begin{aligned} f(x) &= \int dx' f(x') \left[ \frac{1}{2\pi} \int dk e^{ik(x-x')} \right] \\ &= \int dx' f(x') \delta(x - x'). \end{aligned} \quad (4.79)$$

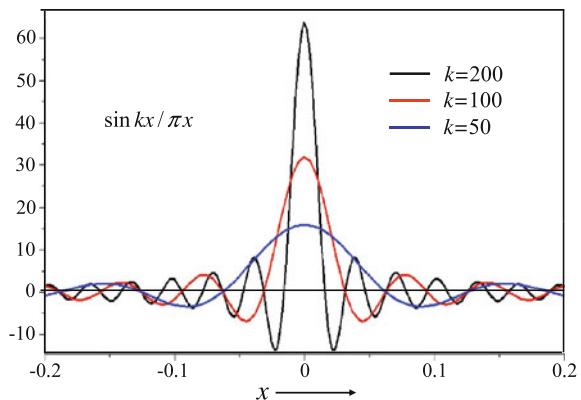
The integral on the right side must again equal the function  $f(x)$ . Under what conditions is that possible? Let us consider the situation a little more in detail.

$\delta(x - x')$  was used in (4.79) to denote the following expression:

$$\begin{aligned} \delta(x - x') &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(x-x')} = \frac{1}{2\pi i} \frac{e^{ik(x-x')}}{(x - x')} \Big|_{-\infty}^{\infty} \\ &= \frac{1}{\pi} \lim_{k \rightarrow \infty} \left[ \frac{\sin k(x - x')}{(x - x')} \right]. \end{aligned} \quad (4.80)$$

The corresponding graph of the function  $\sin kx/\pi x$  in Fig. 4.2 suggests that for increasing  $k$  values ( $k = 50, 100, 200$  in Fig. 4.2) this function approaches an infinitely high peak at  $x = 0$  (for  $k \rightarrow \infty$ ). Simultaneously, the oscillations around the  $x$  axis get higher and higher frequencies. In the limit  $k \rightarrow \infty$ , we expect densely lying nodes on the  $x$  axis, that is, an extremely fast variation of positive and negative functional values. An integral as in (4.79), thus, would be canceled outside the position  $x = 0$ . At  $x = 0$ , the functional value is regained in the integral. The expression

**Fig. 4.2** Graph of the function  $\sin kx/\pi x$  for  $k$  values  $k = 50, 100, 200$ . In the limit  $k \rightarrow \infty$  this functional series represents the delta function (distribution)  $\delta(x)$



(4.80), thus, represents a series of functions (for increasing  $k$  values) for which in the limit  $k \rightarrow \infty$  the following relation holds:

$$f(0) = \int_{-\infty}^{\infty} dx f(x) \delta(x) \quad (4.81a)$$

respectively,

$$f(x) = \int dx' f(x') \delta(x - x'). \quad (4.81b)$$

Only in this sense the identity between left and right side of the relation (4.79) is understood. Dirac has introduced this expression  $\delta(x - x')$  into the physical and mathematical literature. It is called Dirac *delta function*, even though it is not a well defined function in the strict sense. The delta function is only defined by an integral relation as in (4.81a), (4.81b). Apart from (4.80) a number of other functions can fulfill (4.81a), (4.81b), for example,

$$\delta(x - x') = \frac{1}{\sqrt{\pi}} \lim_{\Sigma \rightarrow 0} \left[ \frac{1}{\sqrt{\Sigma}} e^{-\frac{(x-x')^2}{\Sigma}} \right]. \quad (4.82)$$

Beside (4.81a), (4.81b) the delta function obeys the relations:

$$\delta(x - x') = 0 \quad \text{for } x \neq x', \quad (4.83a)$$

$$\int_{-\infty}^{\infty} \delta(x - x') dx' = 1. \quad (4.83b)$$

Expressions which are represented by series of functions, but which obey, as the delta function, integral relations as (4.81a), (4.81b) and (4.83b) are called *distributions*.

The analogy of the delta function for a continuous distribution of eigenvalues with the Kronecker symbol  $\delta_{mn}$  for a discrete distribution of  $m, n$  is obvious from (4.83a),

(4.83b). According to (4.80) the right sides of (4.77a), (4.77b) and (4.80) are exactly the representations of  $\delta(p' - p) = \delta(p - p')$  and  $\delta(x' - x) = \delta(x - x')$ , respectively. Using the delta function, we can write the normalization and completeness conditions for bras and kets with continuous  $p$  and  $x$  eigenvalues as

$$\langle p | p' \rangle = \delta(p - p'), \quad (4.84a)$$

$$\int dp |p\rangle \langle p| = \delta(x - x'). \quad (4.84b)$$

It should again be emphasized that the boundary conditions are important for the fact that the normalization condition for the  $p$  kets is written as in (4.84a) for infinite space or as  $\langle p | p' \rangle = \delta_{pp'}$  for an electron confined in a potential box (Sect. 3.6.1).

As a further example of a continuous eigenvalue spectrum, we consider the position operator  $\hat{x}$ . In Dirac notation, the eigenvalue equation for the position kets  $|x'\rangle$  with eigenvalues  $x'$  is written as

$$\hat{x}|x'\rangle = x'|x'\rangle. \quad (4.85)$$

The orthonormality of the position states is expressed as:

$$\langle x | x' \rangle = \delta(x - x'). \quad (4.86)$$

Equation (4.86) can also be understood as the position representation of a position state  $|x'\rangle$ , namely the projection of  $|x'\rangle$  on the position basis vector  $|x\rangle$ . The delta function implies that in this case of a position measurement [eigenvalue equation (4.85)] the particle position is sharp, well defined, namely exactly at  $x'$ . For an expansion of a general quantum state  $|\psi\rangle$  in terms of position eigenstates  $|x'\rangle$ , we obtain:

$$|\psi\rangle = \int dx' g(x')|x'\rangle = \int dx' \langle x' | \psi \rangle |x'\rangle. \quad (4.87)$$

$g(x') = \langle x' | \psi \rangle$  is the probability amplitude to find an electron exactly at the position  $x'$ , i.e. the familiar wave function  $\psi(x')$  in the common Schrödinger picture. We, thus, learn that the wave function  $\psi(x)$  is nothing else but the representation of a general quantum state in the special Hilbert space of position eigenstates of the  $\hat{x}$  operator. In this context, the time-dependent wave function  $\psi(\mathbf{r}, t)$  of stationary states is written as

$$\psi(\mathbf{r}, t) = \langle \mathbf{r} | \psi, t \rangle, \quad \text{with } |\psi, t\rangle = e^{-iEt/\hbar} |\psi\rangle. \quad (4.88)$$

In Dirac notation the Schrödinger equation reads

$$i\hbar \frac{\partial}{\partial t} |\psi, t\rangle = \hat{H} |\psi, t\rangle. \quad (4.89)$$

In order to make contact to the familiar Schrödinger equation (3.51), we have to transform (4.89) into the position representation, i.e. we must multiply (4.89) from the left with the time-independent bra  $\langle x |$ , the position eigenvector. This means that we select from the general vector  $|\psi, t\rangle$  the particular component in  $\langle x |$  direction by projection on this direction:

$$i\hbar \frac{\partial}{\partial t} \langle x | \psi, t \rangle = \langle x | \hat{H} | \psi, t \rangle. \quad (4.90)$$

The left side of (4.90) the position representation  $\langle x | \psi, t \rangle$  already represents the time-dependent wave function  $\psi(x, t)$ . On the right side, we apply the completeness relation

$$\int dx' |x'\rangle \langle x'| = \hat{1} \quad (4.91)$$

and obtain

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \int dx' \langle x | \hat{H} | x' \rangle \langle x' | \psi, t \rangle, \quad (4.92)$$

$\langle x | \hat{H} | x' \rangle$  is the Hamilton operator in matrix representation in the continuous Hilbert space of position eigenvectors  $|x\rangle$ . How does this matrix element look like? To evaluate the matrix element, we remember that the eigenvectors  $|x'\rangle$  of the position operator  $\hat{x}$  in position representation are given by delta functions  $\delta(x - x')$ . In analogy to the calculation of matrix elements  $\Omega_{mn}$  in discrete Hilbert spaces (4.29), (4.70), we construct the element as

$$\langle x | \hat{H} | x' \rangle = \int d\xi \delta(x - \xi) \hat{H}(\xi) \delta(x' - \xi), \quad (4.93)$$

where  $\hat{H}(\xi)$  is the familiar Hamilton operator (3.50) for a particle. Using the rules for calculations with delta functions (4.81a), (4.81b) we obtain, by inserting (4.93) into (4.92), the following result:

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \psi(x, t) &= \int dx' d\xi \delta(x - \xi) \hat{H}(\xi) \delta(x' - \xi) \psi(x', t) \\ &= \int dx' \delta(x - x') \hat{H}(x') \psi(x', t) = \hat{H}\psi(x, t). \end{aligned} \quad (4.94)$$

As we have expected, this is the familiar Schrödinger equation (3.51). The calculation clearly demonstrates, that the Schrödinger equation is nothing else but the position representation of the equation (4.89) in generalized Dirac notation.

### 4.3.5 Time Evolution in Quantum Mechanics

A physical theory must allow predictions for the future, once the present state of a system is sufficiently well defined. In classical mechanics Newton's dynamical laws or the Hamilton equations (Sect. 3.4) predict the future development of a mechanical system of mass points with known initial conditions (position and velocity) unequivocally. In quantum mechanics the Schrödinger equation (4.89) yields the necessary information about the time evolution of a quantum state  $|\psi, t\rangle$ . Since in (4.89) the first time derivative appears, only the initial state  $|\psi, t = 0\rangle$  must be known as initial condition for the description of future states. Although the wave function allows only statistical statements about the outcome of random physical events, the prediction of the wave function, that is, the basis of the statistical description, is deterministic and well defined by means of the Schrödinger differential equation (4.89). In the Schrödinger picture, also in Dirac notation, the quantum state  $|\psi, t\rangle$  is time-dependent; one could also write  $|\psi(t)\rangle$ .

Assuming a time-independent Hamilton operator  $\hat{H}$  in the Schrödinger equation (4.89), we obtain from

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle \quad (4.95)$$

by formal integration

$$|\psi(t)\rangle = \exp\left(\frac{-i}{\hbar} \hat{H}t\right) |\psi(0)\rangle = \hat{U} |\psi(0)\rangle. \quad (4.96)$$

Hereby, we have treated operators and vectors in Hilbert space as if they were ordinary normal numbers. One can attribute an operator function  $f(\hat{\Omega})$  to a an ordinary function  $f(\omega)$ , if simultaneously the commutation rules of operators are respected. For a time-dependent Hamilton operator  $\hat{H}(t)$ , e.g.,  $\hat{H}(t_1)$  at a later time  $t_1 > t_0$  does not commute with  $\hat{H}(t_0)$ . Under those conditions (4.96) does not solve (4.95). The correspondence between (4.95) and (4.96) is due to the time independence of  $\hat{H}$ . The operator function  $\hat{U} = \exp(-i\hat{H}t/\hbar)$  is defined by the series expansion of the exponential function, that is, by sequential multiple application of the Hamiltonian  $\hat{H}$ :

$$\begin{aligned} \hat{U} &= \exp(-i\hat{H}t/\hbar) \\ &= 1 - \frac{it}{\hbar} \hat{H} + \frac{1}{2!} \left(\frac{-it}{\hbar}\right)^2 \hat{H} \hat{H} + \frac{1}{3!} \left(\frac{-it}{\hbar}\right)^3 \hat{H} \hat{H} \hat{H} \dots \end{aligned} \quad (4.97)$$

Since usually normal functions can be expressed in terms of series expansions, the corresponding operator functions are defined as well by these series, analogously to (4.97). Important examples are the operator functions  $\sin \hat{\Omega}$  and  $\cos \hat{\Omega}$  (Sect. 5.6.2).

The operator  $\hat{U}$  (4.97), which describes the temporal evolution of  $|\psi(0)\rangle$  into  $|\psi(t)\rangle$ , is illustratively called *propagator*. What can we learn about this propagator?

$\hat{H}$  is an Hermitian operator, i.e. according to (4.12a) it obeys the relation

$$\int d^3r \varphi^* \hat{H} \psi = \int d^3r (\hat{H}\varphi)^* \psi. \quad (4.98)$$

The operator  $\hat{U}$ , now, is represented in a series with multiple application of  $\hat{H}$  (4.97). Therefore,  $\hat{U}$  obeys the relation (4.98), too, it is Hermitian. In addition it is obvious that  $\hat{U}^+ \hat{U} = \hat{1}$ , that is, the propagator  $\hat{U}$  is a unitary operator. According to Sect. 4.3.2, the temporal evolution of a quantum state  $|\psi(t)\rangle$  is, thus, described by a rotation of the corresponding state vector in Hilbert space. The vector  $|\psi(0)\rangle$  is rotated during the time  $t$  into  $|\psi(t)\rangle$ , while the length of the vector, the modulus, stays constant:

$$\langle \psi(t) | \psi(t) \rangle = \langle \psi(0) | \psi(0) \rangle. \quad (4.99)$$

In this context, it should be emphasized again, that in spite of a time-independent Hamiltonian  $\hat{H}$ , that is, a stationary physical problem, the state vectors (wave functions) are dependent on time according to (4.88). They are usually represented in “coordinate systems” of the time-independent eigenvectors  $|n\rangle$  of the Hamilton operator. This is a procedure familiar to us from classical dynamics in 3D Euclidean space where motions of particles are described in time-independent coordinate systems. Imagine how complicated the description of the change of a 3D vector would be in a time varying coordinate system.

In order to elucidate the time evolution in quantum mechanics, a little bit more we start with the representation of the propagator  $\hat{U}$  in terms of its matrix elements  $U_{nm}$ :

$$\begin{aligned} \hat{U} &= \sum_{nm} |n\rangle \langle n| e^{-\frac{i}{\hbar} \hat{H} t} |m\rangle \langle m| = \sum_{nm} |n\rangle U_{nm} \langle m| \\ &= \sum_{nm} |n\rangle e^{-\frac{i}{\hbar} E_m t} \langle n|m\rangle \langle m| = \sum_n e^{-\frac{i}{\hbar} E_n t} |n\rangle \langle n|. \end{aligned} \quad (4.100)$$

Hereby  $|n\rangle$  are the eigenkets of the  $\hat{H}$  operator and the orthogonality  $\langle n|m \rangle = \delta_{nm}$  is used.

The time evolution of a quantum state corresponds to a rotation of its state vector in Hilbert space. An appropriate economical description of the time evolution of a state might, therefore, be possible if the coordinate system (eigenstate basis) also rotates, but with the same speed as do the vectors of the quantum states. In such a rotating basis the state vectors would appear as “frozen” or fixed, in contrast to  $|\psi(t)\rangle$ . In contrast to the formalism used so far, the operators attributed to a particular measurement will become time-dependent. This change in the formal description of a measurement must, of course, leave all measurement results, in particular, the expectation values  $\langle \psi | \hat{\mathcal{Q}} | \psi \rangle$  unchanged.

We, therefore, calculate this expectation value by use of (4.96):

$$\langle \psi(t) | \hat{\mathcal{Q}} | \psi(t) \rangle = \langle \psi(0) | e^{\frac{i}{\hbar} \hat{H} t} \hat{\mathcal{Q}} e^{-\frac{i}{\hbar} \hat{H} t} | \psi(0) \rangle. \quad (4.101)$$

In this expression (4.101), we have shifted the formal description of the time evolution from the state vector  $|\psi(t)\rangle$  to the operator which is now time-dependent. The time-dependent operator characterized by a double roof symbol is obtained as

$$\hat{\hat{\Omega}}(t) = e^{\frac{i}{\hbar} \hat{H}t} \hat{\Omega} e^{-\frac{i}{\hbar} \hat{H}t}. \quad (4.102)$$

In this formalism the state vectors  $|\psi(0)\rangle$  do not depend on time any more, they coincide with the initial state vectors (4.96) used so far.

Since  $\hat{H}$  commutes with itself, it also commutes with  $\exp(i\hat{H}t/\hbar)$  (multiple application of  $\hat{H}$ ) and we obtain the time-dependent Hamilton operator as:

$$\hat{H} = e^{\frac{i}{\hbar} \hat{H}t} \hat{H} e^{-\frac{i}{\hbar} \hat{H}t} = \hat{H}. \quad (4.103)$$

The Hamilton operator itself is not affected by the change of formalism, it remains unchanged upon switching from time-dependent state vectors to time-dependent operators. It is not important for the underlying physics if the coordinate system of eigenkets or if the state vectors  $|\psi\rangle$  rotate.

In the formalism of time-dependent state vectors  $|\psi(t)\rangle$  the fundamental dynamic equation is the Schrödinger equation (4.89). The picture of time-dependent operators requires an equivalent dynamic equation, which controls the time evolution in quantum physics. For its derivation, we differentiate a general time dependent operator  $\hat{\hat{\Omega}}$  (4.102) with respect to time:

$$\begin{aligned} \frac{d}{dt} \hat{\hat{\Omega}} &= \frac{d}{dt} (e^{\frac{i}{\hbar} \hat{H}t} \hat{\hat{\Omega}} e^{-\frac{i}{\hbar} \hat{H}t}) = \frac{i}{\hbar} (\hat{H} \hat{\hat{\Omega}} - \hat{\hat{\Omega}} \hat{H}), \\ \boxed{\frac{d}{dt} \hat{\hat{\Omega}} = \frac{i}{\hbar} [\hat{H}, \hat{\hat{\Omega}}].} \end{aligned} \quad (4.104)$$

In case,  $\hat{\hat{\Omega}}$  contains an additional explicit time dependence, for example, due to a time-dependent potential, a further term  $(i/\hbar)\partial \hat{\hat{\Omega}}/\partial t$  has to be added in (4.104).

The dynamical equation (4.104) for time-dependent operators is called *Heisenberg equation*. The time derivative of an operator is given, there, by the commutator of that operator and the Hamilton operator.

An important consequence of the dynamical equation (4.104) concerns the time evolution of an operator which commutes with the Hamilton operator. For this case, we conclude  $\hat{\hat{\Omega}} = \hat{\Omega}$  and  $d\hat{\Omega}/dt = 0$ . The observable belonging to that operator is a *constant of the motion*.

In later sections about many-body systems and the quantization of fields, for example, the electromagnetic field (Chap. 8), we will encounter essentially time-dependent operators and timely fixed state vectors. In this context, the Heisenberg picture of quantum mechanics is of paramount importance. For economical reasons,

we will then denote time-dependent operators (4.102) with only one roof symbol as usually. From the context it is easily seen whether time-dependent Schrödinger state vectors  $|\psi(t)\rangle$  or time-dependent Heisenberg operators are concerned.

## 4.4 Games with Operators: The Oscillator

The *harmonic oscillator* is a model system with great importance in many branches of physics, both in classical mechanics and in quantum physics. Classical oscillating systems are, for example, the pendulum, a vibrating violin string, a vibrating rod fixed at one end or a mass coupled by a spring to a support. In quantum physics, typical examples of an oscillator are atomic vibrations in molecules and in solids or of ions in magnetic and electrostatic potential wells. Furthermore, the dynamic equation of the harmonic oscillator, both of the classical and of the quantum mechanical one, can be solved analytically and the solution is represented in compact analytic form. The oscillator model, thus, yields a simple system, where we can train our ability to apply operators and the commutation rules of quantum mechanics. “Playing” with operators in this context will supply us with some feeling what is it all about with quantum calculations.

### 4.4.1 The Classical Harmonic Oscillator

In its simplest form, the classical harmonic oscillator is realized by an oscillating (in one direction: 1D) mass which is coupled to a fixed position by a spring. For small elongations of the mass Hook’s law yields a proportionality between elongation  $x$  and restoring force  $F$ , that is,  $F = -kx$ . In Newton’s mechanics, the dynamic equation follows as

$$m\ddot{x} = -kx \quad (4.105a)$$

respectively

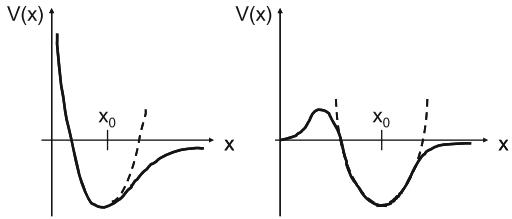
$$\ddot{x} + \omega^2 x = 0. \quad (4.105b)$$

Hereby,  $\omega = \sqrt{k/m}$  is the oscillation frequency of the oscillator. The force law  $F = -kx$  implies a potential  $V = (1/2)kx^2$  for the moving mass  $m$  and thus a Hamilton function

$$H = T + V = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2. \quad (4.106)$$

Equation (4.106) describes the oscillator equally well as (4.105a), (4.105b). Application of the Hamilton equations (3.26a), (3.26b) on (4.106) leads directly back to Newton’s dynamic equations (4.105a), (4.105b). This correspondence of Newton’s and Hamilton’s mechanics was already considered in connection with the pendulum in Sect. 3.4.

**Fig. 4.3** Two different binding potentials  $V(x)$  with minima at  $x_0$ . Around the minima positions the potentials can be approximated by parabolas (broken line)



The great importance of the oscillator model in physics derives from the general applicability of (4.105a), (4.105b) and (4.106) to any problem where a particle is locked (confined) in an arbitrary potential  $V(x)$  near a minimum at  $x_0$  (Fig. 4.3). The particle can oscillate around the potential minimum and for sufficiently small oscillation amplitudes a description in terms of the oscillator model is possible. We can see this quite easily: Every potential with a minimum at  $x_0$  can be expanded in a Taylor series around  $x_0$ :

$$V(x) = V(x_0) + \frac{dV}{dx} \Big|_{x_0} (x - x_0) + \frac{1}{2!} \frac{d^2V}{dx^2} \Big|_{x_0} (x - x_0)^2 + \dots \quad (4.107)$$

The first constant term  $V(x_0)$  can be defined as the zero point of the energy scale and the second term, the first derivative  $dV/dx$ , vanishes at the minimum. Therefore, the first non-negligible term in the Taylor series (4.107) is quadratic in  $(x - x_0)$ , all following terms are significantly smaller and can be neglected to first approximation. The resulting approximated potential is that of the harmonic oscillator (after setting  $x_0 = 0$ ). The harmonic oscillator model thus opens the possibility to approximately treat a wide class of physical problems, namely fluctuations of a particle around a potential minimum.

A short additional remark: The total energy  $E$  contained in an oscillation of the harmonic oscillator is a constant of the motion. With  $x(t) = A \cos(\omega t + \varphi)$  as the general solution of (4.105b) the total energy  $E$  is obtained as:

$$E = T + V = \frac{1}{2} m \dot{x}^2 + \frac{1}{2} m \omega^2 x^2 = \frac{1}{2} m \omega^2 A^2. \quad (4.108)$$

Hereby,  $A$  is the maximum oscillation amplitude, which is a constant for a particular oscillation.

#### 4.4.2 Upstairs-Downstairs: Step Operators and Eigenvalues

For the quantum mechanical treatment of the harmonic oscillator we start, as usually, from the classical Hamilton function (4.106) and substitute the canonically conjugated variables  $p$  and  $x$  by the corresponding operators  $\hat{p}$  and  $\hat{x}$ , to obtain the

Hamilton operator:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2 = \hbar\omega\left(\frac{\hat{p}^2}{2m\hbar\omega} + \frac{m\omega}{2\hbar}\hat{x}^2\right). \quad (4.109)$$

For convenience reasons, the energy is expressed in terms of  $\hbar\omega$ , a familiar notation in quantum physics (particle-wave duality, Sect. 3.1). The direct, but tedious way to solve the time-independent oscillator Schrödinger equation for eigenvalues and eigenvectors, would be the representation of (4.109) in the position basis (Sect. 4.3.4). Then,  $\hat{x}$  would mean multiplication by the  $x$  coordinate and  $\hat{p}$  differentiation with respect to  $x$ . Solution of the Schrödinger equation (3.57) results in solving a differential equation. Qualitatively the outcome of the calculation is obvious. The oscillator potential, at least approximately a parabola with a minimum at the bottom (Fig. 4.3), is a confining potential which locks up a particle. Correspondingly a discrete eigenvalue spectrum is expected, where eigenvalues and eigenfunctions can be numerated with integer numbers  $n$  (Sect. 3.6.1).

Rather than following this standard procedure we prefer an elegant solution method based on a very general commutator algebra, which is furthermore of paramount importance when it comes to field quantization and many body physics (Chap. 8). Only the general representation of state vectors and operators in Hilbert space and the commutation relations between  $\hat{p}$  and  $\hat{x}$  (4.26) are used for the solution of the problem.

Using the binomial formula

$$(\alpha + i\beta)(\alpha - i\beta) = \alpha^2 + \beta^2 \quad (4.110)$$

we factorize the Hamilton operator (4.109) by defining first the following operators being equivalent to  $\hat{x}$  and  $\hat{p}$ :

$$\hat{\alpha} = \sqrt{\frac{m\omega}{2\hbar}}\hat{x} \quad \text{and} \quad \hat{\beta} = \sqrt{\frac{1}{2m\hbar\omega}}\hat{p}. \quad (4.111)$$

Then, the factorization is performed by using (4.111) and defining the new operators  $\hat{b} = \hat{\alpha} + i\hat{\beta}$  and  $\hat{b}^+ = \hat{\alpha} - i\hat{\beta}$  as

$$\hat{b} = \sqrt{\frac{m\omega}{2\hbar}}\hat{x} + i\sqrt{\frac{1}{2m\hbar\omega}}\hat{p}, \quad (4.112a)$$

$$\hat{b}^+ = \sqrt{\frac{m\omega}{2\hbar}}\hat{x} - i\sqrt{\frac{1}{2m\hbar\omega}}\hat{p}. \quad (4.112b)$$

By multiplication and by use of the commutation relation  $[\hat{x}, \hat{p}] = i\hbar$  we conclude from (4.112a), (4.112b):

$$\hat{b}^+ \hat{b} = \frac{1}{\hbar\omega} \frac{\hat{p}^2}{2m} + \frac{1}{\hbar\omega} \frac{m\omega^2}{2} \hat{x}^2 + \frac{i}{2\hbar} [\hat{x}, \hat{p}], \quad (4.113a)$$

that is, by means of (4.109)

$$\hat{b}^+ \hat{b} = \frac{1}{\hbar\omega} \hat{H} - \frac{1}{2}. \quad (4.113b)$$

From the definition of the operators  $\hat{b}$  and  $\hat{b}^+$  (4.112a), (4.112b) and from the commutation relation for  $\hat{x}$  and  $\hat{p}$ , we gain the commutation relation between  $\hat{b}$  and  $\hat{b}^+$ :

$$[\hat{b}, \hat{b}^+] = \hat{b}\hat{b}^+ - \hat{b}^+\hat{b} = 1. \quad (4.114)$$

Since  $\hat{x}$  and  $\hat{p}$  are Hermitian (self-adjoint) operators and since  $(\hat{b}^+)^* = \hat{b}$ ,  $\hat{b}^+$  is the adjoint operator to  $\hat{b}$  (4.71). Furthermore, the Hamilton operator (4.109) is obtained as

$$\hat{H} = \left( \hat{b}^+ \hat{b} + \frac{1}{2} \right) \hbar\omega. \quad (4.115)$$

As the eigensolutions to  $\hat{H}$  span a Hilbert space with discrete numerable eigenstates  $|n\rangle$ , the time-independent Schrödinger equation for the harmonic oscillator must be written as

$$\hat{H}|n\rangle = \hbar\omega \left( \hat{b}^+ \hat{b} + \frac{1}{2} \right) |n\rangle = E_n |n\rangle. \quad (4.116)$$

From the properties of the operators  $\hat{b}$  and  $\hat{b}^+$  (4.112a–4.114) we, now, derive all essential issues about eigenvalues  $E_n$  and eigenstates  $|n\rangle$  of the harmonic oscillator.

- (i) For any general eigenstate  $|n\rangle$  of the eigenvalue equation (4.116), there is a vector  $\hat{b}|n\rangle$ .

As  $\hat{b}^+$  is the adjoint operator to  $\hat{b}$  the modulus (length) of the vector  $\hat{b}|n\rangle$ , namely

$$\langle n | \hat{b}^+ \hat{b} | n \rangle \geq 0 \quad (4.117)$$

must be positive (or zero). This is in direct analogy to 3D space, where the modulus of a vector is  $|\mathbf{a}|^2 = \sum_{i=1}^3 a_i^2 \geq 0$ . From (4.117) and (4.116), we instantly conclude that all eigenvalues  $E_n$  of the harmonic oscillator must be positive.

- (ii) We examine the action of the operator  $\hat{b}^+$  on an eigenstate  $|n\rangle$  of (4.116) by applying  $\hat{b}^+$  on (4.116) from the left:

$$\hat{b}^+ \hat{H} |n\rangle = \hbar\omega \hat{b}^+ \left( \hat{b}^+ \hat{b} + \frac{1}{2} \right) |n\rangle = E_n \hat{b}^+ |n\rangle. \quad (4.118a)$$

On the left side of the equation, we use the commutation relation (4.114), that is,  $\hat{b}^+ \hat{b} + \hat{b} = \hat{b}^+ (\hat{b} \hat{b}^+ - 1)$  and obtain

$$\hbar\omega \left( \hat{b}^+ \hat{b} \hat{b}^+ - \frac{1}{2} \hat{b}^+ \right) |n\rangle = E_n \hat{b}^+ |n\rangle,$$

respectively

$$\hbar\omega \left( \hat{b}^+ \hat{b} - \frac{1}{2} \right) \hat{b}^+ |n\rangle = E_n \hat{b}^+ |n\rangle. \quad (4.118b)$$

Together with (4.115), this yields:

$$\hat{H}(\hat{b}^+ |n\rangle) = (E_n + \hbar\omega)(\hat{b}^+ |n\rangle). \quad (4.119)$$

This is again an eigenvalue equation, analogously to (4.116), but now for the eigenstate  $\hat{b}^+ |n\rangle$  with the eigenvalue  $E_n + \hbar\omega$ . Application of the operator  $\hat{b}^+$  on any eigenstate  $|n\rangle$  transforms this state into the next higher one  $|n+1\rangle$  with an eigenvalue augmented by an amount  $\hbar\omega$  in comparison to the former one. Sequential application of  $\hat{b}^+$  on an eigenstate, thus leads to higher and higher eigenstates. Each time the eigenvalues increase by an amount  $\hbar\omega$  (staircase: upstairs), that is,  $E_{n+1} = E_n + \hbar\omega$ .

- (iii) In analogy, application of the operator  $\hat{b}$  on an eigenstate  $|n\rangle$  with eigenvalue  $E_n$  leads to the next lower eigenstate  $|n-1\rangle$  with eigenvalue  $(E_n - 1)$  (staircase: downsteps), as is easily shown:

$$\begin{aligned} \hat{b} \hat{H} |n\rangle &= \hbar\omega \hat{b} \left( \hat{b}^+ \hat{b} + \frac{1}{2} \right) |n\rangle = E_n \hat{b} |n\rangle, \\ \hbar\omega \left( \hat{b}^+ \hat{b} + \frac{3}{2} \right) \hat{b} |n\rangle &= E_n \hat{b} |n\rangle, \\ \hat{H}(\hat{b} |n\rangle) &= (E_n - \hbar\omega)(\hat{b} |n\rangle). \end{aligned} \quad (4.120)$$

We have again used (4.115) and the commutation relation (4.114). According to (4.120)  $\hat{b} |n\rangle$  is the eigenvector  $|n-1\rangle$  with eigenvalue  $E_{n-1} = E_n - \hbar\omega$ . Starting from an eigensolution  $|n\rangle$  with eigenvalue  $E_n$  all higher and lower eigenvalues  $\dots, E_{n-2}, E_{n-1}, E_n, E_{n+1}, E_{n+2}, \dots$  can be calculated by adding or subtracting each time the energy quantum  $\hbar\omega$ . The spectrum of eigenvalues is a series (ladder) of energetically equidistant levels differing in energy by  $\hbar\omega$ , that is,  $E_{n\pm 1} = E_n \pm \hbar\omega$ . The operators  $\hat{b}$  and  $\hat{b}^+$  are, therefore, appropriately called *step operators* or sometimes jump operators.

- (iv) Upwards, the action of  $\hat{b}^+$  is not limited. The ladder of eigenvalues  $n\hbar\omega$  runs to infinitely high values, since the parabolic potential of the ideal harmonic oscillator  $V(x) \propto x^2$  has no upper limit. Downwards, however, a lower limit must exist for the eigenvalues, since all eigenvalues must be positive according

to (i). Therefore, a lowest eigenvalue  $E_0$  must exist, which belongs to the ground state  $|0\rangle$  (no excitation) of the oscillator. For this ground state the eigenvalue equation (4.116) reads

$$\hat{H}|0\rangle = E_0|0\rangle. \quad (4.121a)$$

Using the expression (4.115) for  $\hat{H}$ , we get

$$\hbar\omega\left(\hat{b}^+\hat{b} + \frac{1}{2}\right)|0\rangle = E_0|0\rangle. \quad (4.121b)$$

For the ground state  $|0\rangle$ , we have  $\hat{b}|0\rangle = 0$ , that is, the lowest ground state energy of the harmonic oscillator amounts to  $E_0 = (1/2)\hbar\omega$ . All higher energy eigenvalues are obtained by sequential addition of a quantum  $\hbar\omega$  each time. The spectrum of the oscillator eigenvalues is, thus, given by

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega. \quad (4.122)$$

- (v) An analytical form of the ground state wave function of the oscillator is easily obtained by representing the corresponding eigenvector equation  $\hat{b}|0\rangle = 0$  in the position vector basis. In this representation  $\langle x|0\rangle$  is the ground state wave function  $\varphi_0(x)$  and  $\hat{b}$  is essentially a multiplication of the wave function by  $x$  and a position derivative ( $\hat{p} = (\hbar/i)d/dx$ ), that is,  $\hat{b}|0\rangle = 0$  is expressed as

$$\left(\sqrt{\frac{m\omega}{2\hbar}}x + \sqrt{\frac{\hbar}{2m\omega}}\frac{d}{dx}\right)\varphi_0(x) = 0. \quad (4.123a)$$

After short calculation, we get

$$\frac{d\varphi_0}{\varphi_0} = -\frac{m\omega}{\hbar}x \, dx, \quad (4.123b)$$

that is, after integration the ground state wave function of the harmonic oscillator:

$$\varphi_0(x) = C \exp\left(-\frac{m\omega}{2\hbar}x^2\right), \quad (4.124)$$

with  $C = (m\omega/\pi\hbar)^{1/4}$  as normalization constant. As expected, this wave function, a Gaussian bell function, describes the localization of the oscillating particle within a spatially limited region due to the confining parabolic potential.

- (vi) Because of their property as step (or jump) operators (4.119, 4.120)  $\hat{b}$  and  $\hat{b}^+$  transform an eigenstate  $|n\rangle$  into the next lower and next higher state, respectively, that is, they obey the relations

$$\hat{b}|n\rangle = c_n|n-1\rangle, \quad (4.125a)$$

$$\hat{b}^+|n\rangle = c'_n|n+1\rangle. \quad (4.125b)$$

Here,  $c_n$  and  $c'_n$  are normalization constants which have some importance in calculations. For their determination, we use that  $\hat{b}^+$  and  $\hat{b}$  are mutually adjoint and derive from (4.125a):

$$\langle n|\hat{b}^+ = \langle n-1|c_n^*. \quad (4.125c)$$

Equations (4.125a)–(4.125c) directly yield:

$$\langle n|\hat{b}^+\hat{b}|n\rangle = \langle n-1|n-1\rangle c_n^*c_n, \quad (4.126a)$$

and because of the normalization  $\langle n-1|n-1\rangle = 1$  and (4.115):

$$\langle n|\frac{1}{\hbar\omega}\hat{H} - \frac{1}{2}|n\rangle = |c_n|^2. \quad (4.126b)$$

A combination of the  $\hat{H}$  operator (4.115) with its eigenvalues (4.122) yields the following expression for the eigenvalue equation (4.116):

$$\hbar\omega\left(\hat{b}^+\hat{b} + \frac{1}{2}\right)|n\rangle = \hbar\omega\left(n + \frac{1}{2}\right)|n\rangle, \quad (4.127)$$

that is, application of  $\hat{b}^+\hat{b}$  on the eigenstate  $|n\rangle$  yields the quantum number  $n$  of the corresponding state:

$$\hat{b}^+\hat{b}|n\rangle = n|n\rangle. \quad (4.128a)$$

$\hat{n} = \hat{b}^+\hat{b}$  is appropriately called *quantum number operator* of the harmonic oscillator.

Furthermore, from (4.126a) the normalization constant  $c_n$  is obtained as

$$\langle n|\hat{b}^+\hat{b}|n\rangle = n\langle n|n\rangle = n = |c_n|^2. \quad (4.128b)$$

An analogous calculation for the application of the operator  $\hat{b}^+$  on the eigenstate  $|n\rangle$  finally leads to important relations for the step operators  $\hat{b}$  and  $\hat{b}^+$ :

$\hat{b} n\rangle = \sqrt{n} n-1\rangle,$	(4.129a)
---	----------

$\hat{b}^+ n\rangle = \sqrt{n+1} n+1\rangle.$	(4.129b)
---	----------

- (vii) By means of (4.129a), (4.129b), a recursion formula for the calculation of a general eigenstate  $|n\rangle$  of the harmonic oscillator can be given. The operator  $\hat{b}^+$   $n$ -times applied on the ground state  $|0\rangle$  leads to the state  $|n\rangle$  [using (4.129b)]:

$$\begin{aligned}|n\rangle &= \frac{1}{\sqrt{n}} \hat{b}^+ |n-1\rangle = \frac{\hat{b}^+}{\sqrt{n}} \frac{\hat{b}^+}{\sqrt{n-1}} |n-2\rangle = \dots \\ &= \frac{1}{\sqrt{n!}} (\hat{b}^+)^n |0\rangle.\end{aligned}\quad (4.130)$$

The oscillator eigen(wave)function  $\varphi_n(x) = \langle x|n\rangle$  for the  $n$ th eigenvalue (4.122) is obtained by representing (4.130) in the position basis and using the position representation of  $\hat{b}^+$  (4.112b):

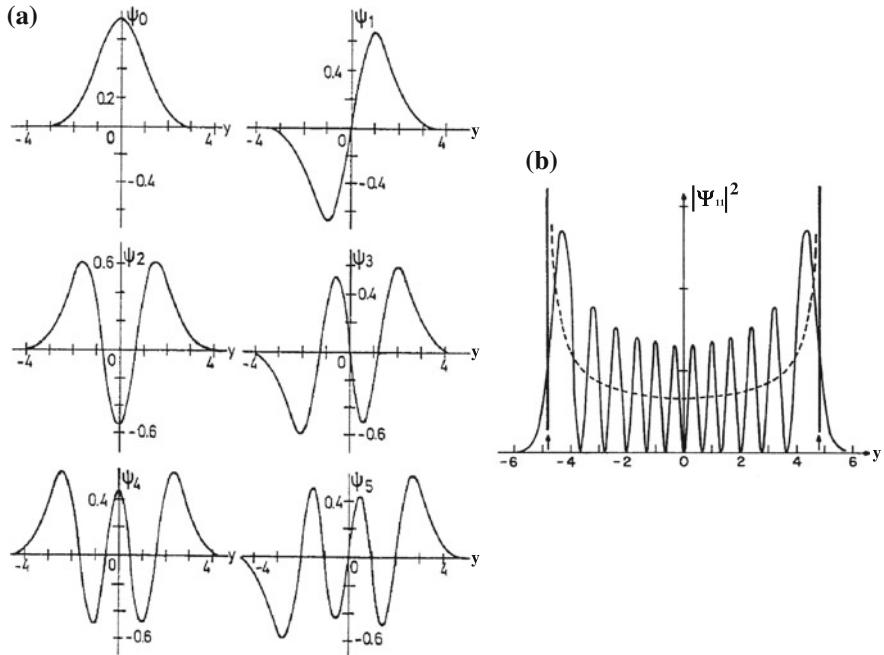
$$\begin{aligned}\langle x|n\rangle &= \varphi_n(x) = \frac{1}{\sqrt{n!}} \left( \sqrt{\frac{m\omega}{2\hbar}} x - \sqrt{\frac{\hbar}{2m\omega}} \frac{d}{dx} \right)^n \varphi_0(x) \\ &= \frac{1}{\sqrt{n!}} \left( \sqrt{\frac{m\omega}{2\hbar}} x - \sqrt{\frac{\hbar}{2m\omega}} \frac{d}{dx} \right)^n \sqrt[4]{\frac{m\omega}{\pi\hbar}} e^{-\frac{m\omega}{2\hbar}x^2}.\end{aligned}\quad (4.131)$$

In this expression, the wave function of the ground state (4.124) was used. Equation (4.131) is a recursion algorithm which allows the calculation of any wave function  $\varphi_n(x)$ . Computational programs as MAPLE can handle this job quite easily. Some low index eigenfunctions of the harmonic oscillator are plotted in Fig. 4.4.

Using the harmonic oscillator as an example we want to elucidate, one times more, the relation between classical and quantum mechanics. For this purpose, we consider a macroscopic classical pendulum, where a mass of 1 g oscillates back and forth one times per second. The oscillation amplitude is  $A = 2$  cm. The corresponding angular frequency  $\omega$ , then, amounts to about  $6\text{ s}^{-1}$ . Using the classical formula (4.108) for the total energy  $E_{\text{class}}$  being contained in the motion, we obtain  $E_{\text{class}} \approx 70 \times 10^{-7}\text{ kg m}^2/\text{s}^2$  from the values  $m = 1\text{ g}$ ,  $A = 2\text{ cm}$ ,  $\omega = 6\text{ s}^{-1}$ . If we compare this value with the energetic distance  $\Delta E = \hbar\omega = 6 \times 10^{-34}\text{ kg m}^2/\text{s}^2$  between two quantum mechanical excitation states (eigenvalues) of the pendulum, we realize the quasi-continuous character of the quantum states for a macroscopic system as the classical pendulum. Accordingly the quantum number  $n$  corresponding to the classical oscillation of the pendulum amounts to

$$n \approx E_{\text{class}} / \Delta E \approx 10^{28}. \quad (4.132)$$

The macroscopic pendulum, thus, is in an extremely high quantum mechanical excitation state. For extremely high quantum numbers quantum mechanics passes over into classical mechanics.



**Fig. 4.4** Wave functions of the eigenstates of the harmonic oscillator [2]. **a** Low index eigenfunctions  $\psi_0(y)$  up to  $\psi_5(y)$  with quantum numbers  $n = 0$  to  $n = 5$ . Instead of a representation as function of the  $x$ -coordinate a renormalized plot versus  $y = x/x_0 = x\sqrt{m\omega/\hbar}$  is chosen. **b** Probability for finding an electron  $|\psi_{11}|^2$  in the oscillator eigenstate with quantum number  $n = 11$ . The classical probability is plotted in *broken line*

#### 4.4.3 The Anharmonic Oscillator

In Sect. 4.4.1, we have discussed that the model of the harmonic oscillator with its potential being parabolic in  $x$ , the elongation, (4.106) can be considered as the first approximation for a general binding potential close to its minimum. In the next higher approximation of such a binding potential (4.107), we have to take into account also terms being cubic in the elongation ( $x - x_0$ ). For this further approximation, called the *anharmonic oscillator*, the Hamilton operator is written as:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 - \frac{1}{3!}g\hat{x}^3. \quad (4.133)$$

In the last term,  $g/3!$  describes the deviation from the parabolic potential, essentially the third derivative of the potential with respect to elongation (4.107). This term being cubic in the elongation is called the *anharmonic term* in the oscillator potential. The Hamiltonian (4.133), of course, is a good approximation only if  $|gx| \ll m\omega^2/2$  holds. In this case, the anharmonic term can be treated as a small perturbation of the

harmonic oscillator and the unperturbed eigenstates  $|n\rangle$  of the harmonic oscillator might be used for an approximate description. Accordingly we substitute the first two terms in (4.133) by the equivalent Hamiltonian of the harmonic oscillator (4.115)

$$\hat{H} = \hbar\omega\left(\hat{b}^+\hat{b} + \frac{1}{2}\right) - \frac{1}{3!}g\hat{x}^3. \quad (4.134)$$

To proceed with the third term in (4.133), we express the position operator  $\hat{x}$  in terms of step operators. From the representation of the step operators  $\hat{b}$  and  $\hat{b}^+$  in terms of  $\hat{x}$  and  $\hat{p}$  (4.112a), (4.112b) we obtain, by addition of (4.112a) and (4.112b), the representation of  $\hat{x}$  in terms of  $\hat{b}$  and  $\hat{b}^+$ :

$$\hat{x} = \frac{1}{2}\sqrt{\frac{2\hbar}{m\omega}}(\hat{b} + \hat{b}^+) = \sqrt{\frac{\hbar}{2m\omega}}(\hat{b} + \hat{b}^+). \quad (4.135)$$

Using (4.135) the Hamiltonian (4.134) is obtained as

$$\hat{H} = \hbar\omega\left(\hat{b}^+\hat{b} + \frac{1}{2}\right) - \frac{1}{3!}g\left(\frac{\hbar}{2m\omega}\right)^{\frac{3}{2}}(\hat{b} + \hat{b}^+)^3 = \hbar\omega\left(\hat{b}^+\hat{b} + \frac{1}{2}\right) - \hat{h}. \quad (4.136)$$

As the anharmonic part  $\hat{h}$  is a small perturbation of the harmonic oscillator, we continue to use the harmonic eigensolutions  $|n\rangle$  (4.130) with its eigenvalues (4.122). How the Hamiltonian (4.136) acts on the  $|n\rangle$  kets is defined by the recipe for the step operators (4.129a), (4.129b).

What is the change in physical behavior being introduced by the small anharmonic perturbation? For the harmonic oscillator the quantum number operator  $\hat{n} = \hat{b}^+\hat{b}$  commutes with the Hamilton operator (4.115) consisting essentially of  $\hat{n}$ . According to Heisenberg's dynamical equation (4.104) this implies that the quantum number  $n$  is a constant of the motion. The eigenstates  $|n\rangle$  are stationary states for the harmonic oscillator. In a harmonic potential a state  $|n\rangle$  once prepared stays unchanged forever. This is different for the anharmonic oscillator (4.136). The number operator  $\hat{n}$  certainly commutes with the first harmonic term in (4.136). The second anharmonic term being proportional to  $(\hat{b} + \hat{b}^+)^3$ , however, does not commute with  $\hat{n} = \hat{b}^+\hat{b}$ . This is easily seen from

$$[\hat{b}, \hat{b}^+\hat{b}] = \hat{b}\hat{b}^+\hat{b} - \hat{b}^+\hat{b}\hat{b} = \hat{b} \quad (4.137)$$

by use of the commutation relations (4.114).

For the anharmonic oscillator, we have  $[\hat{n}, \hat{H}] \neq 0$ , i.e. the quantum number  $n$  is not a constant of the motion. A small amount of anharmonicity in the oscillator potential gives rise to changes of the harmonic eigenstates during time. Eventhough  $|n\rangle$  are still good approximate solutions, they are not stationary anymore. Due to the anharmonicity the oscillator jumps between different harmonic eigenstates  $|n\rangle$ . What jumps are possible?

To answer that question one has to solve the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle \quad (4.138)$$

with the anharmonic oscillator Hamiltonian (4.136) for a general time-dependent state  $|\psi(t)\rangle$ . For this purpose, we expand the state  $|\psi(t)\rangle$  in terms of the time-independent eigenkets  $|n\rangle$  of the harmonic oscillator:

$$|\psi(t)\rangle = \sum_n c_n(t) |n\rangle. \quad (4.139)$$

The time dependence is contained in the amplitudes  $c_n(t)$ . They indicate how the eigenkets  $|n\rangle$  change during time due to the action of the perturbation  $\hat{h}$ . By means of (4.135), we evaluate the anharmonic perturbation  $\hat{h} = g\hat{x}^3/3!$  as a function of the step operators:

$$\begin{aligned} \hat{x}^3 &= \left(\frac{\hbar}{2m\omega}\right)^{3/2} (\hat{b} + \hat{b}^+)^3 \\ &= \left(\frac{\hbar}{2m\omega}\right)^{3/2} (\hat{b}\hat{b}\hat{b} + \hat{b}\hat{b}\hat{b}^+ + \hat{b}\hat{b}^+\hat{b} + \hat{b}\hat{b}^+\hat{b}^+ + \hat{b}^+\hat{b}\hat{b} + \hat{b}^+\hat{b}\hat{b}^+ \\ &\quad + \hat{b}^+\hat{b}^+\hat{b} + \hat{b}^+\hat{b}^+\hat{b}^+). \end{aligned} \quad (4.140)$$

In this calculation, the sequence of order of the step operators  $\hat{b}$  and  $\hat{b}^+$  must strictly be observed because of their commutation rules (4.114). Inserting (4.139) into (4.138) yields

$$i\hbar \sum_n \dot{c}_n(t) |n\rangle = \sum_n c_n(t) \hbar\omega \left( \hat{b}^+ \hat{b} + \frac{1}{2} \right) |n\rangle + \sum_n \hat{h} c_n(t) |n\rangle. \quad (4.141a)$$

Multiplication with the eigenbra  $\langle m|$  from left yields, because of  $\langle m|n\rangle = \delta_{mn}$ :

$$i\hbar \dot{c}_m(t) = \left(m + \frac{1}{2}\right) \hbar\omega c_m(t) + \sum_n \langle m | \hat{h} | n \rangle c_n(t), \quad (4.141b)$$

that is, a system of coupled differential equations, which describes the change of an amplitude  $c_m(t)$  due to its coupling with the other amplitudes  $c_n(t)$ .

By neglecting the anharmonicity  $\hat{h}$ , that is, taking into account only

$$i\hbar \dot{c}_m = \left(m + \frac{1}{2}\right) \hbar\omega c_m = E_m c_m \quad (4.142a)$$

we obtain, as expected, the stationary solution for the harmonic oscillator:

$$c_m(t) = c_m(0) \exp\left(-\frac{i}{\hbar} E_m t\right) \quad (4.142b)$$

with the energy eigenvalue  $E_m = (m + 1/2)\hbar\omega$ .

We, now, calculate the matrix elements  $\langle m | \hat{h} | n \rangle$  appearing in (4.141b). The perturbation operator  $\hat{h}$  is evaluated in terms of step operators according to (4.140):

$$\langle m | \hat{h} | n \rangle = \frac{1}{3!} g \left( \frac{\hbar}{2m\omega} \right)^{3/2} \langle m | \hat{b} \hat{b} \hat{b} + \hat{b} \hat{b} \hat{b}^+ + \hat{b} \hat{b}^+ \hat{b} + \dots | n \rangle. \quad (4.143)$$

The relations (4.129a), (4.129b) indicate how  $\hat{b}$  and  $\hat{b}^+$  act on  $|n\rangle$ . With  $\langle m | n \rangle = \delta_{mn}$  the matrix element  $\langle m | \hat{b} \hat{b} \hat{b} | n \rangle$  does not vanish only if

$$\langle m | \hat{b} \hat{b} \hat{b} | n \rangle = \langle m | n - 3 \rangle \sqrt{n(n-1)(n-2)} = \delta_{m,n-3} \sqrt{n(n-1)(n-2)} \quad (4.144a)$$

that is, if  $m = n - 3$ . Through this particular matrix element the quantum state  $|m\rangle$  with energy  $E_m = (m + 1/2)\hbar\omega$  only couples to a state  $|n\rangle$  whose energy exceeds that of  $|m\rangle$  by three quanta  $\hbar\omega$ . Beside others this state contributes to the decay of  $|m\rangle$  described by the time-dependent amplitude  $c_m(t)$ .

From analogous considerations, the matrix element

$$\langle m | \hat{b}^+ \hat{b}^+ \hat{b} | n \rangle = \langle m | n + 1 \rangle \sqrt{nn(n+1)} = \delta_{m,n+1} \sqrt{nn(n+1)} \quad (4.144b)$$

describes coupling of the state  $|m\rangle$  to a state  $|n\rangle$  whose energy is lower than  $E_m$  by one quantum  $\hbar\omega$  (because of  $m = n + 1$ , i.e.,  $n = m - 1$ ). For transitions into the ground state  $|0\rangle$  of the oscillator, this matrix element is obviously irrelevant.

The number of transitions from excited states into the ground state  $|0\rangle$  ( $m = 0$ ) of the oscillator is limited. From the whole variety of possible  $\hat{b}$  and  $\hat{b}^+$  combinations in (4.140), only the following matrix elements do not vanish:

$$\langle 0 | \hat{b} \hat{b} \hat{b} | 3 \rangle = \langle 0 | 0 \rangle \sqrt{6} = \sqrt{6}, \quad (4.145a)$$

$$\langle 0 | \hat{b} \hat{b}^+ \hat{b} | 1 \rangle = \langle 0 | 0 \rangle = 1, \quad (4.145b)$$

$$\langle 0 | \hat{b} \hat{b} \hat{b}^+ | 1 \rangle = \langle 0 | 0 \rangle \sqrt{4}. \quad (4.145c)$$

The excited states  $|3\rangle$  and  $|1\rangle$ , thus, contribute to the population of the ground state with probability ratios 6:1:4 (squared amplitudes).

The example of the anharmonic oscillator demonstrates how a small perturbation in the Hamiltonian induces transitions between different quantum states. We will encounter this idea in a much broader sense in the context of time-dependent perturbation theory (Sect. 6.4).

## References

1. P.A.M. Dirac, *Principles of Quantum Mechanics*, 4th edn. (Oxford University Press, London, 1958), ISBN 0-198-51208-2
2. F. Schwabl, *Quantenmechanik*, 2nd edn. (Springer, Berlin, 1990), p. 45

# Chapter 5

## Angular Momentum, Spin and Particle Categories

At first sight the motion of a particle on a curved path, in the simplest case a closed circle, might nothing be special; it should be described by the general law of particle dynamics (in Newtonian mechanics:  $K = mv$ ). Nevertheless, there are some peculiarities, even in classical mechanics, which require a stringent formal treatment because of phenomena not matching our everyday experience. The weird behavior of a gyroscope or the accelerating pirouettes of an ice-dancer are good examples.

We will see that on the atomic and subatomic level the *angular momentum*, the most interesting observable related to curved trajectories, is of paramount importance with far reaching consequences for our understanding of matter. The stability of matter, the structure of the periodic table of elements and the classification of elementary particles into two categories, fermions and bosons, are intimately related to properties of atomic angular momenta. Before starting the treatment of quantum mechanical angular momenta classical rotational dynamics shall briefly be reminded.

### 5.1 The Classical Circular Motion

A particle moving on a circular trajectory (Fig. 5.1) underlies scleronomic constraints (Sect. 3.4). These constraints might be due to an attractive potential as the Coulomb potential of the hydrogen nucleus or, in case of a macroscopic circulating body, a string which keeps the mass in constant distance from the pivot (see pendulum, Sect. 3.4). The corresponding force directed in string direction (Fig. 5.1) causes an acceleration of the mass towards the pivot, the so-called centripetal acceleration. The velocity  $v = ds/dt$  along the circular trajectory remains constant since the central force acts normally to the motion. Additionally to the central force there might be forces acting on the mass along the path, which accelerate the mass in direction of motion. Those forces have a component normal to the position vector  $\mathbf{r}$  of the mass, with  $|r|$  as the radius of the circular trajectory. These forces in the direction of motion are adequately described in terms of a *torque*  $\mathbf{D}$ .

**Fig. 5.1** Characteristic variables for the description of a circular motion. Note, that the infinitesimal velocity change  $\mathrm{d}\mathbf{v}$  is normal to the position change  $\mathrm{d}\mathbf{s}$  and directed towards the center of the circle

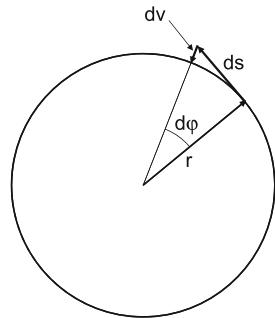


Figure 5.1 explains the calculation of the centripetal acceleration  $a = \mathrm{d}v/\mathrm{d}t$  perpendicular to the circular path and directed to the pivot: The infinitesimal change  $\mathrm{d}\mathbf{v}$  of the velocity is directed parallel to the radius vector  $\mathbf{r}$  and perpendicular to the velocity  $\mathbf{v}$  itself. Thus, the differential displacement along the path is

$$\mathrm{d}s = r \mathrm{d}\varphi = r\dot{\varphi} \mathrm{d}t = r\omega \mathrm{d}t \quad (5.1a)$$

and the velocity

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} = r\omega \quad (5.1b)$$

and finally with  $\mathrm{d}v = v \mathrm{d}\varphi$  (Fig. 5.1) the centripetal acceleration

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = v\omega = r\omega^2, \quad (5.1c)$$

$\omega$  is the angular velocity  $\mathrm{d}\varphi/\mathrm{d}t$ .

For the description in terms of three dimensional vectors it is convenient to attribute a vector  $\boldsymbol{\omega}$  to the angular velocity, which is oriented normal to the plane of the circular motion (upwards for positive angular change).

For the general treatment of rotational dynamics, we start with standard (translational) Newton's dynamic equation

$$\mathbf{K} = \frac{\mathrm{d}}{\mathrm{d}t}(m\mathbf{v}). \quad (5.2)$$

To take into account forces  $\mathbf{K}$  acting normal to the position vector  $\mathbf{r}$ , that is, causing an acceleration along the circular path, we define the *torque*  $\mathbf{D} = \mathbf{r} \times \mathbf{K}$  and obtain from (5.2):

$$\mathbf{D} = \mathbf{r} \times \mathbf{K} = \mathbf{r} \times \frac{\mathrm{d}}{\mathrm{d}t}(m\mathbf{v}). \quad (5.3a)$$

Since the vector cross product of  $\mathrm{d}\mathbf{r}/\mathrm{d}t = \mathbf{v}$  and  $\mathbf{v}$  vanishes, the following relation holds:

$$\frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \frac{d\mathbf{r}}{dt} \times m\mathbf{v} + \mathbf{r} \times \frac{d}{dt}(m\mathbf{v}) = \mathbf{r} \times \frac{d}{dt}(m\mathbf{v}). \quad (5.3b)$$

Thus, (5.3a) can be written as

$$\mathbf{D} = \mathbf{r} \times \mathbf{K} = \frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}). \quad (5.3c)$$

We define  $\mathbf{L} = (\mathbf{r} \times \mathbf{p})$  as the *angular momentum* of the circular motion and obtain the classical fundamental law of rotational dynamics as

$$\mathbf{D} = \mathbf{r} \times \mathbf{K} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \frac{d}{dt}\mathbf{L}. \quad (5.4)$$

The angular momentum  $\mathbf{L}$ , therefore, is a constant of the circular motion as long as no force acts normal to the position vector  $\mathbf{r}$  ( $\mathbf{D} = \mathbf{r} \times \mathbf{K} = \mathbf{0}$ ). This explains why an ice dancer ends up in an accelerating pirouette when she takes her arms close to her body:  $\mathbf{L} = \mathbf{r} \times \mathbf{p}$  must be conserved since no force acts in the direction of the circular motion, apart from the friction between ice and skates (neglected here). The lateral extension of the rotating body decreases (coordinate  $\mathbf{r}$ ), that is, the momentum  $\mathbf{p}$  must increase.

From its definition  $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ , the angular momentum is a vector which is directed normal to the plane of the circular motion similar as  $\boldsymbol{\omega}$ . By means of (5.1b), we can relate the moduli of  $\mathbf{L}$  and  $\boldsymbol{\omega}$  by

$$L = mr^2\omega. \quad (5.5a)$$

The quantity  $mr^2$  is called, in analogy to the inertial mass ( $p = mv$ ) for linear motion, *moment of inertia* of a mass  $m$  which rotates around a pivot at distance  $r$ .

Because of  $v = r\omega$  (5.1b) and (5.5a), the kinetic energy of a mass rotating on a circle of radius  $r$  is derived as

$$E_{\text{kin}} = \frac{1}{2}mv^2 = \frac{L^2}{2mr^2}. \quad (5.5b)$$

This relation will be useful in the following.

## 5.2 Quantum Mechanical Angular Momentum

Circular motion of an atomic or subatomic particle, for example, of an electron in the Coulomb potential of the proton in a hydrogen atom, must be described in terms of operators. The angular momentum operator  $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$  derives from the classical angular momentum  $\mathbf{L} = \mathbf{r} \times \mathbf{p}$  by substituting the classical variables  $\mathbf{r}$  and  $\mathbf{p}$  by their corresponding operators. Since  $\hat{\mathbf{r}}$  and  $\hat{\mathbf{p}}$  do not commute, we expect a non-trivial

commutation algebra for the components of the angular momentum. What do we expect qualitatively?

A circular motion is spatially confined. An electron in the Coulomb potential is “locked-up” as in a potential well. We, thus, expect discrete eigenvalues for the energy and consequently also for the angular momentum according to (5.5b). Furthermore, a circular motion can be decomposed into two mutually perpendicular harmonic oscillations. A similarity between the eigenvalues of the angular momentum and those of the harmonic oscillator (Sect. 4.4.3) is, thus, expected. A discrete eigenvalue spectrum of  $\hat{\mathbf{L}}$  is suggested, maybe even with energetically equidistant levels as for the oscillator.

For the mathematical treatment, we start with the definition of the angular momentum operator  $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$  and the representation of the momentum operator

$$\hat{\mathbf{p}} = \frac{\hbar}{i} \nabla = \frac{\hbar}{i} \left( \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z} \right), \quad (5.6)$$

$\mathbf{e}_x$ ,  $\mathbf{e}_y$  and  $\mathbf{e}_z$  are unity vectors in  $x$ ,  $y$ ,  $z$  direction. This leads directly to the following commutation relations for the vector components of the angular momentum operator:

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z, \quad (5.7a)$$

$$[\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x, \quad (5.7b)$$

$$[\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y. \quad (5.7c)$$

For (5.7c), we check this commutation as follows:

$$\begin{aligned} [\hat{L}_x, \hat{L}_y] &= \left( \frac{\hbar}{i} \right)^2 \left( y \frac{\partial}{\partial z} z \frac{\partial}{\partial x} - z \frac{\partial}{\partial x} y \frac{\partial}{\partial z} \right), \\ &= \left( \frac{\hbar}{i} \right)^2 \left( y \frac{\partial}{\partial x} + yz \frac{\partial^2}{\partial x \partial z} - yz \frac{\partial^2}{\partial x \partial z} \right), \\ &= \left( \frac{\hbar}{i} \right)^2 y \frac{\partial}{\partial x} = -i\hbar \left( y \frac{\hbar}{i} \frac{\partial}{\partial x} \right) = -i\hbar(-\hat{L}_z). \end{aligned} \quad (5.7d)$$

The commutation relations (5.7a)–(5.7d) can be written in a compact formula as

$$\hat{\mathbf{L}} \times \hat{\mathbf{L}} = i\hbar \hat{\mathbf{L}}. \quad (5.8)$$

by using the rules for calculating the vector cross product.

Since an operator always commutes with itself and because of the relation

$$[\hat{L}_x^2, \hat{L}_z] = \hat{L}_x [\hat{L}_x, \hat{L}_z] + [\hat{L}_x, \hat{L}_z] \hat{L}_x, \quad (5.9)$$

we conclude from (5.7a) to (5.7d)

$$\begin{aligned}
[\hat{\mathbf{L}}^2, \hat{L}_z] &= [\hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2, \hat{L}_z] = [\hat{L}_x^2 + \hat{L}_y^2, \hat{L}_z] \\
&= \hat{L}_x[\hat{L}_x, \hat{L}_z] + [\hat{L}_x, \hat{L}_z]\hat{L}_x + \hat{L}_y[\hat{L}_y, \hat{L}_z] + [\hat{L}_y, \hat{L}_z]\hat{L}_y \\
&= -i\hbar\hat{L}_x\hat{L}_y - i\hbar\hat{L}_y\hat{L}_x + i\hbar\hat{L}_y\hat{L}_x + i\hbar\hat{L}_x\hat{L}_y = 0.
\end{aligned} \tag{5.10}$$

The same result is obtained for the other components of the angular momentum:  $[\hat{\mathbf{L}}^2, \hat{L}_x] = [\hat{\mathbf{L}}^2, \hat{L}_y] = 0$ .

We learn that the components of the angular momentum operator do not commute (5.7a)–(5.7d), but that the absolute square  $\hat{\mathbf{L}}^2$  commutes with each single component  $\hat{L}_x, \hat{L}_y, \hat{L}_z$  of the angular momentum. Therefore, the three components of  $\hat{\mathbf{L}}$  separately have the same system of eigenfunctions as  $\hat{\mathbf{L}}^2$ . The following eigenvalue equations, thus, have to be solved:

$$\hat{\mathbf{L}}^2|l, m\rangle = \Lambda(l)\hbar^2|l, m\rangle, \tag{5.11}$$

$$\hat{L}_z|l, m\rangle = m\hbar|l, m\rangle. \tag{5.12}$$

Equations (5.11) and (5.12) are written in a convenient way for calculation: Since the angular momentum has the dimension of an action, the eigenvalues are expressed in units of Planck's constant  $\hbar$ . Furthermore, we need two different discrete quantum numbers  $\Lambda(l)$  or  $l$  and  $m$  (now without dimension) which are attributed to the different eigenvalue spectra of  $\hat{\mathbf{L}}^2$  and  $\hat{L}_z$ . Correspondingly the eigenvectors being the same for the two operators are denoted as  $|l, m\rangle$ .

First, the eigenvalue problem (5.11) is treated. Using (5.12), we can write down the following equations:

$$\hat{\mathbf{L}}^2|l, m\rangle = (\hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2)|l, m\rangle = \Lambda\hbar^2|l, m\rangle, \tag{5.13}$$

$$\hat{L}_z^2|l, m\rangle = \hat{L}_z\hat{L}_z|l, m\rangle = m^2\hbar^2|l, m\rangle. \tag{5.14}$$

By subtraction, one obtains:

$$(\hat{L}_x^2 + \hat{L}_y^2)|l, m\rangle = \hbar^2(\Lambda - m^2)|l, m\rangle. \tag{5.15}$$

We do not explicitly prove that the matrix element  $\langle l, m | \hat{L}_x^2 + \hat{L}_y^2 | l, m \rangle$  is positive, but this can reasonably be guessed because of the squared angular momentum components. We, thus, conclude

$$\langle l, m | \hat{L}_x^2 + \hat{L}_y^2 | l, m \rangle = \hbar^2(\Lambda - m^2)\langle l, m | l, m \rangle \geq 0 \tag{5.16}$$

and the eigenvalues of (5.11) and (5.12) must obey the relation

$$\Lambda(l) - m^2 \geq 0. \tag{5.17}$$

Because of (5.17) the eigenvalues  $m^2$ , respectively  $|m|$  of  $\hat{L}_z$  have an upper limit determined by the eigenvalues  $\Lambda(l)$  of the operator  $\hat{\mathbf{L}}^2$ . This is physically evident, since the angular momentum in a special direction  $z$  can not exceed the total angular momentum  $\sqrt{\hat{L}^2}$ .

In analogy to the harmonic oscillator (4.110)–(4.112b), we factorize the operator  $(\hat{L}_x^2 + \hat{L}_y^2)$  (5.15) by defining two new operators

$$\hat{L}_{\pm} = \hat{L}_x \pm i\hat{L}_y, \quad (5.18a)$$

and obtain

$$\hat{L}_x^2 + \hat{L}_y^2 = \hat{L}_+ \hat{L}_-. \quad (5.18b)$$

From (5.7a) to (5.7d) and (5.10), we derive the following commutation relations for the operators  $\hat{L}_+$  and  $\hat{L}_-$ :

$$[\hat{\mathbf{L}}^2, \hat{L}_{\pm}] = 0, \quad (5.19a)$$

$$[\hat{L}_z, \hat{L}_{\pm}] = \pm \hbar \hat{L}_{\pm}, \quad (5.19b)$$

$$[\hat{L}_{\pm}, \hat{L}_z] = \mp \hbar \hat{L}_{\pm}. \quad (5.19c)$$

By means of (5.19a) the eigenvalue equations (5.11) and (5.13), respectively, allow the conclusion

$$\hat{L}_{\pm} \hat{\mathbf{L}}^2 |l, m\rangle = \hbar^2 \Lambda(\hat{L}_{\pm}|l, m\rangle), \quad (5.20a)$$

$$\hat{\mathbf{L}}^2 (\hat{L}_{\pm}|l, m\rangle) = \hbar^2 \Lambda(\hat{L}_{\pm}|l, m\rangle) \quad (5.20b)$$

that is, if  $|l, m\rangle$  is an eigenstate to  $\hat{\mathbf{L}}^2$ , then the vectors  $\hat{L}_{\pm}|l, m\rangle$  are also eigenstates to the squared angular momentum operator.

A similar relation is derived for the eigenvalue equation (5.12) of the operator  $\hat{L}_z$  by using the action of  $\hat{L}_{\pm}$ :

$$\hat{L}_{\pm} \hat{L}_z |l, m\rangle = \hbar m |l, m\rangle. \quad (5.21a)$$

Because of (5.19c), we conclude

$$(\hat{L}_z \hat{L}_{\pm} \mp \hbar \hat{L}_{\pm}) |l, m\rangle = \hbar m |l, m\rangle, \quad (5.21b)$$

or

$$\hat{L}_z (\hat{L}_{\pm}|l, m\rangle) = \hbar(m \pm 1) (\hat{L}_{\pm}|l, m\rangle). \quad (5.21c)$$

$\hat{L}_{\pm}|l, m\rangle$  are, thus, eigenstates to the operator  $\hat{L}_z$ , the angular momentum in  $z$  direction, if  $|l, m\rangle$  is an eigenstate, but with eigenvalues  $m\hbar$  higher or lower by  $\hbar$  in comparison with those of  $|l, m\rangle$ .  $\hat{L}_{\pm}$  are step operators for the  $z$  component of the

angular momentum. They change the eigenvalues of  $\hat{L}_z$  by  $\pm\hbar$  as do the operators  $\hat{b}^+$  and  $\hat{b}$  for the oscillator eigenstates (Sect. 4.4.2). The dimensionless quantum number  $m$  always changes by an integer:

$$\hat{L}_{\pm}|l, m\rangle = |l, m \pm 1\rangle. \quad (5.22)$$

On the other hand, because of (5.17),  $m^2$  and  $|m|$  have an upper limit determined by the maximum total angular momentum in terms of the eigenvalue  $\Lambda(l)$ . For a fixed  $\Lambda$ , there are maximum and minimum values  $m_{\max}$  and  $m_{\min}$ , which obey the relations

$$\hat{L}_+|l, m_{\max}\rangle = 0, \quad (5.23a)$$

$$\hat{L}_-|l, m_{\min}\rangle = 0. \quad (5.23b)$$

Because of (5.22) ( $m_{\max} - m_{\min}$ ) must be an integer. Furthermore, the algebra of  $\hat{L}_x$ ,  $\hat{L}_y$  (5.7a)–(5.7d) and of  $\hat{L}_{\pm}$  (5.18a), (5.18b) requires:

$$\begin{aligned} \hat{L}_{\mp}\hat{L}_{\pm} &= \hat{L}_x^2 + \hat{L}_y^2 \mp \hat{L}_z\hbar \\ &= \hat{L}^2 - \hat{L}_z(\hat{L}_z \pm \hbar). \end{aligned} \quad (5.24)$$

Since  $|l, m\rangle$  is eigenstate both to  $\hat{L}^2$  and  $\hat{L}_z$ , application of (5.24) on  $|l, m_{\max}\rangle$  and  $|l, m_{\min}\rangle$  yields:

$$\hat{L}_-\hat{L}_+|l, m_{\max}\rangle = [\Lambda(l) - m_{\max}^2 - m_{\max}] \hbar^2 |l, m_{\max}\rangle = 0, \quad (5.25a)$$

$$\hat{L}_+\hat{L}_-|l, m_{\min}\rangle = [\Lambda(l) - m_{\min}^2 + m_{\min}] \hbar^2 |l, m_{\min}\rangle = 0. \quad (5.25b)$$

Both expressions vanish due to (5.23a), (5.23b); but because of the existence of  $|l, m_{\max}\rangle$  and  $|l, m_{\min}\rangle$  the brackets on the right side equal zero and we obtain:

$$\Lambda(l) = m_{\max}(m_{\max} + 1) = m_{\min}(m_{\min} - 1). \quad (5.26)$$

A simple calculation yields

$$m_{\max} = -m_{\min}. \quad (5.27)$$

The reasonable assumption  $l = m_{\max}$  yields the solution to the eigenvalue problems of the operators  $\hat{\mathbf{L}}^2$  and  $\hat{L}_z$  as

$$\hat{L}^2|l, m\rangle = l(l+1)\hbar^2 |l, m\rangle, \quad (5.28a)$$

$$\hat{L}_z|l, m\rangle = m\hbar |l, m\rangle, \quad (5.28b)$$

with  $m = -l, -l+1, -l+2, \dots, 0, \dots, l-1, l$ .

It is, thus, required that

$$l - v = -l,$$

respectively

$$l = v/2 \text{ holds with } v \text{ as integer.} \quad (5.28c)$$

This result needs further analysis. The quantum number  $m$  is called *magnetic quantum number*, since only in a magnetic field states with different orientations of the angular momentum in space (meaning of  $m$ ) can be distinguished with respect to their energy (see Sect. 5.4). A better name is *orientation quantum number*, a term which does only relate to the direct meaning of the orientation of the angular momentum in space. The orientation quantum number  $m$  can change only in integer steps up to maximum and minimum values  $l$  and  $-l$ . Hereby,  $l$  might also have half-integral values according to (5.28c). In this case, the angular momentum  $l = 0$  is excluded, since for integer steps of  $l$  change only  $l \pm 1/2$  and higher half integral values are allowed. The spectrum of eigenvalues  $l$  of the total angular momentum, therefore, divides into two distinct series:

$$l = 0, 1, 2, 3, \dots, \quad (5.29a)$$

$$l = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \dots \quad (5.29b)$$

In order to understand the deeper physical meaning of these eigenvalue spectra of the angular momentum, we refer to the position representation of the operators. For the description of circular motion spherical coordinates are best suited. The definition of spherical coordinates  $(r, \vartheta, \varphi)$  is explained in Fig. 5.2. With the unity vectors  $\mathbf{e}_r$ ,  $\mathbf{e}_\vartheta$ ,  $\mathbf{e}_\varphi$  in  $r$ ,  $\vartheta$  and  $\varphi$  direction the corresponding displacement elements are written as

$$ds_{\mathbf{r}} = dr \mathbf{e}_r, \quad (5.30a)$$

$$ds_\vartheta = r d\vartheta \mathbf{e}_\vartheta, \quad (5.30b)$$

$$ds_\varphi = r \sin \vartheta d\varphi \mathbf{e}_\varphi. \quad (5.30c)$$

The orthogonality of the spherical unit vectors requires

$$\mathbf{e}_r \times \mathbf{e}_\vartheta = \mathbf{e}_\varphi, \quad (5.31a)$$

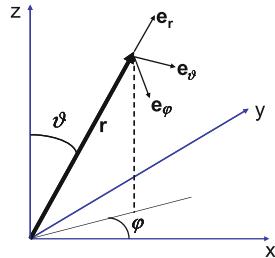
$$\mathbf{e}_r \times \mathbf{e}_\varphi = -\mathbf{e}_\vartheta, \quad (5.31b)$$

$$\mathbf{e}_\varphi \times \mathbf{e}_\vartheta = \mathbf{e}_r. \quad (5.31c)$$

The Nabla operator is, thus, represented in spherical coordinates as

$$\begin{aligned} \nabla &= \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\vartheta \frac{\partial}{\partial s} + \mathbf{e}_\varphi \frac{\partial}{\partial s_\varphi} \\ &= \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\vartheta \frac{1}{r} \frac{\partial}{\partial \vartheta} + \mathbf{e}_\varphi \frac{1}{r \sin \vartheta} \frac{\partial}{\partial \varphi}. \end{aligned} \quad (5.32)$$

**Fig. 5.2** Definition of spherical coordinates by means of their three directional unity vectors  $\mathbf{e}_r, \mathbf{e}_\vartheta, \mathbf{e}_\varphi$



Using (5.31a)–(5.31c) and (5.32), we calculate the angular momentum operator

$$\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}} = \frac{\hbar}{i} (\hat{\mathbf{r}} \times \nabla)$$

in spherical coordinates as follows:

$$\begin{aligned} \mathbf{r} \times \nabla &= \mathbf{r} \times \left( \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\vartheta \frac{1}{r} \frac{\partial}{\partial \vartheta} + \mathbf{e}_\varphi \frac{1}{r \sin \vartheta} \frac{\partial}{\partial \varphi} \right) \\ &= (\mathbf{e}_r \times \mathbf{e}_\vartheta) \frac{\partial}{\partial \vartheta} - (\mathbf{e}_r \times \mathbf{e}_\varphi) \frac{1}{\sin \vartheta} \frac{\partial}{\partial \varphi} \\ &= \mathbf{e}_\varphi \frac{\partial}{\partial \vartheta} - \mathbf{e}_\vartheta \frac{1}{\sin \vartheta} \frac{\partial}{\partial \varphi}, \quad \text{respectively} \\ \hat{\mathbf{L}} &= \frac{\hbar}{i} \left( \mathbf{e}_\varphi \frac{\partial}{\partial \vartheta} - \mathbf{e}_\vartheta \frac{1}{\sin \vartheta} \frac{\partial}{\partial \varphi} \right). \end{aligned} \quad (5.33)$$

According to Fig. 5.2  $(\mathbf{e}_\vartheta)_z = -\sin \vartheta$  and the  $z$  component of the angular momentum is evaluated as

$$\hat{L}_z = \frac{\hbar}{i} (\mathbf{r} \times \nabla)_z = \frac{\hbar}{i} \frac{\partial}{\partial \varphi}. \quad (5.34a)$$

In position representation  $\psi_{m,l} = \langle r | l, m \rangle$  the eigenvalue equation (5.28b) is, thus, obtained as

$$-i\hbar \frac{\partial}{\partial \varphi} \psi_{m,l} = m\hbar \psi_{m,l} \quad (5.34b)$$

that is, the eigenfunctions of  $\hat{L}_z$  in position representation must be

$$\psi_{m,l} \propto e^{im\varphi}. \quad (5.35a)$$

Since the wave function  $\psi_{m,l}$  must be unequivocal on a circle, a change of  $m\varphi$  by multiples of  $2\pi$  leads to the same function, that is, the orientation quantum number  $m$  must be integer. Half integral values for  $m$  are not allowed as eigenvalues to an angular momentum  $\hat{L}_z$  which corresponds to a circular motion of a mass in space. Angular momenta belonging to curved particle trajectories can have as measurement values only  $l(l+1)\hbar^2$  for  $\hat{L}^2$  and  $m\hbar$  ( $m$  integer) for  $\hat{L}_z$ . Hereby the following relations hold:

$$m = -l, -l+1, \dots, 0, \dots, l-1, l; \quad l = 0, 1, 2, \dots \quad (5.35b)$$

A corresponding plot for a total angular momentum  $L^2 = l(l+1)\hbar$  is shown in Fig. 5.3.

The question arises, if the half integral quantum numbers  $l$  of the total angular momentum (5.29b) are meaningless, if they are not realized in nature. We will see that this is not the case. There are angular momenta in nature which obey (5.29b), but they do not belong to motions on curved trajectories (Sect. 5.5). Conclusions from the general commutation algebra of the angular momentum operators are obviously more stringent than relations derived from the position representation of wave functions and operators.

Before we come to a more detailed discussion of these particular types of angular momenta with half-integral quantum numbers  $l$ , an important symmetry property related to the angular momentum shall be considered.

### 5.3 Rotational Symmetry and Angular Momentum; Eigenstates

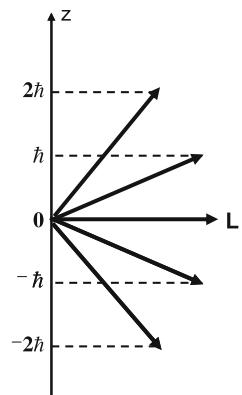
From classical mechanics we know that the angular momentum for motions in a potential with rotational or spherical symmetry is a constant of motion. We expect that this is also true in quantum mechanics.

For the mathematical proof we start with a general function  $f(\mathbf{r})$  defined in 3D space. After performing an infinitesimal displacement  $\delta\mathbf{r}$  the function can be written as

$$f(\mathbf{r} - \delta\mathbf{r}) = f(\mathbf{r}) - \delta\mathbf{r} \cdot \nabla f(\mathbf{r}). \quad (5.36)$$

We now assume the displacement  $\delta\mathbf{r}$  to be a rotation by the infinitesimal angle  $\delta\varphi$  around an axis, i.e.

**Fig. 5.3** Schematic representation of the directional quantization of the angular momentum  $\mathbf{L}$  in units of  $\hbar$  along the  $z$  direction



$$\delta\mathbf{r} = \delta\varphi \times \mathbf{r}. \quad (5.37)$$

Hereby  $\delta\varphi$  is a vector with length  $\delta\varphi$  directed along the rotational axis according to a right handed screw. In this case the infinitesimal change of the function  $f(\mathbf{r})$  is obtained as

$$\delta f(\mathbf{r}) = f(\mathbf{r} - \delta\mathbf{r}) - f(\mathbf{r}) = (\delta\varphi \times \mathbf{r}) \cdot \nabla f(\mathbf{r}). \quad (5.38a)$$

The rule for exchange of vectors in a combined vector cross and scalar product yields

$$\delta f(\mathbf{r}) = -\delta\varphi \cdot (\mathbf{r} \times \nabla f) = -\frac{i}{\hbar} \delta\varphi \cdot \hat{\mathbf{L}} f(\mathbf{r}). \quad (5.38b)$$

Hereby, the definition of the angular momentum operator  $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$  was used. According to (5.38b), a rotation of the coordinate system by  $\delta\varphi$ , thus, changes the function  $f(\mathbf{r})$  by  $\delta f$  and in the rotated coordinate system the modified function  $f'$  is obtained as

$$f'(\mathbf{r}) = f(\mathbf{r}) + \delta f = \left(1 - \frac{i}{\hbar} \delta\varphi \cdot \hat{\mathbf{L}}\right) f(\mathbf{r}). \quad (5.38c)$$

The operator  $[1 - (i/\hbar)\delta\varphi \cdot \hat{\mathbf{L}}]$  generates the corresponding change of the function  $f(\mathbf{r})$  when the coordinate system is rotated by the infinitesimal angle  $\delta\varphi = \mathbf{n}\delta\varphi$  ( $\mathbf{n}$  is unity vector in direction of rotation axis). This operator is called *generator of infinitesimal rotations*. We can easily derive the generator (operator) of finite rotations by building up a finite rotation of angle  $\varphi$  by a large number  $N$  of sequentially performed infinitesimal rotations  $\delta\varphi = \varphi/N$  ( $N \rightarrow \infty$ ,  $\delta\varphi \rightarrow 0$ ):

$$f'(\mathbf{r}) = \lim_{\substack{N \rightarrow \infty \\ \delta\varphi \rightarrow 0}} \left(1 - \frac{i}{\hbar} \frac{\varphi}{N} \mathbf{n} \cdot \hat{\mathbf{L}}\right)^N f(\mathbf{r}). \quad (5.39a)$$

Then, Euler's formula for the exponential function applied on the operator yields

$$f'(\mathbf{r}) = \exp\left(-\frac{i}{\hbar} \varphi \mathbf{n} \cdot \hat{\mathbf{L}}\right) f(\mathbf{r}). \quad (5.39b)$$

The exponential operator in (5.39b) which generates the rotation of the coordinate system is defined by its series expansion, in analogy to (4.97). Instead of rotating the coordinate system we could also rotate the function  $f(\mathbf{r})$ , respectively the vectors  $\mathbf{r}$  in a fixed coordinate system. In that case  $-\delta\mathbf{r}$  in (5.38a) has to be replaced by  $\delta\mathbf{r}$  and a positive sign is obtained in the exponent of (5.39b).

The operator

$$\hat{U}_\varphi = \exp\left(\frac{i}{\hbar} \varphi \mathbf{n} \cdot \hat{\mathbf{L}}\right) \quad (5.40)$$

thus, rotates a quantum state or wave function around a fixed axis  $\mathbf{n}$  (unit vector) in space. Since  $\hat{\mathbf{L}}$  is Hermitian the rotation operator  $\hat{U}_\varphi$  (5.40) is a unitary operator:

$$\hat{U}_\varphi^+ = \exp\left(-\frac{i}{\hbar}\varphi\mathbf{n} \cdot \hat{\mathbf{L}}\right) = \hat{U}_\varphi^{-1}. \quad (5.41)$$

We consider a wave function  $\psi(\mathbf{r})$  and rotate the coordinate system from  $\mathbf{r}$  to  $\mathbf{r}'$ . The wave function is then transformed from  $\psi(\mathbf{r})$  to  $\psi(\mathbf{r}')$ . This transformed wave function is spatially arranged in the transformed coordinate system as the original wave function  $\psi(\mathbf{r})$  in the original non-rotated coordinate system. The question, now, is how a general operator transforms upon rotation of the coordinate system. We describe the action of a general operator  $\hat{\Omega}$  in the non-rotated system by  $\hat{\Omega}\psi(\mathbf{r}) = \varphi(\mathbf{r})$ . Because of  $\hat{U}^+\hat{U} = 1$ , we conclude:

$$\hat{U}\hat{\Omega}\hat{U}^+[\hat{U}\psi(\mathbf{r})] = \hat{U}\varphi(\mathbf{r}), \quad (5.42a)$$

or

$$\hat{U}\hat{\Omega}\hat{U}^+\psi(\mathbf{r}') = \varphi(\mathbf{r}'). \quad (5.42b)$$

For the operator  $\hat{\Omega}'$  in the rotated coordinate system we, thus, obtain

$$\hat{\Omega}' = \hat{U}\hat{\Omega}\hat{U}^+. \quad (5.42c)$$

In these relations, we recognize essential properties of rotations in Hilbert space (Sect. 4.3.2).

By means of (5.38b), we derive interesting properties of potentials  $V(\mathbf{r})$  with rotational symmetry, that is, potentials, which depend only on the radius  $r$  and not on angular changes  $\delta\varphi$  around their center. In that case, we have

$$V(r, \varphi) = V(r, \varphi + \delta\varphi). \quad (5.43a)$$

By means of (5.38b), we can also write:

$$V(r, \varphi + \delta\varphi) = V(r, \varphi) + \frac{i}{\hbar}\delta\varphi \cdot \hat{\mathbf{L}}V(r, \varphi). \quad (5.43b)$$

From this relation we derive the commutation relation between the potential  $\hat{V}(r, \varphi)$  (written as operator) and the angular momentum  $\hat{\mathbf{L}}$ :

$$\begin{aligned} [\hat{V}, \hat{\mathbf{L}}] &= \left[ \hat{V} + \frac{i}{\hbar}\delta\varphi \cdot \hat{\mathbf{L}}\hat{V}, \hat{\mathbf{L}} \right] \\ &= [\hat{V}, \hat{\mathbf{L}}] + \frac{i}{\hbar}[\delta\varphi \cdot \hat{\mathbf{L}}\hat{V}, \hat{\mathbf{L}}], \end{aligned} \quad (5.44a)$$

or

$$\frac{i}{\hbar}[\delta\varphi \cdot \hat{\mathbf{L}} \hat{V}, \hat{\mathbf{L}}] = 0. \quad (5.44b)$$

As  $\hat{\mathbf{L}}$  commutes with itself we conclude that the angular momentum operator  $\hat{\mathbf{L}}$  commutes with each operator  $\hat{V}(r)$  of a potential with rotational symmetry:  $[\hat{V}(r), \hat{\mathbf{L}}] = 0$ .

In order to make far-reaching conclusions about particle motion in a potential with rotational symmetry (electron in Coulomb potential, electron in circular quantum dot, Sect. 5.7.1), we must investigate the commutation of the angular momentum  $\hat{\mathbf{L}}$  with the kinetic energy operator  $\hat{T} = \hat{p}^2/2m$ , too. The Hamilton operator, namely, contains both the kinetic and the potential energy. We are, therefore, interested in the commutation relation  $[\hat{\mathbf{L}}, \hat{p}^2]$ , the essential part of the kinetic energy, that is,

$$[\hat{\mathbf{L}}, \hat{p}^2] = [\hat{\mathbf{r}} \times \hat{\mathbf{p}}, \hat{p}^2]. \quad (5.45a)$$

First, we consider the  $x$  component of the angular momentum:

$$\begin{aligned} [(\hat{\mathbf{r}} \times \hat{\mathbf{p}})_x, \hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2] &= [\hat{y}\hat{p}_z, \hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2] \\ &= [\hat{y}\hat{p}_z, \hat{p}_y^2] \\ &= \hat{y}\hat{p}_z\hat{p}_y^2 - \hat{p}_y^2\hat{y}\hat{p}_z = \hat{p}_z\hat{y}\hat{p}_y^2 - \hat{p}_y^2\hat{y}\hat{p}_z \\ &= \hat{p}_z(\hat{p}_y\hat{y} + i\hbar)\hat{p}_y - \hat{p}_y^2\hat{y}\hat{p}_z. \end{aligned} \quad (5.45b)$$

For the calculation, it was used that the angular momentum components mutually commute and also  $\hat{y}$  with  $\hat{p}_x$  and  $\hat{p}_z$  but not with  $\hat{p}_y$ . By use of the commutation relation  $\hat{y}\hat{p}_y = \hat{p}_y\hat{y} + i\hbar$ , the calculation (5.45b) finally yields:

$$[\hat{L}_x, \hat{p}^2] = 0. \quad (5.45c)$$

An analogous calculation can be performed for the other components  $\hat{L}_y, \hat{L}_z$  of the angular momentum. The important result is that the angular momentum operator  $\hat{\mathbf{L}}$  commutes with the kinetic energy of a moving particle:

$$[\hat{\mathbf{L}}, \hat{p}^2/2m] = 0. \quad (5.46)$$

According to (5.44b) and (5.46) potentials with rotational or spherical symmetry (dependent only on  $r$ , but not on  $\varphi$  and/or  $\vartheta$ ) yield a Hamilton operator  $\hat{H} = \hat{T} + \hat{V}$  which commutes with the angular momentum operator  $\hat{\mathbf{L}}$ . From Heisenberg's dynamical equation we, then, obtain:

$$\frac{d}{dt}\hat{\mathbf{L}} = \frac{i}{\hbar}[\hat{H}, \hat{\mathbf{L}}] = 0. \quad (5.47)$$

In a potential with rotational or spherical symmetry the quantum mechanical angular momentum is a constant of motion. Quantum mechanics confirms the classical result.

The commutation relation  $[\hat{H}, \hat{\mathbf{L}}] = 0$  also implies

$$[\hat{H}, \hat{L}^2] = 0, \quad (5.48)$$

for a Hamiltonian containing a potential with rotational or spherical symmetry. Because of the commutation relations for angular momentum operators (5.9), (5.10), the following commutations are valid for rotational or spherical potentials:

$$[\hat{H}, \hat{L}^2] = [\hat{L}^2, \hat{L}_z] = [\hat{L}^2, \hat{L}_x] = [\hat{L}^2, \hat{L}_y] = 0. \quad (5.49)$$

Consequently  $\hat{L}^2$  and the different components of the angular momentum, for example,  $\hat{L}_z$ , have the same eigenfunction system as the Hamilton operator (with rotational or spherical symmetry).

This does not mean that  $\hat{H}$  and all components  $\hat{L}_i$  have the same eigenfunction system simultaneously ( $[\hat{L}_i, \hat{L}_j] \neq 0$ ).

Nevertheless, for systems with spherical symmetry the eigensolutions for energy and angular momentum can be represented in position representation as

$$\langle r | n, l, m \rangle = R_{n,l}(r) \Upsilon_l^m(\vartheta, \varphi). \quad (5.50)$$

$\Upsilon_l^m(\vartheta, \varphi)$  are the eigenfunctions to the angular momentum operators  $\hat{\mathbf{L}}^2$  and  $\hat{L}_z$ ; they depend only on the angles  $\vartheta$  and  $\varphi$  (spherical coordinates, Fig. 5.2) and are called spherical harmonics (if normalized over the whole solid angle).

The radial part of the eigensolution  $R_{n,l}(r)$  describes the radius dependence of the wave function of an electron in a spherical potential. It depends, of course on the angular momentum quantum number  $l$ , since varying angular momenta imply also different radii of the circular trajectories of the particle. But different radii of the circular motion means simultaneously different total energy (5.5b). The so-called *principal quantum number*  $n$ , therefore, numerates the different energy eigenvalues  $E_{n,l}$  of the Hamilton operator  $\hat{H}$ . The spectrum of  $n$  is discrete because of particle confinement in the spherical binding potential (as assumed here). Neither  $R_{n,l}$  nor  $E_{n,l}$ , depend on the directional quantum number  $m$ , since a spherical potential does not distinguish any particular direction in space. The states for a particular  $n$  and  $l$  are degenerate in  $m$ , they have the same energy for each  $m$  value. An additionally applied magnetic field defines a certain direction in space and quantum states with different directional quantum number  $m$  get different energies, the  $m$  degeneracy is removed. This explains the other name of  $m$ , magnetic quantum number.

We summarize the eigenvalue equations for a spherical potential in position representation as follows:

$$\hat{H} R_{n,l}(r) \Upsilon_l^m = E_{n,l} R_{n,l} \Upsilon_l^m, \quad (5.51\text{a})$$

$$\hat{L}^2 R_{n,l} \Upsilon_l^m = l(l+1)\hbar R_{n,l} \Upsilon_l^m, \quad (5.51\text{b})$$

$$\hat{L}_z R_{n,l} \Upsilon_l^m = m\hbar R_{n,l} \Upsilon_l^m. \quad (5.51\text{c})$$

The radial part of the wave functions  $R_{n,l}(r)$  and the electron energy eigenvalues  $E_{n,l}$  depend on the special shape of the spherical potential, for example, a Coulomb potential in the hydrogen atom or a binding parabolic (oscillator) potential in a semiconductor quantum dot. The angular parts  $\Upsilon_l^m$  of the wave functions, on the other hand, do not depend on details of the particular potential; These eigenfunctions of the angular momentum are always found as part of the eigensolutions to a Hamilton operator containing a potential with rotational or spherical symmetry. They will shortly be presented in the following.

Without presenting the analytical solution of the differential eigenvalue equation (5.51b), the explicit form of some low index spherical harmonics shall be given here:

$$\Upsilon_0^0 = \frac{1}{\sqrt{4\pi}}, \quad (5.52\text{a})$$

$$\Upsilon_0^1 = \sqrt{\frac{3}{4\pi}} \cos \vartheta, \quad \Upsilon_1^1 = -\sqrt{\frac{3}{8\pi}} \sin \vartheta e^{i\varphi}, \quad (5.52\text{b})$$

$$\Upsilon_2^0 = \sqrt{\frac{5}{16\pi}} (3 \cos^2 \vartheta - 1), \quad (5.52\text{c})$$

$$\Upsilon_2^1 = \sqrt{\frac{15}{8\pi}} \sin \vartheta \cos \vartheta e^{i\varphi}, \quad \Upsilon_2^2 = \sqrt{\frac{15}{32\pi}} \sin^2 \vartheta e^{i2\varphi}. \quad (5.52\text{d})$$

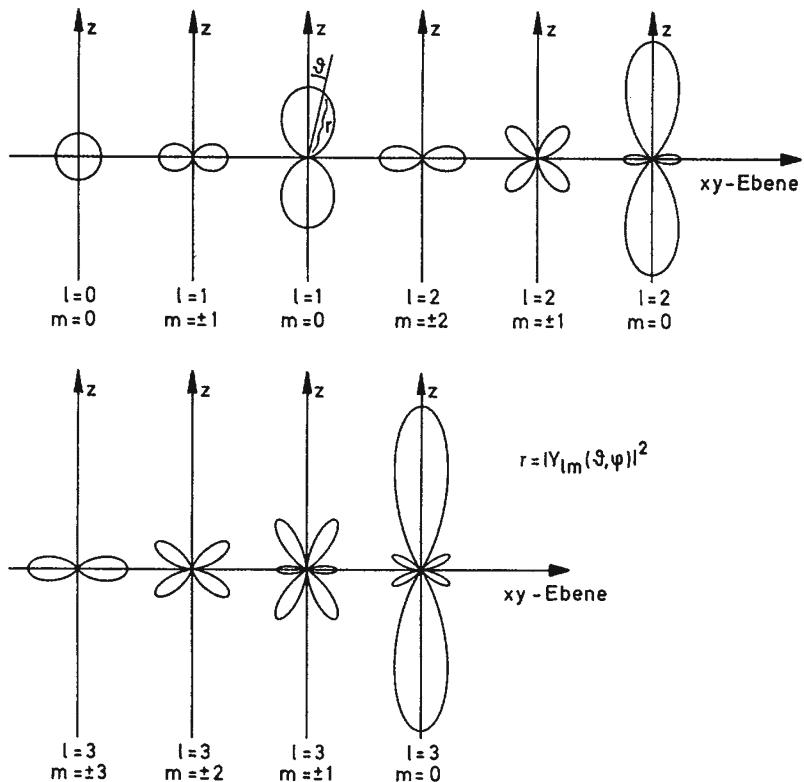
In addition, for negative  $m$  values we have the relation:

$$\Upsilon_l^{-m} = (-1)^m \Upsilon_l^{m*}. \quad (5.53)$$

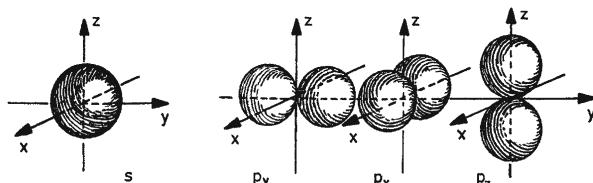
In Fig. 5.4, the spherical harmonics  $\Upsilon_l^m$  with  $l = 0, 1, 2, 3$  are plotted in polar diagrams. Wave functions with  $l = 0$  having spherical symmetry are called *s-orbitals*, those with  $l = 1$  distinguishing a particular direction in space are named *p-orbitals*. Consequently the functions with  $l = 2$  and  $l = 3$  are called *d-* and *f-orbitals*, respectively.

Without distinction of a particular axis in space, for example, by a magnetic field, all angular momentum states with equal quantum number  $l$  but differing  $m$  values are equivalent, they have the same energy. In this situation, linear superpositions of  $\Upsilon_l^m$  with equal  $l$  but differing  $m$  might be more adapted to particular problems than one special function (5.52a)–(5.52d) with a fixed  $m$  value. Thus, by superposition of  $\Upsilon_1^1$  (5.52b) and  $\Upsilon_1^{-1}$  (5.53) we obtain new *p-orbitals*, which are oriented along the  $x, y, z$  axis in space, with equal shape along these axes (Fig. 5.5).

These superposition states are named  $p_x, p_y, p_z$  orbitals, their mathematical expressions are as follows:



**Fig. 5.4** Representation of the angular momentum eigenfunctions  $\gamma_{lm}$  with  $l = 0, 1, 2, 3$  in polar diagrams. The radial distances  $r$  from the center (see  $l = 1, m = 0$ ) are a measure for the value  $|\gamma_{lm}|^2$  as function of  $\vartheta$  [18]



**Fig. 5.5** Three dimensional polar representation of the absolute amounts of the angular momentum wave functions of the  $s$  orbital and of the three  $p_x, p_y, p_z$  orbitals [18]

$$\gamma_{p_x} = \frac{-1}{\sqrt{2}} (\gamma_1^1 - \gamma_1^{-1}) = \sqrt{\frac{3}{4\pi}} \sin \vartheta \cos \varphi, \quad (5.54a)$$

$$\gamma_{p_y} = \frac{-1}{\sqrt{2}i} (\gamma_1^1 - \gamma_1^{-1}) = \sqrt{\frac{3}{4\pi}} \sin \vartheta \sin \varphi, \quad (5.54b)$$

$$\gamma_{p_z} = \gamma_1^0. \quad (5.54c)$$

After the extensive discussion of the quantum mechanical angular momentum, its eigenvalues and eigenfunctions we turn to the question: How can we observe angular momenta in experiment? For this purpose, we must derive a relation between angular momentum of a charged particle and magnetic moment.

## 5.4 Circulating Electrons in a Magnetic Field

### 5.4.1 The Lorentz Force

As we know from classical physics charge carriers (positive charge  $e$ ) moving in a magnetic field are deflected from their linear path. They are subject to the so-called *Lorentz force*

$$m\dot{\mathbf{v}} = \mathbf{F} = e\mathbf{v} \times \mathbf{B}. \quad (5.55)$$

In every moment, this Lorentz force acts perpendicular to the velocity  $\mathbf{v}$  of the charged particle and to the magnetic field  $\mathbf{B}$ . Charge carriers entering a region with a constant magnetic field  $\mathbf{B}$  perpendicularly to the field are forced into a stable circular orbit being normal to the  $B$  field. These so-called cyclotron orbits are characterized by a centrifugal acceleration [outwards of the orbit, inverse direction of centripetal force (5.1c)] which is compensated by the centripetal acceleration due to the Lorentz force (5.55). For the stable cyclotron orbit, we thus obtain the following relation for the rotational frequency, the so-called *cyclotron frequency*:

$$mr\omega_c^2 = mv\omega_c = evB, \quad (5.56a)$$

$$\omega_c = \frac{e}{m}B. \quad (5.56b)$$

The Lorentz force is directed normal to the particle velocity, it does not change the energy of the particle. Thus, it can not be expressed as a potential difference, that is, as a gradient of the potential like the familiar forces in Newtonian mechanics.

The reason for this peculiarity derives from the fact that this force can only be understood on the basis of special relativity [1]. Maxwell's equations of the electromagnetic field are compatible with special relativity, they are Lorentz-invariant. Consequently electric and magnetic fields in inertial reference frames moving with different relative velocity are connected with each other. A static electric charge generates a magnetic field in an inertial frame which moves at a relative velocity with respect to the static charge. Vice versa, a static magnetic field induces an electric field (force on an electric charge) in a second inertial frame moving relative to the first one. While the action of an electric field  $\mathcal{E} = -e \operatorname{grad} \phi$  can be introduced into the Schrödinger equation via its potential  $\phi$  [contained in potential energy  $V(\mathbf{r})$ ], this procedure is not possible for the magnetic field. How, then, can we obtain a reasonable mathematical ansatz, which allows the description of the effect of a magnetic field on particle motion, without using special relativity theory?

The solution to this problem requires an ansatz for the Hamilton operator  $\hat{H}$ , or in classical physics, for the Hamilton function  $H(p, q)$  (3.25), which yields a dynamical equation of the type (5.55) after application of the classical Hamilton formalism (3.26a), (3.26b) or in quantum mechanics the Schrödinger or Heisenberg (4.104) equation.

### 5.4.2 The Hamilton Operator with Magnetic Field

In order to find the correct Hamiltonian with magnetic field, we guess that a good starting point, as often in electrodynamics and in quantum electrodynamics (Sect. 8.2), will be the vector potential  $\mathbf{A}$  of the magnetic field  $\mathbf{B} = \text{curl } \mathbf{A}$ . A charge carrier circulating on a cyclotron orbit in a constant magnetic field  $\mathbf{B}$  has a velocity  $\mathbf{v}$  and respectively, a momentum vector  $\mathbf{p}$  which follows as a tangent the circular trajectory. The same position dependence is given for the vector potential  $\mathbf{A}$  of the magnetic field  $\mathbf{B}$ . For a spatially constant magnetic field  $\mathbf{B} = (0, 0, B_z)$   $\mathbf{A}$  can be chosen as

$$\mathbf{A} = \frac{1}{2} B_z r \mathbf{e}_\varphi = \frac{1}{2} (y B_z \mathbf{e}_x - x B_z \mathbf{e}_y). \quad (5.57)$$

This is easily proven by applying the relations (5.30a)–(5.30c), (5.31a)–(5.31c) between the coordinate unity vectors and the rules for the curl operator. One should remember that the choice of  $\mathbf{A}$  is not unequivocal. One can add any gradient of a scalar function to (5.57) without changing the relation  $\mathbf{B} = \text{curl } \mathbf{A}$  (Sect. 5.4.4).

On a cyclotron orbit the particle momentum  $\mathbf{p}$  has the same spatial dependence as  $\mathbf{A}$  (5.57):

$$\mathbf{p} = m v \mathbf{e}_\varphi = m r \omega \mathbf{e}_\varphi. \quad (5.58)$$

It thus seems reasonable—at this point a try—to introduce the magnetic field in terms of  $\mathbf{A}$  into the Hamiltonian, in parallel to the momentum. Just for having the same units as the momentum  $\mathbf{p}$  we must multiply  $\mathbf{A}$  with an electric charge and can add both quantities. For an electron with negative elementary charge moving in a magnetic field  $\mathbf{B} = \text{curl } \mathbf{A}$  we, thus, make the ansatz:

$$\hat{H} = \frac{1}{2m} (\hat{\mathbf{p}} - e \hat{\mathbf{A}})^2. \quad (5.59)$$

While  $\hat{\mathbf{p}}$  is called *canonical momentum* (canonical variable, Sect. 3.4) the generalized momentum

$$m \dot{\hat{\mathbf{r}}} = \hat{\mathbf{p}} - e \hat{\mathbf{A}} = \hat{\mathbf{p}} \quad (5.60)$$

is named *kinetic momentum* of the motion.

As the operator  $\hat{\mathbf{A}}(\mathbf{r})$  is a vector function, which depends only on  $\mathbf{r}$ , it commutes with  $\mathbf{r}$  but not with  $\hat{\mathbf{p}}$ . With  $\hat{\mathbf{r}} = (x_1, x_2, x_3)$  we derive the following commutation

relations for  $\hat{\pi}$ :

$$[\hat{x}_i, m\dot{\hat{x}}_j] = [\hat{x}_i, \hat{\pi}_j] = i\hbar\delta_{ij}; \quad (5.61)$$

$$\begin{aligned} [\hat{\pi}_i, \hat{\pi}_j] &= [m\dot{\hat{x}}_i, m\dot{\hat{x}}_j] = [\hat{p}_i - e\hat{A}_i, \hat{p}_j - e\hat{A}_j] \\ &= -[\hat{p}_i, e\hat{A}_j] - [e\hat{A}_i, \hat{p}_j] \\ &= i\hbar\left[\frac{\partial}{\partial x_i}, eA_j\right] + i\hbar\left[eA_i, \frac{\partial}{\partial x_j}\right] \\ &= i\hbar e\left(\frac{\partial}{\partial x_i}A_j + A_j\frac{\partial}{\partial x_i} - A_j\frac{\partial}{\partial x_i} + A_i\frac{\partial}{\partial x_j} - \frac{\partial}{\partial x_j}A_i - A_i\frac{\partial}{\partial x_j}\right) \\ &= i\hbar e\left(\frac{\partial}{\partial x_i}A_j - \frac{\partial}{\partial x_j}A_i\right) = i\hbar eB_k \end{aligned} \quad (5.62a)$$

that is,

$$[\hat{\pi}_i, \hat{\pi}_j] = i\hbar eB_k. \quad (5.62b)$$

For the derivation of (5.62a), we have used that an operator product requires differentiation not only of the function  $A_i$  but also of the wave function (on which the operator acts) which is not explicitly expressed in (5.61). This requires the application of the chain rule of differentiation.

We now calculate the time derivative of the kinetic momentum  $\hat{\pi}$  by means of the Heisenberg dynamical equation (4.104). In a first step the commutation relations (5.61) and (5.62b) yield an expression for  $\hat{\pi}$  from which we derive the time derivative:

$$\hat{\pi} = m\dot{\hat{\mathbf{r}}} = \frac{i}{\hbar}[\hat{H}, m\hat{\mathbf{r}}] = \frac{i}{\hbar}\left[\frac{\hat{\pi}^2}{2m}, m\hat{\mathbf{r}}\right] = \hat{\mathbf{p}} - e\hat{\mathbf{A}}, \quad (5.63)$$

$$\dot{\hat{\pi}} = m\ddot{\hat{\mathbf{r}}} = \frac{i}{\hbar}[\hat{H}, m\dot{\hat{\mathbf{r}}}] = \frac{i}{\hbar}[\hat{H}, \pi] = e\dot{\mathbf{r}} \times \mathbf{B}. \quad (5.64)$$

According to (5.64) the kinetic momentum  $\hat{\pi} = \hat{\mathbf{p}} - e\hat{\mathbf{A}}$  indeed obeys a dynamical equation with the Lorentz force as driving force, just as in the classical equation (5.55). Due to this correspondence between quantum and classical mechanics (Sect. 3.4). The Hamilton operator  $\hat{H} = \hat{\pi}^2/2m$  (5.59) must be applied for the solution of problems in which a magnetic field is partially responsible for the particle motion.

### 5.4.3 Angular Momentum and Magnetic Moment

In classical electrodynamics, currents through a closed loop generate a magnetic field which penetrates the loop perpendicularly. At some distance this magnetic field has the same spatial structure as that of a magnetic dipole (Fig. 5.6). In an external magnetic field, the current loop behaves as a magnetic dipole, it is subject to forces, that is, a torque, as a magnetic dipole in an external field. These forces can be derived from the Lorentz force on single electrons which carry the current through the loop. We will see that similar conclusions can be drawn in quantum mechanics, for example, for electrons in atoms or in nanoscopic rings in semiconductor nanoelectronics.

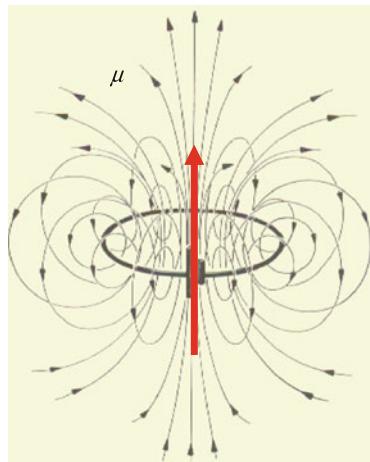
For the mathematical treatment of this problem, we start with the Hamilton operator (5.59) of an electron in a magnetic field:

$$\hat{H} = \frac{1}{2m}(\hat{\mathbf{p}} - e\hat{\mathbf{A}})^2 = \frac{\hat{p}^2}{2m} - \frac{e}{2m}(\hat{\mathbf{p}} \cdot \hat{\mathbf{A}} + \hat{\mathbf{A}} \cdot \hat{\mathbf{p}}) + \frac{e^2 \hat{A}^2}{2m}. \quad (5.65)$$

The choice of the magnetic vector potential (5.57) yields a magnetic field  $\mathbf{B} = \text{curl } \mathbf{A} = B_z \mathbf{e}_z$  in  $z$  direction. For sufficiently small  $\mathbf{B}$  fields the last  $\hat{A}^2$  term can be neglected and only the  $\hat{\mathbf{p}} \cdot \hat{\mathbf{A}}$  and  $\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}$  terms must be considered. Using the chain rule for differentiation, we calculate the action of  $\hat{\mathbf{p}} \cdot \hat{\mathbf{A}}$  on a wave function:

$$\begin{aligned} \hat{\mathbf{p}} \cdot \hat{\mathbf{A}}\psi &= -i\hbar \nabla \cdot (\mathbf{A}\psi) \\ &= -i\hbar[(\nabla \cdot \mathbf{A})\psi + \mathbf{A} \cdot \nabla \psi] \\ &= -i\hbar \mathbf{A} \cdot \nabla \psi = \hat{\mathbf{A}} \cdot \hat{\mathbf{p}}\psi. \end{aligned} \quad (5.66)$$

**Fig. 5.6** Magnetic field produced by an electric current through a closed loop. At some distance from the loop (coil) the magnetic field resembles the field of a magnetic dipole  $\mu$  (schematically indicated by the *central arrow*)



For the derivation, we used  $\nabla \cdot \mathbf{A} = \text{div } \mathbf{A} = 0$ . By means of (5.66), we can write the Hamiltonian (5.65) in linear approximation:

$$\hat{H} = \frac{\hat{p}^2}{2m} - \frac{e}{2m}(2\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}). \quad (5.67)$$

Apart from the kinetic energy of a free particle  $\hat{p}^2/2m$  there is an interaction energy  $\hat{H}_{\text{int}} = e\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}/m$  which can be expressed, by using (5.57), as:

$$\hat{H}_{\text{int}} = -\frac{e}{m}\frac{B_z}{2}(-y\hat{p}_x + x\hat{p}_y). \quad (5.68a)$$

Now the  $z$  component of the angular momentum  $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$  is written as  $\hat{L}_z = x\hat{p}_y - y\hat{p}_x$  and we obtain the interaction operator with the magnetic field (5.68a) as:

$$\hat{H}_{\text{int}} = -\frac{e}{2m}\hat{\mathbf{L}} \cdot \mathbf{B}. \quad (5.68b)$$

The energy of a magnetic dipole  $\mu$  in a magnetic field  $\mathbf{B}$  is  $E = -\mu \cdot \mathbf{B}$  (analogous to an electric dipole in the electric field). According to (5.68b), we attribute a magnetic dipole moment (operator)

$$\hat{\mu} = \frac{e}{2m}\hat{\mathbf{L}} \quad (5.69)$$

to a charged particle (charge  $-e$ , mass  $m$ ) circulating on an orbit with angular momentum  $\hat{\mathbf{L}}$ . The dipole moment (parallel to  $\hat{\mathbf{L}}$ ) is directed normal to the orbit area (Fig. 5.6). It is worth mentioning that at a given angular momentum the mass of the particle determines the magnetic moment. Protons with a mass exceeding that of electrons by a factor of about 2000, therefore, are expected to have a magnetic moment smaller by about the factor 2000 than that of an electron at equal angular momentum. Nuclear magnetism is significantly smaller than electronic magnetism of the electronic shell of atoms. It is furthermore interesting that the quantum mechanical relation (5.59) is equally well derived for classical observables in electrodynamics. We consider a classical circular current loop with radius  $r$  carrying a current  $I$ . According to classical electrodynamics a constant magnetic field  $B$  exerts a torque on this loop as on a magnetic dipole  $\mu$  described by

$$\mu = I \cdot A, \quad (5.70)$$

$A = \pi \cdot r^2$  is the area of the current loop. For one single particle with charge  $q$  circulating along the loop the current (charge per time at a certain position on the loop) amounts to

$$I = qv/2\pi r. \quad (5.71)$$

The magnetic moment (5.70), then, is obtained as

$$\mu = \frac{qv}{2\pi r} \pi r^2 = \frac{q}{2m} mvr = \frac{q}{2m} L, \quad (5.72)$$

by applying the classical formula (5.5a), (5.5b) for the angular momentum  $L = mvr = mr^2\omega$ . Equation (5.72) equals the quantum mechanical expression (5.69) except from the different meaning of classical observables and quantum mechanical operators.

So far we have considered only the terms being linear in  $\mathbf{A}$ , respectively  $\mathbf{B}$ , in the Hamiltonian (5.65). These terms are responsible for the generation of magnetic dipoles due to rotating electrons, which are orientated in an external magnetic field, e.g. electrons in atoms with  $\hat{\mathbf{L}} \neq \mathbf{0}$ . This part of magnetism of atoms or matter in general derives from magnetic dipoles already present, which interact with an external field. It is called *paramagnetism*.

In addition, matter also exhibits the property of *diamagnetism*. This type of magnetism is based on circular currents in atoms which are induced by an external magnetic field. According to Lenz rule the magnetic moment connected with these circular currents is directed opposite to the applied external field. This causes a negative, so-called diamagnetic contribution to the magnetic susceptibility. Since the induced magnetic moment is proportional to the  $\mathbf{B}$  field and, furthermore, the energy of the induced dipole in the external field is also proportional to  $\mathbf{B}$ , the diamagnetic contribution to magnetism is proportional to  $B^2$ . This dependence on the magnetic field arises from the third term  $e^2 \hat{A}^2 / 2m$  in the Hamilton operator (5.65). It is responsible for the diamagnetic properties of atoms molecules and matter in general.

For an estimation of diamagnetism, we assume a magnetic field  $\mathbf{B} = B_z \mathbf{e}_z$  in  $z$  direction and again the representation  $\mathbf{A} = -(1/2)(\mathbf{r} \times \mathbf{B})$  of the magnetic vector potential. Then, the third diamagnetic term (5.65) applied on a wave function yields

$$\begin{aligned} \frac{e^2}{2m} \hat{A}^2 \psi &= \frac{e^2}{8m} (\mathbf{r} \times \mathbf{B})^2 \psi = \frac{e^2}{8m} \{r^2 B^2 - (\mathbf{r} \cdot \mathbf{B})^2\} \psi \\ &= \frac{e^2 B_z^2}{8m} (x^2 + y^2) \psi. \end{aligned} \quad (5.73a)$$

To estimate the order of magnitude of this expression for an atom, we calculate the expectation value as

$$\langle \psi | (x^2 + y^2) e^2 B_z^2 / 8m | \psi \rangle \sim \frac{e^2 a^2 B_z^2}{8m}. \quad (5.73b)$$

The expectation value  $\langle \psi | x^2 + y^2 | \psi \rangle$ , a radius of atomic dimension, is approximated by the Bohr radius  $a \approx 0.05$  nm. It is particularly interesting to compare (5.73b) with the paramagnetic contribution (5.68b). For this purpose, we estimate the expectation value  $\langle \hat{L} \rangle$  in  $\langle \hat{H}_{\text{int}} \rangle$  by  $\hbar$ . The ratio between diamagnetic and paramagnetic contribution is, the, obtained approximately as

$$\left( \frac{e^2 a^2}{8m} / \frac{e\hbar}{2m} \right) B_z \approx 1 \times 10^{-10} \cdot B_z [\text{G}]. \quad (5.74)$$

Even for extremely large magnetic fields in the order of  $B \approx 10^5$  Gauss diamagnetism in atoms is negligibly small in comparison to paramagnetism.

#### 5.4.4 Gauge Invariance and Aharonov–Bohm-Effect

We are used to the fact that energy or potential are determined only apart from an additive constant. Forces measured in an experiment result from a potential difference, that is, a gradient of the potential. Similarly, the magnetic field  $\mathbf{B} = \operatorname{curl} \mathbf{A}$  might be expressed in terms of its vector potential  $\mathbf{A}$  which is not defined unequivocally either. We can generate one and the same magnetic field from a whole variety of different vector potentials  $\mathbf{A}$ , since the gradient of any scalar field  $U(r)$  vanishes under the action of the curl operator for the calculation of the  $\mathbf{B}$  field. Because of  $\nabla \times \nabla U(\mathbf{r}) = \mathbf{0}$  the following relation

$$\mathbf{B} = \operatorname{curl}[\mathbf{A} + \nabla U(r)] = \operatorname{curl} \mathbf{A}, \quad (5.75)$$

is valid. Without changing the magnetic field, we can perform the transformation

$$\mathbf{A} \Rightarrow \mathbf{A}' = \mathbf{A} + \nabla U(\mathbf{r}) \quad (5.76)$$

of the vector potential from  $\mathbf{A}$  to  $\mathbf{A}'$ . Since the vector potential  $\mathbf{A}$  enters the kinetic moment  $\hat{\pi} = (\hat{\mathbf{p}} - e\hat{\mathbf{A}})$  in the Hamiltonian (5.59) the transformation (5.76) certainly modifies the fundamental Schrödinger equation. We will prove, now, that the transformation of the vector potential  $\mathbf{A} \Rightarrow \mathbf{A}'$  (5.76) keeps the Schrödinger equation unchanged if we simultaneously transform the wave function according to

$$\psi(\mathbf{r}, t) \Rightarrow \psi'(\mathbf{r}, t) = \exp\left(\frac{ie}{\hbar} U(\mathbf{r})\right) \psi(\mathbf{r}, t). \quad (5.77)$$

This transformation does not change the probability density either:  $|\psi'|^2 = |\psi|^2$ .

The proof is as follows. We start with the original Schrödinger equation for  $\psi(\mathbf{r}, t)$ :

$$\frac{1}{2m} \left( \frac{\hbar}{i} \nabla - e\mathbf{A} \right)^2 \psi = i\hbar \frac{\partial}{\partial t} \psi. \quad (5.78a)$$

We use the abbreviation  $\gamma = ie/\hbar$  and for simplicity reason the one-dimensional representation:

$$-\frac{\hbar^2}{2m} \left( \frac{\partial}{\partial x} - \gamma A \right)^2 \psi = i\hbar \frac{\partial}{\partial t} \psi. \quad (5.78b)$$

The wave function is transformed ( $\psi \Rightarrow \psi'$ ) according to (5.77) by multiplying (5.78b) from the left with  $\exp[\gamma U(x)]$ :

$$-\frac{\hbar^2}{2m}e^{\gamma U}\left(\frac{\partial}{\partial x} - \gamma A\right)\left(\frac{\partial}{\partial x} - \gamma A\right)\psi = i\hbar\frac{\partial}{\partial t}(e^{\gamma U}\psi). \quad (5.79)$$

On the right side of the equation, the time derivative of the transformed wave function  $\psi'$  appears. For the left side of (5.79), some calculation is necessary to obtain something like the one side of a Schrödinger equation. The factor  $\exp(\gamma U)$  must be “shifted through” the product of operators  $[(\partial/\partial x) - \gamma A]$ , in order to combine it with  $\psi$  for the generation of  $\psi'$ .

The following application of the chain rule of differentiation is helpful:

$$\left(\frac{\partial}{\partial x} - \frac{\partial g(x)}{\partial x}\right)e^{g(x)} = e^{g(x)}\frac{\partial}{\partial x} + e^{g(x)}\frac{\partial g}{\partial x} - e^{g(x)}\frac{\partial g}{\partial x} = e^{g(x)}\frac{\partial}{\partial x}. \quad (5.80)$$

Using this relation, we obtain in a first step from (5.79):

$$-\frac{\hbar^2}{2m}\left(\frac{\partial}{\partial x} - \gamma\frac{\partial U}{\partial x} - \gamma A\right)e^{\gamma U}\left(\frac{\partial}{\partial x} - \gamma A\right)\psi = i\hbar\frac{\partial}{\partial t}(e^{\gamma U}\psi). \quad (5.81a)$$

In a second step, we get

$$-\frac{\hbar^2}{2m}\left(\frac{\partial}{\partial x} - \gamma\frac{\partial U}{\partial x} - \gamma A\right)^2(e^{\gamma U}\psi) = i\hbar\frac{\partial}{\partial t}(e^{\gamma U}\psi). \quad (5.81b)$$

By insertion of  $\gamma = ie/\hbar$  the Schrödinger equation, here written in three dimensions, is obtained for the transformed wave function:

$$\frac{1}{2m}\left[\frac{\hbar}{i}\nabla - e\mathbf{A}(\mathbf{r}) - e\nabla U(\mathbf{r})\right]^2(e^{ieU(\mathbf{r})/\hbar}\psi) = i\hbar\frac{\partial}{\partial t}(e^{ieU(\mathbf{r})/\hbar}\psi). \quad (5.82)$$

With the transformations (5.76) and (5.77), the Schrödinger equation in the transformed system  $\psi'$  and  $\mathbf{A}'$  is obtained as

$$\frac{1}{2m}\left(\frac{\hbar}{i}\nabla - e\mathbf{A}'\right)^2\psi' = i\hbar\frac{\partial}{\partial t}\psi'. \quad (5.83)$$

A simultaneous transformation of the vector potential (by adding the gradient of a function  $U$ ) and the wave function (by multiplication with the factor  $\exp[ieU/\hbar]$ ), thus, conserves the dynamics of a particle, an unexpected and interesting result. This simultaneous transformation of  $\mathbf{A}$  and  $\psi$ , which does not change the dynamical laws, is called *gauge transformation*. Quantum mechanical laws are *gauge invariant*.

We can also go the way in inverse direction. We can require gauge invariance for the Schrödinger equation when a magnetic field shall be introduced into the equation.

This requirement unequivocally leads to the introduction of the kinetic momentum  $\mathbf{p} - e\mathbf{A}$  in the Hamiltonian (Sect. 5.4.1).

This way of using gauge invariance has been proven to be very successful in modern field theories of elementary particle physics to elegantly introduce interactions between different particle fields.

In the following, we want to use gauge invariance to derive an interesting electron interference effect in ring-like nanostructures, the *Aharanov–Bohm-effect*. This effect is important for a fundamental understanding of the interaction of charged particles with magnetic fields but it is also the basis for interesting novel interference devices in quantum electronics.

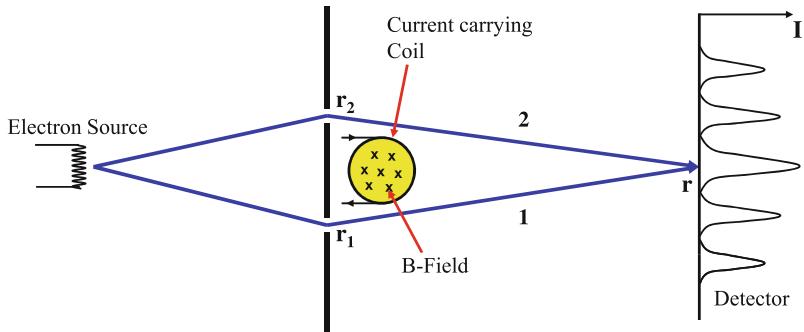
We start with the double-slit experiment for electrons (Sect. 2.4.1) and assume that behind the double-slit, between the two interfering electron beams (1) and (2) a long solenoid is arranged perpendicular to the plane of the beams (Fig. 5.7). A current through the solenoid, then, generates a locally confined magnetic field with field lines normal to the plane of the beam trajectories. The solenoid is sufficiently long that the magnetic stray field outside the coil in the region of the two electron beams can be neglected. The electrons, thus, propagate in a space free of any magnetic field. According to  $\mathbf{B} = \nabla \times \mathbf{A}$  and  $\mathbf{A} = (1/2)Br\mathbf{e}_\phi$  the solenoid, however, is surrounded by a circular magnetic vector potential, through which the electrons propagate.

The electrons, thus, move through a region where

$$\mathbf{B} = \nabla \times \mathbf{A} = \mathbf{0} \quad \text{with } \mathbf{A} \neq \mathbf{0}. \quad (5.84)$$

Because of (5.75), these relations are not modified if we assume  $U(\mathbf{r}) = -W(\mathbf{r})$  and express  $\mathbf{A}$  in terms of  $W(\mathbf{r})$  by

$$\mathbf{A} = \nabla W(\mathbf{r}) \neq \mathbf{0}. \quad (5.85a)$$



**Fig. 5.7** Scheme of an Aharanov–Bohm interference experiment. The magnetic field  $B$  (perpendicular to the image plane), which is produced by an electric current through a coil, produces a phase shift between the two electron waves representing the interfering electron beams (1) and (2) originating from the double slit set-up

Integration along a line between  $r_0$  and  $r$  yields

$$W(\mathbf{r}) = \int_{r_0}^r d\mathbf{s} \cdot \mathbf{A}(\mathbf{s}). \quad (5.85b)$$

Because of gauge invariance we have two possibilities to represent the wave function of an electron and its Schrödinger equation, respectively, in the original form with  $\mathbf{A} \neq \mathbf{0}$

$$\frac{1}{2m} \left( \frac{\hbar}{i} \nabla - e\mathbf{A} \right)^2 \psi = i\hbar \dot{\psi}, \quad (5.86a)$$

or in the transformed version with

$$\begin{aligned} \mathbf{A}' &= \mathbf{A} - \nabla W(\mathbf{r}) = \mathbf{0}: \\ \frac{1}{2m} \left( \frac{\hbar}{i} \nabla \right)^2 \psi' &= i\hbar \dot{\psi}'. \end{aligned} \quad (5.86b)$$

In the latter case, the action of the vector potential  $\mathbf{A}$  is contained in the Schrödinger equation within the transformed wave function  $\psi'$ , that is, in the factor  $\exp(-ieW/\hbar)$ . With  $\psi'$  as the wave function for vanishing magnetic field  $\mathbf{B} = \mathbf{0}$  [in the Schrödinger equation (5.86b)] the vector potential  $\mathbf{A}$  appears in the original wave function  $\psi(\mathbf{r})$ , according to (5.85a), (5.85b), as follows:

$$\psi(\mathbf{r}) = \psi' \exp\left(\frac{ie}{\hbar} W(\mathbf{r})\right) = \psi' \exp\left[\frac{ie}{\hbar} \int_{r_0}^r d\mathbf{s} \cdot \mathbf{A}(\mathbf{s})\right]. \quad (5.87)$$

Equation (5.87) is the adequate representation for the two partial waves which circulate around the spatially confined magnetic field  $\mathbf{B}$  right (1) and left (2) (Fig. 5.7). Their superposition (5.87) generates the interference on the detector screen:

$$\psi(r) = \psi_1 \exp\left(\frac{ie}{\hbar} \int_1 d\mathbf{s} \cdot \mathbf{A}(\mathbf{s})\right) + \psi_2 \exp\left(\frac{ie}{\hbar} \int_2 d\mathbf{s} \cdot \mathbf{A}(\mathbf{s})\right), \quad (5.88)$$

$\psi_1$  and  $\psi_2$  are the wave functions at vanishing vector potential  $\mathbf{A}$ . The path integrals are defined by the slits at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . By passing the two ways (1) and (2) in inverse direction the sum (or difference) of the two integrals in (5.88) can be rewritten into a path integral over a closed loop around the  $\mathbf{B}$  field in the coil and a magnetic flux  $\Phi_B$  through the loop is defined by:

$$\int_1 d\mathbf{s} \cdot \mathbf{A} - \int_2 d\mathbf{s} \cdot \mathbf{A} = \oint d\mathbf{s} \cdot \mathbf{A}(\mathbf{s}) = \int d\mathbf{f} \cdot \text{curl } \mathbf{A} = \int d\mathbf{f} \cdot \mathbf{B} = \Phi_B. \quad (5.89)$$

By taking the exponential factor in the second term of (5.88) in front of the bracket we obtain, by use of (5.89), the superposition wave function on the detection screen as:

$$\psi(\mathbf{r}) = \exp\left[\frac{ie}{\hbar} \int_2 \mathbf{ds} \cdot \mathbf{A}(\mathbf{s})\right] \left( \psi_1 \exp\left[\frac{ie}{\hbar} \Phi_B\right] + \psi_2 \right). \quad (5.90)$$

This relation reminds us of the double slit interference experiment discussed in Sect. 2.4. Two partial waves with nearly equal wave vector  $\mathbf{k}$  but originating from different slits or holes at  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are superimposed and, thus, represented as

$$\psi = \psi_1 + \psi_2 = C' [e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_1)} + e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_2)}]. \quad (5.91a)$$

Because of (5.90) the effect of the magnetic field, i.e. the magnetic flux  $\Phi_B$  is taken into account in the double slit experiment by adding the phase factor  $\exp(ie\Phi_B/\hbar)$  to the wave function  $\psi_1$ . This yields

$$\psi = Ce^{i\mathbf{k} \cdot \mathbf{r}} [e^{-i(\mathbf{k} \cdot \mathbf{r}_1 + \frac{e}{\hbar} \Phi_B)} + e^{-i\mathbf{k} \cdot \mathbf{r}_2}]. \quad (5.91b)$$

Analogously to the calculation in Sect. 2.4 the intensity on the detection screen, i.e. the probability to find an electron at  $\mathbf{r}$  is obtained as:

$$\begin{aligned} \psi^*(\mathbf{r})\psi(\mathbf{r}) &= 2C^*C \left[ 1 + \cos\left\{\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2) + \frac{e}{\hbar} \Phi_B\right\} \right] \\ &= 2C^*C \left[ 1 + \cos\left\{\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2) + 2\pi \frac{\Phi_B}{\phi_0}\right\} \right]. \end{aligned} \quad (5.92)$$

Hereby the so-called *magnetic flux quantum* was introduced as

$$\phi_0 = h/e \approx 4.14 \times 10^{-15} \text{ J A}^{-1}. \quad (5.93)$$

It depends only on fundamental natural constants. We will frequently encounter this flux quantum in problems with current loops and enclosed magnetic fields.

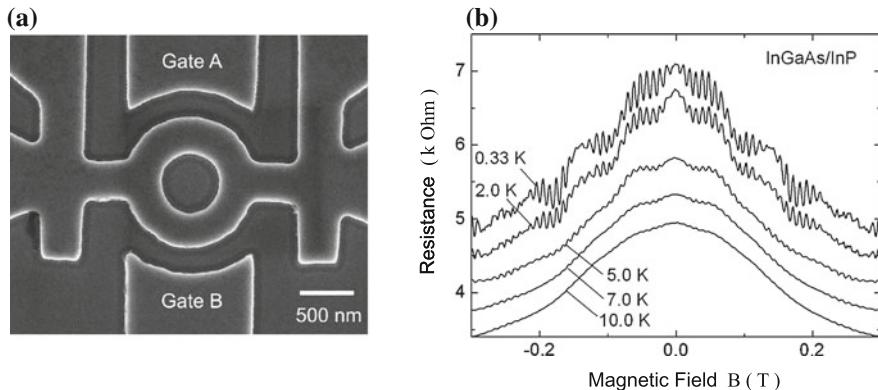
Equation (5.92) shows that the interference pattern arising from the superposition of the two partial waves  $\psi_1$  and  $\psi_2$  is shifted on the screen by a variation of the enclosed magnetic field, that is, the flux through the closed loop of the electron paths. Intensity maxima and minima appearing at vanishing magnetic field at a position determined merely by the phase difference  $\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)$  are shifted according to the value of  $\Phi_B = \int d\mathbf{l} \cdot \mathbf{B}$ . In other words: at a fixed position  $\mathbf{r}$  on the screen maximum or minimum interference intensity is produced by a variation of the magnetic flux  $\Phi_B$ .

The Aharonov–Bohm experiment, as described here, with negligible magnetic field  $\mathbf{B}$  but  $\mathbf{A} \neq \mathbf{0}$  in the region of the electron trajectories was meanwhile made with high precision in electron microscopes [2, 3]. It must be emphasized that such an ideal experiment with  $\mathbf{B} = \mathbf{0}$  outside the coil is not easy to perform. Stray fields must be suppressed very carefully by sophisticated tools. The results of those experiments demonstrate in detail the conclusions (5.92) about the Aharonov–Bohm-effect. This is, in particular, an argument to attribute more importance (reality) to the magnetic vector potential  $\mathbf{A}$  than to the commonly measured magnetic field  $\mathbf{B}$ . We will not dig

into this more philosophical question at this point, but only mention that  $\mathbf{A}$  and  $\mathbf{A}^*$  play a major role in the quantum field theory of the electromagnetic field (Sect. 8.2). These variables will appear as canonically conjugate variables in field quantization (with commutation relations), similarly as  $\mathbf{r}$  and  $\mathbf{p}$  in single particle Schrödinger quantum mechanics.

Aharanov–Bohm interference experiments are meanwhile well established in quantum electronics. By means of electron beam lithography (Appendix B) metallic or semiconductor ring structures with source and drain contacts on opposite sides of the ring are prepared (Fig. 5.8a). The diameter of these rings must be smaller than the mean free path between two electron scattering processes in order to maintain a well defined phase during carrier propagation. For ballistic transport, the mean free path between two inelastic scattering processes on phonons as well as between two elastic processes due to defect scattering must exceed the length of the electron trajectories. But also the weaker condition where elastic scattering in the so-called diffusive transport regime might occur (but not inelastic scattering) allows the observation of Aharanov–Bohm interferences. Roughly speaking, this requires ring diameters in the 500 nm range and below and measurements at low temperature, for example, below 1 K.

Summarizing, the fundamental condition for the observation of Aharanov–Bohm interferences is a well defined phase relation between the two partial waves circulating left and right around the enclosed magnetic flux. This is given in case of ballistic transport or under the weaker condition of only elastic scattering on defects. It is,



**Fig. 5.8** **a, b** Aharanov–Bohm experiment on a semiconductor ring [4, 5]. **a** Mesa ring structure with source (*left*) and drain (*right*) contacts fabricated by electron beam lithography (Appendix B). *Above* and *below* the ring two gate contacts *A* and *B* are prepared as mesas; they allow a change of the electronic Fermi wave vector and, thus, a shift of the phase in the two ring parts with respect to each other. The conducting channel is formed by a 2D electron gas (2DEG) at the interface of an InGaAs/InP heterostructure extended over the whole mesa area (Appendix A). **b** Measured resistance of the ring structure as function of the magnetic field  $B$  penetrating the ring normally. The Aharanov–Bohm oscillations measured at 0.33 and 2 K are suppressed at higher temperatures since inelastic scattering of the electrons on lattice vibrations (Sect. 8.4) destroys phase coherence

furthermore, required that the partial waves arriving at the drain contact have a sufficiently sharp wave vector  $\mathbf{k}$ . Superposition of waves with strongly varying  $\mathbf{k}$  vectors destroys the interference. Thus, the 1D quantum wires forming the ring stricture must have a thickness that only one or at least a very low number of quantum states fit into the confining wire potential (Sect. 3.6.1). The thickness of the wires forming the ring must therefore match the wavelength of the electrons which carry the transport, that is, the Fermi wavelength. In semiconductors electrons in the conduction band (conduction channel) have Fermi wave lengths in the range of 50 nm in comparison to metals with relevant electron wave lengths in the 0.5 nm range.

It is, thus, technologically (lithography, Appendix B) much easier to study Aharonov–Bohm interferences in ring-like semiconductor structures than in metals. Because of the high electron mobilities (large mean free path) rings structured out of 2D electron gases (2DEG) in semiconductor heterostructures as for example, InGaAs/InP, AlGaAs/GaAs or AlGaN/GaN (Appendix A) are particularly suited. An Aharonov–Bohm experiment performed on an InGaAs/InP ring is presented in Fig. 5.8 [4, 5]. A current between left (source) and right (drain) contacts is split into two partial currents through the ring which encloses a magnetic flux due to an applied magnetic field  $B$  normal to the ring (Fig. 5.8a). A variation of the magnetic field strength causes current  $I(B)$  or resistance oscillations as shown in Fig. 5.8b. These oscillations can quantitatively described by Aharonov–Bohm interferences (5.92) due to constructive or destructive superposition of electron waves moving through the two ring arms.

Such experiments on nanoscopic semiconductor or metal rings are, of course, not ideal. The magnetic field penetrating the ring and producing the flux is also present within the current carrying regions and outside the ring. This part of the  $\mathbf{B}$  field does not affect the interferences but, on the other hand, the experiment can not prove the importance of the vector potential  $\mathbf{A}$  rather than the  $\mathbf{B}$  field for the occurrence of Aharonov–Bohm oscillations.

For applications in quantum electronics it is interesting that the relative phase difference between the electron partial waves in the upper and lower part of the ring can also be varied by changing the electric potential in one part of the ring. A potential shift changes the electron occupation within the corresponding part of the wire and, therefore, also the Fermi wave vector  $k_F$  of the current carrying electrons. The phase of the wave function in one arm of the ring is again shifted against that in the other part of the ring, this time by an electric potential change rather than by a change of a magnetic field. For the realization of such an experiment additional gate contacts  $A$  and  $B$  are lithographically structured on the sample in Fig. 5.8a. Different gate voltages at those contacts allow switching between interference maxima and minima of the drain current. An *interference transistor* has been realized which switches between current maxima and minima by changing a gate voltage. Research in this field is often called *phase-based nanoelectronics*.

## 5.5 The Spin

In everyday language, spin has a lot of meanings. In the context of motion, a spinning motion might also be described as whirling. In physics the meaning is more precise. With relation to stars or the earth one speaks of the Earth's spinning motion and means the rotation of the earth (more general of a solid body) around an axis through the earth (body) center. In this sense, spin is also attributed to elementary particles as we will see. A particle thus might have an angular momentum due to a rotation about an internal axis. For massive particles with spatial extension, this sounds as a familiar phenomenon. As we will see, however, in quantum physics even point-like and mass-less particles as photons have a spin. This is again a typically quantum mechanical counter-intuitive phenomenon. In the next section, we will see how the spin of particles was detected and what its peculiarities are in the context of the general rules of angular momenta in quantum physics.

### 5.5.1 Stern–Gerlach Experiment

With the knowledge that electrons moving in an orbit, that is, charged particles with an angular momentum, produce a magnetic moment  $\mu$ , we can understand a fundamental experiment in the early times of quantum mechanics (Stern and Gerlach 1922) [6, 7].

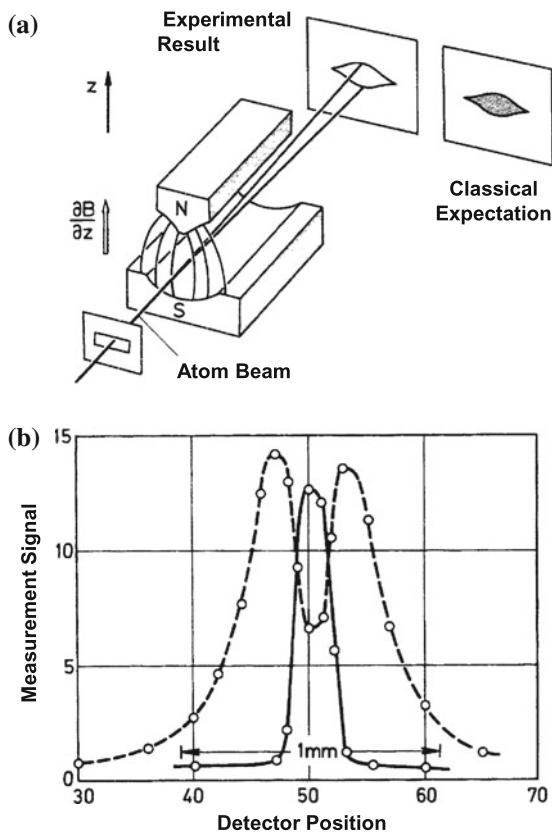
Stern and Gerlach produced a beam of neutral silver (Ag) atoms in a heated oven and sent it through a strongly inhomogeneous static magnetic field  $B(z)$  originating from two asymmetric pole shoes (Fig. 5.9). The silver atoms were collected on a screen where they produced two sharp, clearly separated stripe patterns. This result demonstrates that the atoms exhibit two well defined magnetic dipoles in  $z$  direction. In order to understand the two stripe pattern, we remember that the energy of a magnetic dipole  $\mu$  in a field  $\mathbf{B}$  is  $E = -\mu \cdot \mathbf{B}$  (Sect. 5.4.3). In case of a field being inhomogeneous in  $z$  direction a force in  $z$  direction results according to

$$\mathbf{K} = -\text{grad } E = \nabla(\mu \cdot \mathbf{B}) \simeq \mu_z \frac{\partial B_z}{\partial z} \mathbf{e}_z. \quad (5.94)$$

Depending on the strength of the magnetic dipole the atoms in the beam are deviated more or less from their initial flight direction. Classically a magnetic moment connected to the silver atoms would have a random orientation with respect to the flight direction and a continuous darkening of the screen between two maximum values in  $z$  direction is expected (Fig. 5.9a). The experimental result is in contradiction: The magnetic moment  $\mu_z$  of the silver atoms is quantized, only two sharp well defined values for  $\mu_z$  are derived from the experiment (Fig. 5.9a). This is exactly the result which we would expect for a measurement of the quantized angular momentum. Angular momenta of charged particles are directly related to

**Fig. 5.9 a, b** Stern–Gerlach experiment for the measurement of spin quantization. **a** Scheme of the experimental set-up: an atomic beam passes an inhomogeneous magnetic field and is split into two distinct components. This is in contrast to the classical expectation of a continuous angular distribution.

**b** Measured beam intensity on the detection screen for Ag atoms being used in the experiment described in (a): measurement with magnetic field: *broken line*; measurement without magnetic field: *solid line* [6, 7]



magnetic dipoles (5.69). The experiment obviously demonstrates the quantization of angular momentum. Astonishingly, however, no  $\mu_z = 0$  value is observed nor multiple quantized values. Only two distinct values of the dipole moment, that is, the angular momentum appear. A quantitative analysis of the results yields the two magnetic dipole moments  $\mu_z \approx \pm(e/m_e)\hbar$  which have to be attributed to the silver atoms.

Silver atoms have a nucleus of protons and neutrons and a couple of electrons in their electronic shell. An interpretation of the results of the Stern–Gerlach experiment, thus, requires a detailed knowledge of the electronic structure of silver atoms. A direct, more evident approach to the understanding of the results is derived from a similar experiment by Phipps and Taylor [8], who performed the experiment of Stern and Gerlach, but with neutral hydrogen ( $H$ ) atoms rather than with silver atoms. These authors obtained the same result, qualitatively and quantitatively, with  $H$  atoms as earlier Stern and Gerlach with Ag atoms.  $H$  atoms have only one single electron in their outer shell, which is bounded by the Coulomb potential of the positive nucleus of a proton. Since the magnetic moment, according to (5.69), is inversely

proportional to the particle mass, the dipole moment of the nucleus (proton) can not play a significant role for the observed splitting of  $\mu_z$  into two components (proton mass  $\approx 2000 \times$  electron mass). The observed magnetic dipole moment, thus, has to be ascribed to the electron within the H atom. Under the experimental conditions the electron of the H atom must be in its quantum mechanical ground state with vanishing orbital angular momentum  $L = 0$ . The energies in the experiment did not allow an excitation of the electron into an excited state with  $L \neq 0$ .

The conclusion derived from these experiments is straightforward: An electron of the hydrogen or the silver atom carries an inherent magnetic dipole moment which can be oriented within a magnetic field  $B_z$  parallel or antiparallel to the field. Depending on this orientation, the atoms containing that electron are deviated from their initial path into two well defined directions observed on the screen (Fig. 5.9).

According to the general laws of electrodynamics (Sect. 5.4.3), the magnetic moment of a charged particle must be connected to an angular momentum of the particle. The magnetic dipole moment  $\hat{\mu}$  of an electron on a spatial orbit is related with its angular momentum  $\hat{L}$  via  $\hat{\mu} = (e/2m_e)\hat{L}$ . This relation must not necessarily hold for the inherent magnetic moment of a spinning electron (around internal axis), which does not arise from a circular motion of the particle on an orbit. Therefore, the so-called Lande factor or the gyro-magnetic ratio  $g$  is introduced in order to connect the elementary magnetic moment of an electron itself with its inherent angular momentum, called *spin*:

$$\hat{\mu}_{\text{spin}} = g \frac{e}{2m_e} \hat{\mathbf{S}}. \quad (5.95)$$

Both magnetic moment and spin are quantum-mechanical observables which are described by operators (roof symbols) rather than by normal numbers. The eigenvalues of these operators are found as numbers in an experiment (results of a measurement). By the assumption that the gyro-magnetic ratio  $g$  in (5.95) equals 2 the results of the Stern–Gerlach and the Phipps–Taylor experiment can be explained quantitatively. In particular, the rules for angular momentum quantization (5.28a)–(5.28c), (5.29a), (5.29b) can be fulfilled, including  $\mu_z = 0$  (i.e., also orientation quantum number  $m = 0$ ). We must only assume that the  $z$  component of the spin can have the following eigenvalues:

$$S_z = m_s \hbar \quad \text{with } m_s = \pm \frac{1}{2}. \quad (5.96)$$

The spin orientation quantum number  $m_s$ , then, fulfills the condition of half integrity (5.29b), a condition which is excluded for orbital angular momenta of particles moving on a curved trajectory because of (5.35b). The spin as the inherent angular momentum of an electron can assume only the two measurement values  $S_z = \pm(1/2)\hbar$  according to the Stern–Gerlach and the Phipps–Taylor experiment. However, any multiple half-integer values  $(1/2)\hbar, (3/2)\hbar, (5/2)\hbar, \dots$  of angular momentum (5.29b) are obtained for a particle with spin  $\pm\hbar/2$  which simultaneously has an orbital momentum  $\hat{L}$ . Both angular momenta, then, have to be summed up and the operator of the total angular momentum

$$\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}} \quad (5.97)$$

fulfills the eigenvalue equations

$$\hat{J}_z |l, m, m_s\rangle = (m + m_s)\hbar |l, m, m_s\rangle, \quad (5.98a)$$

$$\hat{\mathbf{J}}^2 |l, m, m_s\rangle = j(j+1)\hbar^2 |l, m, m_s\rangle, \quad (5.98b)$$

where  $m_s \pm (1/2)$  is introduced as spin quantum number. Taking together spin and orbital angular momentum of an electron the total orientation quantum number of the electron  $m_z = m + m_s$  (5.98a) assumes the following values:

$$m_z = -j, -j+1, \dots, j-1, j \quad \text{with } j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots \quad (5.99)$$

This spectrum of quantum numbers, of course, is only obtained for particles with spin one half ( $s = 1/2, m_s = \pm 1/2$ ). Particles with zero spin ( $s = 0$ ), they do exist, the series of integer orientation quantum numbers (5.35b) as for the orbital angular momentum alone is obtained.

At this point it is worth reflecting a little bit on the spin degree of freedom, a quite counter-intuitive property a particle can have. According to present knowledge the electron with a spatial extension of about  $2.8 \times 10^{-13}$  cm is a point-like particle without any internal structure as for example, the proton (contains 3 quarks, Sect. 5.6.4). Nevertheless, this point-like particle carries an inherent angular momentum, the spin, which is related to a magnetic dipole.

The corresponding elementary magnetic dipole moment of the electronic spin is called *Bohr magneton*  $\mu_B$ , its value amounts to

$$\mu_B = e\hbar/2m_e \approx 9.28 \times 10^{-24} \text{ JT}^{-1}. \quad (5.100)$$

The spin is a truly quantum mechanical phenomenon, which can not be derived classically on the basis of charged carriers circulating on an orbit.

Even though counter-intuitive at first glance, the phenomenon of spin and its related elementary magnetic moment instantly follows mathematically from the unification of single particle quantum mechanics with the rules of special relativity, that is, the four dimensional space-time world of relativity theory. In order to make Schrödinger's quantum mechanics compatible with special relativity Dirac invented an extension of the Schrödinger equation, the so-called Dirac equation which requires as solutions pairs of wave functions, so-called spinors, rather than simple scalar wave functions  $\psi(\mathbf{r}, t)$ . The two wave function components of a spinor finally emerge as wave functions related to two different spin orientations. A further success of Dirac's theory of a quantum mechanical relativistic electron is the derivation of the gyro-magnetic ratio (Lande factor)  $g$  of exactly 2 for the spin. As it has been shown later on in quantum field theory (interaction of electrons with quantized electromagnetic field, Sect. 8.2) and also in experiment, the  $g$  value exceeds 2 by a tiny amount

( $g \cong 2.0023 \dots$ ). Experiment and theory agree up to the 7th decimal, a significant success both for theory and experiment.

### 5.5.2 The Spin and Its 2D Hilbert Space

The spin  $\mathbf{S}$  is an angular momentum, that is, a vector defined in 3D Euclidean space. In quantum physics the three vector components, however, must be operators:  $\hat{\mathbf{S}} = (\hat{S}_x, \hat{S}_y, \hat{S}_z)$ . The dimension of the Hilbert space attributed to these operators is given by the number of possible measurement results for these operators. Since only two distinct spin states  $\pm \hbar/2$  are possible the corresponding Hilbert spaces of the operators  $\hat{S}_x, \hat{S}_y, \hat{S}_z$  are 2-dimensional. As for angular momentum operators in general the following commutation rules (5.7a)–(5.7d), (5.8) are valid also for spin operators:

$$[\hat{S}_x, \hat{S}_y] = i\hbar \hat{S}_z, \quad (5.101a)$$

$$[\hat{S}_y, \hat{S}_z] = i\hbar \hat{S}_x, \quad (5.101b)$$

$$[\hat{S}_z, \hat{S}_x] = i\hbar \hat{S}_y, \quad (5.101c)$$

$$\hat{\mathbf{S}} \times \hat{\mathbf{S}} = i\hbar \hat{\mathbf{S}}. \quad (5.101d)$$

Only one single component of  $\hat{\mathbf{S}}$ , e.g.  $\hat{S}_z$  can be measured as a sharp value simultaneously with  $\hat{\mathbf{S}}^2$ , that is,

$$[\hat{\mathbf{S}}^2, \hat{S}_z] = [\hat{\mathbf{S}}^2, \hat{S}_x] = [\hat{\mathbf{S}}^2, \hat{S}_y] = 0. \quad (5.102)$$

$\hat{S}_z$  and  $\hat{\mathbf{S}}^2$  have the same system of eigenstates with the eigenvalues  $\pm \hbar/2$  ( $m_s = \pm 1/2$ ) for  $\hat{S}_z$ . In the abstract Dirac notation, the two spin states related to these eigenvalues are sometimes denoted as  $|+\rangle$  and  $|-\rangle$  or respectively,  $|\uparrow\rangle$  and  $|\downarrow\rangle$  for spin up and spin down. Correspondingly the eigenvalue equations for the spin operators are written as:

$$\hat{S}_z |\uparrow\rangle = +\frac{\hbar}{2} |\uparrow\rangle, \quad (5.103a)$$

$$\hat{S}_z |\downarrow\rangle = -\frac{\hbar}{2} |\downarrow\rangle, \quad (5.103b)$$

$$\hat{\mathbf{S}}^2 |\uparrow\rangle = s(s+1)\hbar^2 |\uparrow\rangle = \frac{3}{4}\hbar^2 |\uparrow\rangle, \quad (5.103c)$$

$$\hat{\mathbf{S}}^2 |\downarrow\rangle = s(s+1)\hbar^2 |\downarrow\rangle = \frac{3}{4}\hbar^2 |\downarrow\rangle. \quad (5.103d)$$

Equations (5.103a)–(5.103d) reflect the general relations for angular momenta (5.28a)–(5.28c). At this point, for the first time, we encounter a Hilbert space with

only two dimensions (2D Hilbert space); it is spanned by only two eigenkets  $|\uparrow\rangle$  and  $|\downarrow\rangle$ . We have to distinguish this 2D spin Hilbert space from the 3D Euclidean space in which the three spin vector components (operators)  $\hat{S}_x$ ,  $\hat{S}_y$ ,  $\hat{S}_z$  are defined. As in our infinite dimensional and continuous Hilbert spaces of the operators  $\hat{H}$ ,  $\hat{\mathbf{p}}$ ,  $\hat{\mathbf{x}}$  etc. (Chap. 4), the spin eigenstates must be orthonormal in the 2D spin Hilbert space, that is,

$$\langle \uparrow | \downarrow \rangle = 0; \quad \langle \uparrow | \uparrow \rangle = \langle \downarrow | \downarrow \rangle = 1. \quad (5.104)$$

The completeness relation, of course, contains only two terms to be summed up:

$$|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow| = \hat{1}. \quad (5.105)$$

It guarantees that every possible spin state can be expressed as a superposition of the two eigenstates  $|\uparrow\rangle$  and  $|\downarrow\rangle$ . A general spin state  $|s\rangle$  is, thus, expressed as:

$$|s\rangle = \alpha_+ |\uparrow\rangle + \alpha_- |\downarrow\rangle, \quad (5.106)$$

where  $\alpha_+$ ,  $\alpha_-$  are probability amplitudes which describe to what extent the two eigenstates  $|\uparrow\rangle$  and  $|\downarrow\rangle$  are contained in the general spin state. The normalization condition  $\langle s|s \rangle = 1$  requires as usually:

$$|\alpha_+|^2 + |\alpha_-|^2 = 1. \quad (5.107)$$

As was discussed in the context of infinite dimensional Hilbert spaces (Sect. 4.3), a representation of spin states in terms of 2-dimensional (2D) vectors of the 2D Hilbert space is possible. The vector components are the probability amplitudes of (5.106) and we get:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_- \\ \alpha_+ \end{pmatrix}, \quad \text{respectively} \quad \boldsymbol{\alpha}^* = (\alpha_+^*, \alpha_-^*). \quad (5.108)$$

We can also say,  $\alpha_+$  and  $\alpha_-$  are the projections of the general spin state  $|s\rangle$  on the two axis of the 2D Hilbert space:

$$|\uparrow\rangle \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |\downarrow\rangle \rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (5.109)$$

that is,

$$\alpha_+ = \langle \uparrow | s \rangle, \quad \alpha_- = \langle \downarrow | s \rangle. \quad (5.110)$$

A general spin operator, for example,  $\hat{S}_z$  is represented in this 2D Hilbert space as a 2D matrix:

$$\underline{\underline{S}}_z = \begin{pmatrix} \langle \uparrow | \hat{S}_z | \uparrow \rangle & \langle \uparrow | \hat{S}_z | \downarrow \rangle \\ \langle \downarrow | \hat{S}_z | \uparrow \rangle & \langle \downarrow | \hat{S}_z | \downarrow \rangle \end{pmatrix}. \quad (5.111)$$

In analogy to the general step operators of the angular momentum  $\hat{L}_\pm$  (5.18a), (5.22), we introduce the following step operators for the spin:

$$\hat{S}_\pm = \hat{S}_x \pm i\hat{S}_y. \quad (5.112a)$$

The inverse operators are

$$\hat{S}_x = \frac{1}{2}(\hat{S}_+ + \hat{S}_-), \quad \hat{S}_y = \frac{1}{2i}(\hat{S}_+ - \hat{S}_-), \quad (5.112b)$$

$\hat{S}_+$  and  $\hat{S}_-$  operate between two states (steps) only, namely the two spin states:

$$\hat{S}_+|\uparrow\rangle = 0, \quad \hat{S}_-|\uparrow\rangle = \hbar|\downarrow\rangle, \quad (5.113a)$$

$$\hat{S}_+|\downarrow\rangle = \hbar|\uparrow\rangle, \quad \hat{S}_-|\downarrow\rangle = 0. \quad (5.113b)$$

From the vector representation (5.109) of  $|\uparrow\rangle$  and  $|\downarrow\rangle$  and from (5.113a), (5.113b) the matrix representation of the spin step operators is derived:

$$\underline{\underline{S}}_+ = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \underline{\underline{S}}_- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \underline{\underline{S}}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.114)$$

According to (5.112b) the three components of the spin can, thus, be expressed as

$$\underline{\underline{S}}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\underline{S}}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{\underline{S}}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.115)$$

Note that this representation involves the arbitrary choice of  $\hat{S}_z$  as the measured spin component. This choice causes the matrix  $\underline{\underline{S}}_z$  to be diagonal with the eigenvalues (diagonal elements)  $\pm\hbar/2$ .

According to (5.115), Pauli has introduced the following matrices (called *Pauli matrices*):

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.116)$$

By means of the Pauli matrices, the spin vector  $\mathbf{S}$  whose components are 2D matrices is represented as

$$\mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma}. \quad (5.117)$$

Pauli matrices are used in many quantum mechanical problems other than spin related ones. All 2D Hilbert spaces can be described by use of Pauli matrices and corresponding 2D vectors. Quantum systems which can be described approximately by

two distinct quantum states only fall into this category. We will discuss these issues more in detail in Sect. 6.5.

The spin degree of freedom can only be probed in an experiment if the Schrödinger equation or the Hamilton operator of the particular problem explicitly contains a term with spin operators. Therefore, a magnetic field or more generally an electromagnetic field must be present. In this case, the Hamiltonian for an electron is obtained as:

$$\hat{H} = \frac{1}{2m} (\hat{\mathbf{p}} - e\hat{\mathbf{A}}(\mathbf{r}, t))^2 + V(\mathbf{r}, t) + \mu_B \hat{\boldsymbol{\sigma}} \cdot \mathbf{B}. \quad (5.118)$$

The potential energy  $V(\mathbf{r}, t)$  contains time dependent electric fields in terms of the electric potential  $e\Phi(\mathbf{r}, t)$ . The magnetic field enters the Hamiltonian two times, in terms of the vector potential  $\mathbf{A}$  and as the scalar product of the  $\mathbf{B}$  field with the Pauli (spin) operator  $\hat{\boldsymbol{\sigma}}$ . This latter term, where the spin operator enters, is the energy of the magnetic dipole ascribed, via the Bohr magneton  $\mu_B$  (5.100), to the electronic spin and the external magnetic field  $\mathbf{B}$ . For vanishing magnetic field the spin operator disappears from the Hamiltonian and a normal Schrödinger equation without spin has to be solved. But note, also in this case two wave functions  $\psi_{\uparrow}(\mathbf{r}, t)$  and  $\psi_{\downarrow}(\mathbf{r}, t)$  exist as solutions according to two spin orientations.

For  $\mathbf{B} = \mathbf{0}$  these two wave functions are identical. The states are degenerate in their spin degree of freedom.

The spin degree of freedom is independent of any spatial degree of freedom. Spin can be measured with absolute accuracy simultaneously with observables as position  $\hat{\mathbf{r}}$ , momentum  $\hat{\mathbf{p}}$  or orbital angular momentum  $\hat{\mathbf{L}}$ . The corresponding operators commute:

$$[\hat{\mathbf{S}}, \hat{\mathbf{r}}] = [\hat{\mathbf{S}}, \hat{\mathbf{p}}] = [\hat{\mathbf{S}}, \hat{\mathbf{L}}] = 0. \quad (5.119)$$

A general state containing spatial degrees of freedom (described by wave function  $\psi$ ) and the spin degrees of freedom, thus, is a product of position and spin states. In a chosen basis  $|\mathbf{r}\rangle|\uparrow\rangle$  and  $|\mathbf{r}\rangle|\downarrow\rangle$ , the probability amplitudes  $\alpha_+$  and  $\alpha_-$  (5.108) must be replaced by the position dependent amplitudes  $\psi_{\uparrow}(\mathbf{r})$  and  $\psi_{\downarrow}(\mathbf{r})$ , the corresponding spin specific wave functions. A general state consisting of a superposition of all possible single states is, then, represented as:

$$|\phi\rangle = \int d^3r [\psi_{\uparrow}(\mathbf{r})|\mathbf{r}\rangle|\uparrow\rangle + \psi_{\downarrow}(\mathbf{r})|\mathbf{r}\rangle|\downarrow\rangle]. \quad (5.120a)$$

The representations in position and spin Hilbert spaces, that is, the corresponding projections of (5.120a) on to the spin and position eigenstates are obtained, as expected:

$$\langle \mathbf{r} | \phi \rangle = \psi_{\uparrow}(\mathbf{r})|\uparrow\rangle + \psi_{\downarrow}(\mathbf{r})|\downarrow\rangle, \quad (5.120b)$$

$$\langle \uparrow | \langle \mathbf{r} | \phi \rangle = \psi_{\uparrow}(\mathbf{r}); \quad \langle \downarrow | \langle \mathbf{r} | \phi \rangle = \psi_{\downarrow}(\mathbf{r}). \quad (5.120c)$$

Equation (5.120b) suggests that in the presence of external magnetic fields the treatment of problems with spin is most conveniently done in the 2D vector and matrix representation. In the position representation, 2D spinors are used:

$$\phi(\mathbf{r}) = \begin{pmatrix} \psi_{\uparrow}(\mathbf{r}) \\ \psi_{\downarrow}(\mathbf{r}) \end{pmatrix}. \quad (5.121)$$

This finally leads to a 2D Schrödinger equation where the spin is represented in terms of Pauli matrices  $\sigma$ :

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} \psi_{\uparrow} \\ \psi_{\downarrow} \end{pmatrix} = \left[ \left\{ \frac{1}{2m} \left( \frac{\hbar}{i} \nabla - e\mathbf{A} \right)^2 + V(\mathbf{r}, t) \right\} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mu_B \boldsymbol{\sigma} \cdot \mathbf{B} \right] \begin{pmatrix} \psi_{\uparrow} \\ \psi_{\downarrow} \end{pmatrix}. \quad (5.122)$$

This is the so-called non-relativistic *Pauli equation*. It replaces the normal scalar Schrödinger equation in case where an external magnetic field induces significant spin effects, that is, where spin degeneracy is removed.

### 5.5.3 Spin Precession

The simplest dynamics of a spin is given for a localized electron in an external magnetic field.

The electron might be confined in a spatially fixed atom (ion trap) or in a 0D quantum dot (Sect. 3.6.1). Translational degrees of freedom can be neglected and for a constant potential, e.g. in the quantum dot, we can assume  $V(\mathbf{r}) = 0$ . The Pauli equation (5.122) for a general spin state  $|s\rangle$  of the electron in a constant external magnetic field  $\mathbf{B}$  is, then obtained as

$$i\hbar \frac{\partial}{\partial t} |s\rangle = \mu_B \boldsymbol{\sigma} \cdot \mathbf{B} |s\rangle = \mu_B \hat{\sigma}_z B_z |s\rangle. \quad (5.123)$$

We have chosen the magnetic field in  $z$  direction, such that only the spin component  $\hat{\sigma}_z$  appears in the Hamiltonian.

The general spin state  $|s\rangle$  is a superposition of  $|\uparrow\rangle$  and  $|\downarrow\rangle$ :

$$|s\rangle = \alpha_+(t)|\uparrow\rangle + \alpha_-(t)|\downarrow\rangle = a_+ e^{-iE_{\uparrow}t/\hbar} |\uparrow\rangle + a_- e^{-iE_{\downarrow}t/\hbar} |\downarrow\rangle. \quad (5.124)$$

Since  $|\uparrow\rangle$  and  $|\downarrow\rangle$  are stationary eigenstates their probability amplitudes must exhibit the characteristic exponential time dependence with the corresponding energy eigenvalues  $E_{\uparrow}$  and  $E_{\downarrow}$  in the exponents.

For the solution of the Schrödinger (Pauli) equation (5.123), we insert the ansatz (5.124) and obtain

$$i\hbar \frac{\partial}{\partial t} [\alpha_+(t)|\uparrow\rangle + \alpha_-(t)|\downarrow\rangle] = \mu_B B_z \hat{\sigma}_z [\alpha_+(t)|\uparrow\rangle + \alpha_-(t)|\downarrow\rangle]. \quad (5.125a)$$

Time differentiation of  $\alpha_+(t)$  and  $\alpha_-(t)$  yields

$$(E_\uparrow \alpha_+ |\uparrow\rangle + E_\downarrow \alpha_- |\downarrow\rangle) = \mu_B B_z \hat{\sigma}_z (\alpha_+ |\uparrow\rangle + \alpha_- |\downarrow\rangle). \quad (5.125b)$$

We multiply these equations from the left side with  $\langle \uparrow |$  and subsequently with  $\langle \downarrow |$  and obtain, by means of the orthogonality relations (5.104):

$$E_\uparrow \alpha_+ = \mu_B B_z \langle \uparrow | \hat{\sigma}_z | \uparrow \rangle \alpha_+ + \mu_B B_z \langle \uparrow | \hat{\sigma}_z | \downarrow \rangle \alpha_-, \quad (5.126a)$$

$$E_\downarrow \alpha_- = \mu_B B_z \langle \downarrow | \hat{\sigma}_z | \downarrow \rangle \alpha_- + \mu_B B_z \langle \downarrow | \hat{\sigma}_z | \uparrow \rangle \alpha_+. \quad (5.126b)$$

For the calculation of the 2D matrix elements  $\langle \uparrow | \hat{\sigma}_z | \uparrow \rangle$  etc., we use the 2D vector representation of spin states (5.109) and obtain for example:

$$\langle \downarrow | \hat{\sigma}_z | \uparrow \rangle = (0, 1) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (0, 1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0. \quad (5.126c)$$

Analogous calculations (5.126a)–(5.126c) yield:

$$E_\uparrow = \mu_B B_z, \quad E_\downarrow = -\mu_B B_z. \quad (5.127a)$$

As expected we find two distinct eigenvalues or energies of the stationary spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  in a magnetic field, corresponding to the two spin orientations in the magnetic field. Using the expression for the Bohr magneton  $\mu_B = e\hbar/2m$  and the frequency  $\omega_0 = eB_z/m$  [identical with cyclotron frequency (5.56b)] we can write these two energy eigenvalues as:

$$E_\uparrow/\hbar = \frac{e}{2m} B_z = \omega_0/2, \quad (5.127b)$$

$$E_\downarrow/\hbar = -\frac{e}{2m} B_z = -\omega_0/2. \quad (5.127c)$$

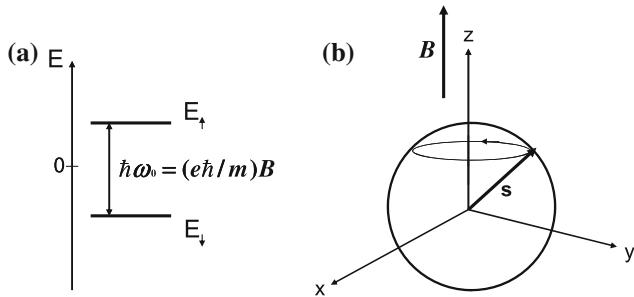
The general spin state (5.124), then, is expressed as

$$|s\rangle = \alpha_+ |\uparrow\rangle + \alpha_- |\downarrow\rangle = a_+ e^{-i\omega_0 t/2} |\uparrow\rangle + a_- e^{i\omega_0 t/2} |\downarrow\rangle. \quad (5.128a)$$

Normalization of the state requires:

$$\langle s | s \rangle = |\alpha_+|^2 + |\alpha_-|^2 = |a_+|^2 + |a_-|^2 = 1. \quad (5.128b)$$

In order to get an intuitive picture of the motion of the spin in a magnetic field, we calculate the expectation values of the spin angular momentum in the three space directions  $\langle (\hbar/2)\hat{\sigma}_x \rangle$ ,  $\langle (\hbar/2)\hat{\sigma}_y \rangle$ ,  $\langle (\hbar/2)\hat{\sigma}_z \rangle$ . These quantities can directly be com-



**Fig. 5.10** **a, b** Precession of a spin ( $\mathbf{s}$ ) in a constant magnetic field  $\mathbf{B}$  in  $z$  direction. **a** Two stable precession modes with spin orientation in  $B_z$  and in negative  $B_z$  direction and with energies  $E_{\uparrow}$  and  $E_{\downarrow}$  are possible. **b** Schematic plot of the spin precession mode with energy  $E_{\uparrow}$

pared with classical dynamical variables. In 2D vector representation, we get

$$\begin{aligned}\left\langle \frac{\hbar}{2}\hat{\sigma}_z\right\rangle &= \frac{\hbar}{2}\langle s|\hat{\sigma}_z|s\rangle = \frac{\hbar}{2}(\alpha_+^*, \alpha_-^*) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_+ \\ \alpha_- \end{pmatrix} \\ &= \frac{\hbar}{2}(\alpha_+^*, \alpha_-^*) \begin{pmatrix} \alpha_+ \\ -\alpha_- \end{pmatrix} = \frac{\hbar}{2}(|\alpha_+|^2 - |\alpha_-|^2) \\ &= \frac{\hbar}{2}(|a_+|^2 - |a_-|^2).\end{aligned}\quad (5.129)$$

According to (5.124), the amplitudes  $a_+$  and  $a_-$  do not depend on time. The  $z$  component of the spin (expectation value) is a dynamical constant of the motion in the magnetic field  $B_z$ . This is not true for the  $x$  and  $y$  components as is easily shown:

$$\begin{aligned}\left\langle \frac{\hbar}{2}\hat{\sigma}_x\right\rangle &= \frac{\hbar}{2}(\alpha_+^*, \alpha_-^*) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_+ \\ \alpha_- \end{pmatrix} = \frac{\hbar}{2}(\alpha_+^*, \alpha_-^*) \begin{pmatrix} \alpha_- \\ \alpha_+ \end{pmatrix} \\ &= \frac{\hbar}{2}(\alpha_+^*\alpha_-^* + \alpha_-^*\alpha_+^*) = a_+a_-\hbar\cos\omega_0t.\end{aligned}\quad (5.130a)$$

Analogously, we get for the  $y$  component

$$\langle S_y \rangle = \left\langle \frac{\hbar}{2}\hat{\sigma}_y \right\rangle = a_+a_-\hbar\sin\omega_0t. \quad (5.130b)$$

According to (5.130a) and (5.130b) the expectation values  $\langle S_x \rangle$  and  $\langle S_y \rangle$  of the spin angular momentum move on a circular path around the magnetic field  $B_z$  ( $z$  direction) with a frequency  $\omega_0 = eB_z/m$  (Fig. 5.10).

When we imagine the quantum mechanical spin as a classical peg top or gyro (rigorously not allowed) this top turns around itself (around internal axis) and simultaneously the internal rotation axis of the peg top circles around the magnetic field

lines as axis (Fig. 5.10). This characteristic spin motion in a magnetic field is called precession with the *precession frequency*  $\omega_0$ . Two stable configurations of this motion are possible,  $z$  component of the spin in  $B_z$  direction and opposite to it, corresponding to the two energy eigenvalues  $E_\uparrow$  and  $E_\downarrow$ . In Sect. 6.5.2, we will learn how small oscillating electromagnetic fields of adequate frequency can flip the spin orientation between these two stable orientations.

## 5.6 Particle Categories: Fermions and Bosons

### 5.6.1 Two and More Particles

So far we have treated quantum mechanical properties only for one single particle, usually the electron. When we consider more than one particle, in the simplest case only two, we encounter totally new unexpected quantum phenomena solely due to the uncertainty in the description of position and momentum of particles. When the wave functions of two identical particles overlap in space we have no chance to distinguish between the two particles, that is, to follow their individual path in that spatial region. The particles are un-distinguishable. This is in contrast to classical particles, they are always distinguishable. We can follow their individual trajectories with absolute accuracy (at least in the abstract ideal case). In addition, we can attribute properties, as for example, color, to classical particles (at least bigger ones) which do not affect their dynamics.

In quantum mechanics, two identical particles are solely described by their wave function  $\psi(\mathbf{r}_1, \mathbf{r}_2, t)$  where  $P(\mathbf{r}_1, \mathbf{r}_2, t) = |\psi(\mathbf{r}_1, \mathbf{r}_2, t)|^2$  is the probability density at time  $t$  to find particle (1) at  $\mathbf{r}_1$  and particle (2) at  $\mathbf{r}_2$ . A measurement of the particle position at a time  $t$  yields  $\mathbf{r}_1, \mathbf{r}_2$  and at a later time  $t'$  the positions  $\mathbf{r}'_1$  and  $\mathbf{r}'_2$ . At each time, only two particles are detected at two differing positions, it is totally undefined in the framework of quantum mechanics what particle of the two is detected here or there. The two-particle wave function  $\psi(\mathbf{r}_1, \mathbf{r}_2, t)$  is defined in a 6-dimensional (6D) position space at a time  $t$ . Together with the spin degree of freedom the wave function must be expressed as  $\psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, t)$  with  $s_i$  as spin quantum numbers. The definition of  $P(\mathbf{r}_1, \mathbf{r}_2)$  as the probability density to find the particles at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , respectively, requires the normalization of the two-particle wave function, this time over the 6D space of the two particles. The normalization guarantees that two particles are found in any case somewhere in the normalization volume. Normalization is expressed as:

$$1 = \langle \psi | \psi \rangle = \int |\langle \mathbf{r}_1, \mathbf{r}_2 | \psi \rangle|^2 d^3 r_1 d^3 r_2 = \int P(\mathbf{r}_1, \mathbf{r}_2) d^3 r_1 d^3 r_2. \quad (5.131a)$$

The two-particle Schrödinger equation and Hamiltonian, respectively, contains the kinetic energies  $\hat{T}_1 = \hat{p}_1^2/2m$  and  $\hat{T}_2 = \hat{p}_2^2/2m$  of the two identical particles and a potential  $V(\mathbf{r}_1, \mathbf{r}_2)$  being dependent on both particle coordinates, since a force

acts on both particles, mostly mutually. Without spin the Hamiltonian for the two identical particles (equal mass) is, thus, obtained as

$$\hat{H} = \frac{\hat{p}_1^2}{2m} + \frac{\hat{p}_2^2}{2m} + V(\mathbf{r}_1, \mathbf{r}_2). \quad (5.131b)$$

It is immediately obvious that the mathematical description of one single particle in two dimensions is identical with that of two particles, each in one dimension. As an example, we consider a system of two identical independent harmonic oscillators (1) and (2). Their Hamiltonian is

$$\hat{H} = \frac{\hat{p}_1^2 + \hat{p}_2^2}{2m} + \frac{1}{2}m\omega x_1^2 + \frac{1}{2}m\omega x_2^2. \quad (5.132)$$

We can assume  $\hat{x}_i$ ,  $\hat{p}_i$  as 1D position and momentum operators of the  $i$ th particle ( $i = 1, 2$ ) or equivalently  $x_1$ ,  $x_2$  and  $\hat{p}_1$ ,  $\hat{p}_2$  as the operators belonging to the two coordinates of one and the same particle.

Furthermore, if we can separate the potential  $V(\mathbf{r}_1, \mathbf{r}_2)$  into two separate potentials of (in this case different) non-interacting particles  $V(\mathbf{r}_1, \mathbf{r}_2) = V(\mathbf{r}_1) + V(\mathbf{r}_2)$ , the Hamiltonian is written as

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = \frac{\hat{p}_1^2}{2m_1} + V_1(\mathbf{r}_1) + \frac{\hat{p}_2^2}{2m_2} + V_2(\mathbf{r}_2). \quad (5.133)$$

It can be separated into two separate operators  $\hat{H}_1$  and  $\hat{H}_2$ , each one acting only on particle (1) and particle (2), respectively. The Schrödinger equation of the two particles, then, is obtained as

$$i\hbar \frac{\partial}{\partial t} |\psi(\mathbf{r}_1, \mathbf{r}_2t)\rangle = (\hat{H}_1 + \hat{H}_2) |\psi(\mathbf{r}_1, \mathbf{r}_2t)\rangle. \quad (5.134)$$

A product ansatz for the two-particle quantum state separates the problem:

$$|\psi(\mathbf{r}_1, \mathbf{r}_2t)\rangle = |\psi_1\rangle |\psi_2\rangle, \quad (5.135)$$

and we obtain from (5.134):

$$i\hbar \frac{\partial}{\partial t} |\psi_1\rangle |\psi_2\rangle = (\hat{H}_1 + \hat{H}_2) |\psi_1\rangle |\psi_2\rangle, \quad (5.136a)$$

$\hat{H}_1$  acts on  $|\psi_1\rangle$ , only, and  $\hat{H}_2$  on  $|\psi_2\rangle$ . The two single particle problems, therefore, allow the familiar ansatz  $|\psi_1 \propto \exp(-iE_1 t/\hbar)\rangle$  respectively,  $|\psi_2 \propto \exp(-iE_2 t/\hbar)\rangle$  and we obtain:

$$i\hbar \frac{\partial}{\partial t} |\psi_1\rangle |\psi_2\rangle = (E_1 + E_2) |\psi_1\rangle |\psi_2\rangle. \quad (5.136b)$$

For two non-interacting particles the single energy eigenvalues  $E_1$  and  $E_2$  are constants of motion and the total energy of the system can be expressed as the sum of the single particle energies:

$$E = E_1 + E_2. \quad (5.137)$$

With  $|E_1\rangle$  and  $|E_2\rangle$  as the time-independent eigenstates which are obtained from the separated eigenvalue equations

$$\hat{H}_1|E_1\rangle = E_1|E_1\rangle, \quad \hat{H}_2|E_2\rangle = E_2|E_2\rangle \quad (5.138)$$

the general state of the two-particle system without particle interaction is expressed as

$$|\psi(t)\rangle = |E_1\rangle e^{-iE_1 t/\hbar} |E_2\rangle e^{-iE_2 t/\hbar}. \quad (5.139)$$

We want to keep in mind: The total energy of two non-interacting particles is the sum of the two single particle energies:  $E = E_1 + E_2$ . Not only  $E$  but also  $E_1$  and  $E_2$  separately are constants of motion. The wave function, respectively the state of the total system (5.139) is the product of the single particle states. The number of dimensions of the new two particle Hilbert space is obtained as the product of the dimensions of the two single particle Hilbert spaces. This is also true for the case of two interacting particles, where the total state of the system can not be expressed as the product of the two single particle states.

A system of two spins (connected to electrons) shall be considered as an example. The Hilbert space of one single spin is spanned by the states  $|\uparrow\rangle$  and  $|\downarrow\rangle$ . A two-spin Hilbert space contains the following four states  $|\uparrow\rangle|\uparrow\rangle$ ,  $|\uparrow\rangle|\downarrow\rangle$ ,  $|\downarrow\rangle|\uparrow\rangle$ ,  $|\downarrow\rangle|\downarrow\rangle$ . These states span the 4D Hilbert space, independent on a possible interaction between the spins. Each two-spin state can be represented as a superposition of these four states.

Since two subatomic particles can not be distinguished except by different quantum numbers, we are lead to another quantum physical peculiarity. Imagine a two-particle state  $|a, b\rangle = |a \rightarrow \mathbf{r}_1, b \rightarrow \mathbf{r}_2\rangle$  in which a position measurement yields particle  $a$  at the position  $\mathbf{r}_1$  and particle  $b$  at  $\mathbf{r}_2$ . We can not distinguish between the particles; the state  $|a, b\rangle$  is, therefore, identical with the state  $|b, a\rangle = |b \rightarrow \mathbf{r}_1, a \rightarrow \mathbf{r}_2\rangle$  where particle  $b$  is detected at  $\mathbf{r}_1$  and particle  $a$  at  $\mathbf{r}_2$ . Both states  $|a, b\rangle$  and  $|b, a\rangle$  solve the Schrödinger equation equally well. Since the Schrödinger equation is a linear differential equation, the most general solutions are the linearly independent superpositions normalized by the factor  $1/\sqrt{2}$ :

$$|\psi_S\rangle = \frac{1}{\sqrt{2}}(|a, b\rangle + |b, a\rangle), \quad (5.140a)$$

$$|\psi_A\rangle = \frac{1}{\sqrt{2}}(|a, b\rangle - |b, a\rangle). \quad (5.140b)$$

A two particle quantum system, thus, has in general two different types of states, a symmetric one  $|\psi_S\rangle$ , which keeps its sign upon exchange of the two particles  $a$  and  $b$ , and an antisymmetric one  $|\psi_A\rangle$ , which changes sign upon particle exchange.

Which type of state is realized in nature, or do they both exist? In the following section, we will relate the property *symmetry* or *antisymmetry* of a quantum state directly with the spin of the particles.

It must be emphasized that the properties derived for two particles are equally found for many particles. For a system of three non-interacting particles, for example, the three-particle wave function (without spin) is written as

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, t) = \psi_1(\mathbf{r}_1, t)\psi_2(\mathbf{r}_2, t)\psi_3(\mathbf{r}_3, t). \quad (5.141)$$

For this product representation of the wave function it is necessary that the total potential can be separated into three single particle potentials:

$$V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = V_1(\mathbf{r}_1)V_2(\mathbf{r}_2)V_3(\mathbf{r}_3). \quad (5.142)$$

We can distinguish between symmetric and antisymmetric wave functions also for many-particle wave functions. It is relevant whether the wave function or the many-body state changes its sign upon exchange of any two particles. An  $N$ -particle state shall be described by  $|\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N\rangle$ , that is, a position measurement yields particle (1), (2), ..., ( $N$ ) at the positions  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N$ . The symmetric  $|S\rangle$  and antisymmetric state  $|A\rangle$ , then, are defined by:

$$|S\rangle = |\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N\rangle_S = |\mathbf{r}_2, \mathbf{r}_1, \mathbf{r}_3, \dots, \mathbf{r}_N\rangle_S, \quad (5.143a)$$

$$|A\rangle = |\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N\rangle_A = -|\mathbf{r}_2, \mathbf{r}_1, \mathbf{r}_3, \dots, \mathbf{r}_N\rangle_A. \quad (5.143b)$$

In (5.143a), (5.143b), each time particle (1) and particle (2) were exchanged, that is, on the left side of the equations particles (1) and (2) were detected at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , respectively, while on the right side particle (1) was detected at  $\mathbf{r}_2$  and particle (2) at  $\mathbf{r}_1$ . In the nomenclature of (5.143a), (5.143b) the order in the series of position coordinates indicates the particle number, while the position coordinates  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$  itself describe the detection site. Many-particle states as in (5.143a), (5.143b) are normalized to one in analogy to (5.131a) in order to guarantee the probability interpretation of the wave function, that is:

$$\begin{aligned} 1 &= \langle S | S \rangle_S = \langle \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N \rangle_S \\ &= \int \psi_S^*(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)\psi_S(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{-N}) d^3r_1 d^3r_2 \dots d^3r_N. \end{aligned} \quad (5.144)$$

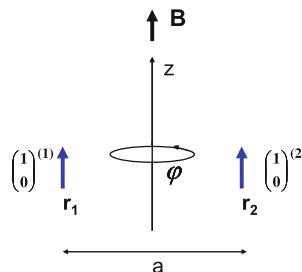
### 5.6.2 Spin and Particle Categories: The Pauli Exclusion Principle

We will see that there is a direct correlation between the spin of a particle and the symmetry property, symmetric or antisymmetric, of the corresponding two (many) particle wave function. For the derivation we use a gedanken experiment. Two electrons are assumed to be localized within a distance  $a$  at the positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , e.g. by confinement in two quantum dots. Due to an external magnetic field  $\mathbf{B}$  in  $z$  direction (normal to  $a$ ) both electrons have the same spin orientation  $\parallel \mathbf{B}$  (Fig. 5.11). A wave function  $\psi(\mathbf{r}_1, \mathbf{r}_2)$  describes the spatial part of the two-particle state. We take into account overlap of the single electron wave functions, which forbids a product representation of the two particle wave function. According to (Sect. 5.5.2) the two-electron state  $|\phi(1, 2)\rangle$  including the spin degree of freedom is, then, written in position representation as:

$$\langle \mathbf{r} | \phi \rangle = \phi(1, 2) = \psi(\mathbf{r}_1, \mathbf{r}_2) |\uparrow\rangle^{(1)} |\uparrow\rangle^{(2)} = \psi(\mathbf{r}_1, \mathbf{r}_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(1)} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(2)}. \quad (5.145)$$

Spin interaction is excluded, that is, the spin states of the two electrons can be written as a product, once in abstract bra-ket notation  $|\uparrow\rangle^{(1)}, |\uparrow\rangle^{(2)}$  and also in 2D Hilbert vector representation (5.109). Note the equal spin orientation due to the magnetic field. Both electrons are identical, also concerning their spin orientation. They can not be distinguished. We, thus, can exchange them without modification of the two-particle quantum state.

Formally this exchange of the two particles can also be performed by a rotation of the two-electron wave function by  $180^\circ = \pi$  around an axis ( $z$  direction) midst between the particles, normal to  $a$  (Fig. 5.11). The wave function originating from the rotation of  $\phi(1, 2)$  (5.145) by the angle  $\pi$ , which is identical with the one after exchanging the two electrons, is denoted by  $\phi^{(\pi)}(1, 2)$ . We will calculate this rotated wave function by using the rotation operator (rotation around the  $z$  axis) as it was



**Fig. 5.11** Schematic representation of a quantum state of two particles (1) and (2) being at the positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . The particles carry spins (see 2D spin vectors), which are oriented equally in  $z$  direction by the constant magnetic field  $\mathbf{B}$  (spin precession not shown). A realization of this two-particle state might be achieved in a double quantum dot. A physically identical state as the one depicted might be prepared by exchange of the two particles or by rotation of the quantum state around the  $z$  axis by an angle  $\varphi = 180^\circ$

introduced in Sect. 5.3 (5.40):

$$\hat{U}_\pi = \exp\left(\frac{i}{\hbar}\pi\hat{J}_z\right) = \exp\left[\frac{i}{\hbar}\pi(\hat{L}_z + \hat{S}_z)\right], \quad (5.146)$$

$\hat{J}_z$  is the total angular momentum operator in  $z$  direction containing the orbital angular momentum  $\hat{L}_z$  and the spin angular momentum  $\hat{S}_z$ . For the two-electron system  $\hat{S}_z$  is expressed by the sum of the two spin matrices  $\sigma_z^{(1)}$  and  $\sigma_z^{(2)}$  which act on electron (1) and electron (2) separately. According to (5.39b), the rotated two-particle wave function is, then, obtained as:

$$\begin{aligned} \phi^{(\pi)}(1, 2) &= \exp\left(\frac{i}{\hbar}\left[\pi\hat{L}_z + \frac{\pi}{2}\hbar\hat{\sigma}_z^{(1)} + \frac{\pi}{2}\hbar\hat{\sigma}_z^{(2)}\right]\right)\psi(\mathbf{r}_1, \mathbf{r}_2)\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(1)}\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(2)} \\ &= \exp\left(\frac{i}{\hbar}\pi\hat{L}_z\right)\psi(\mathbf{r}_1, \mathbf{r}_2)\exp\left(i\frac{\pi}{2}\hat{\sigma}_z^{(1)}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(1)} \\ &\quad \times \exp\left(i\frac{\pi}{2}\hat{\sigma}_z^{(2)}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(2)}. \end{aligned} \quad (5.147)$$

This representation implies that  $\hat{L}_z$  acts on the spatial part of the quantum state, that is, the wave function  $\psi(\mathbf{r}_1, \mathbf{r}_2)$ , while  $\sigma_z^{(1)}$  and  $\sigma_z^{(2)}$  act on the spin parts of the states of electron (1) and electron (2). The  $\pi$  rotation of  $\psi(\mathbf{r}_1, \mathbf{r}_2)$  performed by the action of  $\hat{L}_z$  is identical with an exchange of the two electrons, that is, with  $\psi(\mathbf{r}_2, \mathbf{r}_1)$ .

We, now, have to derive how the application of the spin operator  $\exp(i\pi\hat{\sigma}_z^{(1)}/2)$  modifies the 2D spin Hilbert vector of electron (1) and the same for electron (2). We remember that operator functions (4.97) are represented by series expansions (Sect. 4.3.5), that is,

$$\exp\left(\frac{i}{2}\varphi\hat{\sigma}_z\right) = \cos\frac{\varphi}{2}\hat{\sigma}_z + i\sin\frac{\varphi}{2}\hat{\sigma}_z, \quad (5.148a)$$

and with the expansions of sinus and cosinus functions

$$\cos\frac{\varphi}{2}\hat{\sigma}_z = \sum_{v=0}^{\infty} \frac{(-1)^v}{(2v)!} \frac{\varphi^{2v}}{2} \hat{\sigma}_z^{2v}, \quad (5.148b)$$

$$\sin\frac{\varphi}{2}\hat{\sigma}_z = \sum_{v=0}^{\infty} \frac{(-1)^v}{(2v+1)!} \frac{\varphi^{2v+1}}{2} \hat{\sigma}_z^{2v+1}. \quad (5.148c)$$

Exponential and trigonometric functions of operators are expressed by multiple application of these operators itself on the state or the wave function.

From the spin matrix relations

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{\sigma}_z^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \underline{1}, \quad \hat{\sigma}_z^3 = \hat{\sigma}_z, \quad (5.149)$$

we conclude:

$$\cos \frac{\varphi}{2} \hat{\sigma}_z = \sum_{v=0}^{\infty} \frac{(-1)^v}{(2v)!} \frac{\varphi^{2v}}{2} \stackrel{!}{=} \left( \cos \frac{\varphi}{2} \right) \stackrel{!}{=} \quad (5.150a)$$

and analogously

$$\sin \frac{\varphi}{2} \hat{\sigma}_z = \left( \sin \frac{\varphi}{2} \right) \hat{\sigma}_z. \quad (5.150b)$$

By means of (5.148a), we obtain  $\exp(i\pi\hat{\sigma}_z/2) = i\hat{\sigma}_z$  and finally the  $\pi$ -rotated wave function (5.147) as

$$\begin{aligned} \phi^\pi(1, 2) &= \psi(\mathbf{r}_2, \mathbf{r}_1)(i) \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(1)} (i) \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(2)} \\ &= -\psi(\mathbf{r}_2, \mathbf{r}_1) \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(2)} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(1)} = -\phi(2, 1). \end{aligned} \quad (5.151)$$

Since both electrons are identical, also concerning their spin orientation, we can not distinguish the rotated quantum state (wave function) from the original one, that is, we obtain:

$$\phi_A^{(\pi)}(1, 2) = \phi_A(1, 2) = -\phi_A(2, 1). \quad (5.152)$$

This is a very important result: Particles with half integer spin  $\hbar/2$ , as for example, electrons, have an antisymmetric two-particle wave function and state  $|1, 2\rangle$ , respectively. The quantum state changes its sign upon exchange of the two particles. Because of this antisymmetry, the wave function (5.152) is denoted by the subscript  $A$ .

This interconnection between antisymmetry of the wave function and spin is also valid for particles with spin  $3\hbar/2$ ;  $5\hbar/2$ ;  $7\hbar/2$ ; etc. as is easily shown by a calculation analogously to (5.147)–(5.152).

If one assumes, in the above calculation, integer spin for the two particles, as  $S_z = \pm\hbar$  or multiples of  $\hbar$  including  $S_z = 0$ , the rotation operators in the spin Hilbert space are written as  $\exp(i\pi\hat{\sigma}_z)$  and because of  $\exp(i\pi\hat{\sigma}_z) = -\hat{\sigma}_z$  we obtain a symmetric wave function (subscript  $S$ ):

$$\phi_S^{(\pi)}(1, 2) = \phi_S(1, 2) = \phi_S(2, 1). \quad (5.153)$$

The wave function keeps its sign upon exchange of the two particles.

Depending on whether a particle has a half integer or an integer spin the corresponding two-particle wave functions are *antisymmetric* respectively, *symmetric* upon exchange of the two particles. What we have learnt, here, for two-particle wave functions is generalized to many-particle systems of any kind in relativistic quantum field theories (Chap. 8).

Our world, thus, splits up into two separate categories of particles (electrons, protons, neutrons, photons etc.): Particles with half integer spin  $\hbar/2$ ;  $3\hbar/2$ ;  $5\hbar/2$ ; ... are described by antisymmetric wave functions; they are called *fermions*. Particles

with integer spin 0;  $\hbar$ ;  $2\hbar$ ;  $3\hbar$ ; ... belong to symmetric wave functions, which keep their sign upon exchange of any two particles. These particles are called *bosons*.

A short explanation of the whole variety of possible spin values in nature might be helpful. What is the origin of all these spins  $S = 0; \hbar/2; \hbar; 3\hbar/2; \dots$ ?

The electron has only one absolute spin value  $S = \hbar/2$ . The same is true for protons, neutrons and also for quarks, the constituting particles of proton and neutron (Sect. 5.6.4). Similarly as proton and neutron are built up by three quarks and atoms by protons, neutrons (nucleus) and shell electrons, many atomic and subatomic particles are composed of other particles. The spin of a compound particle, then, is the sum of the spins of the components (no orbital momenta). The proton as a fermion has the spin  $\hbar/2$  since the three constituting quarks (also fermions) have the spin components  $+\hbar/2, +\hbar/2, -\hbar/2$ . In a superconductor current, transport involves so-called Cooper pairs (Sect. 8.4.4), a couple of two paired electrons with opposite spin. These Cooper pairs, thus, have zero spin and bosonic character.

Beside the consideration of the total spin of a compound particle its category—bosonic or fermionic—can also be determined by performing an exchange of two particles in the corresponding two (many) particle wave function. As an example we choose the hydrogen (H) atom, which is composed of a proton  $p$  and an electron  $e$ , both fermions. Now, we consider a system of two H-atoms. Its two-atom (four-particle) wave function is written as  $\psi(p_1, e_1, p_2, e_2)$ . For an exchange of the two H-atoms, we exchange the two protons  $p_1, p_2$  and subsequently the electrons  $e_1$  and  $e_2$ . Because of the fermionic character of the particles this requires:

$$\psi(p_1, e_1; p_2, e_1) = -\psi(p_2, e_1; p_1, e_2) = \psi(p_2, e_2; p_1, e_1). \quad (5.154)$$

In the final expression of the wave function, after exchange of the two H-atoms, the sign is conserved, that is, the H-atom is a boson.

In general, when a particle is composed of an odd number of fermions, this particle itself is a fermion. An even number of fermions, as in the case of the H-atom, set up a bosonic compound particle. Accordingly the  ${}^3\text{He}$  atom with its two protons and one neutron in the nucleus and with its two electrons in the shell is a fermion, while the  ${}^4\text{He}$  atom with two protons and two neutrons in the nucleus (additionally two shell electrons) is a boson.

For non-interacting fermions, the antisymmetry of the many-body wave function can be expressed in a simple way which is called after its inventor (Wolfgang Pauli 1900–1958) the *Pauli exclusion principle*. For its derivation, we consider two electrons (or fermions in general) in a potential box. These two particles, each separately, can assume the energy eigenvalues (3.64) with a discrete series of eigenfunctions  $\psi_n(\mathbf{r}, s)$  (3.62), where  $n$  is the series of integer numbers and  $s$  the spin quantum number. The spin quantum number  $s$  (or  $m_s = \pm 1/2$ ) can also be taken into account in the wave function as  $\psi_{n,s}(\mathbf{r})$ . In the two-particle case, where one particle found at the position  $\mathbf{r}_1$  with spin  $s_1$  has the wave function  $\psi_i(\mathbf{r}_1, s_1)$  and the other one at  $\mathbf{r}_2$  with spin  $s_2$  the wave function  $\psi_j(\mathbf{r}_2, s_2)$ , the anti-symmetric wave function of the non-distinguishable particles must be written as:

$$\begin{aligned}\psi_A(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2) &= \frac{1}{\sqrt{2}} [\psi_i(\mathbf{r}_1, s_1)\psi_j(\mathbf{r}_2, s_2) - \psi_j(\mathbf{r}_1, s_1)\psi_i(\mathbf{r}_2, s_2)] \\ &= \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_i(\mathbf{r}_1, s_1) & \psi_j(\mathbf{r}_1, s_1) \\ \psi_i(\mathbf{r}_2, s_2) & \psi_j(\mathbf{r}_2, s_2) \end{vmatrix}. \quad (5.155)\end{aligned}$$

Since the particles do not interact with each other, the two-particle wave function can be represented as a product of single particle wave functions. Because of the requirement of anti-symmetry a superposition (with inverted sign) of the wave function products with exchanged particles leads to a representation of the two-particle wave function in terms of a determinant.

From the representation (5.155), it is evident that the two-particle wave function  $\psi_A(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2)$  vanishes for  $i = j$  or  $\mathbf{r}_1 = \mathbf{r}_2, s_1 = s_2$ . A determinant with two equal rows or columns is zero. Fermions at one and the same position can not have identical quantum numbers, spin included. In other words: non-interacting fermions can occupy a single particle quantum state only one times; no quantum state can be occupied by more than one fermionic particle. For three non-interacting fermions with single particle wave functions  $\psi_i(\mathbf{r}_1, s_1), \psi_j(\mathbf{r}_2, s_2), \psi_k(\mathbf{r}_3, s_3)$  the normalized antisymmetric three-particle wave function is written as

$$\psi_{ijk}(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \mathbf{r}_3, s_3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \psi_i(\mathbf{r}_1, s_1) & \psi_j(\mathbf{r}_1, s_1) & \psi_k(\mathbf{r}_1, s_1) \\ \psi_i(\mathbf{r}_2, s_2) & \psi_j(\mathbf{r}_2, s_2) & \psi_k(\mathbf{r}_2, s_2) \\ \psi_i(\mathbf{r}_3, s_3) & \psi_j(\mathbf{r}_3, s_3) & \psi_k(\mathbf{r}_3, s_3) \end{vmatrix}. \quad (5.156)$$

According to the rules for calculating a determinant (5.156) vanishes for  $i = j$ ,  $i = k$ , or  $j = k$  as well as for equal position coordinates and spin quantum numbers. The representation (5.156) also guarantees the antisymmetry of the wave function, since exchange of two rows, that is,  $\mathbf{r}_1, s_1 \leftrightarrow \mathbf{r}_2, s_2$ , changes its sign.

The generalization for  $N$  non-interacting fermions is obvious. The corresponding wave function is represented by an  $N$ -dimensional determinant:

$$\begin{aligned}&\psi_{n_1 n_2 \dots n_N}(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \mathbf{r}_3, s_3, \dots, \mathbf{r}_N, s_N) \\ &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{n_1}(\mathbf{r}_1, s_1) & \psi_{n_2}(\mathbf{r}_1, s_1) & \dots & \psi_{n_N}(\mathbf{r}_1, s_1) \\ \psi_{n_1}(\mathbf{r}_2, s_2) & \psi_{n_2}(\mathbf{r}_2, s_2) & \dots & \psi_{n_N}(\mathbf{r}_2, s_2) \\ \vdots & \vdots & & \vdots \\ \psi_{n_1}(\mathbf{r}_3, s_3) & \psi_{n_2}(\mathbf{r}_3, s_3) & \dots & \psi_{n_N}(\mathbf{r}_3, s_3) \end{vmatrix}. \quad (5.157)\end{aligned}$$

This representation is called *Slater determinant*, according to its inventor. For non-interacting fermions the antisymmetry requirement upon exchange of two particles, thus, leads to the Pauli exclusion principle which is formally expressed by the Slater determinant (5.157). According to this principle it is forbidden for two fermions to occupy one and the same single particle quantum state, including spin, at the same position.

It should be emphasized again at this point that the Slater determinant representation can only be applied to a problem of non-interacting fermions. In case of particle interaction, a many-particle wave function is defined, the quantum numbers of which

are ascribed to the total system as a whole rather than to single particle states. Here, only the weaker antisymmetry requirement for the many-body wave function is valid.

### 5.6.3 Two Different Worlds: Fermi and Bose Statistics

Our world consists of large complex many-body systems, which in principle have to be described by coherent many-body wave functions. This is mostly a mathematically un-treatable problem. For the formal description it is, therefore, of eminent importance that in many cases the interaction between particles is weak and spatially as well as timely restricted. In that case we can approximately ascribe single particle states, respectively wave functions with corresponding quantum numbers ( $n_x, n_y, n_z, \mathbf{k}, s, \dots$ ) to the particles, electrons, nuclei, atoms etc. This is even true for particles which are scattered with each other. Usually these particles approach each other from long distances (on the atomic scale) before they hit. Between two scattering processes the particles are essentially free, described by spatially restricted wave packets. Their wave functions overlap only during a short time in a limited spatial region. During this comparatively short scattering time, the two particle wave function can not be factorized into two single particle wave functions. In between the scattering events, however, the description in terms of non-interacting particles is appropriate. We can use well-defined single particle states with quantum numbers  $\mathbf{k}, \mathbf{k}', \dots, s, s', \dots$ .

Now the question arises which single particle states are occupied in an ensemble of fermions or bosons. According to Pauli's exclusion principle, non-interacting fermions can occupy a single particle quantum state only one times. If a state already occupied, a further particle has to find another quantum state. Bosons, on the other hand, can occupy a single particle quantum state in any number. For non-interacting bosons, the most general symmetric many-body wave function is the sum of products of single particle wave functions (with equal sign), in which each time two particles have been exchanged. This sum (superposition) of products results from the fact that quantum particles can not be distinguished except by their quantum numbers. Because of equal signs of the products the general many-particle wave function is symmetric upon exchange of two bosons, as required (Sect. 5.6.2).

We now assume a system consisting of  $N$  bosonic states and we ask the following question: If  $n$  single quantum states are already occupied by bosons, in how many ways can a further boson be added to the system, such that the resulting many-particle wave function is symmetric again. Apart from an occupation of the  $N$  existing states the new particle can be exchanged with the  $n$  particles already present. Therefore,  $(N + n)$  possibilities exist for the new additional boson to be combined with the already present ones to a symmetric many-particle wave function. The possibility to add bosons to a quantum system increases with the number  $n$  of bosons already contained in the system. Bosons obey a "herd instinct"; they collect there where already many of them are found.

In contrast, fermions behave differently: If  $n$  of the given  $N$  single particle states are already occupied by fermions, there are only  $(N - n)$  states which can be occupied by an additional fermion.

On the basis of these considerations we are able to derive an occupation statistics for fermions and bosons. Explicitly said, we ask the question, what the probability is for weakly interacting fermions and bosons to occupy single particle states in a many-body system. The description in terms of single particle states requires long free pathways of the particles and small spatially restricted scattering volumes, that is, our approximation is valid for weak particle interaction and low particle densities.

For the derivation of the particle statistics, we consider a large ensemble of particles, for example, a macroscopic or mesoscopic system (gas) of quantum particles, in which thermal equilibrium is established by collisions. Within the large ensemble we assume two sub-ensembles (1) and (2), also macroscopically large. Dynamical thermal equilibrium, then, implies that the sub-ensembles (1) and (2) undergo energetic fluctuations where the total energies  $E_1$  and  $E_2$  of the two sub-ensembles are permanently equalized by particle collisions back and forth. Since we are concerned with macroscopic systems, both the total system and the sub-ensembles obey the laws of classical physics (correspondence principle, Sects. 3.3 and 3.4). According to classical thermodynamics the probabilities  $w$ , that is, the number of realizations  $v(E_1)$  and  $v(E_2)$  of two systems with energies  $E_1$  and  $E_2$  are related by the Boltzmann factor

$$\frac{w(E_1)}{w(E_2)} = \frac{v(E_1)}{v(E_2)} = \exp\left(-\frac{E_1 - E_2}{k_B T}\right), \quad (5.158a)$$

where  $k_B$  is Boltzmann's constant and  $T$  the temperature of the large total ensemble of which the sub-ensembles (1) and (2) are part of.

Equation (5.158a) describes the dynamical equilibrium of the total system in which the sub-systems are represented with the frequencies  $v(E_1)$  and  $v(E_2)$  or probabilities  $w$ . By dividing the frequencies  $v$  through an equal time interval, we obtain the Boltzmann relation (5.158a) also for the rates  $r$  by which the two sub-systems fluctuate into each other:

$$\frac{r_{1 \rightarrow 2}}{r_{2 \rightarrow 1}} = \exp\left(-\frac{E_1 - E_2}{k_B T}\right). \quad (5.158b)$$

The energy exchange between the two sub-systems can even be due to the transfer of only one particle. The rate equation (5.158b) is, therefore, valid also for single particles. In analogy to the sub-ensemble energies  $E_1$  and  $E_2$ , we denote two single particle energies by  $E_i$  and  $E_j$ . If for the energies  $E_i$  ( $E_j$ ) each time  $N_i$  ( $N_j$ ) quantum states are available and  $n_i$  ( $n_j$ ) states are already occupied, then the number of particles changing their quantum state  $i$  into state  $j$  within the time interval  $\Delta t$  amounts to

$$v_{i \rightarrow j} = r_{i \rightarrow j} n_i (N_j \pm n_j) \Delta t. \quad (5.159a)$$

The particle character, boson or fermion, is essential for the number ( $N_j \pm n_j$ ) of possible final states in (5.159a). Bosons follow a “herd instinct”; here the plus sign is valid. Fermions can only occupy empty states; consequently the minus sign is required. The number of particles, which change their quantum state from  $j$  to  $i$ , is written in analogy as

$$v_{i \rightarrow j} = r_{j \rightarrow i} n_j (N_i \pm n_i) = r_{i \rightarrow j} \left[ \exp\left(-\frac{E_i - E_j}{k_B T}\right) \right] n_j (N_i \pm n_i). \quad (5.159b)$$

Hereby the rate ratio  $r_{j \rightarrow i}/r_{i \rightarrow j}$  was expressed according to (5.158b). In thermal equilibrium the transition rates for collisions from  $i$  to  $j$  must equal those from  $j$  to  $i$ . From (5.59) we, therefore, obtain

$$r_{i \rightarrow j} n_i (N_j \pm n_j) = r_{i \rightarrow j} \left[ \exp\left(-\frac{E_i - E_j}{k_B T}\right) \right] n_j (N_i \pm n_i). \quad (5.160)$$

This equilibrium condition yields for any state  $i$  or  $j$ :

$$\frac{n_i}{N_i \pm n_i} e^{E_i/k_B T} = \frac{n_j}{N_j \pm n_j} e^{E_j/k_B T}. \quad (5.161)$$

Thus, the quantity

$$K = \frac{n}{N \pm n} \exp(E/k_B T). \quad (5.162)$$

Is a constant at a fixed temperature  $T$ . We, therefore, have omitted the indices  $i$ , respectively  $j$  and have denoted the single particle energies by  $E$ . For each energy  $E$  the number of existing quantum states amounts to  $N$ , of which  $n$  states are occupied. The relative occupation number  $n/N$  of the single particle energy level  $E$  is then obtained from (5.162) as

$$w = \frac{n}{N} = \left( \frac{1}{K} e^{E/k_B T} \mp 1 \right)^{-1}. \quad (5.163)$$

The minus sign is obtained for bosons and the plus sign for fermions. For fermions it is comfortable to write the temperature dependent constant  $K$  as

$$K = \exp(E_F/k_B T). \quad (5.164)$$

The quantity  $E_F$  is called *Fermi energy*. With this definition the occupation probability  $w$  of fermionic single particle states is obtained as

$$w = f(E) = \frac{n}{N} = \frac{1}{\exp(\frac{E-E_F}{k_B T}) + 1}. \quad (5.165)$$

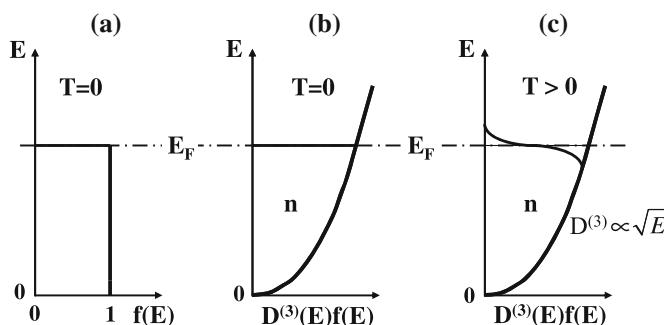
This so-called *Fermi distribution*  $f(E)$  is easily interpreted on the basis that only one single fermion can occupy a quantum state.

As an example consider the problem of free electrons confined in a 3D potential box (Sect. 3.6.1), as it is found, for example, for conduction electrons in a metal. Electrons occupy wave quantum states with discrete wave numbers  $\mathbf{k}$  (3.68) in reciprocal space as quantum numbers (Fig. 3.5). The density of these states  $D(E)$ , that is, the number of states per volume and energy interval  $\Delta E$  at an energy  $E$  (with differing  $\mathbf{k}$  and spin) is a square root function  $D^{(3)}(E) \propto \sqrt{E}$  (Fig. 3.7). How these states are occupied by electrons is controlled by the Fermi distribution  $f(E)$  (5.165). Thus, the density of occupied states at an energy  $E$  is

$$n(E) = D^{(3)}(E)f(E). \quad (5.166)$$

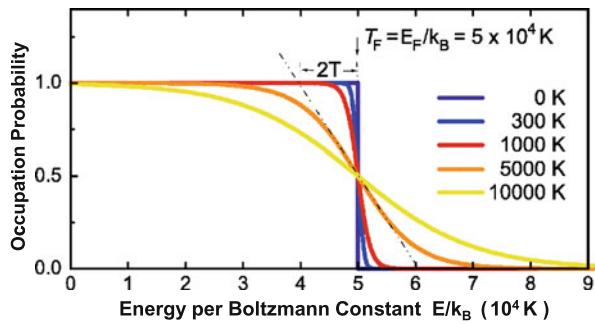
At zero temperature  $T = 0$  the electrons occupy the energetically lowest possible energy states. Because of the Pauli principle they successively fill up all quantum states from  $E = 0$  up to a maximum energy  $E_F$  (Fig. 5.12a).  $E_F$  is just the energy defined above as Fermi energy. It is evident that for  $T = 0$  the Fermi distribution  $f(E)$  is a step function as in Fig. 5.12a. Below  $E_F$  the occupation probability of a state is one while above  $E_F$  it is zero. This is exactly the functional dependence of (5.165) at  $T = 0$ .

It is obvious that the Fermi distribution is a step function at  $T = 0$  as shown in Fig. 5.12a. Below the step at  $E_F$  all states are occupied, that is, the occupation probability is one (certainty) while above  $E_F$  the states are empty and zero occupation probability is given. The square root density of states  $D^{(3)}(E)$  is occupied up to the sharp Fermi energy  $E_F$  (Fig. 5.12b). When the temperature is raised a little bit, electrons from states below  $E_F$  are excited into states closely above  $E_F$ , which were empty before. The density of occupied states given by  $n = D^{(3)}(E)f(E)$  is



**Fig. 5.12 a–c** Occupation of electronic states of a free electron gas in a 3D potential box, whose density of states  $D^{(3)}$  is proportional to  $\sqrt{E}$ . For electrons the occupation is determined by the Fermi statistics  $f(E)$ . **a** Fermi distribution  $f(E)$  at  $T = 0$  K. **b** Electron density  $n$  obtained as product of Fermi function  $f(E)$  and state density  $D^{(3)}$  at  $T = 0$  K. **c** Electron density  $n$  at a finite temperature  $T > 0$  K

**Fig. 5.13** Occupation probability according to Fermi statistics for different temperatures  $T$  as function of electron energy  $E$  (normalized by Boltzmann constant  $k_B$ ).  $T_F = E_F/k_B = 5 \times 10^4 \text{ K}$



plotted in Fig. 5.12c. It reflects the smoothed Fermi function for somewhat elevated temperature. This behavior of the Fermi function (5.165) is quantitatively shown in Fig. 5.13. From this plot, we derive that the temperature dependent energy range, where the Fermi edge is smoothed, amounts to about  $4k_B T$ . Mathematically the properties of the Fermi function (5.165) are easily seen from the limit  $|E - E_F| \gg k_B T$ . For  $E < E_F$  the exponential function in the denominator can be neglected which results in  $n \approx N$  or  $f(E) \approx 1$ . For  $E > E_F$  the exponential function is the dominant part in the denominator and  $n$  becomes negligibly small.

A more general derivation of the Fermi function in the context of statistical thermodynamics shows that the Fermi energy  $E_F$  is nothing else but the chemical potential  $\mu$  of the electrons.

From (5.163), the occupation probability  $g_B$  for bosonic quantum states follows as

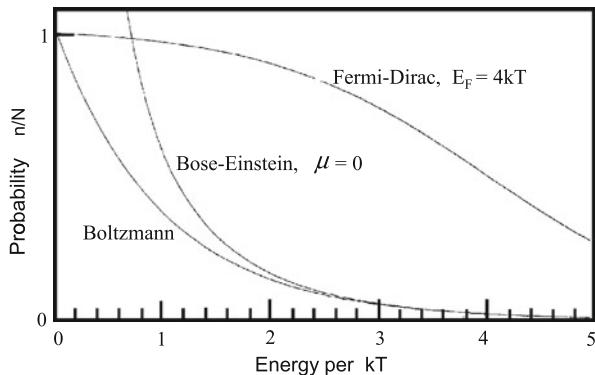
$$w = g_B = \frac{n}{N} = \frac{1}{\exp(\frac{E-\mu}{k_B T}) - 1}. \quad (5.167)$$

In this case,  $\mu$  is the chemical potential of the so-called Bose gas (gas of Bose particles). Bose statistics (sometimes also called Bose–Einstein statistics) is compared with classical Boltzmann and Fermi statistics in Fig. 5.14. It is evident that for elevated energies  $E/kT$  ( $E \gg E_F$  in Fermi statistics) with low occupation of the quantum states also the spatial density of particles is low. In this case, many unoccupied states are available and the different behavior of fermions and bosons does not play an important role. Both Fermi and Bose statistics approach the classical Boltzmann distribution.

An important application of Bose statistics is concerned with the radiation density of the electromagnetic field (Planck's formula). The electromagnetic field is built up by photons (Chap. 8), particles with spin  $\pm\hbar$ , that is, bosons. As we will see, the two spin orientations correspond to clock and counter-clockwise circular polarization with respect to the light wave propagation direction (wave vector). Photons, thus, obey Bose statistics.

Bose statistics (5.167) is the key to the understanding the classical problem of *black-body* radiation. In solving this problem, Planck has opened the door for the

**Fig. 5.14** Comparison of the classical Boltzmann distribution function with the two quantum statistics, the Fermi–Dirac distribution for Fermions and the Bose–Einstein distribution for bosons.  $\mu$  and  $E_F$  are the chemical potentials in both cases



development of modern physics, in particular quantum mechanics. The question is how the possible electromagnetic wave modes are distributed (as a function of temperature) in a cavity the walls of which are in thermodynamic equilibrium with the radiation field. If the cavity has a little hole, it will radiate electromagnetic waves according to this distribution, that is, with corresponding frequencies  $\omega$ . This is the so-called black-body radiation.

Similarly as electrons are confined in a potential box with a discrete spectrum of electron waves as quantum states (Sect. 3.6.1), electromagnetic waves in a metallic cavity form standing waves with similar boundary conditions as electron waves in the box. For periodic boundary conditions (as for electrons in Sect. 6.6.1), a vibrational mode needs a volume  $V_q = (2\pi/L)^3$  in the  $q$  space of the light (photon) wave vectors, with  $L$  as the edge length of the cubic cavity. In analogy to electronic states, we calculate the number of possible photon states in the reciprocal space of wave numbers  $q$  by dividing the volume of an energy shell  $4\pi q^2 dq$  with thickness  $dq$  by the volume of a state  $(2\pi/L)^3$ . Relating to the volume  $L^3$  of the cavity one obtains the number of photon states or vibrational modes of the electromagnetic field per cavity volume:

$$Z(q) dq = 4\pi q^2 dq / (2\pi)^3 = \frac{2q^2}{(2\pi)^2} dq. \quad (5.168)$$

The density of photon states per cavity volume and per energy, respectively frequency interval ( $E = \hbar\omega$ ) is, then, calculated by means of

$$D(\hbar\omega) d\omega = Z(q) dq. \quad (5.169)$$

In contrast to electrons, where the energy-wave number relation is given by  $E = \hbar\omega = \hbar^2 k^2 / 2m$ , photons obey the classical dispersion relation  $\omega = cq$  with  $c$  as light velocity. From (5.168) to (5.169) we, then, obtain the density of photon states as

$$D_{\text{Ph}}(\omega) = \frac{\omega^2}{2\pi^2 c^3}. \quad (5.170)$$

The radiation density of the cavity is calculated from the density of states (5.170) by multiplication with the occupation probability  $g_B(E = \hbar\omega)$  for bosons (photons). We must take into account that each photon state described by a wave vector  $\mathbf{q}$  in reciprocal space contains two spin states represented by opposite circular polarization directions. The radiation density of the cavity is, then, obtained as

$$S(\hbar\omega) = 2g_B(\hbar\omega)D_{\text{Ph}}(\omega)\hbar\omega. \quad (5.171)$$

According to Sect. 2.1 (see also Chap. 8) the particle energy  $\hbar\omega$  has been attributed to the photon. When we refer the single particle energies  $E = \hbar\omega$  to an energy zero-point  $\mu = 0$ , the radiation density of a cavity in thermal equilibrium follows as

$$S(\hbar\omega) = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{\exp(\hbar\omega/k_B T) - 1}. \quad (5.172)$$

This is the famous Planck formula for the electromagnetic radiation density which is emitted by a black body in thermal equilibrium (Fig. 5.15).

As has been discussed before, bosons have a “herd instinct”; they are attracted by high numbers of bosons already present in that state. In a gas of Bose particles, a Bose gas, bosons can occupy one and the same many-particle state in high density at sufficiently low temperature. They condense in a common ground state and the many-particle system is called a Bose-Einstein condensate. In this *Bose-Einstein condensation* process temperature is an essential parameter which determines how close bosons approach each other to form the condensate. For an adequate estimation the average extension of a particle, i.e. the space needed at minimum, is important. A lower limit is certainly the de Broglie wavelength  $\lambda_{DB}$ , which is attributed to the particle at a certain temperature (mean kinetic energy). Below  $\lambda_{DB}$  the wave packet of the particle can not be defined anymore. Because of  $\lambda = 2\pi/k$  the particle wavelength  $\lambda_{DB}$  is calculated from the average wave vector  $\bar{k}$  of particles in a statistical ensemble at a particular temperature:

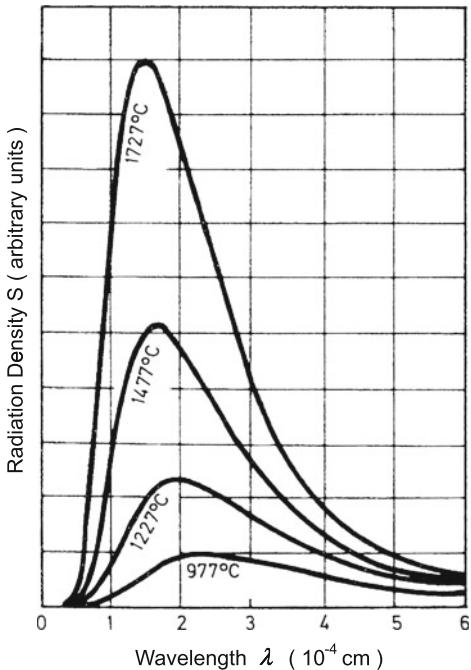
$$\bar{k} = \frac{1}{\hbar} \int_{-\infty}^{\infty} dp \exp(-\beta p^2/2m) = \frac{1}{\sqrt{\hbar}} \sqrt{\frac{2m\pi}{\beta}} = \frac{2\pi}{h} \sqrt{2m\pi k_B T}, \quad (5.173)$$

where  $\beta = 1/k_B T$  and  $p = \hbar k = mv$ , the particle momentum. From (5.173) the so-called *thermal de Broglie wavelength* of the particle is obtained as

$$\lambda_{DB}(T) = 2\pi/\bar{k} = \frac{h}{\sqrt{2m\pi k_B T}}. \quad (5.174)$$

In order to achieve Bose-Einstein condensation the average particle distance must shrink to values in the order of  $\lambda_{DB}$  given by (5.174). This minimum distance ensures enough overlap of the electronic cloud of the particles for the formation of the coherent many-particle condensate. In the approximation of non-interacting Bose

**Fig. 5.15** Planck radiation density  $S$  of a black body as function of the emitted radiation wave length  $\lambda$  for different temperatures



particles the density of the Bose gas can be assumed to be about  $n \approx \lambda_{DB}^{-3}$ . This allows the estimation of a critical temperature  $T_{cr}$  below which Bose-Einstein condensation is expected:

$$T_{cr} \approx 2\pi \frac{\hbar^2 n^{2/3}}{mk_B} . \quad (5.175)$$

Apart from a factor of two (5.175) coincides with the exact expression which is derived by a more rigorous theoretical treatment.

Bose-Einstein condensation was indeed experimentally found by two American groups at the University of Colorado, Boulder [9], and at the MIT, Boston [10]. In Colorado a gas of Rb atoms was cooled down to 170 nK, while at the MIT <sup>23</sup>Na atoms were observed to form the Bose-Einstein condensate. For this work the Nobel prize was awarded to Eric Cornell, Carl Wieman (Boulder) and Wolfgang Ketterle (MIT) in 2001. The new exciting result of this work was the verification of Bose-Einstein condensation with atoms. On the other hand, Bose-Einstein condensation had already been demonstrated by the observation of superconductivity, where Bose type Cooper electron pairs condense in a common many-particle ground state (see Sect. 8.4.4).

### 5.6.4 The Zoo of Elementary Particles

We have seen that in general two types of particles, fermions and bosons, are distinguished because of their inherent spin degree of freedom. Their different spin has paramount consequences for the statistical occupation of single particle states. In this context it is worth describing briefly our present knowledge about number and properties of existing elementary particles which obey the laws of quantum mechanics.

At the beginning of the 20th century, when quantum physics emerged, only the electron and somewhat later proton and neutron as constituents of the atomic nucleus were known as elementary particles. Since about 1970 a clearer picture of the variety of particles has been created in elementary particle physics. For the time being, this picture was completed in the so-called *standard model*.

Without going much into details, we will report some essential results of research in elementary particle physics which are explained in the standard model [11]. This shall provide only a rough impression, down to what level our understanding of the subatomic world reaches and to what smallest entities quantum mechanics is applied so far.

In this context, two fundamental issues are derived from special relativity theory: the equivalence of mass and energy and the existence of antiparticles to each particle (antimatter). Both results have been proven meanwhile by an overwhelming amount of experimental facts.

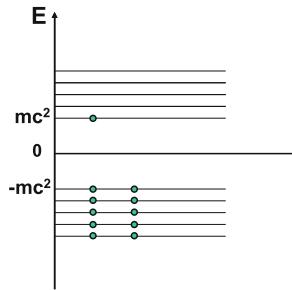
With respect to the equivalence of mass and energy, we remind the expression for the relativistic momentum  $p$  of a particle with mass  $m$  moving with a velocity  $v$ :

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}. \quad (5.176)$$

Note that for  $v^2/c^2 \ll 1$  ( $c$  light velocity) the classical momentum  $mv$  of the particle is obtained. Using (5.176) we calculate, without showing details, the relativistic work  $W$  performed by the force  $F$  to accelerate a mass  $m$  to the velocity  $v$  as

$$W = \int F dx = \int \dot{p} dx = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \approx mc^2 + \frac{mv^2}{2}. \quad (5.177)$$

The second term in (5.177) is the classical kinetic energy of the particle. The first term doubtlessly attributes an energy  $E = mc^2$ , the so-called *rest energy*, to the massive particle. Theory of special relativity tells us that mass and energy are equivalent physical quantities. Energy and mass can mutually be transformed into each other. How this happens in detail, by what mechanisms, does the famous Einstein equation  $E = mc^2$  not explain. Mechanisms of mass-energy transformation have been theoretically developed not before the advent of relativistic quantum field theories (Chap. 8). Nevertheless, we realize, elementary particles must not live forever. After a finite lifetime, they can transform into energy and energy again into another kind



**Fig. 5.16** Schematic representation of the Dirac sea of relativistic electrons with mass  $m$ . In the vacuum state, negative energy states with  $mc^2 < 0$  are occupied and states with  $mc^2 > 0$  are empty. A single electron in vacuum means occupation of one single state of positive energy as shown in the figure

of particle. A particle which transforms into another particle, of course must have a larger mass than the final product. Otherwise, the law of energy conservation would be violated.

Furthermore, it is a familiar habit in elementary particle physics to express the rest mass  $m$  of a particle in terms of energy via the relation  $E = mc^2$ . The GeV ( $=10^9$  eV) scale, hereby, has been proven to yield a suitable order of magnitude.

From (5.176) and (5.177), one derives the relativistic energy-momentum relation for a particle:

$$\begin{aligned} E^2 - c^2 p^2 &= \frac{1}{1 - v^2/c^2} (m^2 c^4 - m^2 c^2 v^2) = m^2 c^4, \\ E^2 &= c^2 p^2 + m^2 c^4. \end{aligned} \quad (5.178a)$$

This equation has solutions with positive and negative energies:

$$E = \pm \sqrt{c^2 p^2 + m^2 c^4}. \quad (5.178b)$$

Dirac, for the first time, has taken both types of solutions for serious and realistic. He did not reject the negative-energy solutions as non-physical but rather postulated the existence of an antiparticle (with negative energy) to the electron carrying the opposite charge of the electron. This at first glance counter-intuitive assumption was made by Dirac for all kinds of fermions. Similarly as in the electronic band scheme of a semiconductor (Sect. 8.3.3) the negative-energy particles are assumed to occupy quantum states with  $E = -mc^2$  in a so-called *Dirac sea* (Fig. 5.16). Under vacuum conditions, all states with positive energy  $E = mc^2$  are empty while all states with negative energy are occupied. The presence of one electron in vacuum is described by one occupied state at  $E = mc^2$ . We remember the semiconductor band scheme with an empty conduction band and a completely occupied valence band (Sect. 8.3.3).

The occupation of the Dirac sea at  $E < 0$  is controlled by Fermi statistics, that is, each single particle state can be occupied only once (spin degree of freedom included). Within this Dirac picture the vacuum state contains an infinitely high charge. This unfamiliar picture does not lead into contradictions, since the completely occupied states with negative energy can not be observed in experiment. Electrons at  $E > 0$  can not interact with those at  $E < 0$ , as no empty states are available in the full Dirac sea (Pauli principle, Sect. 5.6.2). If however a minimum energy of  $2mc^2$  (two times rest energy of an electron (5.178b) with  $p = 0$ ) is transferred to the vacuum, an electron, respectively another fermion of mass  $m$ , can be excited from  $E = -mc^2$  in the Dirac sea to  $E = mc^2$  above. A hole is produced in the Dirac sea at  $-mc^2$  and an electron created in vacuum with its rest energy  $mc^2$ . The hole state in the Dirac sea is called antiparticle of the electron. It has the same mass  $m$  as the electron but opposite (positive) charge. It is called *positron* and was detected experimentally in 1932 by C.D. Anderson in the cosmic radiation [12]. This was the brilliant proof of Dirac's theoretical assumption of antimatter. The existence of antimatter is meanwhile a well established fact, not only for electrons but rather for all kinds of fermions. We will not discuss at this point the generalization to antibosons.

Since electrons have a rest mass (energy) of about 0.5 MeV the described process of electron–positron (pair) production in vacuum can only occur at energy supplies exceeding  $2mc^2 \cong 1$  MeV, that is, at energies being totally irrelevant in solids. We can generally say that processes of particle conversion, annihilation or creation require such high energies, at minimum  $mc^2$ , that they are irrelevant in solid state physics (apart from unstable radio-active nuclei).

The study of elementary particle processes requires particle accelerators which accelerate particles far into the GeV energy range. In earlier days, before the advent of particle accelerators, only the investigation of cosmic radiation has provided information about the high-energy scale of particle physics. In this context, we will briefly describe what elementary particles are known so far, how important they are for our understanding of matter, and how they fit into the categories of bosons and fermions.

As it was already known at the beginning of the 20th century, it is true also on the basis of our present understanding, that matter under normal earth-like conditions as in crystals, liquids, gases, semiconductor nanostructures or in biological systems consists of electrons, protons and neutrons, altogether fermions. Protons and neutrons form the atomic nucleus, which is enclosed by a cloud (shell) of electrons. While the nucleus carries nearly the whole mass of an atom, the electronic shell determines the spatial extension of the atom (nucleus:  $10^{-13}$  cm, shell:  $10^{-8}$  cm). Only electrons and protons are assumed so far to have an infinite lifetime. The electron always appears in experiments as an elementary particle not composed of any sub-particles. In contrast, scattering experiments with high-energy photons ( $\gamma$  particles) reveal a sub-structure of the proton and the neutron. Both particles are composed of three sub-particles. Proton and neutron, called nucleons, are fermions; the sub-particles, therefore, must be fermions, too (Sect. 5.6.2). Their electric charge, added up over three sub-particles for each nucleon, must yield the positive elementary charge  $e_0$  of the proton or zero charge of the neutron.

Within several decades of elementary particle research, these constituents of the nucleons have been identified as really elementary and point-like as the electron. These particles are called *quarks*, a name which has been taken by their theoretical “inventors” Gell-Mann and Zweig from James Joyce’s novel “Finnegans Wake”.

After many years of research, matter has been shown to be built-up by two types of elementary, point-like particles, so-called *leptons* ( $\lambda\varepsilon\pi\tau\circ\varsigma$  = light) as the electron and *quarks* being the constituents of the nucleons. Both types of particles are fermions with half-integer spin. In Table 5.1, all leptons and quarks known so far as the basis of the standard model are listed. All particles, even though elementary, have strongly differing life times  $t$ . While the electron is stable ( $t \rightarrow \infty$ ), the other two leptons with charge  $-e_0$ , the  $\mu$  and  $\tau$ , are short living particles with life times  $2.2 \times 10^{-6}$  s ( $\mu$ ) and  $2.9 \times 10^{-13}$  s ( $\tau$ ), respectively. Both leptons and quarks exist as pairs, in Table 5.1 indicated by brackets: for example, the electron  $e$  together with its neutrino  $\nu_e$  (charge zero, spin 1/2, mass  $< 3$  eV i.e. much smaller than electron mass). There are three families or flavors of leptons and analogously of quarks, which differ by increasing rest masses (Table 5.1). The first family (flavor) of leptons ( $e, \nu_e$ ) and quarks ( $u$ , called up;  $d$ , called down) is encountered in matter on earth under normal conditions. In tri-fold combination, both quarks  $u$  and  $d$  build-up the nucleons: ( $udd$ ) the proton and ( $udd$ ) the neutron. From the quark charges,  $(2/3)e_0$  ascribed to  $u$  and  $-(1/3)e_0$  to  $d$ , the charges  $+e_0$  of the proton and zero of the neutron, are easily derived. Even though scattering experiments with high-energy  $\gamma$  photons doubtlessly demonstrated the threefold substructure of proton and neutron, quarks have never been observed as free single particles. Correspondingly the field theory of quarks, called chromodynamics (see below), is designed that the attracting force between quarks increases (limit infinity) with increasing quark distance. This is in contrast to the electromagnetic Coulomb force between electrons, protons or electrons and protons, which decreases and finally becomes negligibly small for large particle distances.

Already long before the quark model was taken serious as a realistic description of nature, the community of particle physicists argued that proton and neutron might

**Table 5.1** Three families (flavors) of elementary fermions with spin  $\hbar/2$ : the leptons electron ( $e$ ), myon ( $\mu$ ),  $\tau$  particle with their neutrinos  $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ , and the quarks up ( $u$ ), down ( $d$ ); charm ( $c$ ), strange ( $s$ ); top ( $t$ ), bottom ( $b$ ). The masses are given in units of GeV ( $m = E/c^2$ ) in square brackets. Note that the charge of the particles in brackets differs by  $\Delta Q = e_0$  each time. To each particle an anti-particle with equal mass and opposite electrical charge exists ( $\bar{u}$  to  $u$ ,  $\bar{c}$  to  $c$ ,  $\bar{\nu}_e$  to  $\nu_e$  etc.)

	3 Families (flavor)	Charge ( $e_0$ )
Leptons (fermions)	$\left( \begin{array}{c} \nu_e \\ e[0.5 \times 10^{-3}] \end{array} \right) \left( \begin{array}{c} \nu_\mu \\ \mu[0.106] \end{array} \right) \left( \begin{array}{c} \nu_\tau \\ \tau[1.78] \end{array} \right)$	$\left. \begin{array}{l} 0 \\ -1 \end{array} \right\} \Delta Q = e_0$
Quarks (fermions)	$\left( \begin{array}{c} u[0.33] \\ d[0.33] \end{array} \right) \left( \begin{array}{c} c[1.5] \\ s[0.45] \end{array} \right) \left( \begin{array}{c} t[173] \\ b[4.9] \end{array} \right)$	$\left. \begin{array}{l} 2/3 \\ -1/3 \end{array} \right\} \Delta Q = e_0$

be two different quantum states  $|p\rangle$  and  $|n\rangle$  of one and the same particle (Heisenberg 1932: *isospin* model). Major arguments were derived from the similar rest mass of the two particles and from similarly strong interactions in the atomic nucleus in spite of their different charge.

Meanwhile, we know that most elementary particles transform into each other if the energy balance and some other rules are fulfilled. It is, thus, straight forward to describe elementary particles in terms of quantum states by means of wave functions or bras and kets in Dirac notation. The theoretical picture is: matter and energy can assume particular quantum states (as an electron in a potential box, Sect. 3.6.1), namely those of the elementary particles. The nucleon quantum state kets are, then, expressed as

$$|p\rangle = |uud\rangle \text{ for the proton,} \quad (5.179a)$$

$$|n\rangle = |udd\rangle \text{ for the neutron.} \quad (5.179b)$$

We want to keep in mind that the surrounding matter under earth-like conditions, from micro-Kelvin up to thousands of Kelvin, only contains leptons ( $e, \nu_e$ ) and quarks ( $u, d$ ) of the first family (flavor) as constituting particles. In the context of condensed matter physics, chemistry and biology it is sufficient to limit our interest solely to electrons and nucleons, proton and neutron. Quarks as constituents of the nucleons will never appear as single free particles because of quark confinement.

In the following, we want to survey some more results which have been obtained in elementary particle physics by studying particles under extreme and unusual conditions, be it in high-energy accelerators (CERN, DESY, Fermi-Lab etc.) or in cosmological events as supernovae or the big bang. Experiments on accelerators are roughly speaking scattering experiments in which highly accelerated particles as electrons, protons etc. are scattered on other (target) particles, again protons, neutrons etc. At certain well defined energies  $E = mc^2$ , then, an abundance of scattering processes is observed, a band-like structure in the energy spectrum of the scattering processes occurs. The increasing number of scattering events in such a spectral band tells us that at the particular energy the incoming particles are transformed into a new particle. Its mass is determined from the energetic position  $E$  of the spectral band by  $m = E/c^2$ . The observed spectral band in the scattering spectrum has a spectral width  $\Delta E = (\Delta m)c^2$  which is related to a mass uncertainty  $\Delta m$  of the elementary particle. This uncertainty, on the other hand, determines the lifetime of the new elementary particle by means of the uncertainty relation  $t\Delta E \approx \hbar$  (3.23). Excitations in the scattering (elementary particle) spectrum yield direct information about mass and lifetime of particles. In the case of very short lifetimes of particles the term *particle resonance* is mostly used.

Based on a broad experimental basis, mainly scattering experiments of the described kind, the following picture has been developed: Apart from the leptons (Table 5.1) there is a class of particles composed of quarks, which are called *hadrons* ( $\delta\rho\sigma$  = large). As shown in Table 5.2 there are two types of hadrons, one which contains three quarks and the other one which is built-up of quark and anti-quark.

The three-quark hadrons  $|qqq\rangle$  are called *baryons* ( $\beta\alpha\rho\bar{u}\bar{s}$  = heavy), while those containing a quark and an anti-quark  $|q\bar{q}\rangle$  are named *mesons* ( $\mu\varepsilon\sigma o\bar{s}$  = middle). Baryons containing three fermionic quarks are, of course, fermions. The two-quark mesons are bosons for the same reason (Sect. 5.6.2). There is an explanation, why quarks can build-up only these two types of stable hadrons, those containing only three or two quarks. We will make this plausible below.

In Table 5.2, some important hadrons, baryons and mesons, are listed together with their quark content. The table distinguishes between hadrons containing only  $u$  and  $d$  quarks of the first family (flavor) (constituents of surrounding matter) and those which are composed also of charm ( $c$ ), strange ( $s$ ) respectively top ( $t$ ) and bottom ( $b$ ) quarks. The familiar nucleons proton  $|p\rangle$  and neutron  $|n\rangle$  belong to this first family. Only the proton is stable. The neutron has a finite lifetime of 887 s in nuclear reactions. In most atomic nuclei, however, it lives infinitely long, since possible decay products (nuclei) would have a larger mass than the initial nucleus (contradiction to energy conservation). Hereby, it is essential that the binding energy of the nuclei can be expressed as a mass difference.

The  $\Delta^{++}$  baryon ( $\Delta$  resonance) has an extremely short lifetime of only  $5.5 \times 10^{-24}$  s. Lifetimes of baryons, thus, scatter over a large range between infinity and  $10^{-24}$  s. According to a quark content of three  $u$  and/or  $d$  quarks these hadrons have masses around 1 GeV. Minor differences are due to somewhat different internal interactions (mass = energy).

**Table 5.2** Some important hadrons: Baryons consist of three quarks, mesons of quark and anti-quark. For the surrounding matter on earth only proton  $|p\rangle$  and neutron  $|n\rangle$  are important. The names of the quarks  $u, d, s$  are given in Table 5.1. Charge, mass and life-time are experimental values

Some hadrons				
	Quark content	Charge ( $e_0$ )	Mass (GeV)	Life time (s)
Baryons $ qqq\rangle$ (Fermions)	$ uud\rangle =  p\rangle$	+1	0.9383	$\infty$
	$ udd\rangle =  n\rangle$	0	0.9396	887
	$ uuu\rangle =  \Delta^{++}\rangle$	+2	1.232	$\approx 5.5 \times 10^{-24}$
	...			
	$ uds\rangle =  \Lambda\rangle$	0	1.116	$2.6 \times 10^{-10}$
	...			
Mesons $ q\bar{q}\rangle$ (Bosons)	$ u\bar{d}\rangle =  \pi^+\rangle$	+1	0.1396	$2.6 \times 10^{-8}$
	$( d\bar{d}\rangle -  u\bar{u}\rangle)/\sqrt{2} =  \pi^0\rangle$	0	0.1349	$8.4 \times 10^{-17}$
	$ d\bar{u}\rangle =  \pi^-\rangle$	-1	0.1396	$2.6 \times 10^{-8}$
	...			
	$ d\bar{s}\rangle =  K^0\rangle$	0	0.4977	$8.3 \times 10^{-11}$
	$ u\bar{s}\rangle =  K^+\rangle$	+1	0.4937	$1.24 \times 10^{-8}$
...				

With respect to mesons it is remarkable that particles with equal mass but opposite electric charge as the  $\pi$  mesons  $|\pi^+\rangle$  and  $|\pi^-\rangle$  are each time combinations of quark and antiquark  $|u\bar{d}\rangle$ , respectively  $|d\bar{u}\rangle$ .

A detailed analysis of all experimental facts about the  $\Delta^{++}$  resonance revealed a spin of  $3\hbar/2$  for this particle. Its quantum mechanical state including the spin degree of freedom must, thus, be written as

$$|\Delta^{++}\rangle = |uuu\rangle |\uparrow\uparrow\uparrow\rangle. \quad (5.180)$$

With the assumption of negligible angular momentum for the quarks in their ground state ( $s$  state) the representation (5.180) is in contradiction to the Pauli principle, since  $\Delta^{++}$  must be a fermion because of its half integer spin. Exchange of two quarks does not change the state representation (5.180). The particle state  $|\Delta^{++}\rangle$  is symmetric with regard to exchange of two particles, that is, it is not fermionic but rather bosonic, in contradiction to its spin value.

This contradiction, apart from some other important arguments, lead the elementary particle physicists to a totally new and unusual assumption: All quarks have, in addition to their spin, a further internal degree of freedom, i.e. quantum property for distinction. So far particles could be distinguished by their masses, their charge and their spin. Quarks must have, in addition to their charge ( $2e_0/3$  and  $-e_0/3$ ) and spin, a further quantum number, which can assume three different values rather than only two as the spin does.

Using this trinity of new quantum numbers, we can explain, or at least make plausible, why quarks can form only two stable types of hadrons, baryons  $|qqq\rangle$  and mesons  $|q\bar{q}\rangle$ . Physicists have artistic fantasy; they have called this new quantum property of quarks *color*. The fundamental colors red ( $R$ ), green ( $G$ ) and blue ( $B$ ) superimposed with each other yield the neutral color white. As negative and positive charges as well as opposite spins add up to zero, so does in analogy the quantum number color: the three colors  $R$ ,  $G$ ,  $B$ , if superimposed, add up to neutral white, or “zero color”. If  $u$ ,  $d$ ,  $s$  or  $c$  quarks appear in three colors, then the most general states of hadrons must be expressed as superpositions of states having the colors  $R$ ,  $G$ ,  $B$ . This is analogous to the representation of a general spin wave function (5.120a)–(5.120c).

For a bosonic meson of the type  $|q\bar{q}\rangle$  we, thus, write the general quantum state as

$$|q\bar{q}\rangle = \frac{1}{\sqrt{3}}(|R\bar{R}\rangle + |G\bar{G}\rangle + |B\bar{B}\rangle), \quad (5.181)$$

where the prefactor  $1/\sqrt{3}$  guarantees the normalization of the state.

The representation (5.181) is symmetric upon exchange of two quarks  $R \leftrightarrow G$ ,  $R \leftrightarrow B$ ,  $G \leftrightarrow B$ , just what is required for a boson. Furthermore, the superposition of quarks with different color (5.181) is white (neutral); all three colors are contained in the state with the same amount. This state is called a color singlet, in analogy to a spin singlet state, where two opposite spins compensate each other to zero spin.

Just as charged particles form stable composite systems at the lowest possible energy by combining an equal number of opposite charges in a neutral compound, colored quarks, obviously, tend to build up stable, color neutral (white) singlet composite states as mesons of the type (5.181). In a similar way, colored quarks can also form antisymmetric fermionic singlet states (equal number of colors) which are stable. As a short consideration shows the color neutral superposition state with antisymmetry upon exchange of two quarks (Pauli principle), requires, in contrast to bosons (5.181), the superposition of three quarks of the type  $|qqq\rangle$ . Antisymmetry of the state is guaranteed by alternating signs of the elements. We thus arrive at the following representation of a fermionic color neutral singlet hadron state:

$$|qqq\rangle = \frac{1}{\sqrt{6}}(|RGB\rangle - |RBG\rangle + |BRG\rangle - |BGR\rangle + |GBR\rangle - |GRB\rangle). \quad (5.182)$$

As required (5.182) changes its sign upon exchange of two colored quarks. All three colors are represented with equal weight (white singlet). It is evident that the contradiction in the  $\Delta^{++}$  representation (5.180) is removed when the quark state  $|uuu\rangle$  is replaced by its color decomposition, the color singlet (5.182). The  $\Delta^{++}$  resonance is antisymmetric when the color degree of freedom is respected.

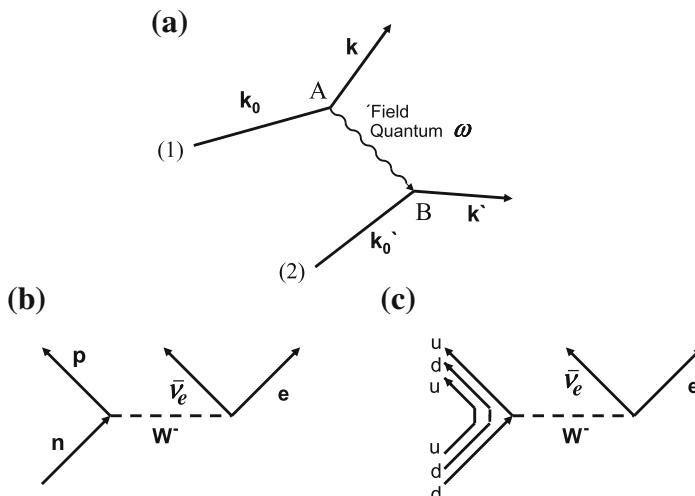
The existence of the color degree of freedom explains all hadrons observed so far. In particular, it explains why only two types of hadrons, mesons and baryons, are found in experiments. It must finally be emphasized again, that only quarks carry color with three quantum numbers  $G, R, B$ . Leptons do not have this internal degree of freedom.

Corresponding to the different elementary particles (Table 5.2) there exist different fundamental interactions in nature between these particles:

- The *electroweak interaction* contains two types of interactions which are separated at low energies in classical physics. On the one hand, this interaction constitutes the electromagnetic Coulomb force between electrically charged particles. The charge is the source of the electric field which is responsible for the force between the particles. On the other hand, this electroweak interaction is also responsible for particle reactions as the radioactive  $\beta$  decay ( $n \rightarrow p + e + \bar{\nu}_e$ ). At sufficiently high energies it, thus, couples to leptons ( $e, \bar{\nu}_e$ ) as well as to quarks ( $|n\rangle = |udd\rangle, |p\rangle = |uud\rangle$ ).
- The *strong interaction* couples only between quarks, and this only on a very short length scale below  $10^{-13}$  cm. On longer distances it vanishes (quark confinement). Strong interaction is inherently related to the quantum property color of the quarks, similarly as the electromagnetic interaction is due to the electric charge. The strong interaction binds two or three quarks together in mesons and baryons as protons and neutrons. Its external forces, outside the proton or neutron (quadrupole or higher moments) constitute the nuclear forces, which bind together the nucleons in an atomic nucleus. This nuclear force, which was not understood for a long time, is thus a Van der Waals type force between color neutral hadrons originating from strong color interactions.

- The *gravitation* is the attracting force between two masses which is solely connected to the property mass of the two bodies. In general, this interaction can only be observed when the much stronger electroweak (Coulomb) interaction is internally compensated by equal numbers of oppositely charged particles in a body (e.g., a planet or a star).

At this point we must briefly describe, without formal mathematical treatment, the outcome of a unified quantum theory of particles and fields (quantum field theory, Chap. 8). A particle in the classical sense with an own identity does not exist in quantum field theory. A particle is rather a local excitation of the extended non-local quantum field (see Sect. 8.3.3). Interactions between particles classically described by fields, as for example, the electromagnetic field between charged particles, are attributed to the exchange of particles, so-called field quanta. This phenomenon is roughly depicted in Fig. 5.17a. Particle (1) propagates with wave vector  $\mathbf{k}_0$  and emits another particle  $\omega$ , a so-called field quantum, near the position A. The underlying reason for the emission might be the decay of particle (1) into another particle. The field quantum  $\omega$  might be absorbed again by a second particle (2) with wave vector  $\mathbf{k}'_0$  at the position B. Emission of particle  $\omega$  at A causes a momentum transfer to particle (1) which changes its momentum from  $\mathbf{k}_0$  to  $\mathbf{k}$ . On the other hand, the absorption of  $\omega$  by particle (2) changes the momentum of that particle into  $\mathbf{k}'$ . If we forget about the



**Fig. 5.17** a–c Representation of the interaction of two particles by exchange of a third particle, a field quantum, in terms of so-called Feynman graphs. **a** General representation of the scattering of two particles (1) and (2) with wave vectors  $\mathbf{k}_0$  and  $\mathbf{k}'_0$  by exchange of a field quantum  $\omega$ . Coulomb scattering of two electrons on each other is mediated by photons as field quanta. **b**  $\beta$ -decay of a neutron  $n$  into a proton  $p$ : During the transformation  $n \rightarrow p$  a  $W^-$  boson is emitted, which decays into an electron  $e$  and an electronic anti-neutrino  $\bar{\nu}_e$ . **c** Representation of the  $\beta$ -decay (**b**) in the quark picture: the process is attributed to the transformation of a  $d$  quark into a  $u$  quark with simultaneous emission of a  $W^-$  boson

exchange of the field quantum (particle)  $\omega$ , the process illustrated in Fig. 5.17a might be interpreted as an interaction between particle (1) and particle (2) in a spatial region which contains the positions  $A$  and  $B$ . Hereby, even the type of the particles might change during the interaction. Leaving the  $\omega$  exchange out of the consideration, thus, leads back to the classical picture of a field mediated interaction between particles (1) and (2).

In the quantum field theoretical description (Chap. 8), the interactions between the fundamental elementary fermions leptons and quarks are described in terms of exchange of other types of elementary particles, so-called field quanta, which are altogether bosons with spin  $\pm\hbar$ .

The Coulomb force between two charged particles, for example, two electrons or an electron and a proton in the hydrogen atom, which is classically described by the electric field, is ascribed to an exchange of *photons* in quantum electrodynamics (Sect. 8.4.4). Photons have zero mass, a spin of  $\pm\hbar$  (bosons) and infinite lifetime (Sect. 8.2).

The weak interaction, which mediates coupling between quarks and leptons and thus causes the  $\beta$  decay ( $n \rightarrow p + e + \bar{\nu}_e$ ), is due to exchange of so-called *heavy  $W^\pm$  bosons* (Fig. 5.17b). These particles have a charge  $\pm e_0$ , a spin  $\pm\hbar$  and a short life time of  $3.1 \times 10^{-25}$  s. By means of the  $W^-$  boson, the  $\beta$  decay is explained in terms of a transformation of a  $d$  quark contained in the neutron into a  $u$  quark and simultaneous emission of the  $W^-$  boson. This boson decays into the products  $e$  and  $\bar{\nu}_e$  while the quark transformation changes the neutron  $|ddu\rangle$  into a proton  $|duu\rangle$  (Fig. 5.17c).

Quarks being the constituents of baryons and mesons (Table 5.2) interact by exchange of so-called *gluons* (derived from “glue”). These bosons being massless as photons with spin  $\pm\hbar$  are supposed to carry in addition the color degree of freedom ( $R, G, B$ ). In this case of the strong interaction between quarks also the field quanta, the gluons; are characterized by color as are the quarks. This is in contrast to the Coulomb interaction (weak interaction) where the mediating particles, the photons as field quanta, are not charged in contrast to the interacting electrons, protons, ions etc. The quantum field theory, which describes the strong interaction between quarks by means of gluons, is called *chromodynamics* ( $\chi\rho\sigma\mu\omega\zeta = \text{color}$ ) because of the characteristic new quantum property color of the involved particles.

Unlike for the electroweak and the strong interaction there does not exist so far a quantum field theory of gravitation, in which the attraction of masses classically described by action at a distance is ascribed to the exchange of field quanta between massive bodies. Nevertheless physicists believe that such a quantum field theory of gravitation will be created in the near future. Based on this belief one has already given the name *gravitons* to these bosonic field quanta of gravitation even though they were neither found in experiment nor does a unified theory of quantum gravitation exist so far.

The standard model discussed in this section is based on the three families of leptons and quarks (Table 5.1) and the force mediating bosons: photons, heavy  $W^\pm$  (and  $Z$ ) bosons and gluons. This standard model had a severe problem so far: it could not explain why elementary particles have a mass (inertia). A solution to this

problem is the so-called *Higgs mechanism*, which was developed as a theoretical model already in 1964 by Higgs [13], Englert (Nobel prize 2013) and Brout [14]. According to this model there exists a field everywhere in space, which is meanwhile called the *Higgs field*. The different particle fields, those of the leptons, the quarks and the heavy bosons ( $W^\pm$ ,  $Z$ ) interact with this field and the corresponding particles (field excitations) feel an inertia upon moving in the Higgs field. They get a finite mass. The mechanism is similar to that of the effective mass, which an electron gains upon moving through a crystal (Sect. 8.3.4). The particular value of the particle mass depends on the interaction strength between the particle field (electron, quark,  $W$  boson etc.) and the background Higgs field.

Up to about 2012 the Higgs mechanism was a purely theoretical model. But between 2011 and 2012 a new particle, the so-called *Higgs boson* was detected at the Geneva CERN accelerator LHC (Large Hadron Collider). This particle has all properties, which are predicted for a particle being the field quantum (excitation) of the Higgs field. It is neutral, has zero spin, a mass of about  $125 \text{ GeV}/c^2$  (approximately  $2.25 \times 10^{-25} \text{ kg}$ ) and a life time of about  $10^{-22} \text{ s}$ . Since the field quantum, the Higgs particle, does exist, the related quantum field must exist, too. It finally explains, why elementary particles can have a finite mass. The experimentally determined masses can not be derived from this general mechanism. So far the interaction strengths of the different particle fields with the Higgs field enter the theory as external parameters. It must also be emphasized, that the mass of common matter, i.e. essentially the mass of atomic nuclei, is determined to a small amount only by the Higgs mechanism. According to the mass-energy equivalence ( $E = mc^2$ ) the nucleon mass results to a high extent from the interaction energy between quarks and gluons being the constituents of the nucleons.

This section could only give a short overview over the elementary particles respectively the quantum states of the matter-field reality known so far. Stringent quantum field theories about particle interaction as quantum electrodynamics and chromodynamics are, of course, based on the fundamental rules of quantum mechanics (particle-wave duality, uncertainty relation etc.), but they are relativistic, that is, consistent with special relativity, in contrast to non-relativistic approaches as treated in Chap. 8. Reactions between elementary particles require extremely high energies such that particle velocities are close to the light velocity. This prohibits non-relativistic treatments in elementary particle physics.

When we restrict, however, our considerations on phenomena related to condensed matter, in particular on nanostructures and its electronic properties, the non-relativistic formulation of quantum mechanics within the frame of the Schrödinger equation is sufficient. There is a considerable gap between characteristic energies in condensed matter physics, typically 1–100 eV, and those in elementary particle physics significantly above MeV. Both fields of physics are completely decoupled on the energy scale. In condensed matter physics, the assumption of electrons and atomic nuclei being the fundamental stable particles of interest is usually sufficient.

## 5.7 Angular Momentum in Nanostructures and Atoms

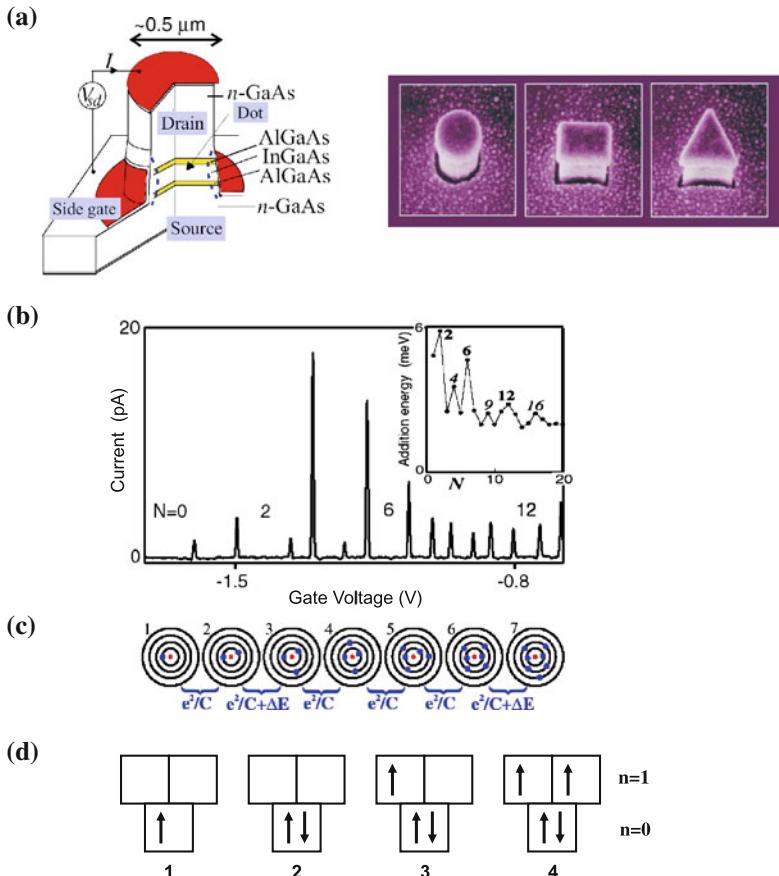
In Sect. 5.6, we have seen that the angular momentum, in particular the spin degree of freedom, represents an important organizing principle in nature, down to the world of elementary particles. In the following, we will learn that the Pauli exclusion principle, which derives from symmetry properties of the spin operators (Sect. 5.6.2), determines essential properties of matter, in particular its stability. Why are atoms composed by certain well defined numbers of electrons and why can they be ordered according to the rules of the periodic table? The answer is given by the Pauli principle as we will see.

It is a benefit of nanotechnology, in particular the modern techniques of nanostructuring semiconductors (Appendix B), that meanwhile nanoscaled device structures can be prepared in which electrons behave similarly as in atoms and molecules found in nature. These artificial semiconductor atoms allow experimental studies which mimic the properties of natural atoms, but under easier conditions for the experiment. Electrical measurements on these semiconductor nanostructures reveal the rules according to which real atoms are built-up. In the following, we will consider an example which clearly demonstrates the importance of the electronic angular momentum for these type of questions.

### 5.7.1 Artificial Quantum Dot Atoms

In natural atoms, electrons are confined to the atomic volume by the action of the attractive Coulomb potential of the positive nucleus. Here, the confining potential has the radius dependence  $e^2/4\pi\epsilon_0 r$ . Nanostructuring techniques, on the other hand, allow the preparation of tiny quasi-one-dimensional (1D) semiconductor structures, so-called quantum dots (Sect. 3.6.1). Depending on details of the shape of the dots and of boundary conditions, the confining potential for an electron might be parabolic or of the hard wall type, that is, much easier to handle than the Coulomb potential of a natural atom.

A common method to prepare such quantum dots is based on semiconductor layer structures such as GaAs/AlGaAs/In<sub>0.05</sub>Ga<sub>0.95</sub>As/AlGaAs/GaAs (Fig. 5.18a) which are epitaxially grown (Appendix B). In this layer structure, the electron is confined to the InGaAs layer (thickness 12 nm) by means of the two AlGaAs neighbouring potential barriers (thicknesses 9 nm and 7.5 nm at the bottom). The InGaAs layer with the lowest forbidden band between conduction and valence band (Sect. 8.3.5) offers the lowest possible energy for an electron in the conduction band in comparison with the surrounding layers. The electron is confined in this layer as in a 2D potential box with finite walls. By means of lithographical techniques (Appendix B), the layer sandwich can be structured vertically into columns with diameters below 500 nm. This leads to a quasi-1D confinement of the electrons in a dot formed by the column dimension and the two AlGaAs barriers on top and bottom (Fig. 5.18a). Tunneling of



**Fig. 5.18** **a–d** Single electron tunneling through artificial quantum dot atoms [15]. **a** Realization of a semiconductor quantum dot by two AlGaAs tunneling barriers surrounding a thin InGaAs region built-in in a GaAs column. The *dot* is formed by the InGaAs region having the lowest potential for conduction electrons in the structure. The potential of the InGaAs dot can be varied in relation to the drain contact by a metallic side gate contact (separated from the current channel by Schottky contact, Appendix A). Drain and source contacts are realized by *n*-doped GaAs regions. Metal contacts are in red colour (left). The lateral structuring of the columns with a diameter of about 500 nm was performed by electron beam lithography (Appendix B). Scanning electron micrographs of three column structures are shown on the right side. **b** Coulomb blockade oscillations measured on the device in (a). Current peaks as function of gate voltage indicate the occupation of electronic states in the quantum dot. Insert Plot of addition energies for addition of 1, 2, 3, 4, ... electrons to the states of the quantum dot. **c** Schematic representation of the successively filled quantum states of the dot in terms of a shell model.  $e^2/C$  represents the Coulomb repulsion upon adding a further electron to the dot.  $\Delta E$  is the energy of a quantum jump from one to the next higher shell. **d** Representation of the shell model by means of boxes which represent electron orbitals: 1st shell (principal quantum number  $n = 0$ ) lower box; 2nd shell (principal quantum number  $n = 1$ ) two boxes on top. Because of the two possible spin orientations each box can be occupied by two electrons (see Table 5.3). According to Hund's rule 4 electrons fill the lower box (1st shell) with 2 electrons and the two upper boxes of the 2nd shell, each with one electron of identical spin orientation

electrons from the bottom (source) GaAs region through the lower AlGaAs barrier into the dot and tunneling through the upper barrier into the top (drain) GaAs region allows the observation of single electron tunneling as described in Sect. 3.7. An external bias between source and drain contact can shift the potential of these contacts against each other. In addition, the potential of the whole dot with respect to source can be varied by means of a side gate contact biased in an adequate way against the source contact (Fig. 5.18a). The arrangement in Fig. 5.18a is a realization of the schematic experimental set-up of Fig. 3.18a for the observation of single electron tunneling effects. As shown in Sect. 3.7 measurement of current, respectively, conductivity  $\sigma$  between source and drain contact as a function of gate voltage  $V_g$  (Fig. 3.18) yields a spectroscopy of the electronic states of the quantum dot. Hereby, we observe the sequential occupation of the states with  $N = 1, 2, 3, 4, \dots$  electrons. Due to the Pauli exclusion principle each quantum state of the dot (resulting from confinement, Sect. 3.6.1) can be occupied only once by an electron tunneling into the dot. In case of spin degeneracy, two electrons with opposite spin, of course, can occupy a state. In addition, an electron tunneling into the dot, where already one or more electrons occupy states, experiences the Coulomb repulsion of the already present electrons. This many-body repulsion energy must be overcome by an additional electron. For successful tunneling into the dot, that is, breaking of the Coulomb blockade, it must carry the so-called charging energy  $e^2/C$  with  $C$  as the capacity of the dot against its surrounding (Sect. 3.7). The addition energy  $\Delta\mu = (e^2/C) + \Delta E$  for adding one further electron to the dot (3.128a), (3.128b), thus, contains two terms: the charging energy (many body repulsion) and the energy difference  $\Delta E$  between the lowest unoccupied quantum state of the dot, which will be occupied by the incoming electron, and the highest already occupied state.

In a first step for understanding single electron tunneling through the quantum dot in Fig. 5.18a, we want to inquire into the spectrum of quantum states originating from electron confinement in a planar dot with rotational symmetry. The electron wave function can be described in this case in the frame of rectangular Cartesian coordinates with  $x$  and  $y$  as coordinates in the plane of the flat box and  $z$  pointing into the direction perpendicular to the heterostructure layer sequence in Fig. 5.18a. But also a description in terms of cylindrical coordinates with  $r$  and  $\varphi$  parallel to the layer sequence is possible. Because of the dimensions of the flat box (thickness 10 nm, diameter approx. 500 nm) cylindrical coordinates are adequate.

The potential  $V(r)$  an electron is exposed to in the box is, thus, two-dimensional (2D) and has a minimum in the center of the box at  $r = 0$ . At the surface of the cylindrical box an electron depletion layer with upwards bending of the conduction band (potential energy of free electrons in the dot) is given, similarly as for a Schottky barrier (Appendix A). In good approximation, the potential is parabolic, centered around  $r = 0$ , and is written as

$$V(r) = \frac{1}{2}m^*\omega_0^2r^2 = \frac{1}{2}m^*\omega_0(x^2 + y^2), \quad (5.183)$$

$m^*$  is the so-called effective mass of the electron, the electronic mass modified by the surrounding crystal lattice (Sect. 8.3.4).  $\omega_0$  is a parameter which describes the curvature of the confining parabolic potential.

In Cartesian coordinates we, thus, obtain the following Hamiltonian for an electron confined in the flat box:

$$\hat{H} = \frac{\hat{p}_x^2 + \hat{p}_y^2}{2m^*} + \frac{1}{2}m^*\omega_0(x^2 + y^2). \quad (5.184)$$

This is the Hamilton operator of two equal oscillators oscillating in directions perpendicular to each other. A circular motion can be reduced to two mutually perpendicular oscillations. The energy eigenvalues of the two oscillators (2D oscillator) in (5.184) are represented as

$$E_{n_x n_y} = E_x + E_y = \left(n_x + \frac{1}{2}\right)\hbar\omega_0 + \left(n_y + \frac{1}{2}\right)\hbar\omega_0 = (n_x + n_y + 1)\hbar\omega_0. \quad (5.185)$$

The quantum numbers  $n_x$  and  $n_y$  separately assume the values  $0, 1, 2, 3, \dots$ . The solution in Cartesian coordinates, thus, yields the following spectrum of electronic states in the cylindrical box ( $n_x + n_y = n$ ):

$$E_n = (n + 1)\hbar\omega_0, \quad (5.186)$$

with  $n$  as integer numbers.

On the other hand, the Hamiltonian of the problem can also be represented in cylindrical coordinates. In this case, we separate the kinetic energy into a radial contribution  $\hat{T}_r$  and an orbital part (5.5b):

$$\hat{H} = \hat{T}_r + \frac{\hat{L}_z^2}{2m^*r^2} + \frac{1}{2}m^*\omega_0^2r^2. \quad (5.187)$$

$\hat{T}_r$  describes the kinetic energy related to a change of the radial component of the position vector and  $\hat{L}_z$  is the angular momentum operator in  $z$  direction. The cylinder symmetry of the potential implicates that the Hamiltonian  $\hat{H}$  commutes with  $\hat{L}_z$ , that is,  $\hat{H}$  and  $\hat{L}_z$  have the same system of eigenstates  $|n, m\rangle$ :

$$\hat{H}|n, m\rangle = E_{n,m}|n, m\rangle, \quad (5.188a)$$

$$\hat{L}_z|n, m\rangle = m\hbar|n, m\rangle. \quad (5.188b)$$

Hereby, the orientation quantum number of the angular momentum assumes the values  $m = 0, \pm 1, \pm 2, \pm 3, \dots$ . Because of the shape of the quantum dot (flat plate) the angular momentum can assume only one orientation, perpendicular to the plate-like dot along  $z$ . The angular momentum quantum number  $l$  is, therefore, identical with the absolute value of the directional quantum number  $|m|$ .

In order to represent the energy eigenvalues  $E_n$  (5.185) as a function of the angular momentum quantum number  $m$ , that is,  $E_{n,m}$ , the Schrödinger equation must be solved in cylindrical coordinates  $r, \varphi$  by using the Hamiltonian (5.187). We skip this mathematical procedure and use the following simple argumentation: The energy of an electron orbiting in the cylindrical box does not depend on the rotational direction, that is, the orientation  $+z$  or  $-z$  of the angular momentum, it therefore depends on  $|m|$ . On the other hand, in any case (5.186) must be one possible representation of the eigenvalues; this is required by the solution of the problem in Cartesian coordinates. The series of integer numbers  $n$  (5.186), thus, must contain the angular quantum number as  $|m|$ . Apart from an increase of the angular momentum, the total energy of the electron can also increase by a change of the kinetic energy  $T_r$ , that is, a variation of the radial component of the position vector. Because of the circular motion, the  $x$  and  $y$  coordinates are equivalent in this variation. Referring to the Cartesian representation (5.185), then, a change of the quantum number  $n_x$  by one unit requires an identical change of the quantum number  $n_y$ . Within the series  $n$  (5.186) of quantum numbers, there must be a sub-set of even numbers  $2k$  ( $k = 0, 1, 2, \dots$ ). We, thus, arrive at the conclusion that the energy eigenvalues  $E_n$  (5.185), (5.186) can also be expressed as

$$E_n = E_{k,m} = (2k + |m| + 1)\hbar\omega_0 = (n + 1)\hbar\omega_0, \quad (5.189)$$

with  $k = 0, 1, 2, 3, \dots$  and  $m = 0, \pm 1, \pm 2, \pm 3, \dots$ . The quantum number  $k$  is attributed to different radial wave functions while the directional quantum number  $m$  describes different angular momenta of the electron orbit, that is, classically different angular velocities of the electron. Both representations (5.186) and (5.189) are equivalent, they only originate from different formalisms of solving the Schrödinger equation.

The quantum number  $n = 2k + |m|$ , which numerates the possible energies  $E_n$  (5.189) of the system, is called *principal quantum number*. For a given energy  $E_n$  of the electron, a maximum angular momentum must exist, since an increase of the angular momentum simultaneously enhances the kinetic energy of the particle. Thus, for a fixed principle quantum number  $n$  a maximum angular momentum quantum number  $|m| = l$  exists, which follows as

$$l = |m| = n - 2k = n, n - 2, n - 4, \dots, 1 \text{ or } 0. \quad (5.190)$$

Equation (5.190) describes the variety of quantum states of a 2D harmonic oscillator, which is a good model for an electron being confined in a flat circular quantum dot.

In Table 5.3, the energetically lowest eigenstates are listed. The occupation of states is governed by the Pauli principle: Each single electron level (spin degenerate) is occupied by at maximum two electrons, one with spin “up” and one with spin “down” ( $s = \pm 1/2$ ). This explains the degrees of degeneracy 2, 4, 6, 8 for the four lowest energy levels  $\hbar\omega_0, \dots, 4\hbar\omega_0$ . 2, 4, 6, 8 electrons can occupy the corresponding energy levels with principle quantum numbers  $n = 0, 1, 2, 3$ . The extension of the scheme to higher quantum numbers is easily done in analogy.

**Table 5.3** Description of the energetically lowest eigenstates of the 2D harmonic oscillator by the different quantum numbers  $n, k, m, s$ . Additionally to principal quantum number  $n$  the angular momentum quantum numbers  $k$  (mostly called  $l$ ) and  $m$  the spin quantum number  $s$  indicates the two possible spin orientations of an electron. The electron configuration corresponding to the energy eigenvalue  $v\hbar\omega_0$  is called  $v$ th shell

Energy eigenvalue $E_n$	Principle quantum No., $n$	Radial quantum No., $k$	Angular mom. q. No., $m$	Spin quantum No., $s$	Degree of degeneracy	
$\hbar\omega_0$	0	0	0	$\pm 1/2$	2	1. Shell
$2\hbar\omega_0$	1	0	$\pm 1$	$\pm 1/2$	4	2. Shell
$3\hbar\omega_0$	2	0	$\pm 2$	$\pm 1/2$	6	3. Shell
		1	0	$\pm 1/2$		
$4\hbar\omega_0$	3	0	$\pm 3$	$\pm 1/2$	8	4. Shell
		1	$\pm 1$	$\pm 1/2$		

The eigenstates of a three-dimensional (3D) oscillator can be obtained in analogy. A 3D oscillator is a superposition of three equal mutually perpendicular 1D oscillators. Accordingly the energy eigenvalues are obtained similarly to (5.189) as

$$E_n^{(3D)} = \left(2k + l + \frac{3}{2}\right)\hbar\omega_0. \quad (5.191)$$

Since in the 3D case the angular momentum can be arbitrarily oriented in space rather than only perpendicular to the plane of a 2D oscillator, the quantum number of the total angular momentum  $l$  appears in (5.191) instead of only the orientation quantum number  $|m|$  in (5.189). We have to take into account, however, that according to different orientations of the angular momentum the quantum number  $m$  assumes integer values between  $-l$  and  $+l$ . This determines finally the degeneracy of each energy level. As for the 2D oscillator, the radial quantum number appears as  $2k$ . In spite of the three-dimensionality of the problem the electron moves on a planar orbit (as in the 2D case) independent of its orientation in space. For the three energetically lowest eigenstates with principal quantum numbers  $n = 0, 1, 2$  we, thus, obtain the following combination of quantum numbers:

$$\begin{aligned} n = 0; \quad l = 0; \quad m &= 0 \\ n = 1; \quad l = 1; \quad m &= 0, \pm 1 \\ n = 2; \quad l = 0, 2; \quad m &= 0, \pm 2, \pm 1. \end{aligned} \quad (5.192)$$

To obtain the degeneracy of an energy level, we must finally take into account also the two possible spin orientations  $s = \pm 1/2$ , that is, for  $n = 1$  the degree of degeneracy amounts to 6.

We now return to the 2D oscillator in order to understand spectroscopic results which have been obtained on the flat disc-like quantum dot of Fig. 5.18. We expect to

find the spectrum of energies (5.189) of Table 5.3. Hereby, each level can be occupied by electrons according to its degree of degeneracy.

An adequate spectroscopy for the study of these effects is single electron tunneling through the quantum dot (Sect. 3.7). Indeed, measurements of the tunneling current at a small source-drain bias as function of gate voltage  $V_g$  (Fig. 5.18b) show sharp spectral bands which are due to Coulomb blockade (Sect. 3.7). Each spectral band indicates a gate voltage at which the quantum dot increases its occupancy by one electron. At around  $-1.6$  V gate voltage the first peak appears, it arises from the occupation of the lowest energy level  $\hbar\omega_0$  (1st shell) by one electron. At a gate voltage of about  $-1.5$  V a second electron can break the Coulomb blockade. It possesses the necessary charging energy  $e^2/C$  to overcome the repulsion of the electrons already present in the dot and occupies the second possible state (with spin opposite to the 1st electron) of the 1st shell ( $n = 0$ ). According to Table 5.3 and Fig. 5.18c the 1st shell, then, is fully occupied. Addition of a further 3rd electron to the dot, now, requires, on top of the charging energy  $e^2/C$  (to overcome the Coulomb repulsion), an additional energy  $\hbar\omega_0 = \Delta E$  for the quantum step to the 2nd shell ( $n = 1$ ). The total addition energy is higher than that for adding the 2nd electron and the voltage distance between 3rd and 2nd peak is larger than that between 2nd and 1st one (Fig. 5.18b).

The 2nd shell with  $n = 1$  can accept four electrons (Table 5.3 and Fig. 5.18c, d). Taking into account the occupation of the 1st shell with two electrons the 3rd shell starts to be filled at the transition from the 6th to the 7th tunneling electron, i.e. peak 6 to 7 in Fig. 5.18b. Indeed, the distance between peaks 7 and 6 is somewhat larger than that between peaks 6 and 5, namely by the excitation energy  $\Delta E = \hbar\omega_0$  between 2nd and 3rd shell (step in principal quantum number).

If the applied gate voltage  $V_g$  would depend linearly on the energetic distances between the energy levels of the dot, the distances between peaks 6 and 7 as well as between 2 and 3 in Fig. 5.18b should be equal, namely  $(e^2/C) + \hbar\omega_0$ . This is not the case. One reason is that a voltage change at the gate does not reflect directly the corresponding shift of the potential on the dot because of various inherent resistances in the semiconductor structure. On the other hand the assumption of the so-called orthodox model (Sect. 3.7) for single electron tunneling is too simple. Different electron occupations of the dot require in principle (slightly) different Coulomb repulsion energies which is neglected in the present simple description. A more profound theoretical description explains the deviations from the experimental results nearly perfectly [15].

A plot of the series of addition energies for adding one electron to the dot (Fig. 5.18b, inset) shows enhanced values where a new shell starts to be filled, i.e. at  $N = 2, 6, 12, \dots$ . This is, of course due to the quantum step to the new shell. But increased addition energies are also found at electron numbers  $N = 4, 9, \dots$ , where each time a shell is half filled. With four electrons, for example, on the dot the 2nd shell is half filled. According to its discoverer this phenomenon is called *Hund's rule*. The underlying reason is again the requirement of antisymmetry of the wave function (Pauli principle): Upon filling a shell (states with equal principle quantum number), the available states are first filled with electrons of equal spin orientation

before subsequently electrons with opposite spins are added to each state of the shell (Fig. 8.15d). As seen in Fig. 8.15d, four electrons in the quantum dot require equal spin orientation of the two electrons in the 2nd shell ( $n = 1$ ). Parallel spins are related to a symmetric two-particle spin wave function. The requirement of antisymmetry for the total wave function, thus, requires an antisymmetric position wave function for the two electrons. Upon exchange of the two electrons, the position wave function must change its sign, that is, it has a node between the two electrons. In comparison with a symmetric one, this antisymmetric wave function forces the electrons to be at largest possible distance from each other. The Coulomb repulsion between the electrons is reduced as compared with a symmetric wave function and the parallel spin orientation of the two electrons in the shell is favored. It is, thus, energetically more favorable to fill up the states of a shell with electrons of equal spin orientation before spins of opposite orientation are added. When a shell is half filled with electrons of equal spins, the addition of a further electron with opposite spin requires a little bit more energy due to the enhanced Coulomb repulsion of electrons being arranged now somewhat closer to each other. This is the straightforward explanation for Hund's rule.

The described experiment of single electron tunneling through a quantum dot excellently demonstrates the importance of angular momentum and spin for the internal electronic structure of many-particle systems. The considered issues also yield the key for an understanding of the structure of atoms and the periodic table, which governs the properties of atoms and their relation with each other.

### 5.7.2 Atoms and Periodic Table

The simplest atom in nature, the *hydrogen* (H) *atom*, also has one confined electron, similarly as the electron in the quantum dot in Sect. 5.7.1. In contrast to the dot, however, the electron of the H atom is confined by the attractive Coulomb potential of the positive proton of the nucleus. Accordingly the Hamilton operator for the valence electron of the H atom is written as ( $m$  free electron mass):

$$\hat{H} = \hat{T}_r + \frac{\hat{L}^2}{2mr^2} + \frac{e}{4\pi\epsilon_0 r}. \quad (5.193)$$

The nucleus (proton) is about 2000 times heavier than the electron; in good approximation it is, therefore, assumed to be at rest. The potential in (5.193) is static.

Because of the 3D character of the problem the total angular momentum  $\hat{L}$  enters the description rather than only its  $z$  component (as in a planar problem). Because of spherical symmetry of the H atom  $\hat{L}_z$  and  $\hat{L}^2$  commute with  $\hat{H}$  (Sect. 5.3). Due to various possible orientations of the electron orbit in space, the  $z$  component of the angular momentum must not coincide with the total momentum. Thus, the following complete system of eigenvalue equations has to be solved:

$$\hat{H}|R_n\rangle|l, m\rangle = E_{n,l,m}|R_n\rangle|l, m\rangle, \quad (5.194a)$$

$$\hat{L}^2|R_n\rangle|l, m\rangle = l(l+1)\hbar^2|R_n\rangle|l, m\rangle, \quad (5.194b)$$

$$\hat{L}_z|R_n\rangle|l, m\rangle = m\hbar|R_n\rangle|l, m\rangle, \quad (5.194c)$$

$|l, m\rangle$  are the angular momentum eigenstates (part of the complete eigenstates), which are given in their position representation (wave functions) and discussed in Sect. 5.3. Between the quantum numbers  $l$  and  $m$  the relation  $m = -l, -l+1, \dots, 0, \dots, l-1, l-2$  holds (5.35b). The angular momentum quantum number  $l$  is bounded above for a particular given energy  $E_{n,l,m}$  of the electron.  $|R_n\rangle$  is the radial part of the complete eigenstate. Different principal quantum numbers  $n$  describe different radial distances of the electron from the nucleus, i.e. also different electron energies  $E_{n,l,m}$ .

A peculiarity of the Coulomb potential is the degeneracy of the energy eigenvalues in the quantum number  $l$ . The energies  $E_{n,l,m}$  coincide for differing angular momentum quantum numbers  $l$ . Without a magnetic field, this degeneracy is then also given for the directional quantum number  $m$ . We do not prove this peculiarity of the Coulomb potential here. But we keep in mind that the energy of the hydrogen valence electron (without any external perturbation) does only depend on the principal quantum number  $n$ :

$$E_n = -\frac{e^4 m}{8\varepsilon_0^2 h^2} \frac{1}{n^2} = -R_y \frac{1}{n^2}. \quad (5.195)$$

The term  $R_y$  being only dependent on natural constants is called Rydberg constant. Equation (5.195) is not formally derived in this book. For this purpose, the eigenvalue problem (5.194a)–(5.194c) had to be solved in spherical coordinates, a rather tedious job. A semiclassical treatment on the basis of Bohr orbits also yields the important result (5.195). One can find the formal solution in every theoretical textbook on quantum mechanics, for example, [16]. The formal treatment, of course, yields the upper limitation of the angular momentum by the electron energy, that is, the upper bound of the quantum number  $l$  by the (energy) principal quantum number  $n$ :

$$l = 0, 1, 2, \dots, n-1. \quad (5.196)$$

In conclusion, the variety of quantum numbers for the electron in the hydrogen atom is given as follows:

Principal quantum number	$n = 1, 2, 3, 4, \dots$
Angular momentum quantum number	$0 \leq l \leq n-1$
Orientation quantum number	$-l \leq m \leq l$
Spin quantum number	$s = \pm 1/2$

In analogy with (5.192), we obtain the following possible combinations of quantum numbers for the hydrogen electron:

$$\begin{aligned} n = 1; \quad l = 0; \quad m = 0; \quad s = \pm 1/2 \\ n = 2; \quad l = 0, 1; \quad m = 0, \pm 1; \quad s = \pm 1/2 \\ n = 3; \quad l = 0, 1, 2; \quad m = 0, \pm 1; \pm 2; \quad s = \pm 1/2. \end{aligned} \quad (5.197)$$

In contrast to (5.192), the spin quantum number with the two possible values  $\pm 1/2$  has explicitly added. The energy levels, thus, have the following degrees of degeneracy (in brackets):

$$E_1(2), \quad E_2(8), \quad E_3(18), \quad \dots \quad (5.198)$$

Notice that the quantized eigenvalues  $E_n$  (5.194a)–(5.194c), (5.195), (5.198) are negative because of the binding character of the potential. The energetically lowest ground state has a binding energy  $E_1 = -13.6$  eV with respect to a potential zero point infinitely far from the atom. We will approximately calculate this value in Sect. 6.2.2.

The ladder of negative energy eigenvalues  $E_n$  (5.195) approaches the limit zero for  $n \rightarrow \infty$ . When an electron is excited from any binding energy  $E_n$  to this limit, the atom is ionized. The electron is no longer bounded, it can evade to infinity.

The radial wave functions  $R_n(r) = \langle \mathbf{r} | R_n \rangle$  (5.194a)–(5.194c) resulting from the solution of the Schrödinger equation with the Hamiltonian (5.193) have a maximum at  $r = 0$  and decay to zero for  $r \rightarrow \infty$ . For principal quantum numbers  $n > 0$ , the radial wave functions oscillate around zero with  $(n - 1)$  zero values (nodes). The probabilities  $2\pi |R_n(r)|^2 r^2$  of finding the electron at a radial distance  $r$  from the nucleus have maxima at values which approximately correspond to the orbits calculated on the basis of Bohr's atomic model. For the ground state this is the so-called Bohr radius (Sect. 6.2.2) of about 0.05 nm [16]. Electronic excitations between the energy levels (5.195) explain the spectrum of sharp absorption and emission lines of atomic hydrogen which had eminent importance for the development of quantum physics.

We now proceed to the consideration of more complex atoms with higher atomic numbers  $Z > 1$ . These atoms contain  $Z$  protons and comparably many neutrons in their nucleus and  $Z$  electrons in their shell. In this case, a many-particle problem with  $Z$  electrons in the attractive Coulomb potential of the  $Z$ -times positively charged nucleus should be solved. This complex procedure is generally avoided by using a single electron approximation: The Schrödinger equation is solved for one single electron in an effective potential which contains the nuclear charge and the other  $Z - 1$  electrons of the shell. The nuclear charge is surrounded by the  $Z - 1$  electrons. This positively charged system is called *atomic core*. The effective potential for the one considered electron is, thus, represented by the screened Coulomb potential of one positive core charge. One-electron eigenstates and energy eigenvalues are approximately calculated in this core potential. A next better approximation is obtained by using the obtained eigenstates (wave functions) for the calculation of a mathematically improved atomic core. This procedure can be repeated self-consistently until deviations between two subsequent solutions are only marginal. The atomic core consisting of the nucleus (charge  $Ze$ ) and  $Z - 1$  shell electrons has spherical symmetry and so has the screened Coulomb potential for the considered

electron. Its eigenstates, therefore, contain the angular momentum eigenstates  $|l, m\rangle$  as factors. The electronic wave functions can, thus, be characterized by the quantum numbers  $l$  and  $m$  and, of course, by the spin quantum number  $s$ . As in the case of the hydrogen atom, there is the additional principal quantum number  $n$  which is related to different radial distances of the electron from the nucleus. Since the potential is not the pure Coulomb potential, there is no degeneracy in the angular momentum quantum number  $l$  as for the hydrogen atom. Different angular momenta of the electron (different  $l$ ) are connected with different energies. A hydrogen energy level  $E_n$  splits into a variety of different levels  $E_{n,l,m}$  where the  $m$  degeneracy is broken only in an applied magnetic field.

The physical reason for the loss of the  $l$  degeneracy in the screened Coulomb potential is a different amplitude of the electronic wave functions with different angular momentum. States with low angular momentum have a large amplitude near the nucleus and the electron samples the nuclear charge to a higher extent than an electron with high angular momentum. States with high angular momentum are suppressed at the origin, they see the nuclear charge shielded by the electrons in the inner orbits. Consequently, at each  $n$  the energy increases with  $l$ . The  $l$  degeneracy is broken.

Pauli principle, now, requires that every single electron level (spin taken into account) is occupied by one single electron only. Consequently we can count how many electrons with equal principal quantum number  $n$  can occupy states in an atom.

For each principal quantum number  $n$  (particular shell of the atom), there are  $n$  possible values of the angular momentum quantum number  $l$ .

- For each quantum number  $l$ , there are  $(2l + 1)$  possible values of the orientation quantum number  $m$ .
- For each pair  $(l, m)$  of quantum numbers, there are two possible values of spin quantum numbers  $s$ .
- Each pair  $(n, l)$  of quantum numbers allows at maximum  $2(2l + 1)$  electrons in the particular  $n$  shell of the atom.

These rules and the degeneracies of the principal quantum numbers (5.197) yield a deeper understanding of the internal structure of atoms. This structure is expressed in the *periodic table of elements* (Fig. 5.19), in which Mendeleev organized the chemical elements for the first time in 1869. At that time, the order was established only on the basis of experimentally observed chemical similarities and reactivities of the elements.

Since in general the electronic energy within a particular shell described by the principal quantum number  $n$  splits up into different levels with quantum numbers  $l$ , a common notation is used for the different angular momentum states (Sect. 5.3): The states with  $l = 0, l = 1, l = 2, l = 3, \dots$  are called  $s, p, d, f, \dots$  orbitals. Accordingly the energy level  $E_{n,l}$  with  $n = 1, l = 0$  is called  $1s$  shell, that with  $n = 2, l = 0$   $2s$  shell, respectively that with  $n = 2, l = 1$   $2p$  shell. In analogy to Fig. 5.18d, chemists got accustomed to using a simple casket scheme (Fig. 5.20). Each casket represents an atomic orbital which can be occupied by two electrons with opposite spin. For each  $p$  level, the three caskets arranged together correspond to

The periodic table is shown with the following annotations:

- Element and Atomic Number:** Brackets indicate the element name and its atomic number.
- Atomic Mass:** In brackets, the mass of the stable isotope.
- Electron Configuration:** Deeper occupied shells are shown in brackets.
- Shells must be added:** Shells must be added to reach the total atomic number.
- Example:** Fe:  $1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^6, 4s^2$

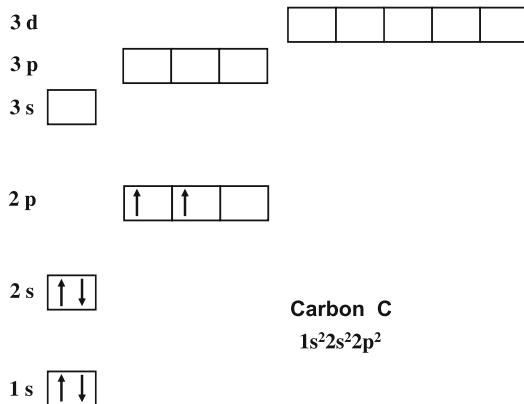
Key orbital occupancy details:

- 1s:** 1s orbital filled with 2 electrons.
- 2s:** 2s orbital filled with 2 electrons.
- 2p:** 2p orbital filled with 6 electrons (2 in each of the three p-orbitals).
- 3s:** 3s orbital filled with 2 electrons.
- 3p:** 3p orbital filled with 6 electrons (2 in each of the three p-orbitals).
- 4s:** 4s orbital filled with 2 electrons.
- 3d:** 3d orbital filled with 10 electrons (2 in each of the five d-orbitals).
- 4p:** 4p orbital filled with 6 electrons (2 in each of the three p-orbitals).
- 5s:** 5s orbital filled with 2 electrons.
- 4d:** 4d orbital filled with 10 electrons (2 in each of the five d-orbitals).
- 5p:** 5p orbital filled with 6 electrons (2 in each of the three p-orbitals).
- 6s:** 6s orbital filled with 2 electrons.
- 5f:** 5f orbital filled with 14 electrons (2 in each of the seven f-orbitals).
- 6p:** 6p orbital filled with 6 electrons (2 in each of the three p-orbitals).
- 7s:** 7s orbital filled with 2 electrons.
- 5g:** 5g orbital filled with 22 electrons (2 in each of the 13 g-orbitals).
- 6f:** 6f orbital filled with 32 electrons (2 in each of the 15 f-orbitals).
- 7p:** 7p orbital filled with 6 electrons (2 in each of the three p-orbitals).

**Fig. 5.19** Periodic table of elements, i.e. of natural atoms. The nomenclature  $1s^2 2s^2 2p^3$  denotes the occupation of the  $1s$  and  $2s$  shell with 2 electrons each and the  $2p$  shell with 3 electrons

the  $p_x$ ,  $p_y$ ,  $p_z$  orbitals of Fig. 5.5. The atom in Fig. 5.20 has occupied  $1s$ ,  $2s$  orbitals and two of the three  $2p$  orbitals are filled with one electron. Note that according to Hund's rule (Sect. 5.7.1), two  $p$  orbitals are filled, each with one electron, rather than a filling of one orbital with two electrons. Equal spin orientation in two orbitals is energetically more favorable than two electrons with opposite spin in one and the same orbital. Comparing Fig. 5.20 with the periodic table in Fig. 5.19, the depicted atom is recognized as the carbon (C) atom with atomic number 6.

Coming back to the periodic table in Fig. 5.19, we begin the description with hydrogen (H), the simplest atom. Corresponding to the atomic number 1 only the  $1s$  orbital is filled with one electron. Filling of the  $1s$  orbital with two electrons of opposite spin leads to helium (He) with atomic number 2. The first shell is now complete. Using arguments based on filling of molecular orbitals (Sect. 6.2.3) a complete electronic shell causes negligible chemical reactivity of the element. Helium, therefore, is a noble gas. The next higher atomic number 3 requires filling of the  $2s$  orbital with one electron (Fig. 5.20). A half filled shell means high chemical reactivity and this is indeed true for the alkali metal lithium (Li). The subsequent atomic numbers are related to partially filled orbitals and the corresponding elements participate in chemical reactions. At an atomic number 10 the  $1s$ , the  $2s$  and the  $2p$  orbitals are completely full; the corresponding unreactive element neon (Ne) is again a noble gas. Along these arguments sequential filling of the atomic orbitals, respectively shells, yields the periodic table of Fig. 5.19.



**Fig. 5.20** Box, respectively shell model of the natural atoms up to the  $d$  shell. As an example the carbon (C) atom with its six electrons is shown. The occupation of the  $2p$  shell with two electrons of equal spin in two different boxes demonstrates again Hund's rule (compare Fig. 5.18d). The notation  $1s^2 2s^2 2p^2$  expresses the total electron occupation of the  $1s$ ,  $2s$  and  $2p$  shells with two electrons each

As is seen in the arrangement of elements, the filling of the  $3p$  orbitals is not followed by occupation of the  $3d$  states but rather of the  $4s$  ones. This is not what one expects from the hydrogen atom where first the  $3d$  orbitals would be filled. The reason for this unexpected anomaly is the non-zero wave amplitude of the  $s$  states at the nucleus. The screening action of the other outer electrons is, thus, reduced and the energy of the  $s$  orbitals is lowered in comparison with the hydrogen model.

Subsequent filling of the  $d$  orbitals generates the  $3d$  transition metals from Sc to Zn. In analogy, there are  $4d$  and  $5d$  transition metals because of initial filling of the  $s$  states each time. The same phenomenon for the  $f$  states is responsible for the existence of the rear earth series Ce to Lu and the actinides Th to Lr.

Quantum mechanics explains the chemical reactivity and all other properties of chemical elements which is reflected in the periodic table having been developed more than hundred years ago solely on the basis of observation.

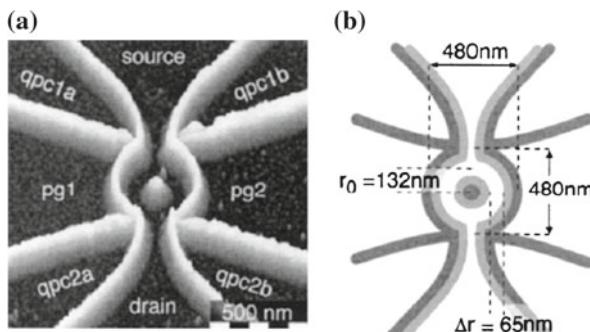
### 5.7.3 Quantum Rings

Instead of using a quantum dot for single electron tunneling spectroscopy (Sect. 5.7.1) we can also study a nanoscopic ring structure. Electrons, then, tunnel through a barrier at one side into the ring (source contact) and leave the ring at the opposite side again through a tunnel barrier (drain contact). The ring as a whole behaves as a quantum dot with respect to charging by addition of an electron from the source. The addition energy  $\Delta\mu = (e^2/C) + \Delta E$  contains as before the charging energy  $e^2/C$  (many-particle repulsion) and the energy difference  $\Delta E$  between the lowest

empty and the highest occupied electronic level in the quantum ring. Single electron tunneling in the Coulomb blockade regime through this ring structure, thus, allows again a spectroscopy of the single electron states in the ring by measuring the  $\Delta E$  term. As in other cases of electrons orbiting in a ring structure, for example, in the Aharonov–Bohm effect (Sect. 5.4.4), measurements in an external magnetic field are of interest.

Such experiments have been performed on quantum rings which were prepared by split-gate technique (Appendix B) in a two-dimensional electron gas (2DEG) at the interface of an AlGaAs/GaAs heterostructure (Appendix A) [17]. The 2DEG positioned about 34 nm below the surface of the layer sandwich was structured laterally by means of local oxidation in a scanning force microscope (Appendix B). Below the oxide traces (thickness in the nanometer range), the metallic 2DEG becomes depleted from electrons (depletion space charge, Appendix A) and the 2DEG is split up into distinct conducting areas separated by insulating barriers (Fig. 5.21). Source and drain contacts are prepared in this way as well as the quantum ring by means of a central dot-like and two ring-like depleted traces (Fig. 5.21). In addition, two side contacts ( $pg1$  and  $pg2$ ) separated from the ring by insulating barriers allow potential variations of the ring as a whole against source by application of a gate voltage. The four contacts  $qpc\ 1a, 2a, 1b, 2b$  arranged rectangular with respect to each other enable a precise control of the quantum point contacts at the entrance to and the exit out of the ring. Adequate applied voltages allow a convenient adjustment of the tunnel barriers between the ring and the source and drain contacts.

Before the presentation of experimental results, we want to get some theoretical insight into the spectrum of electronic states expected for a quantum ring. The ring



**Fig. 5.21** **a, b** Quantum ring prepared by local stripe-wise oxidation of the surface of an AlAs/GaAs heterostructure carrying a high mobility 2D electron gas (2DEG) at the interface [17]. Local surface oxidation depletes the 2DEG below from electrons and creates potential barriers which separate conducting areas within the 2DEG from each other (Appendix B). **a** Scanning electron micrograph of the oxide structures which give rise to potential barriers for electrons within the 2DEG. Source, drain,  $pg1$  etc. are non-oxidized areas where the 2DEG below has full electrical conductance. **b** Schematic plot of the quantum ring with dimensions. *Dark lines* indicate the potential barriers for electrons; *grey lines* show the extension of depletion space charge zones at the borders of the barriers

has a radius of  $r_0 = 132$  nm and the potential  $V$  is assumed to be constant over the wire thickness of  $\Delta r \approx 65$  nm (in spite of parabolic potential in reality, Sect. 5.7.1). Because of the constant potential  $V = \text{const} = 0$  the kinetic energy of the electron is just that of a circular motion:  $E_{\text{kin}} = m^* v^2 / 2 = m^* r_0^2 \omega^2 / 2$ , where  $m^*$  is again, as in Sect. 5.7.1, the effective mass of the electron, i.e. the electronic mass modified by the action of the crystal lattice (Sect. 8.3.5). With  $\omega$  as the angular frequency of the electron its classical angular momentum is  $L_z = m^* r_0^2 \omega$ . We replace all classical quantities by operators and obtain by means of  $E_{\text{kin}} = (1/2)L_z^2/m^*r_0^2$  the following Hamilton operator for an electron moving on a circular orbit with radius  $r_0$ :

$$\hat{H} = \frac{1}{2} \frac{1}{m^* r_0^2} \hat{L}_z^2. \quad (5.199)$$

Because of  $[\hat{H}, \hat{L}] = 0$  the eigensolutions of the problem are identical with those of the angular momentum operator  $\hat{L}_z$ , that is,

$$\hat{L}_z |m\rangle = m\hbar|m\rangle, \quad (5.200)$$

$$\hat{H} |m\rangle = E_m |m\rangle = \frac{\hbar^2}{2m^* r_0^2} m^2. \quad (5.201)$$

Hereby, the directional quantum number assumes the values  $m = 0, \pm 1, \pm 2, \pm 3, \dots$  and the wave function follows as  $\langle r|m\rangle \propto \exp(im\varphi)$  according to (5.35a).

For the experiments, the quantum ring device of Fig. 5.21 was used. Measurements were performed at low temperature with a variable magnetic field oriented perpendicular to the ring ( $z$  direction). The eigensolutions  $E_m$  must, therefore, be calculated in the presence of a magnetic field  $\mathbf{B} = (0, 0, B) = \text{curl } \mathbf{A}$ . In this case, the Hamiltonian (only kinetic energy) is written as  $\hat{H} = (\mathbf{p} - e\mathbf{A})^2/2m^*$ . In analogy to the above case with  $\mathbf{B} = \mathbf{0}$  we express the energy by the angular momentum operator  $\hat{L}_z$  in order to represent the eigenstates in terms of angular momentum states  $|m\rangle$ . Given the magnetic field as  $\mathbf{B} = B\mathbf{e}_z$  the field lines of the vector potential  $\mathbf{A}$  enclose the  $\mathbf{B}$  field lines in the direction of the unity vector  $\mathbf{e}_\varphi$ , that is, the following ansatz for  $\mathbf{A}$  is adequate:

$$\mathbf{A} = \frac{1}{2} Br\mathbf{e}_\varphi. \quad (5.202)$$

The ansatz (5.202) is checked, that is,  $\mathbf{B} = B\mathbf{e}_z = \text{curl } \mathbf{A}$  is derived from (5.202), by using the relations (5.30a)–(5.32) and setting  $\vartheta = \pi/2$  for cylindrical coordinates.

With the mechanical momentum written as

$$\mathbf{p} = m^* \mathbf{v} = m^* r_0 \dot{\varphi} \mathbf{e}_\varphi = m^* r_0 \omega \mathbf{e}_\varphi, \quad (5.203)$$

we obtain the classical kinetic energy as

$$\begin{aligned} E_{\text{kin}} &= \frac{(\mathbf{p} - e\mathbf{A})^2}{2m^*} = \frac{1}{2m^*} \left( m^* r_0 \omega \mathbf{e}_\varphi - \frac{1}{2} e B r_0 \mathbf{e}_\varphi \right)^2 \\ &= \frac{1}{2m^* r_0^2} \left( m^* r_0^2 \omega - \frac{1}{2} e B r_0^2 \right)^2. \end{aligned} \quad (5.204a)$$

We replace the angular frequency  $\omega$  by the angular momentum  $L_z m^* r_0^2 \omega$  and express the classical quantities by operators ( $E_{\text{kin}} \rightarrow \hat{H}$ ) to obtain the corresponding Hamiltonian:

$$\hat{H} = \frac{1}{2m^* r_0^2} \left( \hat{L}_z - \frac{1}{2} e B r_0^2 \right)^2. \quad (5.204b)$$

Since  $B$  is a simple number rather than an operator,  $\hat{H}$  commutes with  $\hat{L}_z$ , i.e.  $[\hat{H}, \hat{L}_z] = 0$ , and the Schrödinger equation is solved again by eigenstates  $|m\rangle$  of the angular momentum operator  $\hat{L}_z$ :

$$\hat{H}|m\rangle = \frac{1}{2m^* r_0^2} \left( \hat{L}_z - \frac{1}{2} e r_0^2 B \right)^2 |m\rangle, \quad (5.205a)$$

$$\hat{H}|m\rangle = \frac{1}{2m^* r_0^2} \left( m\hbar - \frac{1}{2} e r_0^2 B \right)^2 |m\rangle = E_m |m\rangle. \quad (5.205b)$$

The energy eigenvalues  $E_m$  of the orbiting electron indexed by the directional quantum number  $m$  are obtained as

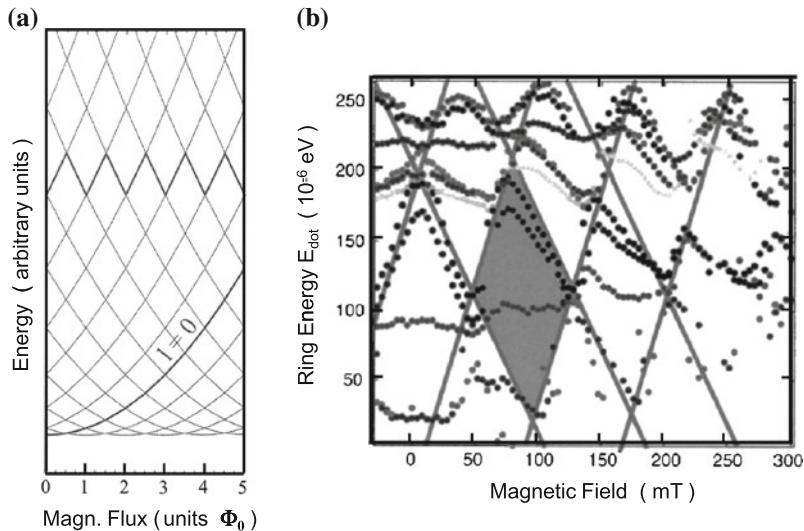
$$E_m = \frac{1}{2m^* r_0^2} \left( m\hbar - \frac{1}{2} e r_0^2 B \right)^2 = \frac{\hbar^2}{2m^* r_0^2} \left( m - \frac{1}{2} \frac{e}{\hbar} r_0^2 B \right)^2. \quad (5.206a)$$

The magnetic flux through the quantum ring is  $\Phi = \pi r_0^2 B$  and we can write

$$E_m = \frac{1}{2m^* r_0^2} \left( m - \frac{e}{\hbar} \Phi \right)^2 = \frac{\hbar^2}{2m^* r_0^2} (m - \Phi/\Phi_0)^2. \quad (5.206b)$$

Hereby,  $\Phi_0 = h/e$  is the magnetic flux quantum introduced in Sect. 5.4.4. For a fixed angular momentum (quantum number  $m$ ), the electron energies as a function of the magnetic field  $B$  or the magnetic flux  $\Phi$  form a parabola (5.206a), (5.206b). Different angular momenta with different quantum numbers  $m$  give rise to a series of parabolas shifted against each other by the flux quantum  $\Phi_0$  (Fig. 5.22a).

In a single electron tunneling experiment the electrons tunneling through the ring must occupy the energy levels (5.206b). When the magnetic field, which penetrates the ring, is varied during the measurement, the electron can not increase its energy to much by following one single parabola with fixed  $m$ . It will rather change from one to the next parabola ( $m \rightarrow m + 1$ ) in order to keep its energy at a minimum (solid zig-zag line in Fig. 5.22a). A corresponding zig-zag pattern should be found exper-



**Fig. 5.22** **a, b** Single electron energy levels of the quantum ring of Fig. 5.21 as function of magnetic field, respectively magnetic flux which penetrates the ring normally. **a** Calculated energy spectrum: the different parabolae shifted with respect to each other correspond to different angular momentum quantum numbers  $l$  ( $\Phi_0$  flux quantum). **b** Experimental spectrum of energy levels of the quantum ring measured by single electron tunneling [17]

imentally when a Coulomb blockade peak is measured as function of the magnetic field penetrating the quantum ring. This is indeed observed as shown in Fig. 5.22b [17]. Beside the zig-zag energy levels additional values are found which depend only weakly on the magnetic field or flux. A more profound theoretical analysis of the experimental data explains these findings by an asymmetry of the potential, that is, deviations from the ideal ring structure. This causes mixing of states with positive and negative angular momentum [17].

## References

1. R. Resnick, *Introduction to Special Relativity* (Wiley, New York, 2002), p.157
2. A. Tonomura, *The Quantum World Unveiled by Electron Waves* (World Scientific, Singapore, 1998)
3. Y. Aharonov, D. Bohm, Phys. Rev. **115**, 485 (1959)
4. J. Appenzeller, T. Schäpers, H. Hardtdegen, B. Lengeler, H. Lüth, Phys. Rev. **51**, 4336 (1995)
5. B. Krafft, A. Förster, A. van der Hart, T. Schäpers, Physica E **9**, 635 (2001)
6. I. Estermann, Recent research in molecular beams, in *A Collection of papers Dedicated to Otto Stern*, ed. by I. Estermann (Academic Press, New York, 1959)
7. H. Kopfermann, *Kernmomente*, 2nd edn. (Akademische Verlagsgesellschaft, Frankfurt, 1956)
8. T.E. Phipps, J.B. Taylor, Phys. Rev. **29**, 309 (1927)
9. M.A. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, E.A. Cornell, Science **269**, 198 (1995)

10. K.B. Davis, M.-O. Mewes, M.R. Andrews, M.J. Van Druten, D.S. Durfee, D.M. Kum, W. Ketterle, Phys. Rev. Lett. **75**, 3969 (1995)
11. C. Berger, *Elementarteilchenphysik—Von den Grundlagen zu den modernen Experimenten* (Springer, Berlin, 2006)
12. C.D. Anderson, Phys. Rev. **43**, 491 (1933)
13. P.W. Higgs, Phys. Lett. **12**, 132 (1964)
14. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
15. L.P. Kouwenhoven, D.G. Austing, S. Tarucha, Rep. Prog. Phys. **64**, 701 (2001)
16. H. Haken, H.C. Wolf, *Atom- und Quantenphysik* (Springer, Berlin, 1980), p. 153
17. A. Fuhrer, S. Lüscher, T. Ihn, T. Heinzel, K. Ensslin, W. Wegscheider, M. Bichler, Nature **412**, 822 (2001)
18. F. Schwabl, *Quantenmechanik*, 2nd edn. (Springer, Berlin, 1990), pp. 104–105

# Chapter 6

## Approximate Solutions for Important Model Systems

Only in the minority of cases exact solutions to a quantum mechanical problem can be obtained. Examples are tunneling of electrons through a rectangular barrier (Sect. 3.6.4), tunneling through double barriers (Sect. 3.6.5), electrons confined in a quantum well with infinitely high walls (Sect. 3.6.1) or an electron in a harmonic oscillator potential (Sect. 4.4.2).

Mostly the potentials in the Schrödinger equation do not allow simple analytical solutions for the dynamics of an electron, even more for more than one particle or for time dependent potentials. Even though modern supercomputers can solve nearly every problem in quantum physics quantitatively with any accuracy, the treatment of approximation methods for solving the Schrödinger equation is of high value. Approximation techniques often yield the algorithm for computer simulations and, furthermore, provide a deeper insight into the underlying physics of a particular complex problem, for example, the relative importance of certain quantities, their interrelation etc. Approximations in physics involve the neglect of certain quantities or functional dependencies in the calculation on the basis of intuitive arguments and estimations. This requires a profound and subtle characterization and analysis of the particular problem. Accordingly specific types of approximation methods have been established which meet particular issues and requirements of physical problems.

For nearly free particles moving in a weak potential, for example, the wave function of a free particle is modified, such that the effect of the weak potential is approximately accounted for.

There are different approximation methods for weakly perturbed stationary quantum states in a time independent potential (slightly different from an already solved problem) and states in a time dependent potential which are no more stationary but can be approximated by transitions between stationary states.

This category of issues also includes approximate solutions to scattering problems of particles. In this case, particles freely propagate in space and undergo a timely limited perturbation upon passing a spatially limited scattering target (potential). Hereby, the initial states of a free particle are transformed into new scattering states.

## 6.1 Particles in a Weakly Varying Potential: The WKB Method

In Sect. 3.6.4, we have considered particles which tunnel through an energy barrier. But also particles which have enough energy to propagate over the barrier are affected. Their wave number  $\mathbf{k}$ , their wavelength  $\lambda$  and their momentum  $\mathbf{p} = \hbar\mathbf{k}$ , respectively, are changed during their path over the barrier.

A free particle with energy  $E$  moving in a constant potential  $V$  along the  $x$  axis is described by the wave function (plane wave):

$$\psi(x) = Ce^{\pm ipx/\hbar} = Ce^{\pm ikx}, \quad (6.1a)$$

where  $C$  is a normalization constant and where momentum  $p$ , respectively, wave number  $k = 2\pi/\lambda$  are given by

$$p = \hbar k = \sqrt{2m(E - V)}. \quad (6.1b)$$

Different potentials  $V$ , thus, shift the phase  $ikx$  of the plane wave by different amounts.

We now assume that the particle moves over a locally varying potential with only slight barrier height variations. The potential  $V(x)$  extending between the positions  $x_0$  and  $x$  can be thought as being composed of infinitesimally small regions of constant potential. Each potential region ( $i$ ), then, shifts the phase of the particle wave by a different amount  $\delta(px)_i$ . The total phase shift of the wave after the particle has passed the potential between  $x_0$  and  $x$  is the sum of all contributions  $\delta(px)_i$ . An approximate solution to the wave function after passage of the weak potential is, therefore a wave with a phase (exponent) which contains the sum of the infinitesimal phase shifts  $\delta(px)_i$ . In the limit, the phase sum  $(i/\hbar) \sum_j \delta(px)_j$  is replaced by an integral and we obtain the approximate wave function as

$$\psi(x) = \psi(x_0) \exp\left[\pm \frac{i}{\hbar} \int_{x_0}^x p(x') dx'\right]. \quad (6.2a)$$

In accord with the spatially varying potential  $V(x')$ , a spatially varying momentum  $p(x')$  respectively, wavelength  $\lambda(x')$  is attributed to the electron wave:

$$p(x') = \sqrt{2m[E - V(x')]} \quad (6.2b)$$

As usual, plus and minus sign of the phase in (6.2a) indicate waves moving to the right and to the left. The most general solution is a linear superposition of both waves.

A more subtle analysis is needed to clarify details of the approximation involved in (6.2a), (6.2b). We have assumed that the wave function remains plane wave like even though this is not true for arbitrary potentials. The plane wave character is

exactly given only for piece-wise constant potentials. In this context, we have to ask if the assumption of a spatially varying momentum, respectively wavelength is physically meaningful. A wavelength is attributed to a spatially extended wave train (packet), it can not be defined for a particular point on the  $x$  axis. The variable  $\lambda(x)$  is only meaningful, if variations  $\delta\lambda$  are negligibly small along the extension of a wavelength  $\lambda$ , i.e.

$$\left| \frac{\delta\lambda}{\lambda} \right| \approx \left| \frac{(\delta\lambda/dx) \cdot \lambda}{\lambda} \right| = \left| \frac{d\lambda}{dx} \right| \ll 1. \quad (6.3)$$

The present approximation method called *WKB approximation* [1, 6] according to their inventors Wentzel, Kramers, Brillouin is, thus, valid for extremely small position derivatives of the wavelength (6.3).

For a formal treatment of the approximation, we solve the time independent Schrödinger equation

$$\left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E \psi(x) \quad (6.4)$$

by the ansatz  $\psi(x) \propto \exp[i\varphi(x)]$ . This ansatz seems convenient because the phase of the wave contains the essential approximation (6.2a). Inserting the ansatz into the Schrödinger equation (6.4) yields

$$[\varphi'(x)]^2 - i\varphi''(x) = \frac{2m}{\hbar^2} [E - V(x)] = k^2(x) = \frac{i}{\hbar} p^2(x). \quad (6.5)$$

Because of the weak position dependence of the potential also wave number  $k$  and momentum  $p$  vary weakly as a function of position. Consequently, we neglect the curvature  $\varphi''(x)$  of the phase. Then, (6.5) is easily integrated:

$$\varphi(x) = \pm \int_{x_0}^x k(x') dx'. \quad (6.6)$$

This relation represents exactly the previously guessed phase integral in (6.2a). Furthermore, the approximation can also be described by the condition  $|\varphi''(x)| \ll |\varphi'(x)|^2$  and by means of the relation  $\varphi'(x) \approx k(x)$  [see (6.5)] we obtain the condition

$$\left| \frac{dk}{dx} \right| \ll k^2, \quad \left| \frac{1}{k} \frac{dk}{dx} \right| \ll |k|, \quad (6.7)$$

an analogous condition to (6.3).

By means of (6.6), equation (6.5) is rewritten into

$$[\varphi'(x)]^2 = k^2(x) + i\varphi''(x) \simeq k^2(x) \pm ik'(x). \quad (6.8a)$$

Taking the square root yields

$$\varphi'(x) \simeq \pm k(x) \sqrt{1 \pm i \frac{k'(x)}{k^2(x)}}. \quad (6.8b)$$

Series expansion of the square root gives

$$\varphi'(x) \simeq \pm k(x) \pm \frac{i k'(x)}{2 k(x)} \quad (6.8c)$$

and finally we obtain the phase angle (6.6) by integration of (6.8c) as

$$\varphi(x) \simeq \pm \int_{x_0}^x k(x') dx' \pm \frac{i}{2} \ln k(x). \quad (6.9)$$

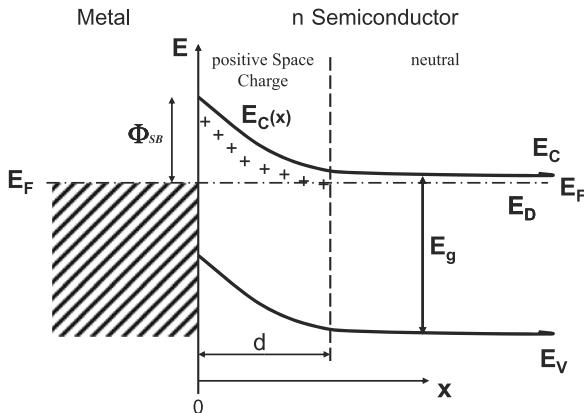
The wave function approximated within the frame of the WKB method, then, is obtained as

$$\psi(x) \simeq \frac{A}{\sqrt{k(x)}} \exp \left[ \pm i \int_{x_0}^x k(x') dx' \right]. \quad (6.10)$$

$A$  is a normalization constant. Apart from the  $1/\sqrt{k(x)}$  term the above wave function (6.2a) is obtained. This pre-factor guarantees current conservation when the particle wave passes the potential  $V(r)$  between  $x_0$  and  $x$ . According to (3.79), the current density of a plane wave is  $j = (\hbar k/m)|\varphi|^2$ . In this way, the WKB approximation ensures that the probability for finding a particle is lower in a region where the particle moves faster. Backscattering of particles from the weak potential (Sect. 3.6.3) is neglected within the approximation.

### 6.1.1 Application: Tunneling Through a Schottky Barrier

Metal-semiconductor junctions are found in every semiconductor device. Such contacts have commonly rectifying properties upon current flow. This phenomenon was already used in the early days of semiconductor electronics in connection with the detection of radio waves. The rectification effect is due to the formation of a so-called Schottky barrier (Fig. 6.1). Electronic interface states (Appendix A) at the metal-semiconductor junction fix the Fermi level (Sect. 5.6.3) at an energy  $\Phi_{SB}$  below the conduction band edge  $E_C$  ( $x = 0$ ) in the forbidden band (gap  $E_g$ ). The Fermi level  $E_F$  is said to be pinned at the interface at an energy  $\Phi_{SB}$  below  $E_C$ . The Schottky barrier height  $\Phi_{SB}$  is characteristic for the particular metal-semiconductor junction; it does neither depend on doping nor on temperature. For an  $n$ -doped semiconductor as in Fig. 6.1, the Fermi level  $E_F$  deep in the bulk ( $x > d$ ) lies, at low tempera-



**Fig. 6.1** Electronic band scheme of a metal/ $n$ -semiconductor junction with Schottky barrier. The metal is described by a potential well, the electronic states of which are occupied (one electron per state) up to the Fermi energy  $E_F$ . In thermal equilibrium  $E_F$  as chemical potential has the same value in the  $n$ -doped semiconductor as in the metal. Due to electronic interface states (Appendix A)  $E_F$  is pinned at the interface near midgap, while in the semiconductor bulk the  $n$ -doping causes  $E_F$  to be located slightly below the conduction band edge  $E_C$ . As a consequence the electronic bands in the semiconductor are bent upwards near the interface causing a depletion of free electrons within a depletion space charge zone of thickness  $d$ . Within this depletion zone the donor atoms are ionized and form a spatially fixed positive space charge.  $E_D$  energy of donors,  $E_V$  valence band edge,  $\Phi_{SB}$  Schottky barrier

ture, between the donor levels (energy  $E_D$ ) and the lower conduction band edge  $E_C$ , that is, close to the conduction band (<30 meV). Since however  $\Phi_{SB}$  amounts to about half the forbidden band energy (for GaAs:  $\Phi_{SB} \cong 0.7$  eV) the electronic bands are bent upwards near the junction. The donor levels slightly below  $E_C$  are emptied from electrons and a positive space charge arising from spatially fixed ionized donors results. The extension  $d$  of the space charge layer depends on the doping level of the material (Appendix A). Within this space charge layer, there are no free electrons in the conduction band anymore, it is called *depletion layer*. Its electrical resistance is high in comparison with the bulk semiconductor. Current flow from the metal into the semiconductor is suppressed. This is in particular true for a negative voltage applied to the metal since electrons then have to overcome the energy barrier  $\Phi_{SB}$  in order to enter the conduction band of the semiconductor. In this *reverse bias* polarity only a very small *reverse current* flows, even at relatively large voltages.

A negative voltage applied to the semiconductor side, on the other hand, shifts the Fermi level upwards in the semiconductor with respect to the metal. The lower conduction band edge is simultaneously lifted and reaches or exceeds the barrier energy  $\Phi_{SB}$ . More and more electrons flow from the semiconductor into the metal and we observe an exponential increase of the so-called *forward current*. In this polarity, the contact is highly conductive whereas under inverse bias the contact blocks the current. The metal-semiconductor junction is a rectifier.

For the calculation of the reverse current (polarity: metal negative) tunneling of electrons through the Schottky barrier (depletion zone) is an important factor beside thermal excitation over the barrier  $\Phi_{\text{SB}}$ . The tunneling contribution can easily be calculated by using the WKB approximation. The tunneling barrier has the height  $\Phi_{\text{SB}}$  and a thickness  $d$  (Fig. 6.1). In between  $x = 0$  and  $x = d$  the conduction band edge  $E_C$  decreases from  $\Phi_{\text{SB}}$  with curved shape down to about  $E_F$ . The functional dependence  $E_C(x)$  is calculated by double integration of Poisson's equation  $d^2V/dx^2 = -\rho/\epsilon\epsilon_0$  with an assumed constant charge density  $\rho = eN_D$  of the ionized donors (density  $N_D$ ) in the space charge region. Hereby a parabolic dependence of the conduction band edge is obtained in the depletion zone:

$$E_C(x) = \Phi_{\text{SB}}[1 - (x/d)^2]. \quad (6.11)$$

The thickness  $d$  of the depletion zone is obtained from the barrier height  $\Phi_{\text{SB}}$  by using the relation

$$\Phi_{\text{SB}} = e^2 N_D d^2 / 2\epsilon\epsilon_0. \quad (6.12)$$

By means of (6.11), the tunneling barrier is completely described mathematically.

For the calculation of the tunneling rate (probability), the probability to find an electron at  $x = d$  after passing the barrier has to be evaluated. For this purpose the wave function (6.2a), (6.2b), respectively (6.1a), (6.1b) has to be squared and we must bear in mind that in these equations the energies  $E$  are positive. In the derivation, the propagating electrons had energies  $E$  which exceed the maximum of the slightly varying potential  $V(r)$ . In the present case of tunneling electrons, however, the electron energies are lower than the potential maximum:  $E < V(x)$ . The real wave vector  $k(x)$  in (6.10) must, therefore, be replaced by an imaginary quantity  $\kappa(x) = \sqrt{2m[V(x) - E]}/\hbar$  (Sect. 3.6.4, (3.91a), (3.91b)). The transmission probability through the barrier is, then, obtained as

$$T = |\psi(x = d)|^2 \sim \exp\left[-2 \int_0^d \kappa(x) dx\right], \quad (6.13a)$$

respectively

$$T \sim \exp\left[-2 \int_0^d \left\{ \frac{2m\Phi_{\text{SB}}}{\hbar^2} \left(1 - \frac{x}{d}\right)^2 \right\}^{1/2} dx\right] = \exp\left(-d\sqrt{\frac{2m\Phi_{\text{SB}}}{\hbar^2}}\right). \quad (6.13b)$$

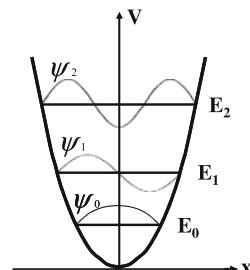
The tunneling probability decays exponentially over a length  $\sqrt{\hbar^2/2m\Phi_{\text{SB}}}$ . For GaAs with a dielectric constant  $\epsilon \approx 10$  and a Schottky barrier height  $\Phi_{\text{SB}} \approx 0.7$  eV, this decay length amounts to about 1 nm. In order to prepare well conducting metal contacts with quasi-ohmic behavior to GaAs, the thickness  $d$  of the Schottky barrier, that is, of the depletion space charge layer must not exceed a value of 1 nm. Such thin depletion layers, however, require extremely high bulk doping levels above

$10^{19} \text{ cm}^{-3}$  [see (6.12)]. Good quasi-ohmic metal-semiconductor contacts are, thus, commonly obtained by preparation of a highly doped layer in the semiconductor just below the metal contact using techniques as epitaxy (Appendix B), diffusion or ion implantation. The thin barriers (depletion zones), then, allow quite high tunneling currents through the contact and the exponential current-voltage characteristics appears linear in a small voltage range. This quasi-linearity offers approximately ohmic behavior (quasi-ohmic contact).

## 6.2 Clever Guess of a Wave Function: The Variational Method

The solution to quantum mechanical problems often benefits from physical intuition. Understanding the essential physical basis of a problem frequently leads to a rough, qualitative picture of the wave function by analogy to similar problems. In a binding 1D rectangular potential box of width  $d$ , e.g., the eigensolutions to the Schrödinger equation are standing sinus waves with  $d = \lambda/2, 2\lambda/2, 3\lambda/2, \dots$  (Sect. 3.6.1). The solutions to the Schrödinger equation with the binding parabolic potential of the harmonic oscillator are qualitatively similar (Sect. 4.4.2). In a binding potential with mirror symmetry with respect to a central plane we expect, in general, a ground state wave function which is convex, downwards open and symmetrical to the mirror plane of the potential (Fig. 6.2). In analogy to the rectangular quantum box potential, the first excited state, also in a general binding potential, will be similar to a standing wave with  $d = \lambda$ , that is, it will change its sign upon reflection on the symmetry plane of the potential (inversion symmetry). The next higher state will have again mirror symmetry with respect to the mirror plane of the potential. But in contrast to the ground state, this excited state exhibits two nodes where  $\psi = 0$  (Fig. 6.2). These statements about wave functions in a binding potential are qualitatively correct, but the exact solutions to a particular problem have to be obtained by solving the Schrödinger equation. We will see, however, that the intuitive knowledge about the solutions of a particular problem provides a pathway to a good quantitative approximate solution of the Schrödinger equation. The first step, hereby, is to guess a wave function based on qualitative similarities with already solved problems.

**Fig. 6.2** General binding potential  $V(x)$  with qualitatively plotted energy eigenfunctions  $\psi_0, \psi_1, \psi_2$  corresponding to the three lowest energies  $E_0, E_1, E_2$



In order to see how a clever guess of a wave function can finally lead to a good quantitative approximation of the problem, we assume we know the exact solution to the problem, that is, the corresponding Schrödinger equation is solved by the kets  $|\varphi_n\rangle$ , that is:

$$\hat{H}|\varphi_n\rangle = E_n|\varphi_n\rangle. \quad (6.14)$$

We now consider a general state  $|\psi\rangle$  which has been guessed as a possible preliminary, trial solution to the Schrödinger equation (6.14) based on qualitative arguments as discussed above. The energy expectation value for this state follows as

$$\langle\psi|\hat{H}|\psi\rangle = \sum_n \langle\psi|\hat{H}|\varphi_n\rangle\langle\varphi_n|\psi\rangle = \sum_n E_n \langle\psi|\varphi_n\rangle\langle\varphi_n|\psi\rangle. \quad (6.15)$$

In the calculation, the completeness condition (Sect. 4.3.3) for the orthonormal system of eigenkets  $|\varphi_n\rangle$  was used. Replacing the series of energy eigenvalues  $E_n$  by the lowest possible ground state energy  $E_0$  we can give a lower limit for (6.15):

$$\langle\psi|\hat{H}|\psi\rangle \geq \sum_n E_0 \langle\psi|\varphi_n\rangle\langle\varphi_n|\psi\rangle = E_0 \langle\psi|\psi\rangle. \quad (6.16a)$$

For an arbitrary guessed (non-normalized) state  $|\psi\rangle$  we, thus, obtain

$$E_0 \leq \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}. \quad (6.16b)$$

The exact energy eigenvalue  $E_0$  of the ground state is obviously smaller than the energy mean value  $\langle\psi|\hat{H}|\psi\rangle/\langle\psi|\psi\rangle$  for any guessed approximate trial wave function  $|\psi\rangle$ .

We discuss the estimation (6.16a), (6.16b) in some more detail by assuming the trial wave function

$$|\psi\rangle = |\varphi_n\rangle + |\delta\varphi\rangle \quad (6.17)$$

for the  $n$ th eigenstate with energy eigenvalue  $E_n$ . The ket  $|\delta\varphi\rangle$  describes a deviation of the trial state vector from the exact solution  $|\varphi_n\rangle$  of the Schrödinger equation. It is orthogonal to  $|\varphi_n\rangle$ , otherwise it could simply be taken into account by the normalization factor of the exact state vector. Equation (6.17) inserted into (6.16b) yields:

$$\begin{aligned} \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} &= \frac{(\langle\varphi_n| + \langle\delta\varphi|)\hat{H}(|\varphi_n\rangle + |\delta\varphi\rangle)}{(\langle\varphi_n| + \langle\delta\varphi|)(|\varphi_n\rangle + |\delta\varphi\rangle)} \\ &= \frac{E_n + \langle\delta\varphi|\hat{H}|\delta\varphi\rangle}{\langle\varphi_n|\varphi_n\rangle + \langle\delta\varphi|\delta\varphi\rangle} = E_n + O(|\delta\varphi|^2). \end{aligned} \quad (6.18)$$

While the guessed trial wave function  $|\psi\rangle$  deviates from the exact solution  $|\varphi_n\rangle$  linearly in  $|\delta\varphi\rangle$ , the error in the energy eigenvalue is quadratic in  $|\delta\varphi\rangle$ . Small errors in the guessed wave function or state vector yield approximate energy eigenvalues which are better than the trial functions. The error in energy is of the second order in the error in the state vector.

A numerical example shall demonstrate this general result. We guess a trial ground state  $|\psi\rangle$  of a particular problem which deviates by 10% of the next excited state  $|\varphi_1\rangle$  from the exact ground state  $|\varphi_0\rangle$ , that is, the guessed ground state wave function exhibits 10% contamination (admixture) of the next excited state:

$$|\psi\rangle = |\varphi_0\rangle + \frac{1}{10}|\varphi_1\rangle. \quad (6.19)$$

The trial wave function (6.19) is inserted into the energy functional (6.16b) and we obtain the approximate energy eigenvalue for the ground state as follows:

$$\begin{aligned} E[\psi] &= \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\langle\varphi_0| + \frac{1}{10}\langle\varphi_1|)\hat{H}(|\varphi_0\rangle + \frac{1}{10}|\varphi_1\rangle)}{\langle\varphi_0 + \frac{1}{10}\varphi_1|\varphi_0 + \frac{1}{10}\varphi_1\rangle} \\ &= \frac{\langle\varphi_0|\hat{H}|\varphi_0\rangle + \frac{1}{100}\langle\varphi_1|\hat{H}|\varphi_1\rangle}{1 + \frac{1}{100}} = \frac{E_0 + 0.01E_1}{1.01} \\ &\simeq 0.99E_0 + 0.01E_1. \end{aligned} \quad (6.20)$$

Although the guessed trial function contains 10% admixture of the wrong excited state wave function, the calculated energy eigenvalue deviates by only 1% of the first excited state energy  $E_1$  from the correct ground state energy  $E_0$ .

So far we have no mean to optimize the trial wave function, that is, to bring it as close as possible to the correct wave function. We can make the approximation method even more effective by improving the trial wave function in a minimization process for the energy functional (6.16b). For this purpose, we parameterize the trial wave function in a clever way by some variables  $\alpha, \beta, \dots$  which have the general features one expects of the true ground state ket. By standard minimization procedure these variables are determined numerically such that the functional  $E[\psi]$  reaches a minimum, that is, approaches the true energy eigenvalue as close as possible. For the ground state we, thus, guess the trial state  $|\psi_0(\alpha, \beta, \dots)\rangle$  which yields a representation of the energy functional  $E[\psi_0]$  as function of the free variables  $\alpha, \beta, \dots$ :

$$E(\alpha, \beta, \dots) = E[\psi_0(\alpha, \beta, \dots)] = \frac{\langle\psi_0(\alpha, \beta, \dots)|\hat{H}|\psi_0(\alpha, \beta, \dots)\rangle}{\langle\psi_0(\alpha, \beta, \dots)|\psi_0(\alpha, \beta, \dots)\rangle}. \quad (6.21a)$$

Minimization of (6.21a) with respect to the variables  $\alpha, \beta, \dots$  finally yields an even better approximation to  $E_0$  in comparison with the use of the simply guessed trial state  $|\varphi_0\rangle$ . The minimization condition for (6.21a) requires:

$$\frac{\partial E}{\partial \alpha} = \frac{\partial E}{\partial \beta} = \dots = 0. \quad (6.21b)$$

Equation (6.21b) determines the values  $\alpha_0, \beta_0, \dots$  at which the energy functional (6.21a) becomes a minimum, that is, the best approximation to the true value  $E_0$ . This optimum approximate energy is obtained by inserting  $\alpha_0, \beta_0, \dots$  into  $E(\alpha, \beta, \dots)$  (6.21a):

$$E_0 \lesssim E(\alpha_0, \beta_0, \dots). \quad (6.21c)$$

The approximation procedure described here for the ground state is easily transferred to higher excited states of the considered system. The first excited state  $|\varphi_1\rangle$  with its true energy eigenvalue  $E_1$  is orthogonal to the ground state  $|\varphi_0\rangle$ . For the calculation of an approximate energy  $E_1$  one, therefore, has to guess a trial excited state  $|\psi_1(\alpha', \beta', \dots)\rangle$  with adjustable variables  $\alpha', \beta', \dots$  which is orthogonal to  $|\varphi_0\rangle$  or its approximation. The functional

$$E(\alpha', \beta', \dots) = \frac{\langle \psi_1(\alpha', \beta', \dots) | \hat{H} | \psi_1(\alpha', \beta', \dots) \rangle}{\langle \psi_1(\alpha', \beta', \dots) | \psi_1(\alpha', \beta', \dots) \rangle} \quad (6.22a)$$

is calculated and minimized by differentiation with respect to  $\alpha', \beta', \dots$ , analogously to (6.21b). The fixed values  $\alpha_1, \beta_1, \dots$  obtained by differentiation of the energy functional (6.22a) and solving equations as (6.21b) yield the optimum approximation for the energy of the first excited state:

$$E_1 \lesssim E(\alpha_1, \beta_1, \dots). \quad (6.22b)$$

In analogy also approximate values for higher energy eigenvalues  $E_2, E_3, \dots$  are calculated. For the calculation of  $E_2$ , of course, the guessed trial wave function for  $|\varphi_2\rangle$  must be orthogonal both to  $|\varphi_0\rangle$  and  $|\varphi_1\rangle$ .

In conclusion, the described variational method is based on the fact that the eigen-solutions (states) of the Hamilton operator  $\hat{H}$  are stationary points of the energy functional  $E[\psi]$ . A change of these kets in first order does not change the functional  $E[\psi]$ . The ground state, furthermore, is the absolute minimum of the functional.

In the following, we present some instructive examples for approximate solutions of the Schrödinger equation.

### 6.2.1 Example of the Harmonic Oscillator

According to Sect. 4.4.2 the problem of the harmonic oscillator can be solved analytically. In a parabolic binding potential ( $\propto x^2$ ), the eigenfunctions of the oscillator Hamiltonian exhibit the described position dependence (Fig. 6.2): the ground state convex and downwards open, without any node (Fig. 4.4,  $n = 0$ ), the first excited

state with one node and inversion symmetry around  $x = 0$  (Fig. 4.4,  $n = 1$ ), all higher excited states with an increasing number of nodes and alternatively with mirror and inversion symmetry around  $x = 0$  (Fig. 4.4).

We will now approximately calculate the ground state energy of the oscillator by means of the variational method even though we know the exact value being  $\hbar\omega/2$  with  $\omega$  as the oscillator frequency (4.122). Having the qualitative shape of the ground state wave function in mind we guess, as a trial wave function, a negative parabola opening downwards (Fig. 6.3):

$$\psi_0 = A \left[ 1 - \left( \frac{x}{a} \right)^2 \right] \propto (1 - \rho^2). \quad (6.23)$$

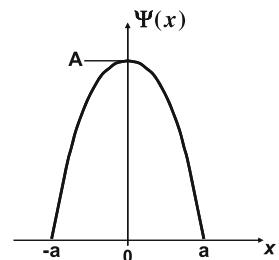
For the minimization of the energy functional we have introduced the free variables  $a$  and  $A$ , where  $A$  is only a normalizing constant. It is already taken into account in the calculation of (6.21a) by the denominator, that is, only one free parameter  $a$  is used for the minimization. With  $\rho = x/a$  the approximate Hamiltonian (energy functional) is obtained as:

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{m}{2} \omega^2 x^2 = -\frac{\hbar^2}{2ma^2} \frac{d^2}{d\rho^2} + \frac{m\omega^2 a^2}{2} \rho^2. \quad (6.24)$$

The numerator  $Z$  of the energy functional (6.21a) is, then, obtained as

$$\begin{aligned} Z &= \langle \psi_0 | \hat{H} | \psi_0 \rangle \\ &= -\frac{\hbar^2}{2m} \int_{-1}^1 (1 - \rho^2) \frac{d^2}{d\rho^2} (1 - \rho^2) a d\rho \\ &\quad + \frac{m\omega^2}{2} \int_{-1}^1 a^2 \rho^2 (1 - \rho^2) a d\rho \\ &= \frac{4}{3} \frac{\hbar^2}{ma} + \frac{16}{105} \frac{m\omega^2 a^3}{2}, \end{aligned} \quad (6.25a)$$

**Fig. 6.3** Downward open parabola as a rough approximation (trial function for variational method) for the ground state wave function of the harmonic oscillator



while the denominator follows as

$$N = \langle \psi_0 | \psi_0 \rangle = a \int_{-1}^1 (1 - \rho^2)^2 d\rho = \frac{16}{15}a. \quad (6.25b)$$

The energy functional (6.21a), thus, results as

$$E[\psi_0(a)] = \frac{Z}{N} = \frac{5}{4} \frac{\hbar^2}{ma^2} + \frac{1}{14}m\omega^2a^2. \quad (6.26)$$

It is minimized by means of the parameter  $a$ :

$$\frac{\partial E}{\partial a} = -\frac{5}{2} \frac{\hbar^2}{ma^4} + \frac{1}{7}m\omega^2 = 0. \quad (6.27a)$$

Resolving (6.27a) for  $a$  yields:

$$a^4 = \frac{35}{2} \frac{\hbar^2}{m^2\omega^2} \quad \text{respectively} \quad a^2 \approx 4.18 \frac{\hbar}{m\omega}. \quad (6.27b)$$

Inserting this expression into (6.26) finally leads to the minimum value of the energy functional

$$E_{\min} \approx 0.597\hbar\omega. \quad (6.27c)$$

That is the best approximation for the ground state energy  $\hbar\omega/2$  of the harmonic oscillator, of course, within the limitations of the particular trial wave function.

Taking into account the quite rough assumptions about the trial function  $\psi_0$  and the simple mathematics the approximate value (6.27c) for  $\hbar\omega/2$  is astonishingly good. Note that the approximate value exceeds the true value a little bit as is expected from (6.21c).

In order to obtain an approximation for the next higher excited state energy  $E_1$ , we must assume a trial function  $\psi_1$  with one node, inversion symmetry around  $x = 0$  and being orthogonal to (6.23). A simple assumption is certainly:

$$\psi_1(b, A) = A \sin \frac{\pi}{b} x, \quad \psi_1 = 0 \quad \text{for } |x| > b. \quad (6.28)$$

This function, correct eigensolution to the rectangular quantum box, fulfills the requirements. With  $b$  as adjustable variable for the minimization of the energy functional (6.22a), a calculation analogous to (6.24)–(6.26) and (6.27a)–(6.27c) yields an approximate energy value for the first excited state of the oscillator ( $= E_1 = 3\hbar\omega/2$ ). The calculation is left to the reader as an exercise.

### 6.2.2 The Ground State of the Hydrogen Atom

The hydrogen (H) atom is the simplest atom. Its most frequent isotope has one proton as nucleus, which binds one electron in the shell by Coulomb attraction. This is in principle a two-body problem. But the proton is about 2000-times heavier than the electron, i.e. for the description of the electron motion the proton can be assumed as static in good approximation. The dynamics of the electron moving around the central nucleus is a problem with spherical symmetry. In analogy to (5.183) and (5.5b) we can, therefore divide the electronic total kinetic energy into a radial part  $\hat{T}_r$  and an orbital contribution  $\hat{L}^2/2m_e r^2$  with  $m_e$  as the electronic mass. Unlike (5.183), where the problem has rotational symmetry (around an axis) our present problem has spherical symmetry with all possible orientations of the angular momentum in space. Instead of  $\hat{L}_z^2$ , therefore, the total angular momentum  $\hat{L}^2$  enters the hydrogen problem. With the Coulomb potential between electron and proton

$$V(r) = \frac{-e^2}{4\pi\epsilon_0 r} \quad (6.29)$$

the Hamilton operator for the electron in the hydrogen atom is obtained as:

$$\hat{H} = \hat{T}_r + \frac{\hat{L}^2}{2m_e r^2} - \frac{e^2}{4\pi\epsilon_0 r}. \quad (6.30)$$

Because of the kinetic energy  $\hat{p}^2/2m_e$  with  $\hat{p} = -i\hbar\nabla$  as momentum operator the radial part  $\hat{T}_r$  of the kinetic energy contains a double differentiation with respect to the radius. A detailed calculation by means of the vector representation (5.30a)–(5.30c), (5.31a)–(5.31c) and (5.32) for spherical coordinates yields (see Problem 4.7):

$$\hat{T}_r = -\frac{\hbar^2}{2m_e} \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right). \quad (6.31)$$

The calculation is not performed here, since the representation of the  $\nabla$  and  $\Delta$  operators in spherical coordinates is given in every textbook of mathematics.

Because of the spherical symmetry of the problem the operators  $\hat{H}$ ,  $\hat{L}^2$ ,  $\hat{L}_z$  commute with each other according to Sect. 5.3; they have the same system of eigenfunctions. The electron wave functions of the hydrogen atom, thus, contain, beside a radial part  $R_{n,l}(r)$ , the spherical harmonics  $\gamma_l^m$  as eigensolutions of the angular momentum operator. The principal quantum number  $n$  numerates the different discrete energy levels of the binding Coulomb potential.

According to (5.28a), the action of the  $\hat{L}^2$  operator on  $\gamma_l^m$  yields  $l(l+1)\hbar^2$  and we obtain the following Schrödinger equation for the electron in the hydrogen atom:

$$\begin{aligned}\hat{H}R_{n,l}\Upsilon_l^m &= \left[ -\frac{\hbar^2}{2m_e} \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{l(l+1)\hbar^2}{2m_e} - \frac{e^2}{4\pi\varepsilon_0 r} \right] R_{nl}\Upsilon_l^m \\ &= E_{n,l,m} R_{nl}\Upsilon_l^m.\end{aligned}\quad (6.32)$$

In this context, we do not discuss the exact eigensolutions in detail (Sect. 5.7.2) but rather determine the ground state energy  $E_0$  approximately by means of the variational method. The ground state has certainly zero angular momentum ( $l = 0$ ), i.e. with the ground state wave function  $\Psi_0 = R_{10}\Upsilon_0^0$  the Schrödinger equation (for the radial part) is written as

$$\hat{H}_r\Psi_0 = \left[ -\frac{\hbar^2}{2m_e} \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) - \frac{e^2}{4\pi\varepsilon_0 r} \right] \Psi_0 = E_0 \Psi_0. \quad (6.33)$$

Because of  $l = 0$  only the radial part of the problem with spherical symmetry must be treated, that is, in the functional (6.16b) the integration over the solid angle can be replaced by a factor  $4\pi \cdot r^2$ . The ground state energy  $E_0$  is then estimated by

$$E_0 \lesssim E[\Psi] = \frac{\int_0^\infty 4\pi r^2 \psi^* \hat{H}_r \psi \, dr}{\int_0^\infty 4\pi r^2 \psi^* \psi \, dr}. \quad (6.34)$$

In the functional  $E[\Psi]$ , a trial wave function is assumed which has spherical symmetry and describes the localization of an electron near the positive nucleus. Outside a certain radial distance  $\rho = 1/\sqrt{a}$  it must vanish. These conditions are certainly fulfilled by a Gauss function which leads to the trial wave function

$$\psi = A \exp(-ar^2), \quad (6.35)$$

$A$  is a normalization constant which leaves  $a$  as a free parameter for the minimization for the energy functional (6.34). Using common integration tables we, thus, obtain the denominator of (6.34) as

$$N = \langle \psi | \psi \rangle = \int_0^\infty 4\pi r^2 e^{-2ar^2} \, dr = 2\pi \int_{-\infty}^\infty r^2 e^{-2ar^2} \, dr = \frac{\pi}{2a} \sqrt{\frac{\pi}{2a}} = \left(\frac{\pi}{2a}\right)^{3/2}. \quad (6.36)$$

The numerator of the functional (6.34) follows as

$$Z = -\frac{4\pi\hbar^2}{2m_e} \int_0^\infty e^{-ar^2} \left( \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} \right) e^{-ar^2} \, dr - \frac{e^2}{\varepsilon_0} \int_0^\infty r e^{-2ar^2} \, dr. \quad (6.37a)$$

After differentiation in the integrand and extension of the integral to minus infinity the numerator is

$$Z = -\frac{6\pi a \hbar^2}{m_e} \int_{-\infty}^{\infty} r^2 e^{-2ar^2} dr - \frac{4\pi a^2 \hbar^2}{m_e} \int_{-\infty}^{\infty} r^4 e^{-2ar^2} dr + \frac{e^2}{\varepsilon_0} \int_0^{\infty} r e^{-2ar^2} dr \\ = \frac{3}{4} \pi^{3/2} \frac{\hbar^2}{m_e} (2a)^{-\frac{1}{2}} - \frac{e^2}{4\pi \varepsilon_0 a}. \quad (6.37b)$$

The energy functional, then, is obtained as

$$E[\psi(a)] = \frac{3}{4} \frac{\hbar^2 (2a)}{m_e} - \frac{e^2 (2a)^{1/2}}{2\varepsilon_0 \pi^{3/2}}. \quad (6.38)$$

The best approximation for the hydrogen ground state, of course within the limitations given by the particular trial function (6.35), follows by minimization of the functional (6.38) with respect to  $a$ :

$$\frac{\partial E}{\partial a} = \frac{6}{4} \frac{\hbar^2}{m_e} - \frac{1}{2} \frac{e^2}{\sqrt{2\varepsilon_0 \pi^{3/2}}} a^{-\frac{1}{2}} = 0. \quad (6.39a)$$

This finally yields:

$$\sqrt{a} = \frac{m_e e^2}{3\sqrt{2\varepsilon_0 \hbar^2 \pi^{3/2}}}, \quad (6.39b)$$

$$a = \frac{m_e^2 e^4}{18\varepsilon_0^2 \pi^3 \hbar^4} [\text{cm}^{-2}]. \quad (6.39c)$$

As  $a$  is positive, the derivative  $\partial E / \partial a$  is also positive:

$$\frac{\partial E}{\partial a} = \frac{e^2}{2\sqrt{2\varepsilon_0 \pi^{3/2}} a^{3/2}} > 0. \quad (6.40)$$

We have indeed found a minimum of the energy functional.

With  $a$  inserted in (6.38) the approximate value for the hydrogen ground state energy is obtained as

$$E_0 \lesssim E = -\frac{1}{12} \frac{m_e e^4}{\pi^3 \varepsilon_0^2 \hbar^2} \simeq -12 \text{ eV}. \quad (6.41)$$

The correct value obtained by analytical solution of the Schrödinger equation and also from experiment amounts to  $-13.6$  eV. As expected, the approximate value is a little bit higher but close to the correct value.

The present approximation also yields some information about the spatial extension of the ground state wave function, that is, the diameter of the hydrogen atom. According to the ansatz (6.35), the half-width of the Gaussian wave function is given by  $1/\sqrt{a}$ . Using (6.39b), this half-width (diameter of H-atom) is calculated to be  $10^{-8} \text{ cm} = 0.1 \text{ nm} = 1 \text{ \AA}$ . This value matches very well the Bohr radius

$a_0 \cong 0.05$  nm of the ground state electron orbit which is calculated semi-classically within the Bohr model of the hydrogen atom.

### 6.2.3 Molecules and Coupled Quantum Dots

In this section, the variational method is applied to a general class of problems in physics which cover the field of covalent bonding in molecules, coupled quantum dots in nano-electronics and the realization of quantum-bits (Q-bit, Sect. 7.5), the information unit in quantum information science.

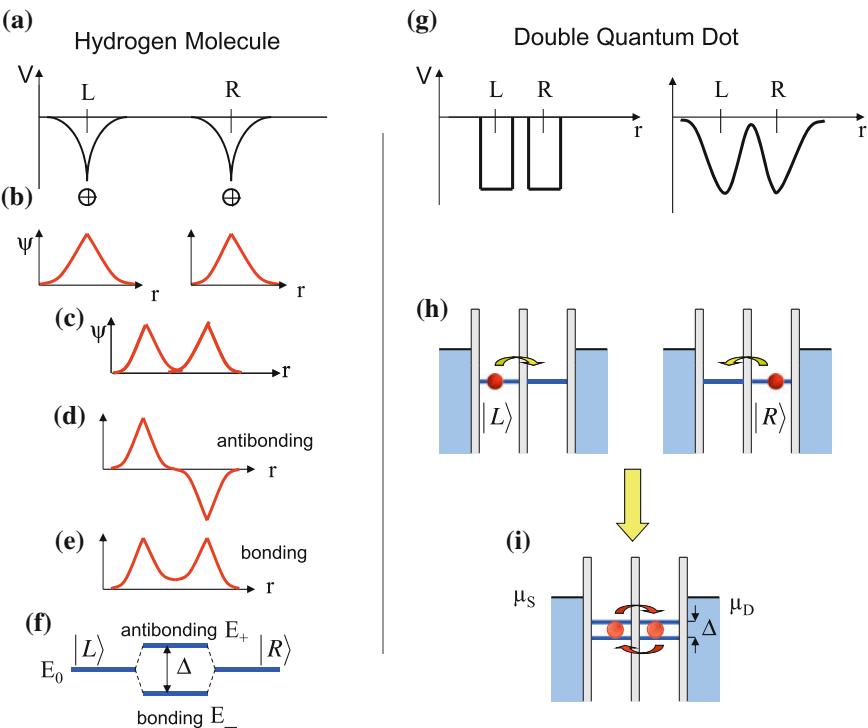
The simplest molecule in nature is certainly the hydrogen molecule ( $H_2$ ) where the two shell electrons of the two H atoms cause the covalent chemical bond. There is an even simpler molecule, namely the  $H_2^+$  ion, in which only one electron binds two positive protons in a covalent bond. The situation is shown schematically in Sect. 6.4. At sufficiently large distance, both protons create two separate Coulomb potentials  $-e/4\pi\varepsilon_0 r$  in their surrounding, which can bind the electron in discrete quantum states. The ground states  $|L\rangle$  and  $|R\rangle$  in the left ( $L$ ) and in the right ( $R$ ) atomic Coulomb potential, respectively, have wave functions  $\psi$  which decay exponentially from each proton (at large distance) with a functional dependence  $\exp(-r/a)$  (Sect. 6.2.2, Fig. 6.4b).

An analogous situation is given for two quantum dots, which might be prepared on a semiconductor chip, laterally by the split-gate technique or vertically by means of mesa lithography (Appendix B). The binding potentials in the two quantum dots (Fig. 6.4g), again, create discrete states in both dots, with ground states  $|L\rangle$  and  $|R\rangle$  left and right. The shape of the binding potentials determines the detailed functional dependence of the wave functions and the energies of the states. In rectangular box potentials, e.g., the ground state wave functions are standing half sine waves (Sect. 3.6.1).

In both cases, the two Coulomb potentials of the protons and the binding potentials of neighbouring quantum dots, the ground state wave functions  $|L\rangle$  and  $|R\rangle$  start to overlap at sufficiently small distances of the protons and the quantum dots, respectively. At small distances the energetic barrier between the two potential wells becomes smaller and smaller and the electron can tunnel between the potentials left and right. As in resonant tunneling (Sect. 3.6.5), one coherent quantum state extending over the two neighbouring protons or quantum boxes is formed. Or in other words, the electron in the  $H_2^+$  ion or the two quantum dots can be found simultaneously both in  $|L\rangle$  and  $|R\rangle$ , the ground states of the decoupled systems.  $|L\rangle$  and  $|R\rangle$  separately solve the Schrödinger equation of the decoupled systems:

$$(\hat{T} + \hat{V}_L)|L\rangle = E_L|L\rangle \quad \text{and} \quad (\hat{T} + \hat{V}_R)|R\rangle = E_R|R\rangle. \quad (6.42)$$

For the approximate solution of the problem by means of the variational method, the linear superposition of the states  $|L\rangle$  and  $|R\rangle$  is assumed as trial wave function. This ansatz corresponds well with the fact that the electron is present in the left or



**Fig. 6.4** a–i Comparison of an  $\text{H}_2^+$  ion with two coupled quantum dots being occupied by one single electron. **a** Potentials of the two protons in the  $\text{H}_2^+$  ion. **b** Qualitative ground state wave functions of an electron in the separate proton potentials. **c** Overlapping wave functions upon approaching of the two protons. **d** Antibonding molecular state (orbital) of the electron in the potential of the coupled atomic nuclei. **e** Bonding molecular state (orbital) with enhanced localization probability the electron between the two nuclei. **f** Splitting of the ground state energies of the free atoms into bonding and antibonding energy levels of the coupled atoms. **g** Two different bonding potentials for the double quantum dot. The similarity with the  $\text{H}_2^+$  potential in (a) is evident. Correspondingly wave functions and energy levels are qualitatively similar with those of the  $\text{H}_2^+$  ion [plots (b)–(f)]. **h** Schematic representation of the interaction between the potentials of the double quantum dot by back and forth tunneling of the electron between the states  $|L\rangle$  and  $|R\rangle$  left and right. **i** Schematic representation of the coupled electronic states with energies split into a bonding and an antibonding level.  $\mu_S$  and  $\mu_D$  are the chemical potentials (Fermi energies) of two metallic contacts (source and drain), left and right, which allow the measurement of the dot energies by single electron tunneling (see also Fig. 5.18)

right binding potential (“whether, or”):

$$|\psi\rangle \simeq c_L|L\rangle + c_R|R\rangle. \quad (6.43)$$

This trial wave function for an electron in the two-center system of two coupled binding potentials describes the partial localization of the electron in the left ( $L$ ) and the right ( $R$ ) potential. The amplitudes  $c_L$  and  $c_R$ , assumed as real numbers in the

simplest case, can be used as adjustable variables for the minimization of the energy functional  $E[\psi(c_L, c_R)]$  (6.21a).

The problem to be solved shall be described briefly again. The Schrödinger equation

$$\hat{H}|\psi\rangle = E|\psi\rangle \quad (6.44a)$$

of the problem contains the kinetic energy  $\hat{T}$  of the electron and the two binding potentials  $V_L(r - r_L)$  and  $V_R(r - r_R)$  left and right. In case of the  $H_2^+$  molecule, both positive nuclei move freely in space and repulse each other. Accordingly the Coulomb repulsion, energy must be taken into account in the Hamiltonian, too:

$$\hat{H} = \hat{T} + \hat{V}_L(r - r_L) + \hat{V}_R(r - r_R) + \frac{e^2}{4\pi\varepsilon_0|r_L - r_R|}. \quad (6.44b)$$

In case of the coupled quantum dots, the two binding potentials are fixed in space at the constant coordinates  $r_L$  and  $r_R$  and the Coulomb term in (6.44b) does not exist or is constant (energy scale normalization).

For the following calculation, it does not matter if we focus onto the  $H_2^+$  molecule or the two coupled quantum dots. The energy functional (6.21a) is calculated by means of (6.43) as follows:

$$\begin{aligned} E[\psi(c_L, c_R)] &= \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{(c_L\langle L| + c_R\langle R|)\hat{H}(c_L|L\rangle + c_R|R\rangle)}{(c_L\langle L| + c_R\langle R|)(c_L|L\rangle + c_R|R\rangle)} \\ &= \frac{c_L^2 H_{LL} + c_R^2 H_{RR} + c_L c_R H_{LR} + c_L c_R H_{RL}}{c_L^2 + c_R^2 + 2c_L c_R S}. \end{aligned} \quad (6.45)$$

Hereby, the following matrix elements are defined:

$$H_{LL} = \langle L|\hat{H}|L\rangle, \quad H_{RR} = \langle R|\hat{H}|R\rangle, \quad (6.46a)$$

$$H_{LR} = \langle L|\hat{H}|R\rangle, \quad (6.46b)$$

$$H_{RL} = \langle R|\hat{H}|L\rangle, \quad (6.46c)$$

$$S = \langle L|R\rangle = \langle R|L\rangle. \quad (6.46d)$$

The problem has been formulated, here, more generally than in the  $H_2^+$  case, namely with two unlike potential wells left and right. A realization might be two quantum dots with different dimensions.

We consider the matrix elements a little bit more in detail: The diagonal elements  $H_{LL}$  and  $H_{RR}$  of the quantum dot system can be represented as

$$H_{LL} = \langle L|\hat{T} + \hat{V}_L + \hat{V}_R|L\rangle = E_L + \langle L|\hat{V}_R|L\rangle = E_L - R_L, \quad (6.47a)$$

$$H_{RR} = \langle R|\hat{T} + \hat{V}_L + \hat{V}_R|R\rangle = E_R + \langle R|\hat{V}_L|R\rangle = E_R - L_R. \quad (6.47b)$$

$E_L$  and  $E_R$  are the ground state energies of the separate uncoupled quantum dots left and right, which are obtained from the Schrödinger equations (6.42).  $R_L$  and  $R_R$  are small corrections to the ground state energies arising from the presence of the neighbouring potential each time. Since the potentials have binding character,  $R_L$  and  $R_R$  have been introduced as positive quantities with negative sign. We define  $S = \langle L|R \rangle = \langle R|L \rangle$  as the so-called overlap integral, a relatively small quantity, which describes the spatial overlap of the two ground state wave functions in the region between  $r_L$  and  $r_R$ . The non-diagonal matrix elements  $H_{LR}$  and  $H_{RL}$  are then written as

$$H_{LR} = \langle L|\hat{T} + \hat{V}_L + \hat{V}_R|R \rangle = E_R \langle L|R \rangle + \langle L|\hat{V}_L|R \rangle, \quad (6.48a)$$

$$H_{RL} = \langle R|\hat{T} + \hat{V}_L + \hat{V}_R|L \rangle = E_L \langle R|L \rangle + \langle R|\hat{V}_R|L \rangle. \quad (6.48b)$$

The terms  $\langle L|\hat{V}_L|R \rangle$  and  $\langle R|\hat{V}_R|L \rangle$  couple the ground states left and right by means of the potentials  $\hat{V}_L$  respectively  $\hat{V}_R$ . The strength of these coupling matrix elements, of course, depends on the overlap of the three involved functions.

In the next section we will show that the matrix elements (6.48a), (6.48b) can be interpreted as probability transition amplitudes for electrons changing their quantum state from  $|R\rangle$  to  $|L\rangle$  respectively  $|L\rangle$  to  $|R\rangle$ . These electronic transitions occur under the action of the potentials  $\hat{V}_L$  (acting on  $|R\rangle$ ) and  $\hat{V}_R$  (acting on  $|L\rangle$ ). They describe in this sense transfer or transition amplitudes of an electron from the right to the left potential well and vice versa. Using the convenient notation  $-t_{\leftarrow}$  and  $-t_{\rightarrow}$  for the transition amplitudes (6.48a), (6.48b) are written as

$$H_{LR} = E_R S - t_{\leftarrow}, \quad (6.48c)$$

$$H_{RL} = E_L S - t_{\rightarrow}. \quad (6.48d)$$

For the further treatment of the energy functional (6.45), we keep in mind that the diagonal elements  $H_{LL}$  and  $H_{RR}$  are, apart from minor corrections  $R_L$  and  $R_R$ , the ground state energies  $E_L$ ,  $E_R$  in the left and the right potential. They are the leading terms in (6.45). In comparison, the non-diagonal elements  $H_{LR}$  and  $H_{RL}$  as well as the overlap integral  $S$  (6.46d) depend on the small overlap of the potentials and wave functions left and right, they are of minor importance in (6.45).

For the approximate calculation of the ground state energy of the two-center problem the energy functional  $E[\psi(c_L, c_R)]$  (6.45) is minimized with respect to the parameters  $c_L$ ,  $c_R$ , that is, we require

$$\frac{\partial E}{\partial c_L} = \frac{\partial E}{\partial c_R} = 0. \quad (6.49)$$

For simplicity reasons, we write (6.45) as

$$E(c_L^2 + c_R^2 + 2c_L c_R S) = c_R^2 H_{LL} + c_L^2 H_{RR} + c_L c_R H_{LR} + c_L c_R H_{RL} \quad (6.50)$$

and perform the differentiation (6.49) in this formula. Because of (6.49), the following relations are obtained:

$$c_L(H_{LL} - E) + c_R\left(\frac{H_{LR} + H_{RL}}{2} - ES\right) = 0, \quad (6.51a)$$

$$c_R(H_{RR} - E) + c_L\left(\frac{H_{LR} + H_{RL}}{2} - ES\right) = 0. \quad (6.51b)$$

For clearer representation of the calculation, the following definition of average values of the non-diagonal elements of the Hamiltonian are made:

$$\bar{H} = (H_{LL} + H_{RR})/2, \quad (6.52a)$$

$$\bar{h} = (H_{LR} + H_{RL})/2. \quad (6.52b)$$

From (6.51a), (6.51b), we then obtain the eigenvalue problem for the energy as

$$\begin{pmatrix} (H_{LL} - E) & (\bar{h} - ES) \\ (\bar{h} - ES) & (H_{RR} - E) \end{pmatrix} \begin{pmatrix} c_L \\ c_R \end{pmatrix} = 0. \quad (6.53a)$$

A non-trivial solution requires a vanishing determinant:

$$E^2 - E \frac{H_{LL} + H_{RR} - 2\bar{h}S}{1 - S^2} + \frac{H_{LL}H_{RR} - \bar{h}^2}{1 - S^2} = 0. \quad (6.53b)$$

As expected the two ground state energies  $E_L, E_R$ , respectively,  $H_{LL}, H_{RR}$  of the two separate potentials change into two new energy levels  $E_{\pm}$  as solutions to (6.53b) due to the coupling of the two systems:

$$E_{\pm} = \frac{\bar{H} - \bar{h}S}{1 - S^2} \pm \frac{1}{1 - S^2} \sqrt{(1 - S^2)(\bar{h}^2 - H_{LL}H_{RR}) + (\bar{H} - \bar{h}S)^2}. \quad (6.54)$$

We first consider the case of equal potential wells shifted against each other by  $(\mathbf{r}_L - \mathbf{r}_R)$ , that is, the two-center system with mirror symmetry. This case is given for the  $\text{H}_2^+$  molecule. Mirror symmetry is not easy to achieve for two coupled quantum dots because of technological limitations in the reproducibility.

With the symmetry requirements

$$\bar{H} = H_{LL} = H_{RR} \quad \text{and} \quad \bar{h} = H_{LR} = H_{RL}. \quad (6.55)$$

Equation (6.54) yields

$$E_{\pm} = \frac{1}{1 \pm S}(H_{LL} \pm H_{RR}) = \frac{1}{1 \pm S}(\bar{H} \pm \bar{h}). \quad (6.56a)$$

This can also be written as

$$E_{\pm} = \frac{1}{1 \pm S} [(E_0 - \delta) \pm (E_0 S - t_{LR})], \quad (6.56b)$$

when we use the representations (6.47a), (6.47b) and (6.48a)–(6.48d) of the diagonal and non-diagonal matrix elements together with the corrections  $R_L = L_R = \delta$  (because of symmetry) and the tunneling amplitudes  $t_{\leftarrow} = t_{\rightarrow} = t_{LR}$ . Because of equal potentials left and right the ground state energies are equal:  $E_L = E_R = E_0$ .

In conclusion, in the  $\text{H}_2^+$  molecule as well as in two coupled quantum dots an electron can occupy two quantum states with energies  $E_+$  and  $E_-$  which are coherently extended over the two-center system. Taking into account that  $\delta$  and  $t_{LR}$  are small in comparison to  $E_0$ , respectively,  $E_0 S$  with  $S$ , the overlap integral being significantly smaller than one, the energies  $E_{\pm}$  are concluded from (6.56a), (6.56b) to lie above and below the ground state energies  $E_0$ , respectively,  $H_{LL}$ ,  $H_{RR}$ , of the separate uncoupled potentials (Fig. 6.4f). For a molecule as  $\text{H}_2^+$  the energetically lower state with energy  $E_-$  is called the *bonding state* while the upper state with the energy  $E_+$  above the original ground state energy  $E_0$  is called the *anti-bonding state*. The lowering of the energy  $E_+$  of the bonding state with respect to the ground state energy  $E_0$  of the free uncoupled potential well favors an approach of the two positive protons and finally the formation of the covalent chemical bond in the  $\text{H}_2^+$  molecule. The formation of bonding and anti-bonding states is the origin of covalent chemical bonding. Anti-bonding states are excited states of a molecule which might lead to dissociation, if they are excited.

The deeper physical reason for the splitting of the ground state energy  $E_0$  into bonding and anti-bonding state energies  $E_-$  and  $E_+$  becomes obvious from a calculation of the eigenvector  $(c_L, c_R)$  (6.53a) by inserting the eigenvalues (6.56a), (6.56b) into (6.53a). In the present case of equal potentials left and right, the eigenvectors are obtained as  $(1, 1)$  and  $(1, -1)$ , that is, the normalized bonding and anti-bonding eigenstates of the coupled potential wells are represented as

$$|\psi_{\text{bond}}\rangle = \frac{1}{\sqrt{2}}(|L\rangle + |R\rangle), \quad (6.57a)$$

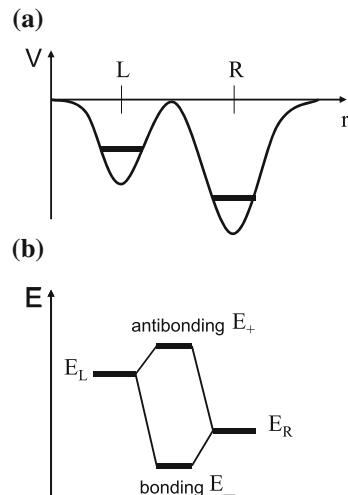
$$|\psi_{\text{antib}}\rangle = \frac{1}{\sqrt{2}}(|L\rangle - |R\rangle). \quad (6.57b)$$

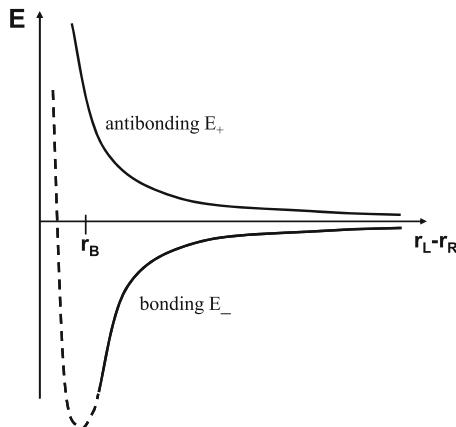
In the bonding state (6.57a), the two wave functions, left and right, are superimposed with positive sign; due to the overlap of the functions negative electronic charge is accumulated between the positive protons (Fig. 6.4e). In comparison to the uncoupled case (Fig. 6.4b) with negligible negative charge between the nuclei, the two protons are now bonded together by means of the negative charge in between (covalent bond). In contrast, in the anti-bonding state (6.57b) there is a wave function node between the two protons (Fig. 6.4d) and electronic charge is removed there in comparison to the decoupled case. The positive proton charge is less screened by electronic charge than in the case of two separate protons. Consequently the energy  $E_+$  exceeds  $E_0$ , the ground state energy of an electron in the separate decoupled proton potential.

For a two-center molecule with unlike partners (no mirror symmetry) as for example, CO, HF or NO or two unlike quantum dots (Fig. 6.5a) the general formula for the energy eigenvalues (6.54) must be discussed more in detail. The two electronic ground state energies  $H_{LL} = E_L - R_L$  and  $H_{RR} = E_R - L_R$  of the separate atoms or quantum dots are, of course, different. Considering the different orders of magnitude of the energy terms  $H_{LL}$ ,  $H_{RR}$ ,  $H_{LR}$ ,  $\bar{H}$ ,  $\hbar S$ , etc. the following result is obtained: The energy  $E_+$  of the anti-bonding state exceeds the energy  $H_{LL}$  or  $H_{RR}$  of the energetically highest state (Fig. 6.5b). Analogously the bonding state energy  $E_-$  lies below the energetically lowest energy  $H_{LL}$  or  $H_{RR}$  (Fig. 6.5b).

In molecules, the centers of the two coupled potentials (protons in  $H_2^+$  and  $H_2$ ) are mobile. In this case, it is interesting to calculate the energies  $E_+$  and  $E_-$  of the anti-bonding and bonding state as functions of the distance ( $r_L - r_R$ ) between the two atomic nuclei. Independent on details of the potentials (Coulomb potential for  $H_2^+$  or screened potentials for bigger atoms) an increase and a decrease of  $E_+$  and  $E_-$ , respectively, is qualitatively obtained as a function of decreasing atomic distance (Fig. 6.6). The energy of the bonding state, however, does not drop to infinitely negative values. In the described calculation the case of extremely close nuclei is not treated. For small distances ( $r_L - r_R$ ) the electron in the  $H_2^+$  becomes spatially more and more confined; because of the momentum/position uncertainty relation (Sect. 3.3) its kinetic energy must increase. An additional contribution to the energy increase is due to the Coulomb repulsion of the two approaching positive nuclei. Both energy contributions give rise to a strong increase of the bonding energy  $E_-$  for small nuclei distances (broken line in Fig. 6.6). The energy minimum between the dropping and increasing part of the  $E_-(r_L - r_R)$  curve determines the equilibrium distance  $r_B$  of the two atoms in the molecule (Fig. 6.6).

**Fig. 6.5** **a** Qualitative plot of the electronic potentials of two unequal coupled quantum dots (no mirror symmetry) or of a two-center molecule consisting of two unequal atoms. Additionally the ground state energy levels of an electron in the uncoupled potential wells are indicated. **b** Splitting of the ground state energies into bonding and anti-bonding levels due to the interaction





**Fig. 6.6** Qualitative dependence of the energies  $E_+$  and  $E_-$  of the antibonding and the bonding state on the nuclear distance  $r_L - r_R$  of the two atoms in a two-center molecule. The part of the bonding potential plotted in *broken line* can not be derived from a consideration of the bonding and antibonding states. The underlying physical reason for this strongly increasing potential branch is the strong spatial confinement of the valence electron at short distances below the bonding distance  $r_B$

The descriptive picture of an electron being localized alternatively once close to the left nucleus (dot) and then near the right one (dot), that is, of an electron which tunnels back and forth between the two potential wells, might also be quantified (Fig. 6.4h,i). For the simple symmetrical case of two equal potentials (6.56b) we calculate the difference between the energies of the bonding and anti-bonding state:

$$E_+ = E_0 - \frac{\delta}{1 + S} - \frac{t_{LR}}{1 + S}, \quad (6.58a)$$

$$E_- = E_0 - \frac{\delta}{1 - S} + \frac{t_{LR}}{1 - S}. \quad (6.58b)$$

In good approximation, we obtain

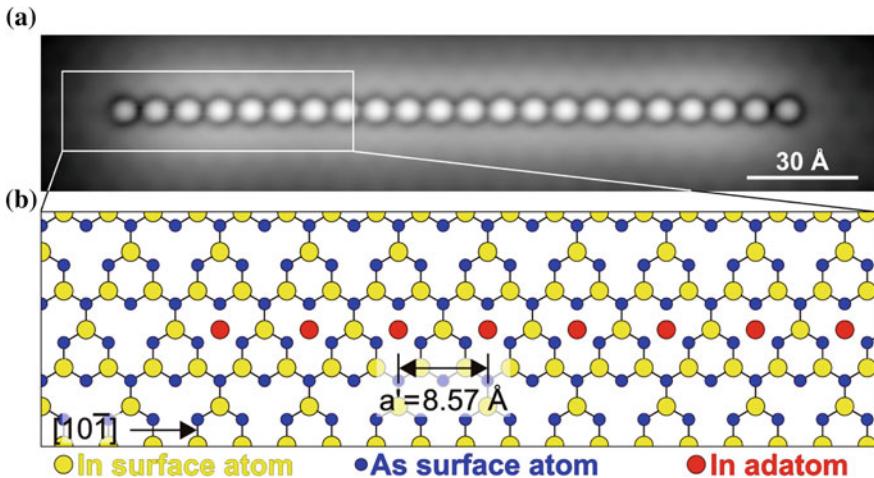
$$\hbar\omega = |E_+ - E_-| \simeq 2|t_{LR}|. \quad (6.58c)$$

The tunneling probability amplitude  $t_{LR} = \langle L | \hat{V}_L | R \rangle = \langle R | \hat{V}_R | L \rangle$ , thus, determines a frequency  $\omega$  with which the electron tunnels between the two potential wells back and forth (Fig. 3.19, Fig. 6.4i, and Fig. 7.1b).

### 6.2.4 Experimental Realisation of a Quantum Dot Molecule

An artificial two-atom molecule consisting of two quantum dots has been prepared experimentally and the two fundamental electronic states, the bonding one  $|\psi_{bond}\rangle$  and the antibonding one  $|\psi_{antib}\rangle$  (6.57a) as described in Sect. 6.2.3, have been observed directly by scanning electron tunnelling microscopy (STM, Sect. 3.6.4) [2]. Standard lithographic techniques (Appendix B) do not allow the preparation of two identical quantum dots because of inevitable variations in size and shape and, consequently, variability in their electronic wave functions and energies. Fölsch *et al.* [2], therefore, have prepared atomically precise quantum dots and couples of them by manipulating single In atoms on a well defined InAs(111)A surface using a scanning electron tunnelling microscope (STM) (Sect. 3.6.4). At a temperature of 5 K they could thus build up atomically well defined quantum dots in an ultra high vacuum (UHV: pressure below  $10^{-8}$  Pa) system. The manipulation process consists of lowering the STM tip over a previously adsorbed In atom, picking it up by the tip (assisted by the tip induced electric field) and then transferring it to a well defined adsorption site on the surface [3]. The latter back transfer process of the In atom to the surface is supported by short range adhesive forces between the surface and the In atom at the tip apex [3]. For the preparation of fully identical nanostructures performing step by step manipulation of single In atoms also well defined adsorption sites for the atoms must be available. This has been achieved in the present experiments by preparing a well defined InAs(111)A surface under UHV conditions using molecular beam epitaxy (MBE, Appendix B). By this technique a thin InAs layer with a thickness of 20 nm has been grown on an InAs substrate of suitable crystal orientation. Such a freshly grown InAs film is terminated by In atoms and a so-called  $(2 \times 2)$  superstructure, where the surface atomic periodicity in two directions is double that of the bulk atomic arrangement (Fig. 6.7b). Semiconductor surfaces usually exhibit atomic arrangements at the surface different from that in the bulk because of energetic reasons due to lacking bonding partners at the vacuum side [4]. The particular InAs surface with  $(2 \times 2)$  superstructure considered here has a well defined defect structure with regularly arranged vacancy sites (Fig. 6.7b). These vacancies are energetically favoured for the adsorption of the In atoms transferred there by the STM tip. This STM manipulation of In atoms is performed in a UHV system different from the MBE growth chamber. The freshly grown InAs film, therefore, had to be protected for transfer through the atmosphere between the two UHV vessels by an evaporated amorphous As film. Within the STM chamber this As film was removed from the sample by thermal desorption. By STM manipulation one finally obtains atomically well defined chains of In atoms, as shown in the constant-current STM image in Fig. 6.7a.

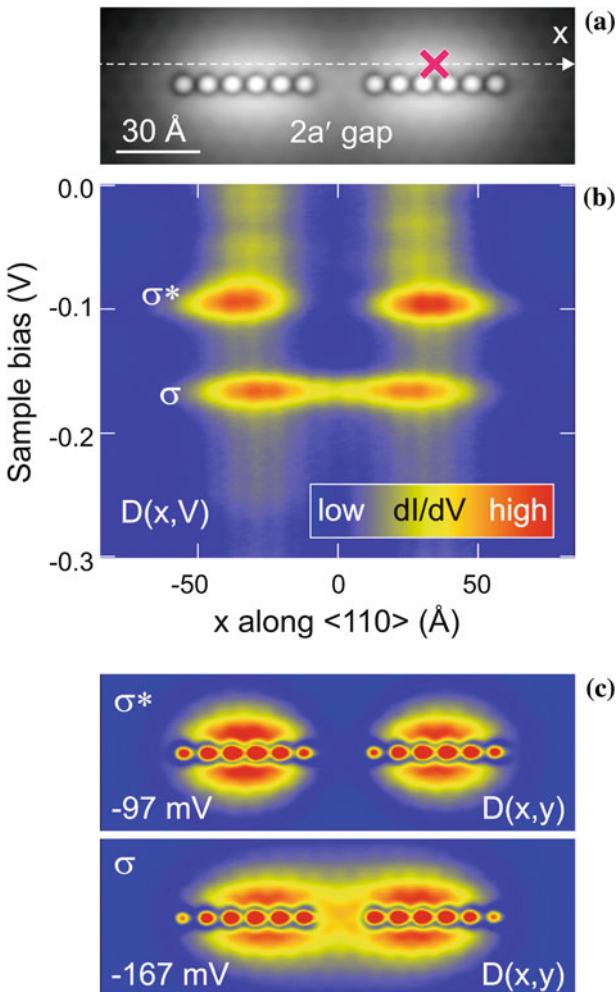
Samples with two six-atomic In chains, separated by different numbers of empty vacancy sites between the chains, were prepared by the described procedure. The study presented in the present context was performed on a sample with two empty vacancy sites between the In atom chains on the InAs surface. These two  $In_6$  chains can be considered as two longish quantum dots being located close to each other



**Fig. 6.7** Chain of In atoms assembled by a low temperature scanning tunnelling microscope (STM) at 5 K on a polar In terminated InAs(111)A surface prepared by MBE growth (Appendix B) [2]. **a** Topographic STM image measured with a sample bias of  $-0.3$  V with respect to the STM tip and at a tunnelling current of  $0.1$  nA. Twenty two In atoms were placed on adjacent In vacancy sites of the  $(2 \times 2)$  reconstructed surface. **b** Atomic structure of the image section in **(a)**. The regular array of In vacancies yields the adsorption sites for the In adatoms (red) of the  $\text{In}_{22}$  chain. The surface consists of In (yellow) and As atoms (blue).  $a$  is the vacancy distance

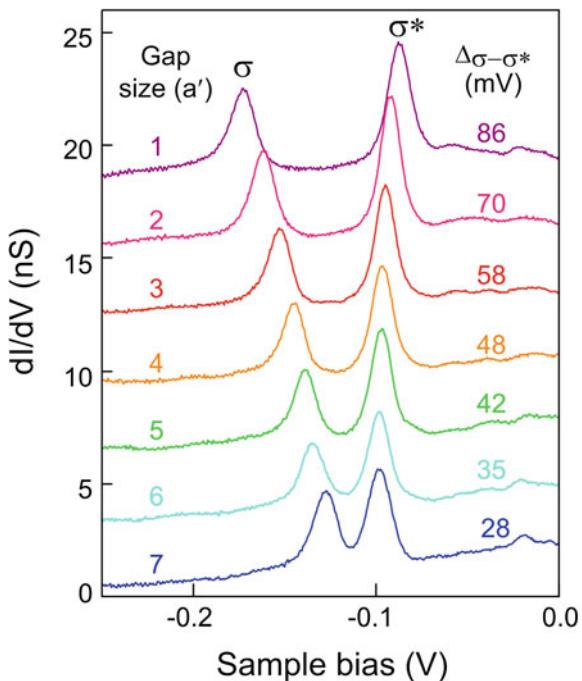
(Fig. 6.8a) such that they can interact by electron tunnelling, or in other words, by spatial overlap of their electronic wave functions. The electronic structure of each  $\text{In}_6$  dot is described by wave functions of an electron confined in this dot. According to Sect. 3.6.1 the ground state wave function ( $\sigma$ ) extends over the six In atoms with one and the same sign; it has no node and leaks into the area around the longish  $\text{In}_6$  dot. The spacing between the two dots is sufficiently small that the wave functions of the two dots overlap. Tunnelling between the dots is possible and the physical situation is exactly that which is mathematically described in Sect. 6.2.3. The two coupled  $\text{In}_6$  dots can be described as an artificial  $\text{In}_6 - \text{In}_6$  or  $(\text{In}_6)_2$  molecule with fixed atomic bonding distance. We expect as the lowest electronic molecular states the bonding  $|\psi_{bond}\rangle$  and the antibonding state  $|\psi_{antib}\rangle$  which were calculated in Sect. 6.2.3.

In the experiment the molecular orbitals have been made visible by the same STM which was used for the preparation of the two-dot molecule (Fig. 6.8). As was explained in Sect. 3.6.4 electronic tunnelling between the STM tip and the surface, respectively the adsorbed In atoms is connected with electron transfer between the electronic states (orbitals) of the tip and the orbitals of the adsorbed  $(\text{In}_6)_2$  molecule. Since the states of the tip are metal states being energetically and spatially continuously distributed and washed out (Sect. 3.6.1) structures which are found in the tunnel spectra or spatial maps are essentially due to the adsorbed quantum dot molecule. Depending on the polarity of the STM bias between tip and sample surface electrons tunnel from occupied  $(\text{In}_6)_2$  states into empty tip states or vice versa



**Fig. 6.8** Artificial double quantum dot molecule consisting of two In<sub>6</sub> chain quantum dots separated by a gap of two In vacancy sites on the InAs(111)A-(2×2) surface [2]. **a** Topographic STM image measured with a sample bias of  $-0.3 \text{ V}$  at  $0.1 \text{ nA}$  at low temperature ( $5 \text{ K}$ ). The dotted arrow line and the cross indicate where the differential tunnelling conductance  $dI/dV$  was measured as a function of position (arrow line) or at fixed tip position (marked by the cross) to obtain a bias-dependent differential tunnelling conductance spectrum (Fig. 6.9). **b** Density of states map  $D(x, V)$  measured along the dotted arrow line in (a). The graded intensities indicate varying probability densities  $|\psi|^2$  of the molecular orbitals along the dotted line. Two orbitals, the bonding one ( $\sigma$ ) and the antibonding one ( $\sigma^*$ ) appear at sample bias voltages of  $-167 \text{ mV}$  and  $-97 \text{ mV}$ , respectively. These voltages indicate the corresponding orbital energies. **c** Probability density  $|\psi^2|$  spatial maps  $D(x, y)$  of the symmetric  $\sigma$  and the antisymmetric  $\sigma^*$  orbitals. The differential tunnelling conductance  $dI/dV$  was measured as a function of position in a 2D scan at the sample bias voltages  $-167 \text{ mV}$  ( $\sigma$ ) and  $-97 \text{ mV}$  ( $\sigma^*$ )

**Fig. 6.9** Differential tunnelling conductance ( $dI/dV$ ) spectra quantifying the  $\sigma - \sigma^*$  splitting  $\Delta_{\sigma-\sigma^*}$  between the energies of the bonding and antibonding orbitals of the  $(In_6)_2$  molecule. The spectra belong to different gap size between the  $In_6$  chains; the gap size is given in numbers (1 – 7) of In vacancies between the  $In_6$  chains [2]



from occupied metallic tip states into empty  $(In_6)_2$  states. At a polarity  $InAs/(In_6)_2$  negative with respect to the STM tip, electrons tunnel from the  $InAs/(In_6)_2$  side into the continuously distributed tip states. By measuring the derivative of the tunnelling current  $dI/dV$  as a function of the sample bias relative to the tip, a spectroscopy of the electronic states of the  $(In_6)_2$  molecule becomes possible. In Fig. 6.9 two characteristic eigenstate energies of the  $(In_6)_2$  molecule are revealed by peaks in the corresponding conductance spectrum shown in magenta line colour, measured with the tip held fixed at the position marked by a cross in Fig. 6.8a. These two energies, denoted  $E_-$  and  $E_+$  in Sect. 6.2.3 belong to the bonding ( $\sigma$  or  $|\psi_{bond}\rangle$ ) or antibonding ( $\sigma^*$  or  $|\psi_{antib}\rangle$ ) states arising from the symmetric or antisymmetric superposition of the two ground states of the  $In_6$  quantum dots left and right  $|L\rangle, |R\rangle$  (6.57a, 6.57b). These  $\sigma$  and  $\sigma^*$  states are occupied electronic states of the quantum dot molecule because the  $dI/dV$  peaks in Fig. 6.9 occur at negative sample-tip bias. As is expected from (6.58c), the energy difference  $|E_+ - E_-|$  decreases with smaller tunnelling amplitude  $t_{LR}$ , i.e. with increasing gap size between the two  $In_6$  chains. The gap size  $a'$  is given in Fig. 6.9 by the number of empty vacancy sites between the  $In_6$  chains.

Additionally, by scanning the STM tip at constant height and constant sample-tip bias along a fixed line across the surface or, alternatively, over the whole area covered by the  $(In_6)_2$  molecule, plots of the density of states of the occupied  $(In_6)_2$  orbitals along the line (Fig. 6.8b) or 2-dimensional (2D) maps of the occupied orbitals (Fig. 6.8c) are obtained. To begin with the first measurement mode, the  $dI/dV$  signal is

probed at fixed bias only along the dashed arrow line indicated as  $x$  in Fig. 6.8a. Starting this procedure at a sample-tip bias of  $-0.3$  V and reducing the bias line-by-line to  $0$  V, the  $x$ -versus-bias  $dI/dV$  map in Fig. 6.8b is obtained. In this representation different colour grades indicate the intensity of the orbital probability density  $|\psi|^2$ , probed along the dashed arrow line close to the two dots in Fig. 6.8a. The bias-dependent line scans show a resonant behaviour at the energies  $E_-$  and  $E_+$  (corresponding to the tip-sample voltages of  $-0.167$  and  $0.097$  V, respectively) revealing the energetically lower bonding  $\sigma$  orbital with finite density in between the two  $\text{In}_6$  chains (dots) and the antibonding  $\sigma^*$  orbital with a node in between the two quantum dots. In the second detection mode of scanning 2D  $dI/dV$  maps the spatial distribution of the  $\sigma$  and  $\sigma^*$  orbitals is directly visualized, as shown in Fig. 6.8c. In this measurement, the sample-tip bias was adjusted to  $-0.167$  and  $0.097$  V, respectively, to record the  $dI/dV$  signal in resonance with the  $\sigma$  and  $\sigma^*$  states. The bonding orbital extends with non-negligible density over the space in between the two quantum dots, whereas the antibonding orbital has a node in between. In contrast to the assumption of a simple hat-like ground state wave function for a single quantum dot in Sect. 6.2.3 the ground state wave function of an electron confined in one single  $\text{In}_6$  chain (dot) is more complex. It is structured with intensity maxima along and on both sides of the atomic chain. This is due to the mixing between electronic and topographic information resulting from the  $dI/dV$  scanning procedure across the chains at constant tip height. Nevertheless, concerning the spatial character and the energetic behaviour of the bonding and antibonding states of the artificial  $(\text{In}_6)_2$  quantum dot molecule the theoretical results of Sect. 6.2.3 are exactly found in the experiment.

### 6.3 Small Stationary Potential Perturbations: The Time-Independent Perturbation Method

Both approximation techniques presented so far, the WKB method and the variational method, treat time-independent problems: A Schrödinger equation with a time-independent Hamiltonian is approximately solved by stationary quantum states, respectively, wave functions.

The same class of stationary problems is also approached by the so-called time-independent perturbation method. This approximation technique can be applied if the problem to be solved is only a slightly varied version of a problem of which the exact solution is known already.

As an example we consider an atom in a stationary electric field  $\mathcal{E}$ . The energies of the quantum states of the unperturbed atom ( $\mathcal{E} = 0$ ) are determined by the confinement of the outer electrons in the atomic potential of the positive nucleus. The responsible electric field strengths being connected with these potentials are in the order of  $10^9$  V/cm, since electrons with binding energies of about  $10\text{--}100$  eV are confined in spatial regions with linear extensions of about  $10^{-8}$  cm. Even the strongest external fields of about  $10^6$  V, as in semiconduc-

tor heterostructures or space charge layers (Sect. 6.1.1, Fig. 6.1, Appendix A) are tiny perturbations of the atomic fields, in the 0.1 % range. For the approximate calculation of atomic states and their energies in an external electric field, therefore, the unperturbed states of the free atom can be taken as the starting point. They are the basis for calculating the modifications of states and energies due to small stationary perturbations of the Hamiltonian. The modified wave functions are most probably only a little bit shifted and deformed with respect to the unperturbed ones. This can be taken care of by mixing into the unperturbed wave functions small contributions of unperturbed wave functions of higher excited quantum states.

For the following formal treatment, we assume the unperturbed problem to be described by the Hamiltonian  $\hat{H}$ , its eigenvalues  $E_n^0$  and  $|n\rangle$  as the eigenkets of the unperturbed Schrödinger equation:

$$\hat{H}|n\rangle = E_n^0|n\rangle. \quad (6.59)$$

The small perturbation of the time-independent potential is now described by an operator  $\lambda \cdot \hat{h}$ , where  $\lambda$  is a numerical parameter ( $0 < \lambda \leq 1$ ) for denoting the order of magnitude of the elements in the series expansion of the state vectors and energy eigenvalues.

The slightly perturbed system [with respect to (6.59)] is then represented by the Schrödinger equation

$$(\hat{H} + \lambda \hat{h})|\psi_n\rangle = E_n|\psi_n\rangle, \quad 0 < \lambda \leq 1. \quad (6.60)$$

The energy eigenvalues  $E_n$  and the state vectors  $|\psi_n\rangle$  are represented as series expansions, where higher elements describe smaller corrections to the unperturbed energy and the state vector, respectively:

$$E_n = E_n^0 + \lambda \varepsilon'_n + \lambda^2 \varepsilon''_n + \dots, \quad (6.61a)$$

$$|\psi_n\rangle = |n\rangle + \lambda |\delta n'\rangle + \lambda^2 |\delta n''\rangle + \dots. \quad (6.61b)$$

These corrections of first, second, third, ... order can now be determined; ordered according to  $\lambda, \lambda^2, \lambda^3, \dots$  they yield ever better solutions to the perturbed problem.

We insert (6.61a), (6.61b) into (6.60) and obtain

$$(\hat{H} + \lambda \hat{h})(|n\rangle + \lambda |\delta n'\rangle + \dots) = (E_n^0 + \lambda \varepsilon'_n + \dots)(|n\rangle + \lambda |\delta n'\rangle + \dots). \quad (6.62)$$

This equation is valid for each  $\lambda$  value, that is, the coefficients of the different  $\lambda$  powers on both sides of (6.62) must be equal. We thus consider the separate equations for  $\lambda^0, \lambda^1, \lambda^2, \dots$ . For  $\lambda^0$ , we get

$$\hat{H}|n\rangle = E_n^0|n\rangle, \quad (6.63a)$$

the Schrödinger equation of the unperturbed problem (6.59), as expected. The next better approximations follow from the equations for  $\lambda^1$ ,  $\lambda^2$ , etc.:

$$\lambda^1 : \hat{h}|n\rangle + \hat{H}|\delta n'\rangle = \varepsilon'_n|n\rangle + E_n^0|\delta n'\rangle, \quad (6.63b)$$

$$\lambda^2 : \hat{h}|\delta n'\rangle + \hat{H}|\delta n''\rangle = E_n^0|\delta n''\rangle + \varepsilon'_n|\delta n'\rangle + \varepsilon''_n|n\rangle. \quad (6.63c)$$

The first order approximation ( $\lambda^1$ ) is obtained from (6.63b) by

$$(\hat{H} - E_n^0)|\delta n'\rangle = (\varepsilon' - \hat{h})|n\rangle. \quad (6.64)$$

We expand the small perturbation  $|\delta n'\rangle$  of the unperturbed state  $|n\rangle$  in terms of the orthonormal system of the unperturbed eigenkets  $|k\rangle$ :

$$|\delta n'\rangle = \sum_k c'_{nk}|k\rangle. \quad (6.65)$$

Hereby the coefficients  $c'_{nk}$  denote the first order ('') approximation to the state  $|n\rangle$ .

Inserting (6.65) into (6.64) yields

$$\sum_k (\hat{H} - E_n^0)c'_{nk}|k\rangle = (\varepsilon'_n - \hat{h})|n\rangle. \quad (6.66)$$

To calculate the perturbation of the eigenvalue  $E_n^0$  in this approximation, we multiply (6.66) from left with the bra  $\langle n|$  and use  $\hat{H}|k\rangle = E_k^0|k\rangle$ :

$$\sum_k c'_{nk}(E_k^0 - E_n^0)\langle n|k\rangle = \varepsilon'_n - \langle n|\hat{h}|n\rangle. \quad (6.67)$$

Because of  $\langle n|k\rangle = \delta_{nk}$  the first order perturbation  $\varepsilon'_n$  of the energy eigenvalue  $E_n^0$  is obtained as

$$\varepsilon'_n = \langle n|\hat{h}|n\rangle. \quad (6.68)$$

In case of  $\lambda \neq 1$  the corresponding  $\lambda$  value must be included in (6.68). The simple result in (6.68) expresses the first order perturbation  $\varepsilon'_n$  of the energy eigenvalue as the diagonal matrix element of the perturbation operator  $\hat{h}$  in the the unperturbed state  $|n\rangle$ .

To obtain the first expansion coefficients  $c'_{nk}$ , that is, the perturbation induced modification of the state  $|n\rangle$ , we multiply (6.66) from the left with the bra  $\langle m|$  where  $m \neq n$ . The result is

$$\begin{aligned} \sum_k c'_{nk} (E_k^0 - E_n^0) \langle m|k \rangle &= \varepsilon' \langle m|n \rangle - \langle m|\hat{h}|n \rangle, \\ c'_{nm} (E_m^0 - E_n^0) &= -\langle m|\hat{h}|n \rangle, \\ c'_{nm} &= \frac{\langle m|\hat{h}|n \rangle}{E_n^0 - E_m^0}. \end{aligned} \quad (6.69)$$

The first order perturbation of the quantum state, thus, follows as

$$|\delta n'\rangle = \sum_{k \neq n} \frac{\langle k|\hat{h}|n \rangle}{E_n^0 - E_k^0} |k\rangle. \quad (6.70)$$

For the next approximation step, the calculation of  $E_n''$  and  $|\delta n''\rangle$ , the terms in  $\lambda^2$  are collected and compared:

$$(\hat{H} - E_n^0) |\delta n''\rangle = (\varepsilon'_n - \hat{h}) |\delta n'\rangle + \varepsilon''_n |n\rangle. \quad (6.71)$$

The perturbations  $|\delta n'\rangle$  and  $|\delta n''\rangle$  are expanded in terms of the unperturbed states, that is, according to (6.65) and

$$|\delta n''\rangle = \sum_k c''_{nk} |k\rangle, \quad (6.72)$$

we obtain

$$\sum_k c''_{nk} \hat{H} |k\rangle - \sum_k c''_{nk} E_n^0 |k\rangle = \varepsilon'_n \sum_k c'_{nk} |k\rangle - \sum_k c'_{nk} \hat{h} |k\rangle + \varepsilon''_n |n\rangle. \quad (6.73)$$

Both sides of (6.73) are multiplied by the bra  $\langle m|$  and because of  $\langle m|k \rangle = \delta_{mk}$ , we obtain

$$c''_{nm} \langle m|\hat{H}|m \rangle - c''_{nm} E_n^0 = \varepsilon'_n c'_{nm} - \sum_k c'_{nk} \langle m|\hat{h}|k \rangle + \varepsilon''_n \delta_{mn}, \quad (6.74a)$$

$$c''_{nm} (E_m^0 - E_n^0) = c'_{nm} \varepsilon'_n - \sum_k c'_{nk} \langle m|\hat{h}|k \rangle + \varepsilon''_n \delta_{mn}. \quad (6.74b)$$

For calculating the second correction  $\varepsilon''_n$  to the energy eigenvalue  $E_n^0$  from (6.74a)–(6.74c)  $m = n$  is assumed in (6.74b) and we obtain

$$\varepsilon''_n = \sum_{k \neq n} c'_{nk} \langle n|\hat{h}|k \rangle + c'_{nn} \langle n|\hat{h}|n \rangle - c'_{nn} \varepsilon'_n. \quad (6.74c)$$

Because of (6.68), the two last terms cancel and the perturbation follows as

$$\varepsilon''_n = \sum_{k \neq n} c'_{nk} \langle n|\hat{h}|k \rangle. \quad (6.75a)$$

After inserting (6.69), we finally obtain

$$\varepsilon_n'' = \sum_{k \neq n} \frac{|\langle n | \hat{h} | k \rangle|^2}{E_n^0 - E_k^0}. \quad (6.75b)$$

For the calculation of the second order perturbation  $|\delta n''\rangle$  of the state vector (6.72), that is, the coefficients  $c_{nm}''$ , we start with (6.74b) but assume  $m \neq n$ . This derivation is skipped here and more extended books on quantum mechanics are recommended. For practical applications, the approximations considered here are mostly sufficient.

We summarize: A small stationary potential  $\hat{h}$  being assumed as a perturbation to an already solved problem with known energy eigenvalues  $E_n$  and known eigenkets  $|\psi_n\rangle$  gives rise to the following slightly modified energies and eigenstates:

$$E_n = E_n^0 + \langle n | \hat{h} | n \rangle + \sum_{k \neq n} \frac{|\langle n | \hat{h} | k \rangle|^2}{E_n^0 - E_k^0} + \dots, \quad (6.76a)$$

$$|\psi_n\rangle = |n\rangle + \sum_{k \neq n} \frac{\langle k | \hat{h} | n \rangle}{E_n^0 - E_k^0} |k\rangle + \dots. \quad (6.76b)$$

The approximation has the following implications:

- The requirement that the perturbation operator  $\hat{h}$  is sufficiently small means in detail that the series expansions (6.76a), (6.76b) must converge sufficiently fast. For this purpose, the matrix elements  $\langle n | \hat{h} | k \rangle$  must be significantly smaller than the energetic distances between the unperturbed energies  $E_k^0$ .
- A problem appears when the energy denominators in (6.76a), (6.76b) vanish. In this case of degeneracy of the considered state  $|n\rangle$  with other states  $|k\rangle$ , we have to use a modified perturbation method which is presented in the next section (Sect. 6.3.1).
- The first order perturbation  $\langle n | \hat{h} | n \rangle$  might have a positive or negative sign. On the other hand, considering a perturbation of the ground state energy  $E_0^0$  we realize that all denominators  $(E_0^0 - E_k^0)$  are negative. Since the numerators in (6.76a) are all positive, perturbations of the considered type always lower the ground state energy in second order.

### 6.3.1 Perturbation of Degenerate States

The described approximation method breaks down if in (6.70) two or more unperturbed energy levels are equal and the energy denominators vanish. This situation appears in the case of degeneracy, where one and the same energy level belongs to several different quantum states.

Degeneracies mostly arise from symmetry properties of the potential in the Schrödinger equation. In a spherical potential, for example, the angular momentum states are degenerate for different directional quantum numbers  $m$ . The states with different  $m$  all have the same energy. This degeneracy is lifted by an external

magnetic field  $\mathbf{B}$  which breaks the spherical symmetry of the problem. Different orientations of the angular momentum  $m$  in the  $\mathbf{B}$  field cause different energies.

Degenerate states arising from a certain symmetry belong to the same unperturbed energy eigenvalue  $E_n^0$ , they have equal weight in the representation of a general state. A symmetry breaking perturbation now causes a different weight of the unperturbed (degenerate) states in the representation of the new perturbed state. This new weighting of the degenerate states is not known at the beginning, it must be evaluated within the frame of the perturbation calculation. For simplicity reasons, we assume two-fold degeneracy of the unperturbed state with energy  $E_n^0$ . Two orthogonal states  $|n_1\rangle$  and  $|n_2\rangle$ , then, solve the unperturbed Schrödinger equation with one and the same energy eigenvalue  $E_n^0$ . The series expansion (6.61a) of the perturbed energy must incorporate the fact that the small perturbation operator  $\hat{h}$  causes a splitting of the unperturbed energy  $E_n^0$  into two new levels  $E_{n1}$  and  $E_{n2}$ :

$$E_{n1} = E_n^0 + \lambda \varepsilon'_{n1} + \lambda^2 \varepsilon''_{n1} + \dots, \quad (6.77a)$$

$$E_{n2} = E_n^0 + \lambda \varepsilon'_{n2} + \lambda^2 \varepsilon''_{n2} + \dots. \quad (6.77b)$$

Similarly, we have to take into account that in the expansion of the perturbed eigenkets (6.61b) two new states  $|\psi_{n1}\rangle$  and  $|\psi_{n2}\rangle$  must occur due to the two-fold degeneracy. Expansion coefficients (amplitudes)  $c_{11}$ ,  $c_{12}$ ,  $c_{21}$  and  $c_{22}$  describe the new weighting of the unperturbed eigenkets  $|n_1\rangle$  and  $|n_2\rangle$  in these expansions:

$$|\psi_{n1}\rangle = c_{11}|n_1\rangle + c_{12}|n_2\rangle + \lambda|\delta n'_1\rangle + \dots, \quad (6.78a)$$

$$|\psi_{n2}\rangle = c_{21}|n_1\rangle + c_{22}|n_2\rangle + \lambda|\delta n'_2\rangle + \dots. \quad (6.78b)$$

The determination of the amplitudes  $c_{ij}$  yields the correct combination of eigenstates  $|n_1\rangle$  and  $|n_2\rangle$  for the description of the broken symmetry due to the perturbation.

For the further calculation, we proceed analogously to (6.62) and insert (6.77a), (6.77b) and (6.78a), (6.78b) into the perturbed Schrödinger equation:

$$\begin{aligned} & (\hat{H} + \lambda \hat{h})(c_{11}|n_1\rangle + c_{12}|n_2\rangle + \lambda|\delta n'_1\rangle + \dots) \\ &= (E_n^0 + \lambda \varepsilon'_{n1})(c_{11}|n_1\rangle + c_{12}|n_2\rangle + \lambda|\delta n'_1\rangle + \dots), \end{aligned} \quad (6.79a)$$

$$\begin{aligned} & (\hat{H} + \lambda \hat{h})(c_{21}|n_1\rangle + c_{22}|n_2\rangle + \lambda|\delta n'_2\rangle + \dots) \\ &= (E_n^0 + \lambda \varepsilon'_{n2})(c_{21}|n_1\rangle + c_{22}|n_2\rangle + \lambda|\delta n'_2\rangle + \dots). \end{aligned} \quad (6.79b)$$

Comparison of the expansion elements belonging to  $\lambda^0 = 1$  yields, as in (6.63a), the solution to the unperturbed problem:

$$\hat{H}(c_{11}|n_1\rangle + c_{12}|n_2\rangle) = E_n^0(c_{11}|n_1\rangle + c_{12}|n_2\rangle), \quad (6.80)$$

respectively, the same relation with the amplitudes  $c_{21}$  and  $c_{22}$ . These are all possible linear combinations if  $|n_1\rangle$  and  $|n_2\rangle$  are the eigenkets with the energy eigenvalue  $E_n^0$ .

For the calculation of the first approximation we compare, in analogy to (6.63b), the elements belonging to  $\lambda^1$ . From (6.79a), (6.79b) we obtain:

$$c_{11}\hat{h}|n_1\rangle + c_{12}\hat{h}|n_2\rangle + \hat{H}|\delta n'_1\rangle = c_{11}\varepsilon'_{n1}|n_1\rangle + c_{12}\varepsilon'_{n1}|n_2\rangle + E_n^0|\delta n'_1\rangle, \quad (6.81a)$$

$$c_{21}\hat{h}|n_1\rangle + c_{22}\hat{h}|n_2\rangle + \hat{H}|\delta n'_2\rangle = c_{21}\varepsilon'_{n2}|n_1\rangle + c_{22}\varepsilon'_{n2}|n_2\rangle + E_n^0|\delta n'_2\rangle. \quad (6.81b)$$

The perturbations  $|\delta n'_1\rangle$  and  $|\delta n'_2\rangle$  must be normal to the state vectors  $|n_1\rangle$  and  $|n_2\rangle$ , of course, otherwise they could be represented in terms of normalization factors in the unperturbed states, that is:

$$\langle n_1|\hat{H}|\delta n'_1\rangle = E_n^0\langle n_1|\delta n'_1\rangle = 0, \quad (6.82a)$$

$$\langle n_2|\hat{H}|\delta n'_1\rangle = E_n^0\langle n_2|\delta n'_1\rangle = 0. \quad (6.82b)$$

Analogous relations are valid for  $|\delta n'_2\rangle$  and by multiplication of (6.81a) with the bras  $\langle n_1|$  and  $\langle n_2|$  from left we obtain

$$c_{11}\langle n_1|\hat{h}|n_1\rangle + c_{12}\langle n_1|\hat{h}|n_2\rangle = c_{11}\varepsilon'_{n1}, \quad (6.83a)$$

$$c_{11}\langle n_2|\hat{h}|n_1\rangle + c_{12}\langle n_2|\hat{h}|n_2\rangle = c_{12}\varepsilon'_{n1}. \quad (6.83b)$$

With the matrix elements  $h_{ij} = \langle n_i|\hat{h}|n_j\rangle$  of the perturbation operator the following secular equation system is obtained:

$$\begin{pmatrix} (h_{11} - \varepsilon'_{n1}) & h_{12} \\ h_{21} & (h_{22} - \varepsilon'_{n1}) \end{pmatrix} \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (6.84a)$$

The solution yields the perturbation  $\varepsilon'_{n1}$  of the energy eigenvalue and the amplitudes  $c_{11}$  and  $c_{12}$  which represent the relative contributions of the unperturbed states  $|n_1\rangle$  and  $|n_2\rangle$  to the perturbed wave function.

The same treatment of (6.81b) yields in analogy to (6.84a):

$$\begin{pmatrix} (h_{11} - \varepsilon'_{n2}) & h_{12} \\ h_{21} & (h_{22} - \varepsilon'_{n2}) \end{pmatrix} \begin{pmatrix} c_{21} \\ c_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (6.84b)$$

The two eigenvalue problems for the determination of  $\varepsilon'_{n1}$  and  $\varepsilon'_{n2}$  are identical. To obtain non-trivial solutions, the determinant of the matrix must vanish. By setting the eigenvalue  $\varepsilon = \varepsilon'_{n1} = \varepsilon'_{n2}$  it is required that

$$\begin{vmatrix} (h_{11} - \varepsilon) & h_{12} \\ h_{21} & (h_{22} - \varepsilon) \end{vmatrix} = (h_{11} - \varepsilon)(h_{22} - \varepsilon) - h_{12}h_{21} = 0. \quad (6.85)$$

The solution of this quadratic equation yields two energy corrections:

$$\varepsilon_{\pm} = \frac{h_{11} + h_{22}}{2} \pm \sqrt{h_{12}h_{21} + \frac{1}{4}(h_{11} - h_{22})^2}. \quad (6.86)$$

As expected, the symmetry breaking perturbation potential  $\hat{h}$  splits the two-fold degenerate energy level  $E_n^0$  into two new energy levels:

$$E_{n1} = E_n^0 + \varepsilon_+ \quad \text{and} \quad E_{n2} = E_n^0 + \varepsilon_- . \quad (6.87)$$

From a calculation of the eigenvectors  $(c_{11}, c_{12})$  and  $(c_{21}, c_{22})$ , we obtain a symmetrical and antisymmetrical superposition of the state vectors  $|n_1\rangle$  and  $|n_2\rangle$ .

One should not be astonished about the mathematical similarity of these approximate solutions to the problem of two coupled quantum dots or of the  $H_2^+$  molecule (Sect. 6.3.2). The coupled quantum dots and the  $H_2^+$  molecule could also be treated in the frame work of the time-independent perturbation method. In this treatment the neighbouring nucleus (proton) in the molecule is assumed as a perturbation potential for the states of an H atom (unperturbed system).

The extension of the two-dimensional calculation presented here to the general problem of  $d$ -fold degeneracy is straightforward. When  $E_n^0$  is an energy level with  $d$ -fold degeneracy in the unperturbed system, there are  $d$  orthogonal states  $|n_1\rangle, |n_2\rangle, \dots, |n_d\rangle$  with this same energy. In analogy to (6.84a), (6.84b), we construct a  $d$ -dimensional secular equation system with the perturbation matrix elements  $\langle n_i | \hat{h} | n_j \rangle$ . By setting the determinant to zero  $d$  different solutions are obtained which represent the first order corrections to the energy  $E_n^0$ .

### 6.3.2 Example: The Stark Effect in a Semiconductor Quantum Well

We consider a GaAs quantum well (thickness  $2L = 10$  nm) embedded between two AlAs regions. The whole layer structure is prepared by epitaxy (Appendix B). Even though the conduction band discontinuities (Appendix A) between GaAs and AlAs amount to about 0.4 eV only, we assume, for simplicity reasons, infinitely high potential walls of the well. The Schrödinger equation for an electron in the conduction band (Sect. 8.3.4), that is, the well, is one-dimensional with the coordinate  $x$  being the epitaxial growth direction normal to the layer sequence. The corresponding representation of the well potential is written as:

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < 2L, \\ \infty & \text{elsewhere.} \end{cases} \quad (6.88)$$

With  $m^*$  as effective mass of the conduction electron (Sect. 8.3.4) the Schrödinger equation

$$\hat{H}|n\rangle = \left( \frac{p^2}{2m^*} + V(x) \right)|n\rangle = E_n^0|n\rangle \quad (6.89)$$

has the solution (Sect. 3.6.1):

$$\psi_n(x) = \langle x|n\rangle = \frac{1}{\sqrt{L}} \sin\left(\frac{n\pi x}{2L}\right), \quad n = 1, 2, \dots \quad (6.90)$$

These wave functions have mirror symmetry around the center of the quantum well (symmetric) for odd  $n$  values, for even  $n$  they change their sign (anti-symmetric) upon mirror reflection at a central plane through the well. The energy eigenvalues obtained from (6.89) are

$$E_n^{(0)} = \frac{\hbar^2 k_n^2}{2m^*} = \frac{\hbar^2 \pi^2 n^2}{8m^* L^2} \quad \text{with } k_n = \frac{n\pi}{2L}. \quad (6.91)$$

We now assume an electric field applied normal to the layer structure along the  $x$  direction. It generates an electric field strength  $\mathcal{E}$  in the quantum well between  $x = 0$  and  $x = L$ . The system is, thus, perturbed by the energy operator

$$\hat{h} = ex|\mathcal{E}|. \quad (6.92)$$

Voltages in the Volt range applied over a quantum well with a length of 100 nm typically produce fields in the  $10^5$  V/cm range. These fields are small in comparison to atomic fields in the  $10^9$  V/cm range and the time-independent perturbation method can be applied to calculate field induced changes  $\varepsilon'_n$  (6.68) of the energy eigenvalues (6.61a). According to (6.68) and (6.90), we obtain

$$\varepsilon'_n = \langle n|\hat{h}|n\rangle = \langle n|e|\mathcal{E}|x|n\rangle = \frac{e|\mathcal{E}|}{L} \int_0^{2L} x \sin^2\left(\frac{n\pi x}{2L}\right) dx. \quad (6.93)$$

Using the general relation  $2 \sin x \sin y = \cos(x - y) - \cos(x + y)$  and  $x = y = nx\pi/2L$  one obtains from (6.93):

$$\varepsilon'_n = \frac{e|\mathcal{E}|}{2L} \int_0^{2L} x \left[ 1 - \cos\left(\frac{n\pi x}{L}\right) \right] dx = \frac{e|\mathcal{E}|}{2L} \frac{x^2}{2} \Big|_0^{2L} = eL|\mathcal{E}|. \quad (6.94)$$

This shift linear in the electric field strength was detected by Stark on atomic systems [5]. Accordingly it is called Stark effect.

For the considered AlAs/GaAs/AlAs quantum well with the thickness  $2L = 10$  nm an applied electric field of  $10^5$  V/cm changes the energy levels by  $\varepsilon'_n = 50$  meV.

At this point, a peculiarity of the problem must be noted. In case that the quantum well had been assumed symmetrical around the zero point of the  $x$  axis, i.e.  $V(x) = 0$  for  $-L < x < L$ , the wave functions  $\psi_n$  would have been cosine-like rather than

sine-like. An analogous calculation, then, yields vanishing matrix elements  $\langle n|\hat{h}|n\rangle$  (6.93); a *linear Stark effect* does not exist anymore.

How can we understand this physically different behavior even though only the coordinate system was shifted? In the first case of a quantum well only on the positive  $x$  axis a non-vanishing moment of electric charge with respect to the zero point exists. In the second symmetrical case, the moment of charge vanishes. This case is always given when the electric field is applied symmetrical to the quantum well, that is, by electrical contacts symmetrical to the potential well. This is the common experimental situation: The linear Stark effect vanishes. The calculation of field induced energy level shifts requires the next step of perturbation calculation up to quadratic order in the electric field. This is called the *quadratic Stark effect*.

In the mathematical treatment, we analyse the expression  $\varepsilon_n''$  (6.75a). This shall be done only for the ground state with  $n = 1$ , where the following matrix elements must be evaluated:

$$\begin{aligned}\langle 1|\hat{h}|k\rangle &= \frac{e|\mathcal{E}|}{L} \int_0^{2L} x \sin\left(\frac{\pi x}{2L}\right) \sin\left(\frac{k\pi x}{2L}\right) dx \\ &= \frac{e|\mathcal{E}|}{2L} \int_0^{2L} x \left[ \cos\frac{(k-1)\pi x}{2L} - \cos\frac{(k+1)\pi x}{2L} \right] dx.\end{aligned}\quad (6.95a)$$

Partial integration yields

$$\begin{aligned}\langle 1|\hat{h}|k\rangle &= \frac{e|\mathcal{E}|}{2L} \left[ \frac{4L^2\{\cos(k-1)\pi - 1\}}{\pi^2(k-1)^2} - \frac{4L^2\{\cos(k+1)\pi - 1\}}{\pi^2(k+1)^2} \right] \\ &= -\frac{4eL}{\pi^2} |\mathcal{E}| \left[ \frac{1}{(k+1)^2} - \frac{1}{(k-1)^2} \right] = -\frac{16eL}{\pi^2} |\mathcal{E}| \frac{k}{(k^2-1)^2}.\end{aligned}\quad (6.95b)$$

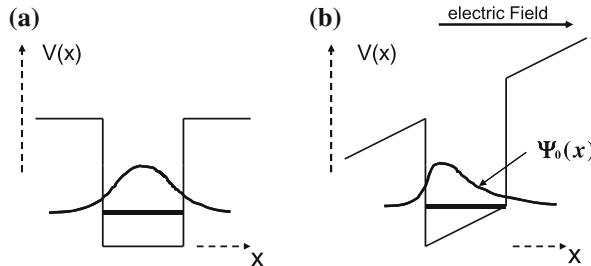
For  $(k \pm 1)$  the matrix element  $\langle 1|\hat{h}|k\rangle$  vanishes because of symmetry reasons. The perturbation potential mixes into the symmetrical ground state only antisymmetrical higher states. It can generally be shown that for higher states with  $n > 1$  only states with opposite parity to  $|n\rangle$  contribute to the perturbation matrix elements. For the first approximation, that is, the lower limit of the matrix element (6.95b), we take into account only a contribution of the first excited state with  $k = 2$  to the ground state:

$$\langle 1|\hat{h}|2\rangle = -\frac{16e}{9\pi^2} (2L)|\mathcal{E}|. \quad (6.96)$$

According to (6.75b), the quadratic part of the Stark effect in the ground state  $|1\rangle$  is then obtained as

$$\varepsilon_1'' \simeq \frac{|\langle 1|\hat{h}|2\rangle|^2}{E_1^0 - E_2^0} = -\frac{256}{243\pi^4} \cdot \frac{(e2L)^2}{E_1^0} |\mathcal{E}|^2. \quad (6.97)$$

A more detailed calculation of the matrix elements  $\langle 1|\hat{h}|k\rangle$  taking into account also contributions of higher mixed-in states modifies the matrix element  $\langle 1|\hat{h}|2\rangle$  (6.96)



**Fig. 6.10 a, b** Qualitative representation of the ground state wave function  $\psi_0(x)$  of an electron in a 1D potential well with finite potential wall heights;  $V(x)$  is the potential curve. **a** Unperturbed system without electric field. **b** System with externally applied electric field

only in the percentage range. For the evaluation of the quadratic Stark effect (6.75b) we can, thus, restrict the approximate calculation to admixtures of the next higher states.

Figure 6.10a qualitatively shows a potential well with finite wall height. In contrast to the idealized solutions (6.90) for infinitely high walls, the wave functions penetrate the potential walls right and left. Application of an electric field  $E$  in  $x$  direction (Fig. 6.10b) tilts the potential lines, that is, the lower edge of the conduction band  $E_c$  in the semiconductor, and the center of the ground state wave function is shifted due to admixtures of the next higher state. This shift of the center of electronic charge against the positive background charge, for example, holes in the valence band of the quantum well, generates an electric dipole moment. This field induced dipole moment causes a shift of the energy eigenvalues in the electric field. The quadratic Stark effect is, thus, related to field induced dipoles ( $\propto E$ ) which cause an energy change ( $\propto E^2$ ) in the electric field. The linear Stark effect, on the other hand, requires the presence of an electric dipole already for vanishing electric field. This existing dipole is directed in the field.

## 6.4 Transitions Between Quantum States: The Time-Dependent Perturbation Method

The approximation techniques described so far were all focused on a solution of the Schrödinger equation with time-independent potential. The stationary eigensolutions of the problem were, then, slightly modified by small time-independent potential perturbations. We encounter a different type of problem when a time dependent perturbation acts on stationary states of a system. We expect that time-dependent perturbation potentials in the Schrödinger equation excite the system, that is, induce transitions from one stationary state into other states. States might be excited or might decay. These types of problems can also be treated if the time-dependent perturbation potentials are sufficiently small in comparison with the large built-in stationary potential which determines the original stationary states.

We consider a Schrödinger equation with time-independent  $\hat{H}$  operator:

$$i\hbar|\dot{\psi}\rangle = \hat{H}|\psi\rangle. \quad (6.98)$$

Its solutions  $|\psi\rangle$  are assumed to be known. Because of the time-independent potential a state with eigenenergy  $E_n$  has a time dependence  $\exp(-iE_nt/\hbar)$  as all stationary wave functions (3.58). We introduce the time-dependent perturbation into (6.98) by means of a small potential contribution  $\hat{h}(t)$ :

$$i\hbar|\dot{\psi}\rangle = (\hat{H} + \hat{h}(t))|\psi\rangle. \quad (6.99)$$

The perturbation calculation, now, must yield a probability (amplitude) for the transition of an initial state  $|i\rangle$  into a final state  $|f\rangle$  due to the action of the time dependent perturbation potential  $\hat{h}$ .  $|i\rangle$  and  $|f\rangle$  belong to the variety of stationary eigenstates  $|n\rangle$  of the unperturbed time-independent Hamiltonian (6.98). Typical examples of a perturbation  $\hat{h}$  are a short electric pulse or a harmonically (sine) varying electric field acting on an electron in an atom or in a quantum dot.

The solution  $|\psi\rangle$  to the Schrödinger equation (6.99) can, thus, be expanded in a series of orthonormal eigenvectors  $|n\rangle$  of the unperturbed time-independent problem (6.98):

$$|\psi(t)\rangle = \sum_n a_n(t) e^{-iE_nt/\hbar} |n\rangle. \quad (6.100)$$

In this representation, the time dependence of the solution originating from  $\hat{h}(t)$  is expressed in terms of time-dependent probability amplitudes  $a_n(t)$ . They describe the probability for the emergence of stationary states  $|n\rangle$  other than the initial state as a consequence of the time-dependent perturbation. To perform the first order perturbation calculation, the series expansion (6.100) is inserted into the time-dependent Schrödinger equation (6.99):

$$\begin{aligned} i\hbar \left[ \sum_n \dot{a}_n(t) e^{-iE_nt/\hbar} - i \frac{E_n}{\hbar} a_n(t) e^{-iE_nt/\hbar} \right] |n\rangle \\ = \sum_n a_n(t) E_n e^{-iE_nt/\hbar} |n\rangle + \sum_n a_n(t) e^{-iE_nt/\hbar} \hat{h} |n\rangle. \end{aligned} \quad (6.101a)$$

This yields

$$i\hbar \sum_n \dot{a}_n(t) e^{-iE_nt/\hbar} |n\rangle = \sum_n a_n(t) e^{-iE_nt/\hbar} \hat{h} |n\rangle, \quad (6.101b)$$

an equation, where the unperturbed Hamilton operator appears solely in terms of its eigenvalues  $E_n$ . Only the time-dependent potential  $\hat{h}(t)$  determines the time development of the probability amplitudes  $a_n(t)$ . To answer the question how a particular stationary final state  $|f\rangle$  is reached, in (6.101b) the projection on this state has to be

calculated. We thus multiply this equation from the left with the bra  $\langle f | \exp(iE_f t/\hbar)$  and keep in mind  $\langle f | n \rangle = \delta_{fn}$ :

$$i\hbar \dot{a}_f(t) = \sum_n \langle f | \hat{h}(t) | n \rangle e^{i\omega_{fn} t} a_n(t),$$

with  $\omega_{fn} = (E_f - E_n)/\hbar$ . (6.102)

In this relation, the time change (derivative) of the final state amplitude  $a_f(t)$  at time  $t$  depends, via the perturbation matrix elements  $\langle f | \hat{h} | n \rangle$ , on all other amplitudes at time  $t$ . In general, at this time all  $a_n(t)$  are non-vanishing. Their detailed form depends on the initial conditions of the problem and the type of the perturbation.

When we consider the zero approximation of the solution to (6.102),  $a_n$  is assumed to vanish and we obtain, as expected, stationary behavior; the state  $|f\rangle$  does not change in time.

In first approximation, we take into account from the whole variety of  $a_n$  only one single fixed initial state  $|i\rangle$  with  $a_i(t=0) = 1$ , as in zero order only time-independent states with  $a_n = \text{const}$  are given. Then, the first approximation is obtained as

$$\dot{a}_f(t) = \frac{-i}{\hbar} \langle f | \hat{h}(t) | i \rangle e^{i\omega_{fi} t}, \quad \text{with } \hbar\omega_{fi} = E_f - E_i, \quad (6.103)$$

and after integration

$$a_f(t) = \delta_{fi} - \frac{i}{\hbar} \int_0^t \langle f | \hat{h}(t') | i \rangle e^{i\omega_{fi} t'} dt'. \quad (6.104)$$

$\delta_{fi}$  is required as integration constant since in a stationary situation  $|i\rangle$  does not change and equals  $|f\rangle$ .

We will not treat higher order approximations in the present context but will rather consider one of the most important applications of the time-dependent perturbation method, namely the excitation of an atomic system by a periodic perturbation.

#### 6.4.1 Periodic Perturbation: Fermi's Golden Rule

A light wave incident on an atom, a molecule or a solid can excite electronic transitions between stationary quantum states if the photon energy matches the energetic difference between these states. The treatment of this problem is a predominantly important application of the time-dependent perturbation method.

In the simplest case, the electric field of the light wave causes a time and eventually position dependent perturbation of the potential energy which varies in time periodically as

$$\hat{h}(t) = \hat{h}_0 e^{-i\omega t}. \quad (6.105)$$

Here,  $\hat{h}_0$  might be a constant or a position-dependent operator  $\hat{h}_0(\mathbf{r})$ . According to (6.104), the change of the final state amplitude  $a_f(t)$  in time is calculated for a given initial state  $|i\rangle$  as:

$$a_f(t) = \frac{-i}{\hbar} \int_0^t \langle f | \hat{h}_0 | i \rangle e^{i(\omega_{fi} - \omega)t'} dt'. \quad (6.106a)$$

It is assumed that the system is in contact with the perturbation at  $t = 0$ . We then obtain

$$a_f(t) = -\frac{i}{\hbar} \langle f | \hat{h}_0 | i \rangle \frac{e^{i(\omega_{fi} - \omega)t} - 1}{i(\omega_{fi} - \omega)}. \quad (6.106b)$$

Using the relation

$$\begin{aligned} |e^{i\varphi} - 1|^2 &= |e^{i\varphi/2}(e^{i\varphi/2} - e^{-i\varphi/2})|^2 \\ &= \left(2 \sin \frac{\varphi}{2}\right)^2. \end{aligned} \quad (6.107)$$

Equation (6.106b) yields (by squaring) the transition probability from  $|i\rangle$  to  $|f\rangle$  at the time  $t$ :

$$W_{i \rightarrow f} = |a_f|^2 = \frac{1}{\hbar^2} |\langle f | \hat{h}_0 | i \rangle|^2 \left[ \frac{\sin((\omega_{fi} - \omega)t/2)}{(\omega_{fi} - \omega)t/2} \right]^2 t^2. \quad (6.108)$$

The function in squared brackets resembles the representation (4.80) of the  $\delta$  function in Sect. 4.3.4 if the time variable would be defined both for positive and negative values and the limes  $t \rightarrow \infty$  is considered.

In (6.108), however, only positive values  $t \geq 0$  are meaningful. Nevertheless, it must be noted that the function has a narrow peak at  $t = 0$ . The width of this peak might be estimated from the first zero point of the numerator appearing at  $(\omega_{fi} - \omega)t/2 = \pi$ . It is thus concluded that essentially those final states  $|f\rangle$  are reached from  $|i\rangle$  for which the following relations hold:

$$|(\omega_{fi} - \omega)t/2| \leq \pi, \quad \text{i.e.} \quad (6.109a)$$

$$E_f - E_i = \hbar\omega \pm 2\pi\hbar/t, \quad \text{or} \quad (6.109b)$$

$$E_f - E_i = \hbar\omega(1 \pm 2\pi/\omega t). \quad (6.109c)$$

For large times  $t$ , that is, long acting harmonic perturbations with frequency  $\omega$  (oscillation quantum  $\hbar\omega$ , Chap. 8) a transition between the states  $|i\rangle$  and  $|f\rangle$  is induced. For this transition, it is required:

$$E_f - E_i = \hbar\omega. \quad (6.109d)$$

The quantum energy  $\hbar\omega$  of the exciting vibration (perturbation) is consumed for overcoming the energy difference between initial and final state of the system.

It must be emphasized that (6.109d) is valid only for large time intervals during which the perturbation is in action. For small times  $t$  the system exhibits no particular preference for the level  $E_f = E_i + \hbar\omega$  (6.109d). The reason is simple: In the beginning, the system does not know that it is dealing with a periodic perturbation. It must wait a few cycles to get the message.

In order to derive an expression for the transition rate from  $|i\rangle$  to  $|f\rangle$ , we consider the action of a harmonic perturbation with long duration more in detail. The perturbation  $\hat{h}$  is assumed to interact with the system during the time interval  $-\tau/2 < t < \tau/2$  with  $\tau \rightarrow \infty$ , i.e. the probability amplitude  $a_f$  for the final state follows as

$$a_f = \frac{-i}{\hbar} \lim_{\tau \rightarrow \infty} \int_{-\tau/2}^{\tau/2} \langle f | \hat{h}_0 | i \rangle e^{i(\omega_{fi} - \omega)t} dt. \quad (6.110a)$$

The transition probability is obtained as the square of (6.110a):

$$W_{fi} = |a_f|^2 = \frac{1}{\hbar^2} |\langle f | \hat{h}_0 | i \rangle|^2 \lim_{\tau \rightarrow \infty} \int_{-\tau/2}^{\tau/2} e^{i(\omega_{fi} - \omega)t} dt \int_{-\tau/2}^{\tau/2} e^{i(\omega_{fi} - \omega)t'} dt'. \quad (6.110b)$$

This probability, of course, grows with increasing time  $\tau$  during which the system is exposed to the perturbation. The time-independent transition rate (probability per time)  $W_{fi}/\tau$  is, therefore, the more interesting quantity. For its calculation, we must evaluate, how the right side of (6.110b) depends on the time variable  $\tau$ .

Because of the representation of the  $\delta$  function (Sect. 4.3.4) the first integral in (6.110b) is identical, for  $\tau \rightarrow \infty$ , with  $\delta(\omega_{fi} - \omega)$ , apart from a factor  $2\pi$ . This  $\delta$  function has non-vanishing values only for  $\omega_{fi} = \omega$ , that is, if we set  $\omega_{fi} = \omega$  in the integral (approximation for the  $\delta$  function), we obtain a factor  $\tau$  in front of the second integral by integration. The second integral approaches the function  $\delta(\omega_{fi} - \omega)$  just as well. The approximate calculation of the product of the two  $\delta$  functions in (6.110b), therefore, yields the following transition rate  $R_{fi}$  from state  $|i\rangle$  into state  $|f\rangle$ :

$$R_{fi} = \frac{W_{fi}}{\tau} = \frac{2\pi}{\hbar^2} |\langle f | \hat{h}_0 | i \rangle|^2 \delta(\omega_{fi} - \omega). \quad (6.111a)$$

Expressing the transition frequency  $\omega_{fi}$  in terms of the energies  $E_f$  and  $E_i$  of the initial and final states, respectively, and using the relation  $\delta(ax) = a^{-1}\delta(x)$  for  $\delta$  functions we obtain for the transition rate

$$R_{fi} = \frac{W_{fi}}{\tau} = \frac{2\pi}{\hbar} |\langle f | \hat{h}_0 | i \rangle|^2 \delta(E_f - E_i - \hbar\omega). \quad (6.111b)$$

This relation for the calculation of transition rates between stationary quantum states is of predominant importance for application; according to its inventor it is called *Fermi's golden rule* [7]. One must not worry about the  $\delta$  function with its infinite

values. In all applications, integral expressions of (6.111b) enter the calculation and the result will depend on the finite area under the  $\delta$  function.

For sufficiently long perturbation time ( $\tau \rightarrow \infty$ ), the  $\delta$  function in (6.111b) guarantees the equivalence of the quantum energy  $\hbar\omega$  of the exciting harmonic perturbation (oscillation) with the difference between final and initial state energies. This energy difference  $E_f - E_i$  which is necessary for the electronic transition is supplied by the exciting light field. As we will see in Chap. 8, the electromagnetic field is described in quantum field theory as being built-up by light particles, the photons, which carry the photon energy  $\hbar\omega$ . According to (6.111a), (6.111b) one photon is able to induce the transition and the energy of the total system electron plus light field is conserved. In this quantized field picture (Chap. 8), the energy conservation during the electronic transition can, then, be expressed in terms of a  $\delta$  function  $\delta(E_f^{\text{tot}} - E_i^{\text{tot}})$  where  $E_f^{\text{tot}}$  and  $E_i^{\text{tot}}$  are the final and initial energies of the total system electron plus photon.

### 6.4.2 Electron–Light Interaction: Optical Transitions

Fermi's golden rule (6.111a), (6.111b) shall now be applied to the calculation of light induced transition rates between electronic quantum states in matter. In the simplest case, we consider an electron in an atom or in a quantum dot which interacts with electromagnetic radiation (light). The electron is in a potential  $V(r)$ , be it an atomic potential of the nucleus or the well potential of a quantum dot. The electron can occupy the discrete energy levels of a binding potential. According to Sect. 5.4.3 (5.67) the Hamiltonian of the electron in the surrounding light field described in terms of its vector potential operator  $\hat{\mathbf{A}}(\mathbf{r}, t)$  is, thus, given by

$$\hat{H} = \frac{\hat{\mathbf{p}}^2}{2m} + V(\mathbf{r}) - \frac{e}{2m}(2\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}). \quad (6.112)$$

In this approximate expression, only terms linear in  $\hat{\mathbf{A}}$  (small perturbation) are taken into account. With a positive elementary charge  $e > 0$  the electron carries the negative charge  $-e$ . For common electromagnetic fields, the last term in (6.112) is considered as a time-dependent perturbation:

$$\hat{h} = -\frac{e}{m}\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}. \quad (6.113)$$

The exciting light wave is described by its electric field

$$\mathcal{E}(\mathbf{r}, t) = 2e\mathcal{E}_0 \cos(\mathbf{q} \cdot \mathbf{r} - \omega t), \quad (6.114a)$$

with  $\mathbf{e}$  as unit vector oriented perpendicular to the light wave vector  $\mathbf{q}$ .

Because of  $\mathcal{E} = -\partial \mathbf{A} / \partial t$  the vector potential of the light field is

$$\mathbf{A}(\mathbf{r}, t) = (2e\mathcal{E}_0/\omega) \sin(\mathbf{q} \cdot \mathbf{r} - \omega t). \quad (6.114b)$$

Using the exponential representation of the sine function, we obtain the perturbation (6.113) as

$$\hat{h} = -\frac{e\mathcal{E}_0}{im\omega} [\mathrm{e}^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} - \mathrm{e}^{-i(\mathbf{q} \cdot \mathbf{r} - \omega t)}](\mathbf{e} \cdot \hat{\mathbf{p}}). \quad (6.115)$$

For the derivation of Fermi's golden rule, the perturbation had been assumed to have a time dependence  $\hat{h}(t) = \hat{h}_0 \exp(-i\omega t)$ . Accordingly, (6.115) yields two different perturbation terms, one with  $\omega > 0$  and the other one with  $\omega < 0$ . Two different transition rates, are, therefore obtained:

$$R_{fi}^{(1)} = \frac{2\pi}{\hbar} \left( \frac{e\mathcal{E}_0}{m\omega} \right)^2 |\langle f | \mathbf{e} \cdot \hat{\mathbf{p}} | i \rangle|^2 \delta(E_f - E_i - \hbar\omega), \quad (6.116a)$$

$$R_{fi}^{(2)} = \frac{2\pi}{\hbar} \left( \frac{e\mathcal{E}_0}{m\omega} \right)^2 |\langle f | \mathbf{e} \cdot \hat{\mathbf{p}} | i \rangle|^2 \delta(E_f - E_i + \hbar\omega). \quad (6.116b)$$

Both rates have the same absolute value but the  $\delta$  functions describing the energy conservation during the transition are different. In (6.116a), the energy balance reads

$$E_f = E_i + \hbar\omega, \quad (6.117a)$$

while (6.116b) yields

$$E_f = E_i - \hbar\omega. \quad (6.117b)$$

In the first case (6.117a), the final state  $|f\rangle$  is reached by addition of one photon energy quantum  $\hbar\omega$  to the initial state  $|i\rangle$ . The light quantum (photon) is destroyed during the transition  $|i\rangle \rightarrow |f\rangle$ ; it is absorbed by the electron. We are concerned here with *optical absorption* by excitation of the state  $|f\rangle$ . In the second case (6.117b), the initial state energy  $E_i$  is decreased by the photon energy  $\hbar\omega$  to reach the final state with energy  $E_f$ . The initial state  $|i\rangle$  is de-excited into  $|f\rangle$  and light with photon energy  $\hbar\omega$  is emitted by the electron. This is the inverse effect to optical absorption; it is called *optical emission*.

Note that this emission process is stimulated by irradiated photons, that is, a perturbation by the external light field. In Sect. 8.2.2 we will see that the classical description of the light field in terms of field variables  $\mathbf{A}(\mathbf{r}, t)$  and  $\mathcal{E}(\mathbf{r}, t)$  is not complete in the quantum mechanical sense. At a fixed position in space, the light field appears as harmonic oscillations of the field vectors. These oscillations obey the laws of quantum mechanics and must be quantized according to the rules for the harmonic oscillator (Sect. 4.4). The energies of the oscillations assume discrete values including a non-vanishing ground state energy. As a consequence, an electron in a surrounding radiation field interacts also with the ground state of the field (ground state means: no macroscopic external field) and photon emission is induced even without stimulation by an external light field. The electron being in an excited state is de-excited into the ground state of the field by emission of a photon. This process is called *spontaneous*

*emission* since no external light field is necessary. More details about these emission processes are presented in Chap. 8 in the general context of field quantization.

Light emission by de-excitation of excited electronic quantum states is the physical basis for such important applications as lasers and light emitting diodes (LED).

For many applications, a more descriptive representation of the transition matrix elements  $\langle f | \mathbf{e} \cdot \hat{\mathbf{p}} | i \rangle$  in the transition rates (6.116a), (6.116b) is useful. For its derivation, we use the quantum mechanical dynamic equation in the Heisenberg representation (Sect. 4.3.5). In this formalism, the time derivative of the position operator  $\hat{\mathbf{r}}$ , that is, essentially the momentum is given by

$$i\hbar \dot{\hat{\mathbf{r}}} = i\hbar \hat{\mathbf{p}}/m = [\hat{\mathbf{r}}, \hat{H}] = -[\hat{H}, \hat{\mathbf{r}}], \quad (6.118a)$$

that is,

$$\hat{\mathbf{p}} = i\frac{m}{\hbar}[\hat{H}, \hat{\mathbf{r}}]. \quad (6.118b)$$

The essential term of the matrix element (6.116a), (6.116b) is thus obtained as

$$\begin{aligned} \langle f | \hat{\mathbf{p}} | i \rangle &= i\frac{m}{\hbar} [\langle f | \hat{H} \hat{\mathbf{r}} | i \rangle - \langle f | \hat{\mathbf{r}} \hat{H} | i \rangle] \\ &= i\frac{m}{\hbar} (E_f - E_i) \langle f | \hat{\mathbf{r}} | i \rangle \\ &= im\omega_{fi} \langle f | \hat{\mathbf{r}} | i \rangle. \end{aligned} \quad (6.119)$$

The transition rates for absorption (6.116a), respectively, stimulated emission (6.116b) of a light quantum (photon) can then be written as

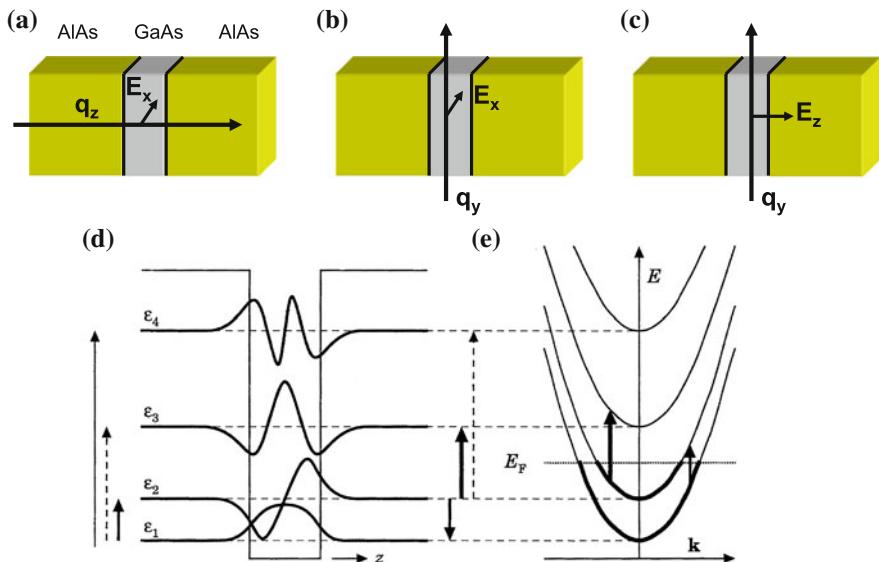
$$R_{fi}^{(1)} = \frac{2\pi}{\hbar} \left( \frac{\omega_{fi}}{\omega} \right)^2 |\langle f | \mathbf{e} \mathcal{E}_0 \mathbf{e} \cdot (\mathbf{e} \hat{\mathbf{r}}) | i \rangle|^2 \delta(E_f - E_i - \hbar\omega). \quad (6.120)$$

In this representation, the perturbation operator emerges as the product of the electric field amplitude  $\mathbf{e} \mathcal{E}_0$  and the dipole moment  $e\hat{\mathbf{r}}$  related to the oscillating electron (between final and initial state). This is nothing else but the energy of a dipole in an oscillating electric field. We could have guessed this energy perturbation in the Hamiltonian already by classical analogy, without the quantum mechanical derivation via the magnetic vector potential operator. Quantum state transitions induced by an electromagnetic field, thus, require the existence of electric dipole moments being connected with the transitions. This is at least true for the so-called *dipole approximation* considered here.

### 6.4.3 Optical Absorption and Emission in a Quantum Well

A simple interesting example of optical transitions in nanostructures is excitations between confined electronic states in a semiconductor quantum well. Such a quantum well can be realized by a GaAs layer (band gap  $\approx 1.4$  eV) with a thickness between 10 and 100 nm epitaxially grown between two AlAs layers (band gap  $\approx 2.2$  eV). The structure is qualitatively shown in Fig. 6.11a. According to Sect. 3.6, the eigen-solutions of the Schrödinger equation within the 2D quantum well are products of confined bound states  $\varphi_i(z)$  and plane waves [see, e.g., (3.72)] which describe the free motion of an electron in the quantum well parallel to the layer sequence of the structure. With  $z$  as the coordinate normal to the layer sequence and  $x$  and  $y$  the coordinates parallel to the quantum well plane (Fig. 6.11), the eigensolutions are represented as

$$\psi_{i\mathbf{k}} = \langle \mathbf{r}|i\mathbf{k}\rangle = C\varphi_i(z)e^{i\mathbf{k}\cdot\mathbf{r}_{\parallel}}, \quad (6.121)$$



**Fig. 6.11** a–e Optical absorption due to electronic transitions in a 1D quantum well of GaAs embedded between AlAs barriers. Electronic transitions between occupied and empty states (energetically separated by Fermi energy  $E_F$ ) of the 2D sub-bands are considered. a–c Possible relative orientations of light wave vector  $\mathbf{q}$  and electric field  $\mathbf{E} = (E_x, E_y, E_z)$  of the light wave, respectively, with regard to the layer sequence of the AlAs/GaAs heterostructure, which give rise to optical transitions allowed in the dipole approximation. d Wave functions along the  $z$  coordinate, i.e. the layer sequence of the quantum well, and corresponding energy levels  $\varepsilon_1$  to  $\varepsilon_4$ . The line thickness of the arrows is a qualitative measure of the oscillator strength; arrows in broken line indicate forbidden transitions. e Sub-band structure of the electronic states with  $\mathbf{k}$  vector parallel to the quantum well layer, i.e.  $\parallel x, y$  plane. Allowed transitions are perpendicular in this plot

$\mathbf{r}_{\parallel}$  is a position vector in the plane of the quantum well. In case of infinitely high energetic walls  $\varphi_i(z)$  are sine functions of the type (6.90) and according to Sect. 3.6.1 the energies  $E_i(k)$  of the electronic states in the well are

$$E_i(\mathbf{k}_{\parallel}) = \varepsilon_i + \frac{\hbar^2 k_{\parallel}^2}{2m}, \quad (6.122)$$

$\varepsilon_i$  are the energies of the confined bound states and the second term on the right side describes the kinetic energy ( $\propto k^2$ ) of an electron moving freely along  $\mathbf{r}_{\parallel}$ .  $E_i(\mathbf{k}_{\parallel})$  form a sequence of parabolas along an electronic wave vector  $\mathbf{k}_{\parallel} = (k_x, k_y)$  of the 2D reciprocal space parallel to the plane of the quantum well (Fig. 6.11e). These energy parabolas are called sub-bands of the quantum well.

Assuming the AlAs barriers to be doped to such an extent, that the Fermi level is located between the energies  $\varepsilon_2$  and  $\varepsilon_3$  (Fig. 6.11e), an electron can be excited from the occupied levels  $\varepsilon_1$  and  $\varepsilon_2$  into higher empty levels  $\varepsilon_3, \varepsilon_4$  etc. Subsequently, the electron might be de-excited into energetically lower states. These processes correspond to absorption and emission, respectively, of light quanta. The related transition rates are calculated according to (6.116a), (6.116b) and (6.120), respectively.

It is straightforwardly seen that the calculation of the transition matrix elements  $\langle j, \mathbf{k}' | \mathbf{e} \cdot \hat{\mathbf{p}} | i, \mathbf{k} \rangle$  between final and initial states  $|j, \mathbf{k}'\rangle, |i, \mathbf{k}\rangle$  sensitively depends on the orientation of the electric field  $\mathcal{E}$ , that is, the unity vector  $\mathbf{e}$  of the light polarization. It is important whether the light polarization, that is, the vector  $\mathbf{e}$  is oriented parallel or perpendicular to the plane of the quantum well (Fig. 6.11a–c).

At first, we assume a polarization direction parallel to  $x$ , that is, parallel to the quantum well plane (Fig. 6.11a, b). The polarization unity vector is, then, written as  $\mathbf{e} = (1, 0, 0)$ . The light wave can propagate in this case in  $y$  or in  $z$  direction, that is, within the plane of the well or perpendicular to the layer sequence. The matrix element (6.116a), (6.116b) is represented as

$$\mathbf{e} \cdot \hat{\mathbf{p}} = -i\hbar \frac{\partial}{\partial x}. \quad (6.123a)$$

The momentum operator  $\hat{\mathbf{p}}$  is essentially a derivation with regard to the  $x$  coordinate; it does not affect the bound part  $\varphi_i(z)$  of the wave function (6.121). From the orthogonality condition, follows

$$\langle j \mathbf{k}' | \mathbf{e} \cdot \hat{\mathbf{p}} | i \mathbf{k} \rangle = \hbar k_x \langle j \mathbf{k}' | i \mathbf{k} \rangle = 0. \quad (6.123b)$$

This relation has an important consequence: Light irradiation with a polarization direction in the plane of the quantum well, that is, parallel to the layer sequence of the AlAs/GaAs double-heterostructure, is not connected with any absorption or stimulated emission of light. This is true for light propagation (direction of wave vector  $\mathbf{q}$ ) both parallel and perpendicular to the layer sequence (Fig. 6.11a, b).

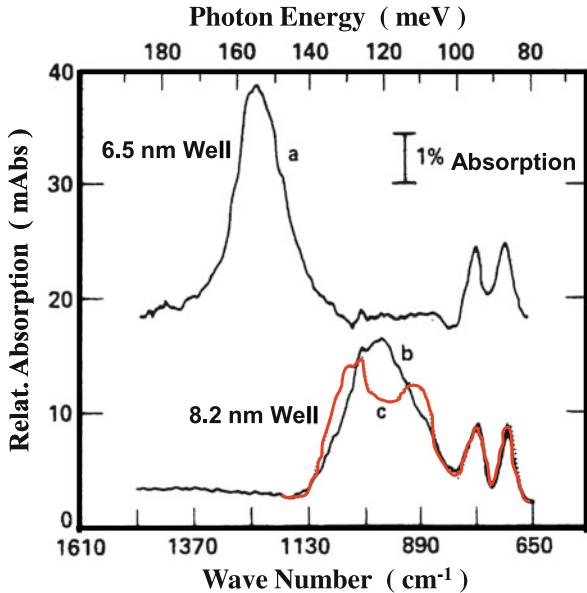
Let us now consider the case of light polarization normal to the layer sequence [ $\mathcal{E} = (0, 0, E_z)$ ] and propagation direction parallel to the well (Fig. 6.11c). The transition matrix element is, then, obtained as

$$\begin{aligned} \langle j\mathbf{k}'|\mathbf{e} \cdot \hat{\mathbf{p}}|i\mathbf{k}\rangle &= |C|^2 \int dz \int dx dy \varphi_j^*(z) e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{r}_\parallel} \hat{p}_z \varphi_i(z) \\ &= |C|^2 \int dz \varphi_j^*(z) \hat{p}_z \varphi_i(z) \int dx dy e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{r}_\parallel}. \end{aligned} \quad (6.124)$$

The second integral on the right side over  $dx$  and  $dy$  yields a  $\delta$  function, that is, it vanishes for all electron wave vectors except for  $\mathbf{k} = \mathbf{k}'$ . The electron wave vector in the final state must equal that in the initial state.

We summarize: optical transitions between electronic states in a quantum well, that is, light absorption or emission do only occur with light polarization having an electric field component  $E_z$  normal to quantum well plane. Furthermore, the electron wave number  $\mathbf{k}$  is conserved in the transition. These are “vertical” transitions in the band scheme of the electronic sub-bands (Fig. 6.11e). The light photon energies (frequencies), thus, equal the energetic distance between the sub-bands. In spite of the continuous spectrum of sub-bands (parabolas) the optical absorption and emission spectra of a quantum well consist of discrete sharp bands (Fig. 6.12). Their energetic location is directly related to the width of the quantum well (Sect. 3.6.1).

**Fig. 6.12** a–c Optical absorption measured at  $T = 300$  K on a sequence of 50 quantum wells of GaAs embedded between AlAs. The electronic transitions occur between sub-bands of conduction band quantum wells in GaAs [10]. **a** Quantum well thickness 6.5 nm, **(b)** and **(c)** quantum well thickness 8.2 nm



We can rewrite the remaining matrix element in (6.124) as

$$\langle j | \hat{p}_z | i \rangle = \int dz \varphi_j^*(z) \hat{p}_z \varphi_i(z) = im\omega_{ji} \langle j | \hat{\mathbf{z}} | i \rangle. \quad (6.125a)$$

We have used the dipole representation (6.119) with  $\hbar\omega_{ji}$  as the energy difference between the sub-bands  $E_j$  and  $E_i$ . From the representation of the matrix element

$$\langle j | \hat{\mathbf{r}} | i \rangle = \int dz \varphi_j^*(z) z \varphi_i(z) \quad (6.125b)$$

a further restriction for the occurrence of optical transitions follows. We assume the zero point of the  $z$  axis in the integral of (6.125b) to be in the center of the quantum well. The integral, then, separates into two parts, one where  $\varphi_i(z)$  is multiplied by positive  $z$  values, the other one where  $\varphi_i(z)$  is multiplied by negative  $z$  values. The integral in (6.125b) does not vanish, therefore, only if  $\varphi_i(z)$  or  $\varphi_j(z)$  change their sign upon reflection at the zero point of the  $z$  axis. This property is given for example, for the  $\sin(z)$  function. Wave functions which keep their sign upon this reflection operation are called functions of *even parity*. Functions which change their sign have *odd parity*. According to (6.125a), (6.125b), light induced optical transitions are only possible between electronic states of different parity, for example, from  $i = 1$  to  $j = 2, 4, \dots$ , but not to  $j = 3, 5, \dots$  (Fig. 6.11d). This result, of course, is only valid for quantum wells with a symmetric potential  $V(z) = V(-z)$ , in which the eigenstates separate into those with even and those with odd parity. In order to lift this *parity selection rule*, one must prepare asymmetric quantum wells.

For better measurability of the optical transitions, the absorption is enhanced by repetition of one and the same quantum well structure in the epitaxial multi-layer sandwich. In Fig. 6.12, results of an optical absorption measurement on a layer stack of 50 GaAs quantum wells, each single one with a thickness of 6.5 nm or 8.2 nm are shown. Relatively sharp absorption bands as expected from theory are observed. Furthermore, the absorption line at higher photon energy results from the thinner quantum well. The thinner the well the more the sub-bands are energetically spread.

#### 6.4.4 Dipole Selection Rules for Angular Momentum States

According to Sect. 6.4.3, the symmetry of the final and initial states as well as the light polarization direction are essential factors for the intensity of optical transitions. They determine whether particular optical transitions actually couple to the electromagnetic radiation. So-called *selection rules* tell us whether under particular state symmetries and irradiation geometries the transition matrix elements  $\langle f | \mathbf{r} | i \rangle$

(6.125b) vanish, that is, that the corresponding optical transitions are forbidden. In approximations higher than the dipole approximation considered here quadrupole moment transitions, of course, can occur.

Within the frame of the dipole approximation, it is useful to distinguish between allowed and forbidden transitions solely on the basis of selection rules. This saves calculation work in obtaining the transition rates. We have done this in Sect. 6.4.3 already for the potential box.

An important class of problems reaching from atoms to quantum dots (Sect. 5.7.1) and quantum rings (Sect. 5.7.2) is based on potentials with rotational or spherical symmetry (Sect. 5.3). In this case, the eigensolutions of the problem in position representation are obtained as a product of a radial part  $R_{n,l}(r)$  and the spherical harmonics  $\Upsilon_l^m(\vartheta, \varphi)$ , the eigenfunctions of the angular momentum operators  $\hat{L}^2, \hat{L}_z$  (Sect. 5.3):

$$\langle \mathbf{r}|n, l, m\rangle = R_{n,l}(r)\Upsilon_l^m(\vartheta, \varphi). \quad (6.126a)$$

The spin degree of freedom is not considered here. It is therefore interesting to find out eventual selection rules, which allow or forbid particular optical transitions between angular momentum states with different quantum numbers  $m$  and  $l$ . For this purpose, we consider dipole transitions between the angular momentum states

$$\langle \mathbf{r}|l, m\rangle = \Upsilon_l^m(\vartheta, \varphi), \quad (6.126b)$$

and calculate transition matrix elements of the type  $\langle l', m'|\mathbf{r}|l, m\rangle$ .

To derive selection rules for the orientation quantum number  $m$ , that is, for transitions between states with different orientation of the angular momentum in space, we must calculate the matrix element  $\langle l', m'|z|l, m\rangle$ , where  $z$  is the particular direction in space determined for example, by a magnetic field. The interesting eigenstates are therefore those of the operator  $\hat{L}_z$ . It seems useful in this context to consider the commutator  $[\hat{L}_z, \hat{z}]$  which contains both the angular momentum and the position operator. From

$$\hat{L}_z = (\mathbf{r} \times \hat{\mathbf{p}})_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x \quad (6.127a)$$

and

$$[\hat{x}\hat{p}_y - \hat{y}\hat{p}_x, \hat{z}] = 0 \quad (6.127b)$$

we conclude

$$[\hat{L}_z, \hat{z}] = 0. \quad (6.127c)$$

The matrix element of the commutator (6.127c), then, is obtained as

$$\begin{aligned} 0 &= \langle l', m' | [\hat{L}_z, \hat{z}] | l, m \rangle = \langle l', m' | \hat{L}_z \hat{z} | l, m \rangle - \langle l', m' | \hat{z} \hat{L}_z | l, m \rangle \\ &= m' \langle l', m' | \hat{z} | l, m \rangle - m \langle l', m' | \hat{z} | l, m \rangle. \end{aligned} \quad (6.127d)$$

This yields the following interesting relation for the interesting transition matrix element:

$$\langle l', m' | \hat{z} | l, m \rangle (m' - m) = 0. \quad (6.128)$$

Consequently, only for  $m' = m$  the transition matrix elements for an electronic dipole moment in  $z$  direction does not vanish. Optical transitions in which the electronic transition dipole oscillates in  $z$  direction require conservation of the orientation quantum number  $m$ .

For optical transitions with electronic dipole orientation normal to  $z$ , that is, parallel to  $x$ ,  $y$ , we analogously consider the commutator  $[\hat{L}_z, \hat{x}]$ :

$$\begin{aligned} [\hat{L}_z, \hat{x}] &= -[\hat{y} \hat{p}_x, \hat{x}] = -\hat{y} \hat{p}_x, \hat{x} + \hat{x} \hat{y} \hat{p}_x \\ &= -\hat{y} (\hat{x} \hat{p}_x + i\hbar) + \hat{x} \hat{y} \hat{p}_x \\ &= -i\hbar \hat{y}. \end{aligned} \quad (6.129a)$$

Taking into account also the  $y$  component yields

$$[\hat{L}_z, \hat{x} \pm i\hat{y}] = \pm(\hat{x} \pm i\hat{y})\hbar. \quad (6.129b)$$

The calculation of the matrix elements of this commutator in the basis of the angular momentum states leads to the relation

$$\langle l', m' | [\hat{L}_z, \hat{x} \pm i\hat{y}] | l, m \rangle = \langle l', m' | \pm(\hat{x} \pm i\hat{y}) | l, m \rangle \hbar, \quad (6.130a)$$

and finally to

$$\begin{aligned} \langle l', m' | \hat{L}_z (\hat{x} \pm i\hat{y}) | l, m \rangle - \langle l', m' | \pm(\hat{x} \pm i\hat{y}) \hat{L}_z | l, m \rangle \\ = (m' - m)\hbar \langle l', m' | (\hat{x} \pm i\hat{y}) | l, m \rangle = \langle l', m' | \pm(\hat{x} \pm i\hat{y}) | l, m \rangle \hbar. \end{aligned} \quad (6.130b)$$

Comparing the last two equations in (6.130b) we must require  $m' - m = \pm 1$  for non-vanishing matrix elements of dipoles in  $x$ , respectively,  $y$  direction.

We keep in mind: For optical dipole transitions between angular momentum states  $|l, m\rangle$  the orientation quantum number  $m$  must obey the following selection rules:

$$m' - m = 0 \quad \text{for dipole in } z\text{-direction}, \quad (6.131a)$$

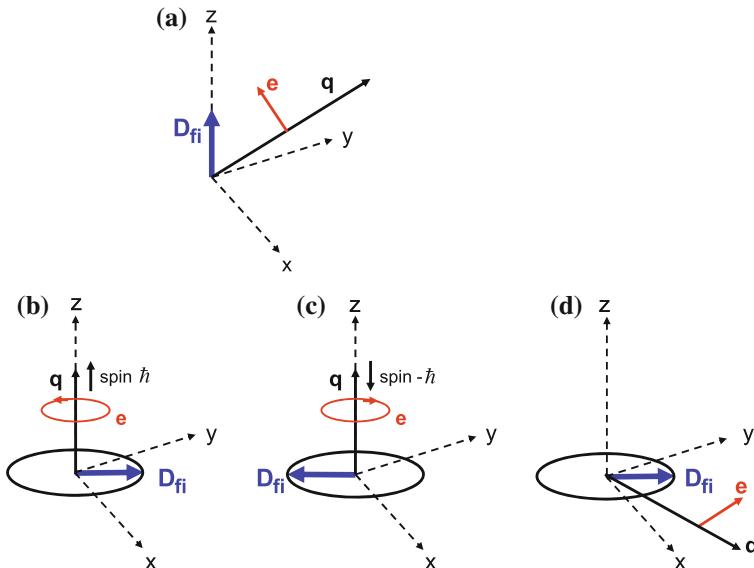
$$m' - m = \pm 1 \quad \text{for dipole in } x, y\text{-direction}. \quad (6.131b)$$

The orientation quantum number must be conserved or can change by  $\pm 1$ . Both cases belong to different emission or absorption geometries regarding light polarization and propagation direction.

According to (6.120), the transition rate  $R_{fi}$  is different from zero only if the scalar product of light polarization direction  $\mathbf{e}$  (oscillation direction of electric field) and electronic transition dipole  $\mathbf{D}_{fi} = \langle f | e\mathbf{r} | i \rangle$  does not vanish.

Optical transitions with conservation of the orientation quantum number ( $m' = m$ ) and dipole moment  $\mathbf{D}_{fi}$  oriented in  $z$  direction, thus, require a light polarization with an electric field component parallel to  $\mathbf{D}_{fi}$ , that is, along  $z$ . The light propagation direction (emitted or absorbed) described by the wave vector  $\mathbf{q} \perp \mathbf{e}$  can not be directed parallel to  $\mathbf{D}_{fi}$ , that is, along  $z$ ; or in other words: Light irradiated along the  $z$  direction can not excite dipole transitions with an electric dipole moment parallel  $z$  and such transitions can not cause light emission in  $z$  direction (Fig. 6.13a).

Let us now consider the case of transition matrix elements  $\mathbf{D}_{fi} = \langle f | e\mathbf{r}| i \rangle$  with orientation in the  $x$ - $y$  plane (Fig. 6.13b–d). Here, the selection rule  $m' - m = \pm 1$  is valid. In this geometry, the light polarization  $\mathcal{E}_0\mathbf{e}$  must have a component in the  $x$ - $y$  plane. This polarization might be given both for light propagation in  $z$  direction ( $\mathbf{q} \parallel z$ ) (Fig. 6.13b, c) and for propagation within the  $x$ - $y$  plane (Fig. 6.13d). In any case the electric field  $\mathcal{E}_0\mathbf{e}$  of the light wave must have a component parallel to the transition dipole  $\mathbf{D}_{fi}$  oriented within the  $x$ - $y$  plane.



**Fig. 6.13 a–d** Schematic representation of the dipole selection rules for optical transitions between angular momentum states. Optical absorption and emission in dipole approximation is possible, **a** if the electric field of the light with wave vector  $\mathbf{q}$  and oscillation direction  $\mathbf{e}$  (unity polarization vector) has a vector component in the direction of the transition dipole  $\mathbf{D}_{fi}$  mediating between initial and final states of the electron. **b, c** If a circularly polarized light field propagates with a wave vector  $\mathbf{q}$  normal to the transition dipole  $\mathbf{D}_{fi}$ . Depending on the light polarization direction, left hand or right hand ( $\mathbf{e}$  unity polarization vector) the photon spin is directed in positive or negative  $z$  direction. **d** If a linearly polarized light field propagates with a wave vector  $\mathbf{q}$  in the  $x$ ,  $y$  plane containing the transition dipole  $\mathbf{D}_{fi}$  and if its electric field (directional unity vector  $\mathbf{e}$ ) has a component in  $\mathbf{D}_{fi}$  direction

In the first case of light propagation parallel to the  $x$  axis, we conclude from (6.130b):

$$\langle l', m' | \hat{x} | l, m \rangle = \pm i \langle l', m' | \hat{y} | l, m \rangle, \quad (6.132a)$$

and finally for the orientation of the dipole matrix element:

$$\mathbf{D}_{fi} \propto \begin{pmatrix} 1 \\ \pm i \\ 0 \end{pmatrix}. \quad (6.132b)$$

Due to (6.132a), (6.132b) and due to the dependence of the transition rate  $R_{fi}$  (6.120) on the scalar product  $\mathbf{e} \cdot \mathbf{D}_{fi}$ , the emitted light has left-hand or right-hand circular polarization for light propagation along  $z$  (Fig. 6.13b, c). From this discussion, we learn an important fact: Since for the described transition the electronic angular momentum in  $z$  direction changes by  $\Delta m = \pm 1$ , and since the total angular momentum, however, is conserved, the emitted or absorbed photon participating in the process must carry the spin (angular momentum)  $\hbar$ . This result, that is, the bosonic character of photons, was already mentioned in Sect. 6.4.1 without any detailed proof. The conclusion is, of course, based on the fact that the electron being involved in the transition does not change its spin state. A photon with right-hand circular polarization carries a spin  $+\hbar$  in propagation direction while left-hand polarization is identical with the photon spin  $-\hbar$ , that is, a spin direction opposite to the propagation direction. Linearly polarized light waves are described by a superposition of waves having left-hand and right-hand circular polarization, that is, of photons with spins  $+\hbar$  and  $-\hbar$ .

In case of a transition matrix element  $\mathbf{D}_{fi}$  oscillating in the  $x-y$  plane and light propagation in this plane, too, the propagation direction (light wave vector  $\mathbf{q}$ ) and the vector  $\mathbf{D}_{fi}$  must have a certain finite angle between each other for a non-vanishing scalar product  $\mathbf{e} \cdot \mathbf{D}_{fi}$  (Fig. 6.13d). The light interacting with the electron, be it in absorption or emission, is then polarized in the  $x-y$  plane ( $\mathbf{e} \perp \mathbf{q}$ ). Light waves emitted under a certain finite angle with the  $z$  axis and the  $x-y$  plane are elliptically polarized.

It should be mentioned again, that linearly polarized light as in Fig. 6.13a, b can be represented as a superposition of right- and left-hand circularly polarized light. According to Fig. 6.13b,c this is a superposition state of photon spins  $\pm \hbar$ , that is, a photon state with zero spin. In interaction processes as in Fig. 6.13a, therefore, the spin state of the electron is conserved because of the requirement  $\Delta m = 0$ .

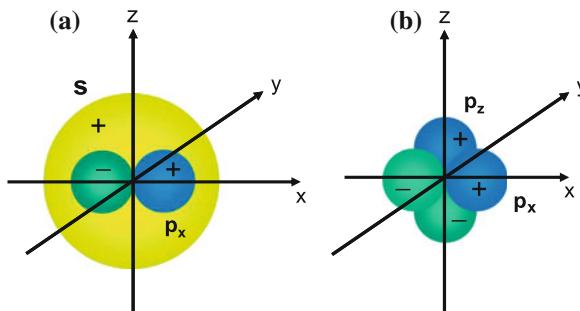
The considerations so far were concerned only with changes of the orientation quantum number  $m$ , that is, with changes of the angular momentum orientation of the electron upon optical dipole transitions. The question arises if also the total angular momentum of an electron described by the operator  $\hat{L}^2$  obeys certain selection rules in optical dipole transitions. To answer this question we could, in analogy to (6.127a)–(6.130b), calculate commutators between the operators  $\hat{L}^2$  and  $\mathbf{r}$ . This would yield information about vanishing or non-vanishing of the matrix elements  $\langle l', m' | \hat{\mathbf{r}} | l, m \rangle$  upon certain changes of the quantum number  $l$ . One could also calculate the dipole matrix elements with the angular momentum eigenfunctions  $\gamma_l^m$

to check their behavior for particular  $l, l'$  combinations. These procedures lead into complex and tedious formal calculations.

Instead, we present an illustrative example of optical transitions between  $s(l=0)$  and  $p(l=1)$  states. The graphic description yields an illustrative picture of the angular momentum selection rules and it can be transferred to many other similar problems. We assume a transition between a spherical  $s$  state with  $l=0$  and a  $p_x$  state with  $l=1$  which is induced by a dipole matrix element  $\mathbf{D}_{fi}$  oriented along the  $x$  axis (e.g., for light propagation along  $z$ ). The situation is shown in Fig. 6.14a. In the relevant matrix element  $\langle l=1, m|x|l=0, m \rangle$ , the initial state  $|s, m\rangle = |l=0, m\rangle$  has spherical symmetry while the final state  $|p_x, m\rangle = |l=1, m\rangle$  has rotational symmetry around the  $x$  axis with opposite sign of the wave function for positive and negative  $x$  values. We split the transition matrix element into two parts, one with positive  $x$  values right of the  $z$  axis, the other one left of it with negative  $x$  values (Fig. 6.14a):

$$\begin{aligned}\langle p_x, m|x|s, m\rangle &= \int_{x>0} \psi_{px}^* x \psi_s d^3r + \int_{x<0} \psi_{px}^* x \psi_s d^3r \\ &= \int_{x>0} \psi_{px}^* x \psi_s d^3r + \int_{x>0} (-\psi_{px}^*)(-x) \psi_s d^3r \\ &= 2 \int_{x>0} \psi_{px}^* x \psi_s d^3r.\end{aligned}\quad (6.133)$$

In the second partial integral, we have transformed  $x \rightarrow -x$  and have used the antisymmetry relation  $\psi_{px}(-x) = -\psi_{px}(x)$ . According to (6.133) the matrix element does not vanish even if the total wave function including also the radial part



**Fig. 6.14 a, b** Qualitative three-dimensional representation of the energetically lowest  $s$ ,  $p_x$ ,  $p_y$ ,  $p_z$  orbitals with sign of the wave functions in the different spatial regions (see also Fig. 5.5). **a** The symmetry of the orbitals  $s$  ( $l=0$ ) and  $p_x$  ( $l=0$ ) allows optical transitions  $s \leftrightarrow p_x$  in dipole approximation. These transitions correspond to a change of the angular momentum quantum number  $\Delta l = \pm 1$ . **b** Because of the symmetry of the  $p_x$  ( $l=1$ ) orbitals transitions of the type  $p_x \leftrightarrow p_z$  are forbidden in dipole approximation. In dipole approximation transitions with  $\Delta l = 0$  have a negligible transition dipole moment  $\mathbf{D}_{fi}$

$R_{n,l}(r)$  (6.126a) exhibits several radial node lines due to higher principal quantum numbers  $n$ . The optical dipole transition  $s \leftrightarrow p$ , be it in absorption or in emission, is not forbidden.

Now we consider a transition with dipole moment along  $x$  between two states with equal angular momentum quantum number  $l$ , but differing angular momentum orientation  $p_x$  and  $p_z$  (Fig. 6.14b). The electron orbitals both of the initial and final state are antisymmetrical upon the transformations  $z \rightarrow -z$  and  $x \rightarrow -x$ , respectively. Consequently we split the relevant matrix element into four partial integrals of equal amount:

$$\begin{aligned} \langle l, p_z | x | l, p_x \rangle &= \int_{x,z>0} \psi_{pz}^* x \psi_{px} d^3r + \int_{x>0, z<0} \psi_{pz}^* x \psi_{px} d^3r \\ &\quad + \int_{x<0, z>0} \psi_{pz}^* x \psi_{px} d^3r + \int_{x,z<0} \psi_{pz}^* x \psi_{px} d^3r. \end{aligned} \quad (6.134a)$$

In the second, the third and the fourth integral the following transformations are performed:  $z \rightarrow -z$ ,  $x \rightarrow -x$  and  $x, z \rightarrow -x, -z$ . Because of the symmetry properties of the  $p_x$  and  $p_z$  orbitals (Fig. 6.14) we, then, obtain:

$$\begin{aligned} \langle l, p_z | x | l, p_x \rangle &= \int_{x,z>0} \psi_{pz}^* x \psi_{px} d^3r + \int_{x,z>0} (-\psi_{pz}^*) x \psi_{px} d^3r \\ &\quad + \int_{x,z>0} \psi_{pz}^*(-x)(-\psi_{px}) d^3r + \int_{x,z>0} (-\psi_{pz}^*)(-x)(-\psi_{px}) d^3r \\ &= 0. \end{aligned} \quad (6.134b)$$

According to Fig. 6.14b the absolute amounts of the four integrals are equal but different signs make the dipole matrix element for transitions between  $p_x$  and  $p_z$  states (equal  $l$  quantum number) vanish.

The same arguments hold for transitions with dipole moments along  $z$  or  $y$ , that is, the transition matrix element between an  $s$  state and  $p_z$  and  $p_y$  states is non-zero, while the matrix elements between equal  $p$  states vanish. It is also obvious by similar symmetry arguments that transitions between two  $s$  orbitals ( $l = 0$ ) with spherical symmetry having different principal quantum numbers  $n$  (6.126a), (6.126b) are forbidden. In the corresponding transition matrix elements, we split the total integral in analogy to (6.133) into two parts with  $x < 0$  and  $x > 0$ , respectively,  $y, z < 0$  and  $y, z > 0$ . The transformations  $x \rightarrow -x$ ,  $y \rightarrow -y$ ,  $z \rightarrow -z$  keep the amount of the wave functions in a partial integral ( $x < 0$ ,  $y < 0$ ,  $z < 0$ ) unchanged, they change, however, the sign in front of the integral and the transition matrix element vanishes.

Without an extended formal proof, we generalize these results by means of the following dipole selection rule

$$l' - l = \pm 1. \quad (6.135)$$

Only those dipole transitions are allowed in which the angular quantum number  $l$  changes by the amount of one.

This selection rule for changes of the total electronic angular momentum in optical dipole transitions also shows that the light quanta, the photons, participating in the transition must carry an angular momentum, the spin, of an amount of  $\pm\hbar$ . This conclusion requires that the participating electron keeps its spin during the transition. Spin conservation is indeed given, since the considered dipole interaction operator acts only on the position part of the wave function rather than on the spin part.

## 6.5 Electronic Transitions in 2-Level Systems: The Rotating Wave Approximation

The treatment of electronic transitions between quantum states was concerned so far only with the determination of transition probabilities, respectively transition rates [Fermi's golden rule (6.111a), (6.111b)]. To analyse the internal dynamics of those transitions induced by a timely varying perturbation potential, we consider a simple quantum mechanical model system which contains only two quantum states, a so-called 2-level system. Such 2-level systems can be realized in numerous ways and they are of great practical importance. An obvious example is the electron or nuclear spin which can assume two energetically different orientations (states) in an external magnetic field. Another example is a double quantum dot system, each dot containing one confined electronic state (if decoupled). Electronic interaction via a thin energetic barrier between the two quantum dots couples the dots and induces two new separate states, the binding and the antibonding state (Sect. 6.2.3). A similar situation is given for two atoms which are covalently bound in a molecule by overlap of particular atomic orbitals (Sect. 6.2.3).

But also for free atoms the 2-level model system might be applied in many cases. If two energy levels are sufficiently far separated from the other levels, transitions between these two levels might be treated without considering coupling to other levels.

### 6.5.1 2-Level Systems in Resonance with Electromagnetic Radiation

We consider a 2-level system, for example, two coupled quantum dots with the two electronic states  $|e\rangle$  (excited) and  $|g\rangle$ , the ground state. An external electromagnetic field couples, through its electric field component, to the electron which can occupy the states  $|g\rangle$  and  $|e\rangle$ . According to Sect. 6.4, transitions between the states are possible for incident radiation which fulfills the resonance condition

$$\hbar\omega = E_e - E_g \quad (6.136)$$

$E_e$  and  $E_g$  are the energies of the excited state  $|e\rangle$  and the ground state  $|g\rangle$ , respectively. In case that the spatial extension of the system is small in comparison with the wavelength of the incident radiation, the action of the electric field on the electron can be described in terms of the dipole approximation (Sect. 6.4.2). The electric field  $\mathcal{E}(t)$  at the position of the electron exerts a force  $\mathbf{F} = (-e)\mathcal{E}$  on the electron. This is described in terms of the potential  $V(\mathbf{x}, t) = e\mathbf{x} \cdot \mathcal{E}$  with  $\mathbf{D} = e\mathbf{x}$  as the corresponding dipole moment. The interaction potential is thus written as

$$V(\mathbf{x}, t) = e\mathbf{x} \cdot \mathcal{E} = \mathbf{D} \cdot \mathcal{E}. \quad (6.137)$$

For the analogous spin dynamics, the corresponding expression contains the spin magnetic moment and the oscillating external magnetic field  $\mathbf{B}$ .

For the electronic 2-level system of the coupled quantum dots, the Schrödinger equation follows as

$$i\hbar\dot{\psi} = (\hat{H} + \hat{\mathbf{D}}\mathcal{E})\psi, \quad (6.138)$$

with  $\hat{\mathbf{D}}$  as the dipole operator.  $\hat{H}$  is the Hamilton operator of the unperturbed 2-level system, it has the eigenvalues  $E_g$  and  $E_e$ . We assume the polarization direction of the perturbing oscillating electric field to be parallel to the connecting line between the two quantum dots. The problem (6.138) is then 1-dimensional with scalar quantities  $\hat{D}$  and  $\mathcal{E}$ . The time dependence of the electric field might be assumed as

$$\mathcal{E}(t) = \mathcal{E}_A \cos \omega t = 2\mathcal{E}_0 \cos \omega t, \quad (6.139)$$

with the frequency  $\omega$  obeying the resonance condition (6.136).

As the electron in our system can only occupy the two states  $|g\rangle$  and  $|e\rangle$ , the most general solution to (6.138) is a linear superposition of these two states:

$$\psi = c_g|g\rangle + c_e|e\rangle. \quad (6.140)$$

While  $|g\rangle$  and  $|e\rangle$  are stationary states of the Hamiltonian  $\hat{H}$  with the characteristic time dependences  $\exp(-iE_g t/\hbar)$  respectively  $\exp(-iE_e t/\hbar)$ , the superposition (6.140) has two different energy exponents. It is, therefore, not a stationary solution to the problem nor is this expected because of the time-dependent perturbation. In the sense of a small perturbation of the stationary states, however,  $|g\rangle$  and  $|e\rangle$  might be taken as the constituting elements of the general solution and we are interested in the time evolution of  $\psi(t)$  respectively  $|g\rangle$  and  $|e\rangle$ . This question concerns the time dependence of the amplitudes  $c_g$  and  $c_e$  in (6.140). To find the solution, we make the following time-dependent ansatz:

$$c_g = g(t)e^{-iE_g t/\hbar}, \quad (6.141a)$$

$$c_e = f(t)e^{-iE_e t/\hbar}. \quad (6.141b)$$

The exponential functions take into account the time dependences of the corresponding stationary solutions.

Inserting (6.140) into (6.138) yields

$$i\hbar(\dot{c}_g|g\rangle + \dot{c}_e|e\rangle) = (\hat{H} + \hat{D}\mathcal{E})(c_g|g\rangle + c_e|e\rangle). \quad (6.142)$$

Because of the orthogonality  $\langle e|g\rangle = 0$  we obtain after multiplication with  $\langle g|$  respectively,  $\langle e|$  from the left

$$i\hbar\dot{c}_g(t) = E_g c_g + \langle g|\hat{D}|g\rangle\mathcal{E}c_g + \langle g|\hat{D}|e\rangle\mathcal{E}c_e, \quad (6.143a)$$

$$i\hbar\dot{c}_e(t) = E_e c_e + \langle e|\hat{D}|g\rangle\mathcal{E}c_g + \langle e|\hat{D}|e\rangle\mathcal{E}c_e. \quad (6.143b)$$

With the assumption that the 2-level system does not have a static dipole moment neither in the ground nor in the excited state, the matrix elements  $\langle g|\hat{D}|g\rangle$  and  $\langle e|\hat{D}|e\rangle$  vanish. This is evident for two equal coupled quantum dots which represent a potential with inversion symmetry for the electron.

The non-diagonal elements of the dipole operator

$$\langle g|\hat{D}|e\rangle = \langle e|\hat{D}|g\rangle^* = D_{ge} \quad (6.144)$$

do not vanish because of the different parity of  $|g\rangle$  and  $|e\rangle$ . They describe the dipole moment during the simultaneous presence of the ground and the excited state.

Inserting the probability amplitudes (6.141a), (6.141b) into (6.143a), (6.143b), thus, yields

$$i\hbar e^{-iE_g t/\hbar} \dot{g} + E_g g e^{-iE_g t/\hbar} = E_g g e^{-iE_g t/\hbar} + D_{ge} f e^{-iE_e t/\hbar} \mathcal{E}_0 (e^{i\omega t} + e^{-i\omega t}). \quad (6.145)$$

Because of the resonance condition (6.136), we obtain

$$\begin{aligned} i\hbar\dot{g} &= D_{ge} \mathcal{E}_0 e^{-i\omega t} (e^{i\omega t} + e^{-i\omega t}) f(t) \\ &= D_{ge} \mathcal{E}_0 (1 + e^{-2i\omega t}) f(t). \end{aligned} \quad (6.146)$$

An analogous calculation for (6.143b) leads to a coupled system of differential equations:

$$i\hbar\dot{g}(t) = D_{ge} \mathcal{E}_0 (1 + e^{-2i\omega t}) f(t), \quad (6.147a)$$

$$i\hbar\dot{f}(t) = D_{ge}^* \mathcal{E}_0 (1 + e^{2i\omega t}) g(t). \quad (6.147b)$$

Without the  $\exp(\pm 2i\omega t)$  terms the differential equations would be quite simple giving solutions  $g(t)$  and  $f(t)$  which certainly have oscillatory behavior. With the assumption that the time changes of  $g(t)$  and  $f(t)$  are slow as compared with the frequency  $\omega$  of the perturbing light field, we are lead to the so-called *rotating wave approximation*. This name originates from the spin 2-level system where an external per-

turbing high frequency electromagnetic field changes the orientation of spins in a static magnetic field (Sect. 6.5.2). In accord with the requirement for the rotating wave approximation the spin orientation switches slowly in comparison with the frequency  $\omega = (E_e - E_g)/\hbar = (E_\uparrow - E_\downarrow)/\hbar$  in this case. The requirement of a fast oscillating  $\exp(\pm 2i\omega t)$  term in (6.147a), (6.147b) means that this term changes sign frequently during the transition of the 2-level system between  $|g\rangle$  and  $|e\rangle$ . In the integration of (6.147a), (6.147b), the many contributions of the exponential function with different sign cancel each other in good approximation. This part of the integral is therefore neglected in the rotating wave approximation and we obtain as relevant equations:

$$\dot{g}(t) = -i \frac{1}{\hbar} D_{ge} \mathcal{E}_0 f(t), \quad (6.148a)$$

$$\dot{f}(t) = -i \frac{1}{\hbar} D_{ge}^* \mathcal{E}_0 g(t). \quad (6.148b)$$

Further differentiation of (6.148a) allows the elimination of  $f(t)$  from (6.148a) by using

$$\ddot{g}(t) = -i \frac{1}{\hbar} D_{ge} \mathcal{E}_0 \dot{f} \quad \text{and} \quad (6.149a)$$

$$\dot{f}(t) = \frac{1}{-i} \frac{\hbar}{D_{ge} \mathcal{E}_0} \ddot{g}. \quad (6.149b)$$

An analogous treatment of  $f(t)$  in (6.148b) finally yields the following differential equations:

$$\ddot{g}(t) = - \left| \frac{D_{ge} \mathcal{E}_0}{\hbar} \right|^2 g(t); \quad (6.150a)$$

$$\ddot{f}(t) = - \left| \frac{D_{ge} \mathcal{E}_0}{\hbar} \right|^2 f(t). \quad (6.150b)$$

These are familiar oscillator differential equations, whose solutions  $g(t)$  and  $f(t)$  oscillate with a frequency  $\Omega = |D_{ge} \mathcal{E}_0 / \hbar|$  between their maximum and minimum values. The extreme values are determined from the normalization of the wave function  $\psi$  (6.140), that is, the requirements

$$|c_g|^2 + |c_e|^2 = 1, \quad (6.151a)$$

$$|f|^2 + |g|^2 = 1. \quad (6.151b)$$

The oscillation ansatz

$$g(t) = \alpha \cos \Omega t + \beta \sin \Omega t \quad (6.152a)$$

for a general solution  $g(t)$ , together with the boundary condition (6.151b), yields

$$f(t) = i\beta \cos \Omega t - i\alpha \sin \Omega t. \quad (6.152b)$$

As initial condition we assume that the system is in its ground state  $|g\rangle$  at  $t = 0$ , that is,  $\beta = 0$ . Then, the probability amplitudes for the ground and the excited states are obtained as

$$c_g(t) = \alpha(\cos \Omega t)e^{-iE_g t/\hbar}, \quad (6.153a)$$

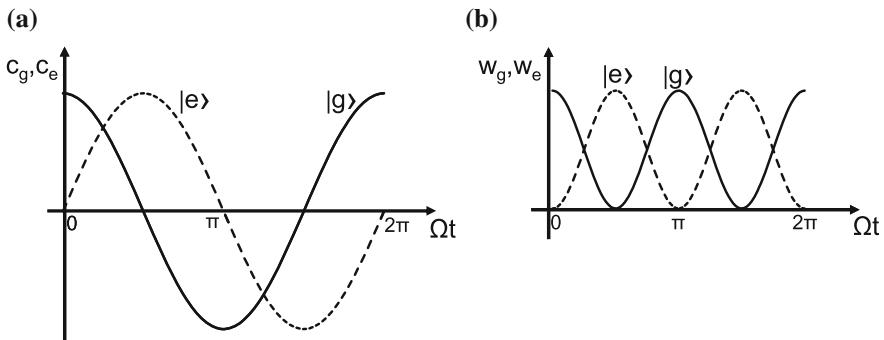
$$c_e(t) = -i\alpha(\sin \Omega t)e^{-iE_e t/\hbar}. \quad (6.153b)$$

The representations of the amplitudes (6.53a), (6.53b) and the occupation probabilities  $w_g = |c_g|^2$  and  $w_e = |c_e|^2$ , respectively, show that the 2-level system being in resonance with an external electromagnetic field oscillates between its ground and excited state (Fig. 6.15). The oscillation frequency

$$\Omega = \frac{1}{\hbar}|D_{ge}|\mathcal{E}_0 = \frac{1}{2\hbar}|D_{ge}|\mathcal{E}_A \quad (6.154)$$

is called *Rabi frequency*.

As example we use the system of the two coupled quantum dots and we assume that the electron initially occupies the bonding ground state  $|g\rangle$  (6.57a). The system can then be excited into the excited (antibonding) state  $|e\rangle$  by irradiation of electromagnetic radiation with the resonance frequency  $\omega = (E_e - E_g)/\hbar$ . To bring



**Fig. 6.15** **a, b** Rabi oscillation of a 2-level system having the ground state  $|g\rangle$  and the excited state  $|e\rangle$ . For such a system being in resonance with an electromagnetic field of frequency  $\omega (E_e - E_g = \hbar\omega)$  the occupation of the states  $|g\rangle$  and  $|e\rangle$  varies periodically. **a** Occupation amplitudes  $c_g$  and  $c_e$  of the states  $|g\rangle$  and  $|e\rangle$  as function of time. **b** Occupation probabilities  $w_g$  and  $w_e$  of  $|g\rangle$  and  $|e\rangle$  as function of time.  $\Omega$  is the Rabi frequency, which is determined by the coupling strength between the 2-level system and the light field, i.e., essentially by the transition dipole moment  $\mathbf{D}_{ge}$

the system exactly into  $|e\rangle$  the interaction time with the radiation (light pulse) has a well-defined value  $t_\pi$ . According to Fig. 6.15, this light pulse has a length which is given by

$$\begin{aligned}\Omega t_\pi &= \frac{1}{2\hbar} |D_{ge}| \mathcal{E}_A t_\pi = \pi/2, \\ t_\pi &= \frac{\pi\hbar}{|D_{ge}| \mathcal{E}_A}.\end{aligned}\quad (6.155)$$

Such a pulse is called  $\pi$  pulse. The name is derived from the spin problem, where an electromagnetic pulse of this length switches the spin orientation in a static magnetic field by  $\pi = 180^\circ$ , that is, from  $|\uparrow\rangle$  into  $|\downarrow\rangle$  and vice versa.

According to Fig. 6.15, the application of a  $\pi/2$  pulse on the ground state  $|g\rangle$  transfers the system into a quantum state in which both states  $|g\rangle$  and  $|e\rangle$  are contained with equal contribution. The pulse length is in analogy

$$t_{\pi/2} = \frac{\pi\hbar}{2|D_{ge}| \mathcal{E}_A}. \quad (6.156)$$

and the obtained final state has the representation

$$|S\rangle = \frac{1}{\sqrt{2}} (|g\rangle + |e\rangle). \quad (6.157)$$

This is a so-called *superposition state* which is characteristic for quantum mechanics. The electron simultaneously occupies the states  $|g\rangle$  and  $|e\rangle$ . This is, of course, impossible in classical physics. We encounter here one of the strange phenomena of quantum physics. It must be emphasized that the superposition state is destroyed by an energy measurement on the system. The outcome of the measurement is always one of the states  $|g\rangle$  or  $|e\rangle$ , respectively their energies  $E_g$  or  $E_e$ . This weird, counter-intuitive behavior of quantum systems is considered more in detail in the next chapter.

### 6.5.2 Spin Flip

Ideal 2-level systems are represented by the spin of particles as electrons, protons, neutrons or atomic nuclei in a static external magnetic field. In the magnetic field  $\mathbf{B}$  the spin can assume the two different states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  with parallel and anti-parallel orientation to the field  $\mathbf{B}$ . In these two states the spin, a gyro itself, performs precessions around the magnetic field direction with the precession frequency  $\omega_0 = eB_0/m$ . According to the two spin orientations the spin moment has its z-component in the  $\mathbf{B}$  field direction or opposite to it (Sect. 5.5.3). An additionally irradiated electromagnetic field with its magnetic field component normal to the static  $\mathbf{B}$  field interacts with the magnetic spin moment  $\mu_B$ . The interaction is analogous to that of an

oscillating electric field acting on a charge dipole (Sect. 6.5.1). We expect transitions between the spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , that is, periodically occurring spin flips.

For the formal description of the process, we use the spin Schrödinger equation, the Pauli equation (5.122), (5.123). We replace the static magnetic field in (5.123) by a superposition of a constant magnetic field  $\mathbf{B}_0$  and an oscillating field  $\mathbf{B}_{\text{osc}}(t)$  oriented normal to  $\mathbf{B}_0$ :

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_{\text{osc}}(t), \quad (6.158a)$$

$$\mathbf{B}_0 = (0, 0, B_z^0), \quad (6.158b)$$

$$\mathbf{B}_{\text{osc}} = (B_x(t), B_y(t), 0). \quad (6.158c)$$

The most general spin state, as in (5.124), is a superposition of the two spin eigenstates of the  $\sigma_z$  operator:

$$|s\rangle = \alpha_+ |\uparrow\rangle + \alpha_- |\downarrow\rangle = a_+ e^{-i\omega_0 t/2} |\uparrow\rangle + a_- e^{-i\omega_0 t/2} |\downarrow\rangle = \begin{pmatrix} \alpha_+ \\ \alpha_- \end{pmatrix}. \quad (6.159)$$

The exponential terms express the stationary character of the spin states with energy eigenvalues:

$$E_{\uparrow}/\hbar = \frac{e}{2m} B_z = \omega_0/2, \quad (6.160a)$$

$$E_{\downarrow}/\hbar = -\frac{e}{2m} B_z = -\omega_0/2. \quad (6.160b)$$

According to Sect. 5.5.3  $\mu_B = e\hbar/2m$  is the Bohr magneton. The energetic distance between the two spin levels  $E_{\uparrow}$  and  $E_{\downarrow}$  is, therefore, given by the spin precession frequency  $\hbar\omega_0$ .

For the calculation of the time change of the probability amplitudes  $\alpha_+(t)$  and  $\alpha_-(t)$ , we insert (6.159) together with the magnetic field (6.158a)–(6.158c) into the Pauli equation (5.123) and obtain:

$$\begin{aligned} \hat{\sigma} \cdot \mathbf{B} &= \hat{\sigma}_x B_x + \hat{\sigma}_y B_y + \hat{\sigma}_z B_z \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} B_x(t) + \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} B_y(t) + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} B_z^0 \\ &= \begin{pmatrix} B_z^0 & B_x - iB_y \\ B_x + iB_y & -B_z^0 \end{pmatrix}, \end{aligned} \quad (6.161)$$

$$i\hbar \frac{\partial}{\partial t} |s\rangle = \mu_B \hat{\sigma} \cdot \mathbf{B} |s\rangle, \quad (6.162a)$$

$$i\hbar \begin{pmatrix} \dot{\alpha}_+ \\ \dot{\alpha}_- \end{pmatrix} = \mu_B \begin{pmatrix} B_z^0 & B_x - iB_y \\ B_x + iB_y & -B_z^0 \end{pmatrix} \begin{pmatrix} \alpha_+ \\ \alpha_- \end{pmatrix}. \quad (6.162b)$$

To facilitate the calculation we assume a circularly polarized oscillating magnetic field  $\mathbf{B}_{\text{osc}}$ :

$$B_x(t) = A \cos \omega_0 t, \quad (6.163a)$$

$$B_y(t) = A \sin \omega_0 t. \quad (6.163b)$$

We expect spin flip transitions under the resonance condition  $E_{\uparrow} - E_{\downarrow} = \hbar\omega_0$ . Consequently, the oscillation frequency of the magnetic field has been assumed to be identical with the spin precession frequency  $\omega_0$  (6.160a), (6.160b).

With  $\mu_B = e\hbar/2m$  as the Bohr magneton and by using (6.160a), (6.160b), we obtain from (6.162b):

$$i\hbar \begin{pmatrix} \dot{\alpha}_+ \\ \dot{\alpha}_- \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\hbar\omega_0 & \mu_B(B_x - iB_y) \\ \mu_B(B_x + iB_y) & -\frac{1}{2}\hbar\omega_0 \end{pmatrix} \begin{pmatrix} \alpha_+ \\ \alpha_- \end{pmatrix}. \quad (6.164)$$

With the relation

$$B_x \pm iB_y = A(\cos \omega_0 t \pm i \sin \omega_0 t) = Ae^{\pm i\omega_0 t} \quad (6.165)$$

equation (6.164) yields

$$i\hbar\dot{\alpha}_+ = \frac{1}{2}\hbar\omega_0\alpha_+ + \mu_B Ae^{-i\omega_0 t}\alpha_-, \quad (6.166a)$$

$$i\hbar\dot{\alpha}_- = \mu_B Ae^{i\omega_0 t}\alpha_+ - \frac{1}{2}\hbar\omega_0\alpha_-. \quad (6.166b)$$

We now use the representations  $\alpha_+(t) = a_+ \exp(-i\omega_0 t/2)$  and  $\alpha_-(t) = a_- \exp(i\omega_0 t/2)$  of (6.159) and obtain from (6.166a), (6.166b):

$$i\hbar\dot{\alpha}_+ = \frac{1}{2}\hbar\omega_0 a_+ e^{-i\omega_0 t/2} + \mu_B A a_- e^{-i\omega_0 t/2}, \quad (6.167a)$$

$$i\hbar\dot{\alpha}_- = \mu_B A a_+ e^{i\omega_0 t/2} - \frac{1}{2}\hbar\omega_0 a_- e^{i\omega_0 t/2}. \quad (6.167b)$$

After performing the differentiation of  $\alpha_+$  and  $\alpha_-$  on the left side by using the representation (6.159) one finally gets:

$$i\hbar\dot{a}_+ = \mu_B A a_-, \quad (6.168a)$$

$$i\hbar\dot{a}_- = \mu_B A a_+. \quad (6.168b)$$

These equations for  $a_+$  and  $a_-$  are identical with those for  $g(t)$  and  $f(t)$  (6.148a), (6.148b) in the last Sect. 6.5.1. Correspondingly, the analogous [to (6.153a), (6.153b)] solutions to the spin problem are then:

$$a_-(t) \propto (\cos \Omega t) e^{-iE_{\downarrow} t/\hbar}, \quad (6.169a)$$

$$a_+(t) \propto i(\sin \Omega t) e^{-iE_{\uparrow} t/\hbar}. \quad (6.169b)$$

Hereby the *Rabi frequency* for the spin system is defined as

$$\Omega = \mu_B A / \hbar. \quad (6.170)$$

Again, electromagnetic radiation having the resonance frequency  $\omega_0$  for transitions between the spin split states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  in a static magnetic field causes periodic oscillations between these states, that is, a periodic sequence of spin flips.

The Rabi frequency of these spin flip oscillations is proportional to the amplitude  $A$  of the oscillating magnetic field and to the magnetic spin moment  $\mu_B$ . This is in complete analogy to the electronic 2-level system where the Rabi frequency (6.154) depends on the electric field amplitude and the electronic transition dipole  $D_{ge}$ .

As in the case of an electron, which can occupy the two states  $|g\rangle$  and  $|e\rangle$  (Sect. 6.5.1), we can ask, how long the oscillating magnetic field must interact for one single spin flip, that is, a transition from  $|\uparrow\rangle$  to  $|\downarrow\rangle$ . Because of (6.170) this so-called  $\pi$  pulse has a duration

$$t_\pi = \frac{\pi\hbar}{2\mu_B A}. \quad (6.171)$$

Accordingly a  $\pi/2$  pulse of time duration

$$t_{\pi/2} = \frac{\pi\hbar}{4\mu_B A} \quad (6.172)$$

flips the spin from a “normal” state  $|\uparrow\rangle$  (eigenstate to  $\sigma_z$ ) into a superposition state

$$|S_{\text{superp}}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle). \quad (6.173)$$

In this state, the spin pivot is directed normal to the static magnetic field  $\mathbf{B}_0$ .

These descriptive results might be substantiated by a calculation of the average spin coordinates as functions of time. Only the results of these calculations are given here:

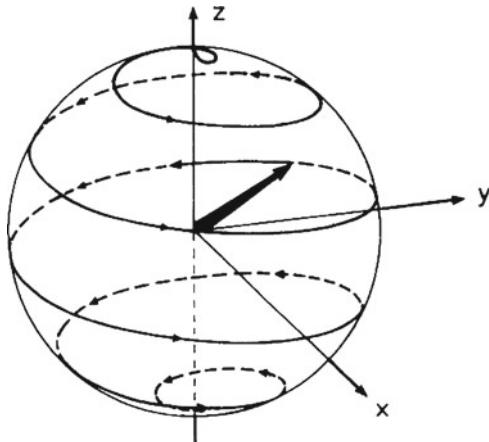
$$\langle \hat{S}_z \rangle = -\frac{\hbar}{2} \cos(2\Omega t), \quad (6.174a)$$

$$\langle \hat{S}_x \rangle = -\frac{\hbar}{2} \sin(2\Omega t) \sin(\omega_0 t), \quad (6.174b)$$

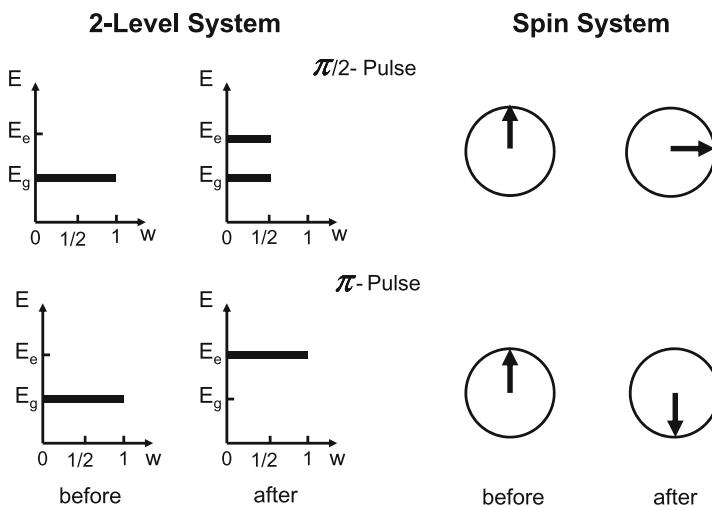
$$\langle \hat{S}_y \rangle = -\frac{\hbar}{2} \sin(2\Omega t) \cos(\omega_0 t). \quad (6.174c)$$

The flip motion of the  $z$ -component is described solely by the Rabi frequency (6.170). The  $x$ - and  $y$ -components contain in addition a fast orbital motion of the spin with the precession frequency  $\omega_0$ . While the spin changes its orientation from  $+z$  to  $-z$ , it simultaneously performs a precession around the  $z$ -axis (direction of the static  $\mathbf{B}_0$  field) (Fig. 6.16).

We keep in mind: A circularly polarized electromagnetic field with its magnetic component  $\mathbf{B}_{\text{osc}}$  rotating with the spin precession frequency  $\omega_0$  around the direction of the static magnetic field  $\mathbf{B}_0$  induces a spin flip. In most experiments, a linearly polarized electromagnetic field is applied rather than a circularly polarized one. The linearly polarized magnetic field  $\mathbf{B}_{\text{osc}}$  can then be thought to be built up by two circularly polarized fields rotating in opposite directions. One of these fields rotates, as described here, with the spin's precession direction. The other field rotates in opposite direction with the double precession frequency  $2\omega_0$  as seen from the rotating spin



**Fig. 6.16** Schematic representation of a spin (plotted as *solid arrow*) flip process in a constant magnetic field directed along the  $z$  axis. The spin flip is induced by an additionally irradiated  $\pi$  pulse of RF radiation. During the flip from  $+z$  to  $-z$  the spin gyro performs a precession motion around the  $z$  axis, the direction of the magnetic field



**Fig. 6.17** Comparison of a general 2-level system having the quantum states  $|g\rangle$  and  $|e\rangle$  with a spin in a constant magnetic field. Both systems are exposed to a  $\pi/2$  and a  $\pi$  pulse of exciting electromagnetic radiation. For the 2-level system (*left*) the occupation probabilities  $w$  of the quantum states are given, while for the spin system (*right*) the spin orientation is plotted

system. In the dynamic equations (6.168a), (6.168b), this field produces additional  $\exp(\pm i\omega_0 t)$  terms as in (6.147a), (6.147b). Neglecting these fast oscillating terms, as already done in Sect. 6.5.1, explains the name “rotating wave approximation”. In Fig. 6.17, the close analogy between an electronic 2-level system and the 2-level spin system under the action of a  $\pi$  and a  $\pi/2$  radiation pulse is schematically shown.

### 6.5.3 Nuclear Spin Resonance in Chemistry, Biology and Medicine

Experiments where in a static magnetic field particle spins are flipped by the action of an additional electromagnetic high frequency RF (radio frequency) field (Sect. 6.5.2) are called spin resonance experiments. The energy consumed by the spin transition is taken from the RF field and a corresponding absorption band is observed in the RF radiation spectrum at the photon energy  $\hbar\omega_0$ , respectively, the spin precession frequency  $\omega_0$ . The RF field is in resonance with the spin flip transition. A common observation technique for these spin resonance transitions is purely electrical: The spin flip is connected with an orientation change of the spin magnetic dipole. This induces an electric AC signal (pulse) in a coil (called RF probe) surrounding the sample under study.

Such spin resonance experiments with electrons, so-called *electron spin resonance (ESR)*, are of great importance in solid state physics. Defects in solids are often accompanied by additional electronic spins depending on details of the atomic surrounding of the defect. ESR thus provides a powerful tool for the study of defects in solids.

Also atomic nuclei being built up from fermionic protons and neutrons (Sect. 5.6.4) have a spin. Correspondingly there are *nuclear magnetic resonance (NMR)* experiments (nuclear spin flips) which have gained paramount importance in material science, chemistry, biology and medicine. Both spectroscopic and imaging (tomography) techniques are meanwhile based on the nuclear spin flip (NMR) mechanism [8].

There is a simple reason that nuclear spins have become such an important tool in many condensed matter applications. In condensed matter, the cloud of valence electrons (their spins included) fills space and is responsible for the chemical bonding and most of the condensed matter properties. In contrast, the atomic nucleus with tiny dimensions as compared with the Bohr radius is kind of isolated from the behavior of the electronic part of matter. There is only a weak coupling between both systems. The precession motion of the nuclear spins is essentially unaffected by translatory and rotatory motions of the nuclei. Fluid flow or thermal motion of the atoms do not influence a measurement of the nuclear resonance. This property accounts for the high analytical power of the NMR method in the study of condensed matter up to living cells in biology and medicine.

Because of its angular momentum  $J$  an atomic nucleus has a magnetic moment  $\mu_K$ , in analogy to the spin (Sect. 5.5.1) of an electron (5.69), (5.95):

$$\boldsymbol{\mu}_N = g_K \frac{e}{2m_K} \mathbf{J}. \quad (6.175)$$

As usually the angular momentum can assume the values  $|J| = \sqrt{j(j+1)}\hbar$  with  $j$  as integer quantum numbers. The nuclear mass  $m_K$ , however, in particular that of the proton  $m_P$ , is higher than that of the electron by a factor 1836. Consequently, nuclear magnetic moments are smaller than those of electrons by at least a factor of 2000.

In analogy to the electron spin (gyromagnetic ratio  $\cong 2$ ) the gyromagnetic ratio of the nucleus is  $g_K$ . It is a quantity characteristic for each nucleus. Analogously to the electron the nuclear magnetic moment is measured in units of the *nuclear magneton*:

$$\mu_K = \mu_B / 1836 = \frac{e\hbar}{2m_p} = 0.505 \cdot 10^{-26} \text{ A m}^2. \quad (6.176)$$

In contrast to the electron the gyromagnetic ratio  $g_K$  of a nucleus can not be derived from the orbital angular momentum by means of the rules of electrodynamics. As proton and neutron, the constituents of an atomic nucleus, are composed of quarks (Sect. 5.6.4), only chromodynamics, the theory of the strong interaction, can yield this important physical quantity. It must be noted that also for nuclei the total angular momentum  $\mathbf{J}$  is composed of the orbital momentum  $\mathbf{L}$  and the nuclear spin moment  $\mathbf{S}$ . In their ground state nuclei do not have an orbital momentum such that the total angular momentum contains only the spin nuclear momenta of the constituents. Both protons and neutrons have half-integral spin  $\hbar/2$ , that is, nuclei with even mass number  $A$  have an even-valued spin. For odd mass numbers  $A$  the nuclear spin has a multiple value of  $\hbar/2$ . In the nucleus, both protons and neutrons tend to arrange with opposite spins (Pauli principle). Therefore, nuclei with an even number of protons and neutrons have vanishing spin and zero magnetic moment. They can not be used in NMR experiments. Examples are the (even–even) isotopes  $^{12}\text{C}$  and  $^{16}\text{O}$  of carbon and oxygen which are abundant in nature.

Table 6.1 shows a compilation of some important nuclear spins together with their magnetic moments in units of the nuclear magneton (6.176). Furthermore, the spin precession frequency  $\omega_0/2\pi$  (NMR frequency) is given for a magnetic field of  $B_0 = 2.3487$  T. This field strength has been chosen because the NMR frequency of the proton, then, amounts to 100 MHz. According to Sect. 6.5.2, this is a consequence of the resonance condition:

$$E_{\uparrow} - E_{\downarrow} = \hbar\omega_0 = g_K \mu_K B_z = g_K \frac{e\hbar}{2m_p} B_z. \quad (6.177)$$

The sign of the magnetic moment in Table 6.1 indicates the orientation of the spin magnetic moment relative to the spin angular momentum: For positive sign, both vectors have the same orientation; negative sign indicates antiparallel orientation. While for the proton  $p$  ( $^1\text{H}$ ) magnetic moment and angular momentum have the same orientation, the electron's spin magnetic moment and angular momentum have opposite orientation. We could intuitively expect this property because of the opposite charge of proton and electron. The relative abundance of the isotopes in nature tells us what isotopes are particularly interesting for biological and medical applications.

In NMR, studies compact condensed matter (water, biological cells, parts of human body etc.) is investigated rather than single spins in a magnetic field. In an NMR experiment based on the hydrogen isotope  $^1\text{H}$  the sample, thus, contains about  $10^{23}$  protons per cubic centimeter. In thermal equilibrium, there will be many spins oriented in the direction of the magnetic field  $\mathbf{B}_0$  according to the energy state

**Table 6.1** Properties of some atomic nuclei being of interest for nuclear magnetic resonance (NMR) in biology and medicine [8]

Nucleus	Spin	Magn. moment (Nucl. magnetons)	Natural appearance (%)	NMR-frequency in 2.3487 T (MHz)
$^1\text{H}$	$\frac{1}{2}$	2.79	99.98	100.00
$^2\text{H}$	1	0.86	0.015	15.35
$^3\text{H}$	$\frac{1}{2}$	2.98	0	106.68
$^{13}\text{C}$	$\frac{1}{2}$	0.70	1.11	25.14
$^{14}\text{N}$	1	0.40	99.6	7.22
$^{15}\text{N}$	$\frac{1}{2}$	-2.83	0.4	10.13
$^{17}\text{O}$	$\frac{5}{2}$	-1.89	0.04	13.56
$^{19}\text{F}$	$\frac{1}{2}$	2.63	100	94.08
$^{31}\text{P}$	$\frac{1}{2}$	1.13	100	40.48

$m_s = 1/2$  and many others are oriented opposite to the magnetic field direction according to  $m_s = -1/2$ . The occupation ratio of the two states, of course, depends on temperature and is determined by the Boltzmann factor (Sect. 5.6.3):

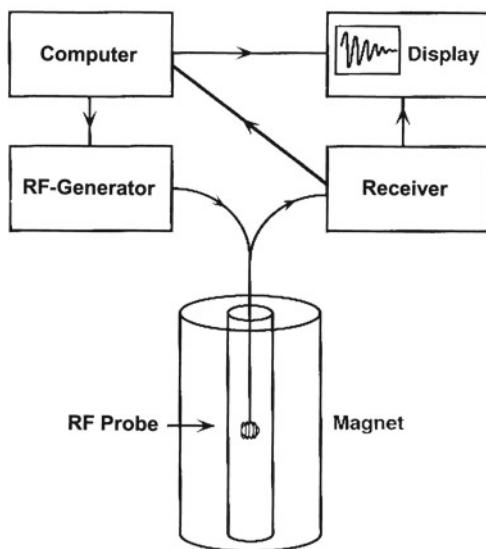
$$\exp\left(\frac{2\mu_P B_0}{kT}\right) \approx 1 + \frac{2\mu_P B_0}{kT}. \quad (6.178)$$

Only a tiny portion of the proton spins (moment  $\mu_P$ ) oriented in field direction  $\mathbf{B}_0$  (energetically lowest state) exceeds the total number of spins with both orientations. At room temperature, this tiny surplus of parallel spins in a field of 1 T amounts to only  $8 \cdot 10^{-6}$ . This explains the approximation of the exponential function in (6.178). The surplus of spins parallel to  $\mathbf{B}_0$  creates a macroscopic magnetic moment, the nuclear magnetization  $M_N$  of the sample. This nuclear magnetization responds to an additionally irradiated RF field with spin precession frequency (Larmor frequency)  $\omega_0$ . Under the action of a  $\pi/2$  pulse (Sect. 6.5.2) the nuclear magnetization oriented in  $\mathbf{B}_0$  direction in thermal equilibrium turns into a direction normal to the magnetic field  $\mathbf{B}_0$ . After switching off the pulse the spins tend again into thermal equilibrium, the nuclear magnetization relaxes back in  $\mathbf{B}_0$  direction. According to the law of induction the flipping nuclear magnetization induces an electric voltage oscillating with the Larmor frequency (NMR frequency) in the probe coil surrounding the sample (Fig. 6.18).

This decaying RF signal after application of the  $\pi/2$  pulse is called *NMR signal* (Fig. 6.18). Sometimes also the decay of the non-equilibrium magnetization after application of a  $\pi$  pulse is measured. Such a  $\pi$  pulse flips the equilibrium magnetization from  $M_K$  (in  $\mathbf{B}_0$  direction) into  $-M_K$ . Also in this case the magnetization turns back into field direction after switching off the pulse, of course, according to an exponential decay:

$$M(t) = M_K (1 - 2e^{-t/\tau}). \quad (6.179)$$

**Fig. 6.18** Schematic plot of the experimental set-up for measuring magnetic nuclear spin resonance (NMR)



$\tau$  is the relaxation time after which thermal equilibrium is reached again. For protons in clean water  $\tau$  amounts to about 3 s at room temperature.

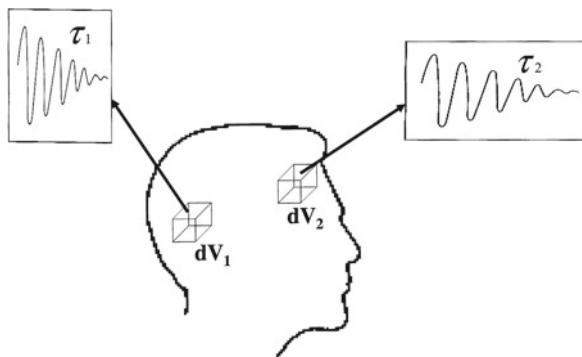
We now have to elucidate the basic mechanism which causes the spins to relax into thermal equilibrium after switching off the  $\pi$  or  $\pi/2$  pulse. An essential effect is due to spins in neighbouring atomic nuclei which generate a tiny magnetic field at the location of the considered spin. This magnetic field fluctuates in time more or less strongly depending on the medium in which it originates, a crystalline solid, a liquid or biological tissue. In water, for example, each proton is exposed to another proton with spin at a distance of about  $1.5 \cdot 10^{-8}$  cm. Each neighbouring spin generates a fluctuating field of about  $\pm 5G = \pm 5 \cdot 10^{-4}$  T at the position of the first proton spin. The time fluctuations of this tiny additional field occur on a time scale of  $10^{-11}$  s which is given by the molecular motion in the water.

A spectral analysis, that is, a Fourier analysis, of this random motion of the water molecules, respectively their proton spins yields a constant frequency spectrum up to  $10^{11}$  Hz, the inverse of  $10^{-11}$  s. All frequencies up to  $10^{11}$  Hz are found in this spectrum including the NMR frequency, for example, 100 MHz, which induces spin flips in the resonance experiment. Thus, in addition to the exciting RF (NMR) field which induces excited states by  $\pi$  or  $\pi/2$  pulses there is the perturbing fluctuating field which also acts on the spins, in particular, by de-exciting from the excited into the ground state. The relaxation process of one nuclear spin after switching off the  $\pi$  or  $\pi/2$  pulse, therefore, sensitively depends on the dynamics of all other surrounding nuclear spins. Water is a good example: It exists as a liquid and as a well ordered crystalline solid (ice). In liquid water, the dynamics of the proton spins runs on the  $10^{-11}$  s time scale with a relaxation time  $\tau$  of about 3 s. In crystalline ice,

the proton spins are much more fixed to their position because of the rigid crystal lattice. Accordingly, the relaxation time is higher, in the order of 600 s. Because of the uncertainty relation  $\Delta E \times \tau \approx \hbar$  or  $\Delta\omega \times \tau \approx 1$  extremely different frequency half-widths of the NMR signal result. For ice, the frequency peak width of the NMR signal is in the order of 50 kHz while in liquid water a sharp band with a width of about 0.1 Hz is found.

We want to emphasize again that the relaxation time of nuclear spins for reaching the equilibrium distribution after their excitation by  $\pi$  or  $\pi/2$  pulses sensitively depends on the dynamics of the surrounding spins. The local behavior of biological tissue containing a number of quite different components as cell membranes, protein conglomerates, compartments with different liquid content etc. differs dramatically concerning the nuclear spin relaxation times. Depending on type of tissue and position within the cell, relaxation times ranging between 50 and 1000 ms are observed.

Imaging NMR as used in biology and medicine is based on the local variation of relaxation time  $\tau(\mathbf{r})$  which depends on the tissue structure at the position  $\mathbf{r}$  (Fig. 6.19). As a function of position  $\mathbf{r}$ , the exponentially decaying NMR signal is measured after irradiation of a  $\pi$  or  $\pi/2$  pulse of RF radiation. The envelope of the NMR signal (decaying oscillation) determines the relaxation time  $\tau$  at this particular position  $\mathbf{r}$  (Fig. 6.19). The measured quantity  $\tau$  is transferred into a black–white contrast or color scheme and plotted as a function of  $\mathbf{r}$ . The position  $\mathbf{r}$  at which a particular color or contrast, that is, a particular relaxation time  $\tau$ , appears in the image, is evaluated by the use of an inhomogeneous large static magnetic field  $\mathbf{B}(\mathbf{r})$  rather than a homogeneous one in non-imaging NMR experiments. This large static magnetic field (0.02–5 T) varies in a well defined manner over the animal's or human organ (head, leg or whole body). Correspondingly, the NMR frequency  $\omega_0$  (excitation frequency



**Fig. 6.19** Image generation at nuclear magnetic spin tomography (imaging NMR) in biology and medicine. The patient's head is exposed to a strong, spatially inhomogeneous magnetic field. Different NMR frequencies in the volume elements  $dV_1$  and  $dV_2$ , thus, define the position of these elements  $dV_1$  and  $dV_2$  in space. The NMR decay times  $\tau_1$  and  $\tau_2$  measured at those positions after irradiation of a  $\pi/2$  or  $\pi$  RF radiation pulse depend on the special type of tissue. The locally measured  $\tau_1$ ,  $\tau_2$ , etc. values are transformed into degrees of brightness or color by a computer and are used to build up the image pixel by pixel

of spins) varies locally. From the known magnetic field variation, then, the measured precession frequency value  $\omega_0(\mathbf{r})$  determines the position  $\mathbf{r}$  by using (6.160a), (6.160b). According to Fig. 6.19, a particular volume element  $dV_i$  is characterized by a certain NMR frequency  $\omega(\mathbf{r}_i)$ . Its local measurement allows the determination of  $\mathbf{r}_i$  at which on the other hand  $\tau(\mathbf{r}_i)$  is evaluated from the envelope of the NMR signal. Elaborate computer calculations generate an image of the biological sample or part of the human body by correlating particular  $\tau(\mathbf{r}_i)$  values (pixels) represented by different colors or contrasts with the positions  $\mathbf{r}_i$  of their occurrence. The generated image shows biological tissue and/or human organs with areas weighted according to their different nuclear spin relaxation times  $\tau$ . Sophisticated scanning techniques allow even 3-dimensional images via sheet-wise recording of the NMR signals. Tuning the NMR frequency to particular nuclear spin flip frequencies (spin transition energies) allows the recording of images of different nuclei, that is, different chemical elements. Most important in practice are proton ( $^1\text{H}$ ) based images but also images based on  $^{7}\text{Li}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{15}\text{N}$  or  $^{23}\text{Na}$  are used depending on the desired information about a particular disease. As an example Fig. 6.20 shows an NMR image (magnetic resonance tomogram, MRT) of a human head as it is now commonly recorded in brain research.



**Fig. 6.20** NMR tomographic image of a human brain; from *left* to *right* axial, sagittal, and coronal view. The 3D image has been recorded by means of a clinical whole-body scanner in a magnetic field of 1.5 T. The image resolution of 0.6 mm has been obtained by averaging over 10 separate raster images [12]

To produce such large area images of human organs or the whole human body in medicine, big superconducting magnets are used for the generation of high magnetic fields up to 5 or even 10 T. The patient is shifted into the opening of the magnet and exposed to pulses of the adequate RF radiation. Additional coils generate well-defined static magnetic field gradients which locally vary the NMR frequency  $\omega_0$  in order to define the position  $\mathbf{r}_i$  in the image. Furthermore, the body part under study is surrounded by an RF coil which produces the RF  $\pi$  or  $\pi/2$  pulses. The subsequently

decaying NMR signal is measured by a probe coil which is sometimes identical with the coil for the RF pulse generation.

It is evident that the generation of an NMR image (MR tomogram) in medicine requires enormous computer power for scanning and calculating the NMR frequency  $\omega_0(\mathbf{r}_i)$  position selectively as well as attributing the measured decay times  $\tau[\omega_0(\mathbf{r}_i)]$  locally to image pixels in real time. NMR as an important imaging technique in medicine became possible after the development of powerful computers.

Finally, we emphasize again that NMR used as a spectroscopy rather than an imaging tomography has gained enormous importance for the elucidation of complex molecular structures up to biological molecules as proteins of medium length or nucleic acids. In these applications, simpler experimental set-ups as for example, in Fig. 6.18 are used to measure the frequency spectrum with high resolution. The spectral width of the observed absorption bands yields information about life times, that is, decay times  $\tau$ . In addition, the exact nuclear resonance frequency depends on the chemical surrounding of the nucleus. The NMR frequency  $\omega_0$  at which the proton resonance appears in a fixed static magnetic field depends on the chemical bond, be it that the proton is bonded in a hydroxyl or an ethyl group etc. Slight shifts of the resonance line are due to extra-nuclear electron spins which cause weak magnetic screening. Meanwhile, there exists a broad literature of measured chemical shifts of particular NMR lines which allows detailed conclusions about the corresponding chemical surrounding of the particular nucleus, that is, its chemical bond to other structural groups. NMR has, thus, developed into an indispensable technique for the elucidation of complex molecular structures.

## 6.6 Scattering of Particles

In physics and in material science scattering of particles as electrons, photons, neutrons etc. has become a predominantly important technique to get information about the scattering sample, the target. In solid state physics and in material science, the elastic scattering of photons (X-rays) or electrons has developed into a standard technique for the study of the crystalline structure of materials, surfaces or nanostructures. An example in high energy particle physics is the elucidation of the substructure of nucleons (proton and neutron) as a triad of quarks. This important breakthrough in our understanding of elementary particles (Sect. 5.6.4) was achieved by scattering of high energy  $\gamma$  particles on protons.

In general, scattering of particle waves on a target, for example, a macroscopic crystal or a composed particle as a nucleon, is a complex many body problem. The way a scattering experiment is performed, however, suggests a mathematically feasible approximate description in terms of a single particle problem.

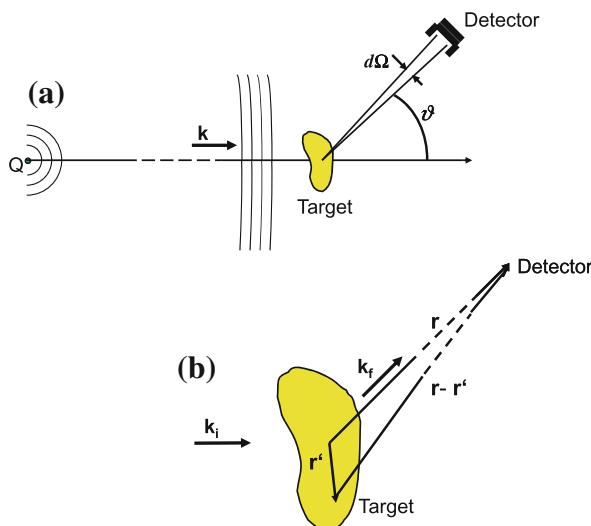
Firstly, we describe the scattering target by a spatially fixed scattering potential. This potential acts on the incident particle, or more precisely, on the particles belonging to the ensemble of the incident primary particle beam. This primary beam is described by a spherical wave which originates from a source  $Q$  (Fig. 6.21a). At

the target position at a large distance from  $Q$  the wave front is nearly flat and the primary wave can be represented by a plane wave with wave vector  $\mathbf{k}_0$ .

In case of a spatially fixed scattering potential, no energy can be transferred from the scattering particle to the target and the scattering process is elastic. In case of inelastic scattering, the target must have internal degrees of freedom, e.g. vibrations of the atoms constituting the target, which can exchange energy with the scattering particle.

Secondly, in a scattering experiment (Fig. 6.21) we detect particles scattered from a spatially restricted target within a small limited detection angle  $d\Omega$ . The detector is located at a large distance from the target and its position with respect to the direction of the incident primary beam ( $z$  axis) is described by the angles  $\vartheta$  and  $\varphi$  (around  $z$  axis). The spatial distribution of the scattered particles (scattered radiation) is, thus, measured at a distance  $r$  from the target as a function of  $\vartheta$  and  $\varphi$  (Fig. 6.21).

The model scenario of a typical scattering experiment for solving approximately the Schrödinger equation is, thus, characterized by an incoming plane wave describing a stationary flux of a particle ensemble. These primary particles interact with the spatially limited target during a limited time while they pass the target. After having passed the target the particles on their way to the detector occupy again quantum states which correspond to free motion and are not subject to any interaction.



**Fig. 6.21 a, b** Scheme of a scattering experiment with a particle beam on a scattering target. **a** A particle wave originating from source  $Q$  moves towards a target; near the target at large distance from  $Q$  the spherical wave can be approximated by a plane wave with wave vector  $\mathbf{k}$ . Again, at large distance from the target the scattered particles are detected under the scattering angle  $\vartheta$  by a detector with an angle resolution  $d\Omega$ . **b** Explanation of position coordinates and wave vectors used for the description of the scattering experiment

This approximate description of the scattering process which is intuitively derived from the experimental set-up is the basis for the following so-called scattering theory.

### 6.6.1 Scattered Waves and Differential Scattering Cross Section

In the described scattering experiment scattering intensities, that is, numbers of scattered particles detected per time under a certain angle at the detector position are measured. The detector position is described by the angles  $\vartheta$  and  $\varphi$  with respect to the direction of the incoming primary particle beam. In order to characterize the scattering strength of a target potential, the scattering intensity has to be related to the rate of incoming primary particles and to the angle opening  $d\Omega$  of the detector. Hereby, the most important quantity for the description of a scattering experiment is defined as the *differential scattering cross section*:

$$\frac{d\sigma(\vartheta, \varphi)}{d\Omega} = \frac{\text{number of particles scattered into } d\Omega/\text{sec}}{\text{number of incident particles/sec}}. \quad (6.180)$$

Based on the simplifying assumptions about the scenario of a scattering experiment, the wave function of a scattered particle is represented as a sum of the incident plane wave (primary particle beam propagating along  $z$ ) and the scattering solution  $\psi_{sc}$ :

$$\psi_k = Ae^{ikz} + \psi_{sc}(r, \vartheta, \varphi). \quad (6.181)$$

Note that a stationary flow of particles in an ensemble is described such that  $\psi_k$  is a solution of the stationary time-independent Schrödinger equation.

Within our approximate description of the scattering experiment the scattered particles propagate as free particles between target and detector, since the scattering target potential is spatially limited and the distance between target and detector is large as compared to the target extension. The wave function after scattering  $\psi_{sc}$  can, thus, be represented as a spherical wave of free particles starting at the target with an amplitude decaying as  $1/r$  with  $r$  as the distance from the target. The probability to find a scattered particle at a distance  $r$  from the target decreases as  $1/r^2$  as expected. The internal structure of the scattering potential determines the angle dependent scattering amplitude  $f(\vartheta, \varphi)$  of the spherical scattered wave:

$$\psi_k \xrightarrow[r \rightarrow \infty]{} Ae^{ikz} + f(\vartheta, \varphi) \frac{e^{ikr}}{r}. \quad (6.182)$$

Within the frame of this solution to the scattering problem all information about the scattering target is contained in the function  $f(\vartheta, \varphi)$  which consequently also determines the differential scattering cross section (6.180). For the calculation of the scattered particle flux measured in the experiment, the quantum mechanical current

density operator (3.79) must be applied to  $\psi_k$  (6.181). Hereby a problem arises: The scattered wave  $\psi_{sc}$  (6.181), (6.182) has a significantly smaller amplitude than the primary wave  $A \exp(ikz)$ . Since the current is quadratic in  $\psi_k$ , it contains mixed elements of the sum (6.180) in which the primary wave dominates the scattered wave.

For an analysis of the scattered radiation, therefore, the primary wave must be omitted. This approximation is justified because of the following reason. Although the primary particle beam is described by a plane wave with flat extended wave fronts at the target position, the beam extension is limited by apertures such that only the target is illuminated over its cross section. A large part of the primary particles passes the target without scattering and appears as a sharp high intensity spot at the center of the detection plane. This part of the signal can be omitted from the detection of scattered particles since for the scattering experiment only particles being scattered into various other directions than the direction of the incident beam are of interest. Only by masking the major part of the incident primary beam a sensitive detection of scattered particles (waves) is possible.

Current densities of incident beam and scattered waves are, therefore, separately calculated. For the primary beam, we obtain by means of (3.79):

$$|\mathbf{j}_{inc}| = \frac{\hbar}{2mi} (e^{-ikz} \nabla e^{ikz} - e^{ikz} \nabla e^{-ikz}) = \frac{\hbar k}{m}. \quad (6.183)$$

As expected, this is the particle velocity of a single particle, that is, its current density. For the calculation of the scattered wave at large distance from the target

$$\psi_{sc} = f(\vartheta, \varphi) \frac{e^{ikr}}{r} \quad (6.184)$$

we must use the nabla operator  $\nabla$  in spherical coordinates (5.32):

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\vartheta \frac{\partial}{r \partial \vartheta} + \mathbf{e}_\varphi \frac{1}{r \sin \vartheta} \frac{\partial}{\partial \varphi}. \quad (6.185)$$

Since the scattered particles are detected far away from the target ( $r \rightarrow \infty$ ) the derivatives with respect to  $\vartheta$  and  $\varphi$  are irrelevant because of the prefactor  $1/r$ . Only the derivative in  $r$  direction determines the current density. Neglecting  $1/r^2$  terms, we get

$$\frac{\partial}{\partial r} f(\vartheta, \varphi) \frac{e^{ikr}}{r} \simeq f(\vartheta, \varphi) ik \frac{e^{ikr}}{r} \quad (6.186)$$

and finally for the current density of scattered particles:

$$\mathbf{j}_{sc} = \mathbf{e}_r |f(\vartheta, \varphi)|^2 \frac{1}{r^2} \frac{\hbar k}{m}. \quad (6.187)$$

The probability for a particle to be detected within the solid angle element  $d\Omega$  is

$$W(d\Omega) = \mathbf{j}_{sc} \cdot \mathbf{e}_r r^2 d\Omega. \quad (6.188a)$$

The differential scattering cross section  $d\sigma/d\Omega$  is, thus, derived as

$$\frac{\partial\sigma}{\partial\Omega} d\Omega = \frac{W(d\Omega)}{j_{inc}} = |f(\vartheta, \varphi)|^2 d\Omega, \quad (6.188b)$$

$$\frac{\partial\sigma}{\partial\Omega} = |f(\vartheta, \varphi)|^2. \quad (6.188c)$$

Knowledge of the scattering amplitude  $f(\vartheta, \varphi)$  provides all possible information about the target, which can be obtained by the measurement of the differential scattering cross section.

### 6.6.2 Scattering Amplitude and Born Approximation

For the calculation of the important quantity  $f(\vartheta, \varphi)$ , the scattering amplitude, the time-independent Schrödinger equation for primary and scattered particles has to be solved. Since both primary and scattered particle beams consist of freely propagating particles, the energy eigenvalues are those of free particles:  $E_k = \hbar^2 k^2 / 2m$ . But, of course, the Schrödinger equation contains the scattering potential  $V(\mathbf{r})$ , that is,

$$\left( \frac{-\hbar^2}{2m} \Delta + V(\mathbf{r}) \right) \psi_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m} \psi_{\mathbf{k}}, \quad \text{respectively} \quad (6.189a)$$

$$(\nabla^2 + k^2) \psi_{\mathbf{k}} = \frac{2m}{\hbar^2} V(\mathbf{r}) \psi_{\mathbf{k}}. \quad (6.189b)$$

The time-independent solution  $\psi_{\mathbf{k}}$  is assumed as in (6.182) but with a  $\mathbf{k}$  vector of the primary wave in arbitrary direction:

$$\psi_{\mathbf{k}} \xrightarrow[r \rightarrow \infty]{} C e^{i\mathbf{k}\cdot\mathbf{r}} + f(\vartheta, \varphi) \frac{e^{ikr}}{r}. \quad (6.190)$$

Note that the general solution contains also the time dependent factor  $\exp(-E_{\mathbf{k}}t/\hbar)$ . This factor fixes the spherical wave as originating from the target. It is evident that in regions of vanishing scattering potential  $V(\mathbf{r}) = 0$  the Schrödinger equation (6.189b) is solved by the incident primary plane wave. The scattered wave essentially determined by the scattering amplitude  $f(\vartheta, \varphi)$  is calculated by plugging  $\psi_{\mathbf{k}}$  (6.190) into (6.189b).

For the solution of the differential equation (6.189b), we remind you of the Poisson equation of electrostatics:

$$\nabla^2 \phi = -\frac{\rho(\mathbf{r})}{\epsilon_0}. \quad (6.191)$$

It allows the the calculation of the electric potential  $\phi(\mathbf{r})$  from the spatial distribution of charge  $\rho(\mathbf{r})$ , the charge density. Note the close similarity between (6.189b) and (6.191). For the solution of (6.191) one imagines the charge density  $\rho(\mathbf{r})$  to be composed of a collection of point-like charges  $e\delta(\mathbf{r} - \mathbf{r}')$  at the positions  $\mathbf{r}'$  or alternatively of infinitesimal space elements with the charge  $\rho(\mathbf{r}') d^3 r'$ . The potential of a point charge is the familiar Coulomb potential:

$$d\phi(\mathbf{r}) = \frac{e\delta(\mathbf{r} - \mathbf{r}')}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}'|} \simeq \frac{\rho(\mathbf{r}') d^3 r'}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}'|}. \quad (6.192)$$

Since (6.191) is a linear differential equation, the solving potential  $\phi(\mathbf{r})$  is a linear superposition (sum) of the contributions  $d\phi(\mathbf{r})$  (6.192). The solution of (6.191) in case of a point charge is commonly called after its inventor *Green function*  $G(\mathbf{r} - \mathbf{r}')$ , that is, the definition of the Green function is given by

$$\nabla^2 G(\mathbf{r} - \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (6.193a)$$

The Green function for the point charge potential is therefore:

$$G(\mathbf{r} - \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}. \quad (6.193b)$$

The potential  $\phi(\mathbf{r})$  is obtained as the sum, respectively the integral over the point charge Green functions, i.e.

$$\phi(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}') \frac{\rho(\mathbf{r}')}{\epsilon_0} d^3 r' = \int \frac{1}{4\pi\epsilon_0} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \quad (6.194)$$

By inserting this expression into (6.191) one easily demonstrates that the potential solves the Poisson equation:

$$\nabla^2 \phi = \int \nabla^2 G(\mathbf{r} - \mathbf{r}') \frac{\rho(\mathbf{r}')}{\epsilon_0} d^3 r' = - \int \delta(\mathbf{r} - \mathbf{r}') \frac{\rho(\mathbf{r}')}{\epsilon_0} d^3 r' = - \frac{\rho(\mathbf{r})}{\epsilon_0}. \quad (6.195)$$

An analogous procedure leads to the solution of the differential equation (6.189b). We firstly determine the Green function from the corresponding equation

$$(\nabla^2 + k^2) G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \quad (6.196)$$

In analogy to the calculation of the electric potential the term on the right side of (6.189b) must be assumed to be the equivalent to the charge density in the Poisson equation. The solution  $\psi_{\mathbf{k}}$  is then calculated by multiplying the right term of (6.189b) with the Green function obtained from (6.196) and integrating over the spatial region of the scattering potential  $V(\mathbf{r})$ .

In principle, we can guess the Green function of a point-like scattering center, which is assumed to be the constituting element of the spatially extended scattering potential  $V(\mathbf{r})$ : We expect a spherical wave  $\exp(ik|\mathbf{r} - \mathbf{r}'|)/|\mathbf{r} - \mathbf{r}'|$  originating from the scattering center  $\mathbf{r}'$ . For the formal calculation we generalize the ansatz (6.193b) by a function  $u(\mathbf{r} - \mathbf{r}')$ :

$$G(\mathbf{r}, \mathbf{r}') = G(\mathbf{r} - \mathbf{r}') = \frac{u(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (6.197)$$

Because of (6.196), the function  $u(\mathbf{r} - \mathbf{r}')$  is then determined for  $\mathbf{r} - \mathbf{r}' \neq \mathbf{0}$  by

$$(\nabla^2 + k^2) \frac{u(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = 0. \quad (6.198a)$$

For simplicity reasons, we use the abbreviation  $R = |\mathbf{r} - \mathbf{r}'|$  and apply the delta operator in spherical coordinates:

$$(\nabla_R^2 + k^2) \frac{u(R)}{R} = \left[ \frac{1}{R^2} \frac{\partial}{\partial R} \left( R^2 \frac{\partial}{\partial R} \right) + k^2 \right] \frac{u(R)}{R} = 0. \quad (6.198b)$$

Then, the following simple differential equation is obtained

$$u''(R) = -k^2 u(R) \quad (6.198c)$$

which has the solutions  $u(R) = \exp(\pm ikR)$ . The Green function of (6.196), thus, follows as

$$G|\mathbf{r} - \mathbf{r}'| = A \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} + B \frac{e^{-ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|}. \quad (6.199a)$$

Only spherical waves running out from the scattering center (within the target) are physically meaningful. This determines the time-dependent exponential factor as  $\exp(-iE_k t/\hbar)$  which requires  $B = 0$ . As guessed, we obtain the Green function of a point-like scattering center as

$$G(\mathbf{r} - \mathbf{r}') = A \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|}. \quad (6.199b)$$

The solution to the scattering problem (6.189a), (6.189b) can now be formulated as a superposition, that is, integration over the term  $2mV\psi_{\mathbf{k}}/\hbar^2$  in (6.189b) multiplied with the Green function (6.199b):

$$\begin{aligned} \psi_{\mathbf{k}} &= Ce^{i\mathbf{k}\cdot\mathbf{r}} + \frac{2m}{\hbar^2} \int G|\mathbf{r} - \mathbf{r}'| V(\mathbf{r}') \psi_{\mathbf{k}}(\mathbf{r}') d^3 r' \\ &= Ce^{i\mathbf{k}\cdot\mathbf{r}} + A \frac{2m}{\hbar^2} \int \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} V(\mathbf{r}') \psi_{\mathbf{k}}(\mathbf{r}') d^3 r'. \end{aligned} \quad (6.200)$$

We will not go into details of the determination of the normalization constants  $A$  and  $C$ . It should only be mentioned that both constant have a well defined ratio which determines the relative phase between incident and scattered wave. By plugging (6.199b) into (6.196) we obtain  $A/C = -1/4\pi$  by using the limit  $|\mathbf{r} - \mathbf{r}'| \rightarrow 0$  and the normalization of the 3D  $\delta$  function.

For the following it is important to consider the integral term in the scattering solution (6.200) more in detail. In contrast to the case of the Poisson equation (6.193a), (6.193b) this integral contains the solution  $\psi_{\mathbf{k}}$  to the problem itself rather than only a spatially fixed charge density  $\rho$  as in the electrostatic problem (6.194). Actually the relation (6.200) is not a real solution to the scattering problem but rather an integral equation for the calculation of  $\psi_{\mathbf{k}}$ . The solution of this integral equation is easily performed by iteration: We start with a known (maybe guessed) bad solution  $\bar{\psi}_{\mathbf{k}}(\mathbf{r}')$  to the problem and plug it into the integral on the right side of (6.200) to calculate a better solution  $\psi_{\mathbf{k}}$ . This better solution is then inserted again into the integral of (6.200) instead of  $\bar{\psi}_{\mathbf{k}}$ . Again an even better solution is calculated and so on. The iteration process can be performed up to arbitrary high order.

What, now, is a convenient function to start the iteration process? The zero order approximation is certainly given as the solution for vanishing scattering potential  $V = 0$ . In this case the incident primary plane wave alone solves the problem. The first order approximation of  $\psi_{\mathbf{k}}$  is obtained by inserting this incident plane wave into the integral in (6.200). This first order approximation to the scattering problem is called *Born approximation*:

$$\psi_{\mathbf{k}}(\mathbf{r}) \propto e^{ik\cdot\mathbf{r}} - \frac{2m}{4\pi\hbar^2} \int \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} V(\mathbf{r}') e^{ik\cdot\mathbf{r}'} d^3 r'. \quad (6.201)$$

The scattering potential  $V(\mathbf{r}')$  is spatially limited by the extension of the target; its extension is small in comparison with the distance  $r$  between detector and target location. Consequently within the integral of (6.201) there is always  $|\mathbf{r}| \gg |\mathbf{r}'|$  and we can approximate  $|\mathbf{r} - \mathbf{r}'| \approx r$  in the denominator. This approximation can not be made in the exponent of the exponential function under the integral, because the directional dependence of the scattering would be lost in that case. The exponent is approximated somewhat more carefully:

$$\begin{aligned} k|\mathbf{r} - \mathbf{r}'| &= kr \left| \mathbf{e}_r - \mathbf{e}_{r'} \frac{\mathbf{r}'}{r} \right| \\ &= kr \sqrt{\left( \mathbf{e}_r - \mathbf{e}_{r'} \frac{\mathbf{r}'}{r} \right) \left( \mathbf{e}_r - \mathbf{e}_{r'} \frac{\mathbf{r}'}{r} \right)} \\ &\approx kr \sqrt{\left( 1 - 2 \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right)} \approx kr \left( 1 - \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right) \\ &= kr - (k\mathbf{e}_r) \cdot \mathbf{r}' \\ &= kr - \mathbf{k}_f \cdot \mathbf{r}'. \end{aligned} \quad (6.202)$$

In the last step, the wave vector  $k\mathbf{e}_r$ , in the direction of detection has been denoted  $\mathbf{k}_f$  (final), as it describes the final state of the particle after being scattered. Analogously, we can denote the initial state before scattering by  $\mathbf{k}_i$ , the wave vector of the incident particles. The wave function (6.201) is, then, written as

$$\psi_{\mathbf{k}}(r) \propto e^{ik_i \cdot \mathbf{r}} - \frac{2m}{4\pi\hbar^2} \frac{e^{ikr}}{r} \int e^{-ik_f \cdot \mathbf{r}'} V(\mathbf{r}') e^{ik_i \cdot \mathbf{r}'} d^3 r'. \quad (6.203)$$

This wave function has exactly the form of the before introduced scattering solution for large distances from the target (6.182), (6.190): A spherical wave runs out from the target; its amplitude is spatially modulated as a function of the scattering angles  $\vartheta$  and  $\varphi$ . The scattering amplitude  $f(\vartheta, \varphi)$  determined essentially by the integral in (6.203) contains all possible information about the internal structure of the target in terms of the scattering potential  $V(\mathbf{r}')$ .

The scattering amplitude

$$f(\vartheta, \varphi) = -\frac{m}{2\pi\hbar^2} \int e^{-ik_f \cdot \mathbf{r}'} V(\mathbf{r}') e^{ik_i \cdot \mathbf{r}'} d^3 r' \quad (6.204)$$

is easily interpreted in the following way: The incident primary wave with wave vector  $\mathbf{k}_i$  excites, in proportion to the potential value  $V(\mathbf{r}')$  at the position  $\mathbf{r}'$ , a scattered wave  $\exp(-ik_f \cdot \mathbf{r}')$  which runs out from the volume element  $d^3 r'$  at the position  $\mathbf{r}'$ . This partial wave contributes to the total scattered wave corresponding to the potential strength at the particular position  $\mathbf{r}'$ . The total scattered wave is obtained by interference, i.e. superposition (integral over  $\mathbf{r}'$ ) of all partial waves originating from the different locations in the target (Fig. 6.21b). This result is completely analogous to the so-called Fresnel description of interference phenomena in classical optics. Furthermore, the described Born approximation includes only single scattering events of the incident wave. The next better approximation, where the scattering solution  $\psi_{\mathbf{k}}$  obtained from Born approximation is again plugged into the integral of (6.200), takes into account twofold scattering on the target potential. Also those waves which are scattered a second time in the target after having undergone a first scattering event already, contribute to the interference pattern of the total scattered wave. Further improved approximations by multiple iteration steps allow also for 3-fold, 4-fold etc. scattering events in the target. Higher order contributions, of course, contribute less to the total scattered wave. In the following we will consider only single scattering events in the frame of the Born approximation. This approximation of the scattering problem is often called *kinematic scattering theory*. Higher approximations taking into account multiple scattering are called *dynamic scattering theory*.

The change of the particle wave vector upon scattering determines a so-called wave vector or momentum transfer to the particle:

$$\mathbf{K} = \mathbf{k}_f - \mathbf{k}_i, \quad \text{or} \quad (6.205a)$$

$$\hbar\mathbf{K} = \mathbf{p}_f - \mathbf{p}_i. \quad (6.205b)$$

In case of elastic scattering, (6.205a), (6.205b) means a change of propagation direction. Using (6.205a), we can write the scattering amplitude (6.204) as

$$f(\vartheta, \varphi) = f(\mathbf{K}) = -\frac{m}{2\pi\hbar^2} \int e^{-i\mathbf{K}\cdot\mathbf{r}'} V(\mathbf{r}') d^3 r'. \quad (6.206)$$

In Born approximation, the scattering amplitude is nothing else but the Fourier transform of the scattering potential with respect to the wave vector, respectively, the momentum transfer to the scattered particle.

Again the analogy to classical optics is obvious. In the Fresnel interpretation of light diffraction on spatially extended structures, the diffraction pattern is a superposition of partial waves originating from points of the diffracting structure.

From the representation of the scattering amplitude as a Fourier transform (6.206), we can conclude that spatially far extended scattering potentials scatter particles in narrow angles while small targets (in comparison to the wave length  $\lambda = 2\pi/k$  of the primary beam) scatter into large solid angles.

A last remark on Born approximation: The integral in the scattering amplitude (6.204) can also be written as

$$f(\vartheta, \varphi) \propto \langle \mathbf{k}_f | \hat{V} | \mathbf{k}_i \rangle; \quad (6.207)$$

with  $\langle k_f |$  and  $| k_i \rangle$  as bra and ket of final and initial state of the scattered particle, that is, the scattered and the incident wave. Within this view the scattering potential,  $\hat{V}(\mathbf{r})$  transfers an initial state into a final state, a typical problem of time dependent perturbation theory (Sect. 6.4.1). Fermi's golden rule ((6.111a), (6.111b)) is evidently recognized in (6.207). We could indeed have derived all results about Born's approximation by using Fermi's golden rule.

### 6.6.3 Coulomb Scattering

One of the most important examples for the application of Born's approximation is the scattering of a charged particle, for example, an electron, on the Coulomb potential of another charged particle (charge  $Ze$ ). We encounter this problem again and again, from the beginning of atomic physics up to modern high energy particle physics or the calculation of electrical resistance in solid state physics. In semiconductors, an essential contribution to the electrical resistance is due to scattering of conduction electrons or holes on the charged cores of ionized donor or acceptor atoms, for example, As (donors) or B cores (acceptors) incorporated in Si.

In the considered case, the interaction between the spatially fixed, charged dopant core, the target, and the scattered particle is described by the Coulomb potential  $Ze^2/4\pi\epsilon_0 r'$ , where  $r'$  is the instantaneous distance between electron and charged target (charge  $Ze$ ). Because of the spherical symmetry of the scattering Coulomb potential [ $V(\mathbf{r}') = V(r')$ ], the scattering amplitude (6.206) can be written as

$$f(\vartheta, \varphi) = -\frac{m}{2\pi\hbar^2} \int e^{-i\mathbf{K}\cdot\mathbf{r}'} V(\mathbf{r}') d^3 r' \quad (6.208)$$

with  $\mathbf{K} = \mathbf{k}_f - \mathbf{k}_i$  as scattering vector or wave vector transfer. For elastic scattering ( $|\mathbf{k}_i| = |\mathbf{k}_f|$ ), one derives from Fig. 6.22:

$$\frac{1}{2} K = k \sin(\vartheta/2), \quad \text{respectively} \quad (6.209a)$$

$$K^2 = 4k^2 \sin^2(\vartheta/2). \quad (6.209b)$$

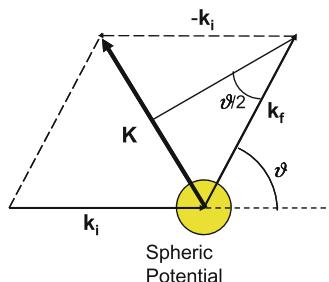
Because of the spherical symmetry of  $V(\mathbf{r})$ , we choose spherical coordinates for calculating the volume integral and for simplicity reasons we put the  $z$  axis into the direction of the scattering vector  $\mathbf{K}$ . We then obtain:

$$\mathbf{K} \cdot \mathbf{r}' = Kr' \cos \vartheta', \quad (6.210a)$$

$$d^3 r' = r'^2 dr' \sin \vartheta' d\vartheta' d\varphi' = -r'^2 dr' (\cos \vartheta') d\varphi'. \quad (6.210b)$$

The integration over  $\varphi'$  yields a factor  $2\pi$  and the integration over  $\vartheta'$  is transformed into an integration over  $\cos \vartheta'$ . This modifies the integration limits into  $\cos \vartheta' = -1$  ( $\vartheta' = \pi$ ) and  $\cos \vartheta' = 1$  ( $\vartheta' = 0$ ) and with  $\cos \vartheta' = \xi$  we obtain from (6.208):

$$\begin{aligned} f(\vartheta, \varphi) &= \frac{m}{\hbar^2} \int_{\xi=1}^{\xi=-1} d\xi dr' e^{-iKr'\xi} V(\mathbf{r}') r'^2 \\ &= \frac{m}{\hbar^2} \int \frac{1}{-iKr'} [e^{iKr'} - e^{-iKr'}] V(\mathbf{r}') r'^2 dr' \\ &= -\frac{2m}{\hbar^2} \int \frac{\sin Kr'}{K} V(\mathbf{r}') r' dr' = f(\vartheta). \end{aligned} \quad (6.211)$$



**Fig. 6.22** Explanation of symbols for the description of particle scattering on a spherical potential, e.g., a Coulomb potential. The masses of the particles are assumed to be very different, i.e., the much heavier target particle can be assumed to be at rest.  $\mathbf{k}_i$  and  $\mathbf{k}_f$  are the wave vectors of the incident and the scattered particle, respectively.  $\mathbf{K}$  wave vector transfer,  $\vartheta$  scattering angle

Note that after the integration over  $\vartheta'$  (not  $\vartheta$ ) the scattering amplitude is still a function of the scattering angle  $\vartheta$ .

Rather than plugging the singular Coulomb potential into (6.211) we prefer to use the the screened so-called *Yukawa potential*:

$$V(r') = g \frac{e^{-\alpha r'}}{r'} \quad \text{with } g = \frac{Ze^2}{4\pi\epsilon_0}. \quad (6.212)$$

This potential which is named after its inventor does not get singular at  $r' \rightarrow 0$  and is easier to handle in the integration. Equation (6.211) then yields

$$\begin{aligned} f(\vartheta) &= -\frac{2mg}{\hbar^2 K} \int_0^\infty dr' \frac{1}{2i} [e^{(iK-\alpha)r'} - e^{-(iK+\alpha)r'}] \\ &= -\frac{2mg}{\hbar^2 K} \frac{1}{2i} \left[ \frac{-1}{iK - \alpha} + \frac{-1}{iK + \alpha} \right] = \frac{-2mg}{\hbar^2(K^2 + \alpha^2)}. \end{aligned} \quad (6.213)$$

By use of (6.209b) and (6.212) and by squaring the scattering amplitude  $f(\vartheta)$  we obtain the differential scattering cross section

$$\frac{d\sigma}{d\Omega} = \frac{4m^2 g^2}{\hbar^4(K^2 + \alpha^2)} = \frac{m^2(Ze^2)^2}{4\pi^2\hbar^4\epsilon_0^2} \frac{1}{[\alpha^2 + 4k^2 \sin^2(\vartheta/2)]^2}. \quad (6.214)$$

In the limit of a negligible reciprocal Yukawa screening length  $\alpha$  scattering on the Coulomb potential is described. Representing the wave vector  $k$  of the scattered electron by its energy  $E = \hbar^2 k^2 / 2m$  finally yields the following expression for the differential scattering cross section:

$$\frac{d\sigma}{d\Omega} = \frac{m^2(Ze^2)^2}{16\pi^2\hbar^4\epsilon_0^2 k^4 \sin^4(\vartheta/2)} = \frac{(Ze^2)^2}{64\pi^2\epsilon_0 E^2 \sin^4(\vartheta/2)}. \quad (6.215)$$

An essential issue of (6.215) is the fact that the scattering cross section for scattering of charged particles on a Coulomb potential decreases with the kinetic energy  $E$  of the particles as  $E^{-2}$ . The faster the particles pass the target the less they are affected by the scattering process—a very plausible conclusion.

We want to use this important property of Coulomb scattering to estimate the temperature dependence of the mobility  $\mu$  of free conduction electrons in a doped semiconductor, where the resistance is partially determined by scattering of the carriers on ionized dopant atoms. In order to control the conductivity of a semiconductor dopant atoms with a chemical valency different from that of the host material are incorporated during epitaxial growth (Appendix B) or afterwards by diffusion or ion implantation. In Si, for example, pentavalent As atoms are substitutionally built in on tetravalent Si sites of the lattice. Only four valence electrons of the As, then, participate in the chemical bonding; the fifth valence electron is only weakly bound to the As core. At room temperature, a large amount of these electrons (according

to Fermi statistics, Sect. 8.3.4) is released from the bond to the core (ionized) and can contribute as quasi-free electrons to an electrical current. Upon their motion through the crystal lattice due to an applied electric field, the electrons are scattered on the ionized, monovalent positive As donors. These scattering processes essentially contribute to the electrical resistance, in particular, at temperatures below room temperature (300 K).

In the electrical conductivity  $\sigma = \rho^{-1} = en\mu$ , the scattering processes are taken into account in the mobility  $\mu$ . This quantity  $\mu$  describes the average drift velocity  $\langle v \rangle$  of the carriers in the electric field  $\mathcal{E}$ :

$$\mu = \langle v \rangle / \mathcal{E}. \quad (6.216)$$

The drift velocity is essentially determined by scattering processes, which limit the carrier velocity originating from acceleration in the external electric field. From simple Drude theory and from semi-classic Boltzmann theory of electrical transport [9], the carrier mobility is essentially given by the mean free time of flight  $\tau$  of a carrier between two scattering processes:

$$\mu \propto \frac{e\tau}{m}, \quad (6.217)$$

$m$  is the carrier mass, the effective mass in a solid (Sect. 8.3.5). More precisely,  $\tau$  is the relaxation time, the average time during which the non-equilibrium distribution of the carriers induced by the electric field relaxes into the equilibrium distribution after switching off the field. This time is close to the mean free time of flight and for certain kinds of collisions both times are identical.

The mean free time of flight  $\tau$  is directly related to the scattering cross section, while the temperature dependence of the average drift velocity can be obtained from a consideration of the carrier statistics. Note that the free electron concentration in a semiconductor (conduction band, Sect. 8.3.5) is considerably lower, typically  $10^{17} \text{ cm}^{-3}$ , than the total electron concentration ( $\approx 10^{23} \text{ cm}^{-3}$ ). Free electrons in a semiconductor must, therefore, originate from occupied electronic states being located far above the Fermi level [ $(E - E_F) \gg 2kT$ ] where the Fermi distribution (Sect. 5.6.3) assumes very small values ( $\ll 1$ ). There, the Fermi function can be approximated by the so-called Boltzmann approximation:

$$f(E, T) = \frac{1}{\exp(\frac{E-E_F}{kT}) + 1} \approx \exp\left(-\frac{E - E_F}{kT}\right) \ll 1. \quad (6.218)$$

Conduction electrons in a semiconductor are, thus, approximately described by the Boltzmann distribution as classical particles of an ideal gas. Correspondingly, the average value of the velocity exhibits a temperature dependence

$$\langle v \rangle \propto \sqrt{T}. \quad (6.219)$$

For the estimation of the scattering cross section  $\Sigma$  we imagine an electron approaching a positive donor core. Significant scattering should occur when the electron penetrates an area  $\Sigma$  perpendicular to the flight direction. By this definition,  $\Sigma$  is proportional to the above differential scattering cross section (6.215):

$$\Sigma \propto E^{-2} \propto \langle v \rangle^{-4}. \quad (6.220)$$

With  $\lambda$  as the mean free path between two scattering events on the average one scattering process occurs within a volume  $\Sigma\lambda$ . With  $N_D$  as the density of donor scattering centers, we obtain

$$\Sigma\lambda \propto N_D^{-1}. \quad (6.221)$$

Expressing the mean free path  $\lambda$  by the mean free time of flight ( $\tau$ )

$$\lambda = \langle v \rangle \tau, \quad (6.222)$$

yields the following relation:

$$\tau \langle v \rangle \Sigma \propto N_D^{-1} = \text{const}, \quad (6.223a)$$

respectively with (6.220):

$$\tau \propto \frac{1}{\langle v \rangle \Sigma} \propto \langle v \rangle^3. \quad (6.223b)$$

Because of (6.219) we, then, obtain the temperature dependence of the electron mobility  $\mu$  (6.217) due to Coulomb scattering of the carriers on ionized donor cores as

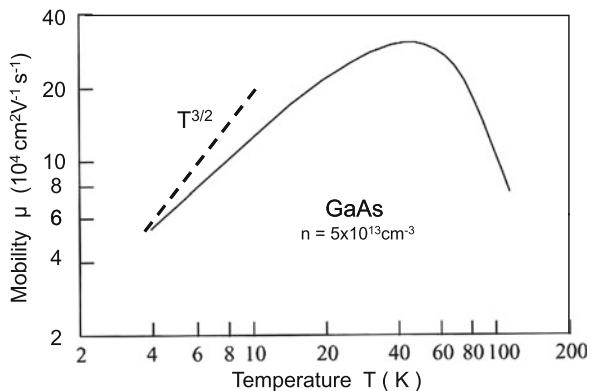
$$\mu(T) \propto T^{3/2}. \quad (6.224)$$

Figure 6.23 shows experimental results for  $n$ -doped GaAs. For temperatures below about 30 K, the overall trend of the temperature dependence is found but there is no exact agreement between theory and experimental curve. An important reason for the deviation is certainly the presence of additional scattering mechanisms. In particular, at higher temperatures above 50 K the drop of the mobility with increasing temperature is ascribed to scattering of the carriers on lattice vibrations (phonons, Sect. 8.4).

#### 6.6.4 Scattering on Crystals, on Surfaces and on Nanostructures

In the following, we want to apply scattering theory in its simplest form, the Born approximation, to crystalline solids including also systems with reduced dimensions as crystalline surfaces and crystallites with nanoscaled dimensions.

**Fig. 6.23** Experimentally determined electron mobility  $\mu$  of  $n$ -doped GaAs. In broken line the  $T^{3/2}$  dependence of  $\mu$  is plotted, which is obtained from a simple estimation of Coulomb scattering of the free electrons on ionized donor impurities [11]



As simplest example of a macroscopic crystal target with dimensions in the centimeter range, we assume a crystal with one atom in the elementary cell. In the ideal case equal atoms build up a three-dimensional (3D) arrangement with translational symmetry, the crystal lattice. As scattering particles X-ray photons, electrons or neutrons are most interesting for application. These particles are frequently used in analysis techniques in condensed matter physics. The scattering potentials which describe the interaction of the scattering particles with the target atoms are, of course, different for the different particles as they depend on the interaction mechanism between the partners. Neutrons, for example, are scattered solely by the atomic nuclei, that is, the scattering potential is extended over an extremely small space in comparison with the atomic dimension. This causes scattering into large angles as compared with scattering on more extended potentials with atomic radii. Such scattering potentials are given for the interaction of X-ray photons and electrons with atoms. Both types of particles interact with the electronic shell which determines the spatial extension of an atom. The interaction of photons and electrons with the electronic shell, that is, the corresponding scattering potential is, however, different. Electrons are scattered directly by electron-electron collisions while photons interact via dipole moments of oscillating electrons (Sect. 6.4.2).

In this section, we do not focus on details of the elementary scattering process but rather concentrate on the effect of the crystal lattice, the periodic arrangement of atoms. Independent on the type of the scattering particles we, therefore, use the general atomic scattering potential  $v(\mathbf{r} - \mathbf{r}_n)$  for the description of all three types of particles. Hereby,  $\mathbf{r}_n$  describe the time-independent positions of the atoms, that is, the coordinates of the atomic nuclei in the lattice. The vectors  $\mathbf{r}_n$  span the 3D periodic crystal lattice. Be  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  the unity vectors of the elementary cell, that is, in a primitive lattice the vectors connecting one atom with its three next neighbors, then  $\mathbf{r}_n$  can be written as

$$\mathbf{r}_n = m\mathbf{a} + n\mathbf{b} + p\mathbf{c}, \quad m, n, p \text{ integer numbers.} \quad (6.225)$$

The triple of integer numbers  $m, n, p$  is formally written as a subscript vector  $\mathbf{n}$  in this representation. The total scattering potential  $V(\mathbf{r})$  of the whole crystal is represented as the sum over the potentials of the single atoms:

$$V(\mathbf{r}) = \sum_{\mathbf{n}} v(\mathbf{r} - \mathbf{r}_{\mathbf{n}}). \quad (6.226)$$

This idealized potential is time-independent. The atomic sites  $\mathbf{r}_{\mathbf{n}}$  assumed to be fixed in space can not move and, therefore, can not interchange energy with the scattering particle. Only elastic scattering can be described by this time-independent potential.

For the calculation of the scattering amplitude (6.204), we insert (6.226) into (6.204) respectively, (6.207) and obtain by use of (6.205a), (6.205b), apart from a constant prefactor:

$$\begin{aligned} f(\vartheta, \varphi) &\propto \langle f | V(\mathbf{r}) | i \rangle \\ &= \int d^3 r' e^{-i\mathbf{k}_f \cdot \mathbf{r}'} V(\mathbf{r}') e^{i\mathbf{k}_i \cdot \mathbf{r}'} \\ &= \sum_n \int d^3 r' e^{-i\mathbf{K} \cdot (\mathbf{r}' - \mathbf{r}_n) - i\mathbf{K} \cdot \mathbf{r}_n} v(\mathbf{r}' - \mathbf{r}_n) \\ &= \sum_n e^{-i\mathbf{K} \cdot \mathbf{r}_n} \int d^3 r' e^{-i\mathbf{K} \cdot (\mathbf{r}' - \mathbf{r}_n)} v(\mathbf{r}' - \mathbf{r}_n). \end{aligned} \quad (6.227)$$

The integral within the sum runs over the spatial range of one single atom. This integral is identical for all atoms of the lattice. It is called *atomic scattering factor*  $f_A$ . We can take it out of the sum:

$$f(\vartheta, \varphi) \propto f_A \sum_n e^{-i\mathbf{K} \cdot \mathbf{r}_n} = f_A \sum_{m, n, p} (e^{-i\mathbf{K} \cdot \mathbf{a}})^m (e^{-i\mathbf{K} \cdot \mathbf{b}})^n (e^{-i\mathbf{K} \cdot \mathbf{c}})^p. \quad (6.228)$$

Since the atomic scattering potential has spherical symmetry, we can evaluate the atomic scattering factor  $f_A$  in spherical coordinates in analogy to Sect. 6.6.3:

$$\begin{aligned} f_A &= 4\pi \int v(\mathbf{r}') \frac{\sin Kr'}{K} r' dr' \\ &= 4\pi \int v(\mathbf{r}') \frac{\sin[4\pi r' \{\sin(\vartheta/2)\}/2]}{4\pi \sin(\vartheta/2)/\lambda} r' dr'. \end{aligned} \quad (6.229)$$

Hereby, we have assumed the atomic position as  $\mathbf{r}_{\mathbf{n}} = (0, 0, 0)$  and we have used (6.209a), (6.209b) as well as  $k = 2\pi/\lambda$  for the description of the scattered particle wave vector. In (6.229), the atomic scattering factor still depends on the scattering angle  $\vartheta$  and on the radiation wavelength  $\lambda$ .

Now the contribution of the periodic crystal lattice [sums in (6.228)] to the scattering amplitude is evaluated. By means of the sum rule for the geometrical series

$$\sum_{n=0}^N a^n = \frac{1 - a^{N+1}}{1 - a}, \quad a < 1 \quad (6.230)$$

the sums in (6.228) are calculated as

$$\sum_m^M (e^{-i\mathbf{K} \cdot \mathbf{a}})^m = \frac{1 - (e^{-i\mathbf{K} \cdot \mathbf{a}})^{M+1}}{1 - e^{-i\mathbf{K} \cdot \mathbf{a}}} = \frac{e^{-i\mathbf{K} \cdot \mathbf{a}} - e^{-i\mathbf{K} \cdot \mathbf{a}M}}{e^{-i\mathbf{K} \cdot \mathbf{a}} - 1}, \quad (6.231)$$

and analogously the sums over  $n$  and  $p$ .

The scattering amplitude  $f$  (6.228) enters the scattering cross section (6.188c) quadratically. The contribution of the crystal lattice to the scattering cross section is, thus, obtained by the absolute square of (6.231) (sum over  $m$ , only one direction):

$$\left| \sum_m^M (e^{-i\mathbf{K} \cdot \mathbf{a}})^m \right|^2 = \frac{1 - \cos((M+1)\mathbf{K} \cdot \mathbf{a})}{1 - \cos(\mathbf{K} \cdot \mathbf{a})} = \frac{\sin^2[(M+1)\mathbf{K} \cdot \mathbf{a}/2]}{\sin^2(\mathbf{K} \cdot \mathbf{a}/2)}. \quad (6.232)$$

The sums over  $n$  and  $p$  yield analogous expressions.

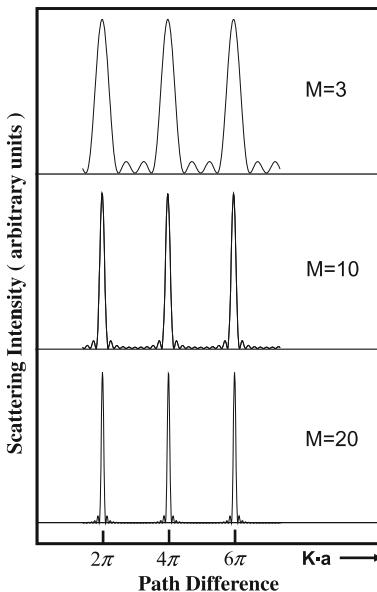
The expression (6.232) is well known as the so-called *grating function* from light diffraction on an optical grating. According to l'Hopital's rule (differentiation of numerator and denominator with respect to  $\mathbf{K} \cdot \mathbf{a}$ ), the expression stays finite; it has main maxima for vanishing denominators, that is, each time for  $\mathbf{K} \cdot \mathbf{a} = 2\pi h$  ( $h$  integer). Between these main maxima, the function oscillates with zeros each time for  $(M+1)\mathbf{K} \cdot \mathbf{a} = 2\pi$  ( $M$  integer), i.e. where the numerator vanishes (Fig. 6.24). Between two main maxima, there are  $M$  zeros of the function. The larger  $M$  is, the closer is the first zero point located to the main maximum and the sharper is the spectral band around the main maximum (Fig. 6.24).

For a macroscopic crystal with centimeter dimensions,  $M$ ,  $N$  and  $P$  are of the order of  $10^8$  and the bands around the main maxima become extremely sharp, nearly like a  $\delta$  function. Intensity in the scattered particle beams is observed, then, only for

$$\begin{aligned} \mathbf{a} \cdot \mathbf{K} &= 2\pi h, \\ \mathbf{b} \cdot \mathbf{K} &= 2\pi k, \quad h, k, l \text{ integer} \\ \mathbf{c} \cdot \mathbf{K} &= 2\pi l. \end{aligned} \quad (6.233)$$

These important conditions for the appearance of scattered particle peaks are called after their discoverer *Laue equations* (Max von Laue, Nobel price 1914). The scattering peaks are also called *Bragg (diffraction) peaks*. A simple interpretation of the Laue equations is possible by means of the so-called *reciprocal space* (reciprocal with respect to the real space of the crystal lattice). A general vector of the reciprocal space  $\mathbf{G}_{hkl}$  is represented by means of the basis vectors  $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$  as

$$\mathbf{G}_{hkl} = h\mathbf{g}_1 + k\mathbf{g}_2 + l\mathbf{g}_3. \quad (6.234)$$



**Fig. 6.24** Grating function for the description of diffraction (or scattering) of plane waves from a linear arrangement of equidistant narrow slits having a distance  $\mathbf{a}$ . The quantity  $\mathbf{K} \cdot \mathbf{a}$  is the path difference between diffracted and incident beam (wave) with  $\mathbf{K}$  as wave number transfer. The main diffraction peaks become sharper with increasing number  $M$  of scattering centers (slits)

By definition the reciprocal basis vectors obey the following conditions:

$$\mathbf{g}_1 \cdot \mathbf{a} = 2\pi, \quad \mathbf{g}_2 \cdot \mathbf{b} = 2\pi, \quad \mathbf{g}_3 \cdot \mathbf{c} = 2\pi, \quad (6.235)$$

$$\mathbf{g}_1 \cdot \mathbf{b} = \mathbf{g}_3 \cdot \mathbf{b} = 0, \quad \mathbf{g}_2 \cdot \mathbf{a} = \mathbf{g}_3 \cdot \mathbf{a} = 0, \quad \mathbf{g}_1 \cdot \mathbf{c} = \mathbf{g}_2 \cdot \mathbf{c} = 0.$$

These relations define the reciprocal lattice and by means of (6.225) we obtain immediately

$$\mathbf{G}_{hkl} \cdot \mathbf{r}_n = hma \cdot \mathbf{g}_1 + knb \cdot \mathbf{g}_2 + lpc \cdot \mathbf{g}_3 = 2\pi\nu. \quad (6.236)$$

Furthermore, it is easily seen that the equations for the definition of the reciprocal basis vectors (6.235) are fulfilled by

$$\mathbf{g}_1 = 2\pi \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \quad \text{and cyclic relations.} \quad (6.237)$$

Given a certain crystal lattice in real space we, therefore, construct the related reciprocal lattice according to the following recipe: The reciprocal basis vector  $\mathbf{g}_1$  is normal to the plane spanned by the vectors  $\mathbf{b}$  and  $\mathbf{c}$  of real space and it has a length  $|\mathbf{g}_1| = 2\pi/[\cos(\mathbf{g}_1, \mathbf{a})]$ . For a rectangular lattice the length of  $\mathbf{g}_1$  is just  $2\pi/a$ , essentially (apart from  $2\pi$ ) the reciprocal real space basis vector length. Figure 6.25

shows how the reciprocal basis vectors  $\mathbf{g}_1$  and  $\mathbf{g}_2$  of a lattice with oblique angles are constructed by means of the real space vectors  $\mathbf{a}$  and  $\mathbf{b}$ .

Using (6.235) and (6.236), we derive an alternative way to describe the occurrence of scattered intensity (Bragg) peaks, beside the Laue equations (6.233):  $\mathbf{K}$ , the wave vector transfer upon scattering must equal a reciprocal lattice vector, i.e.:

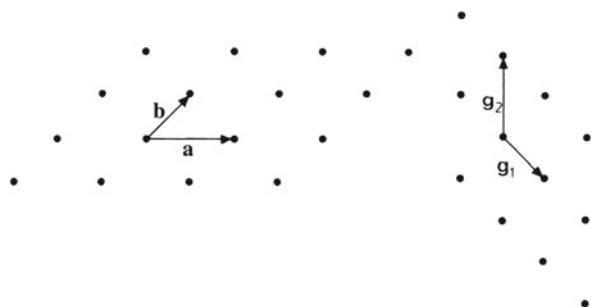
$$\mathbf{K} = \mathbf{G}_{hkl}. \quad (6.238)$$

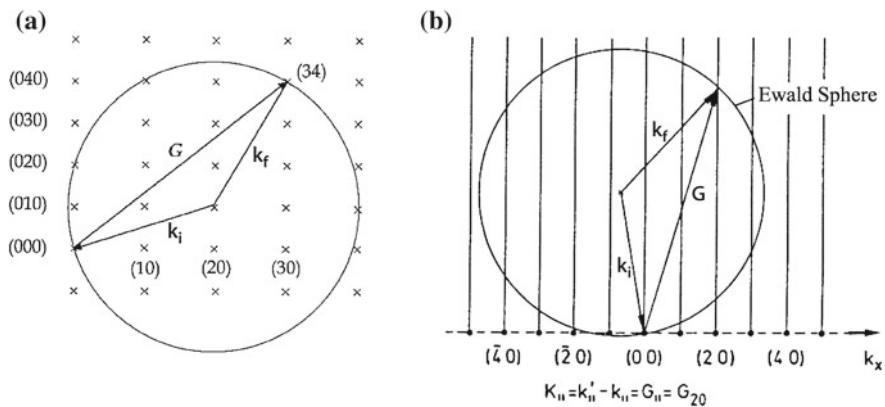
The scattering, respectively, diffraction peaks can be denoted according to vectors of the reciprocal space, that is, to the numbers  $h, k, l$  in the representation (6.234). These numbers are called *Miller indices*.

There is a simple geometrical construction called after its inventor *Ewald construction* which provides an overview over the possible diffraction (scattering) peaks generated by a particular crystal lattice (Fig. 6.26a). According to the recipe (6.235), respectively Fig. 6.25 one constructs the reciprocal lattice to the real space lattice. Then, the wave vector of the incident primary particles  $\mathbf{k}_i$  is plotted into the reciprocal lattice with its top pointing to the (arbitrarily chosen) origin (0 0 0) of the reciprocal lattice. Around the origin of the vector  $\mathbf{k}_i$  a sphere or in two dimensions a circle is drawn, the Ewald sphere or the Ewald circle. Because of elastic scattering, that is,  $k_f = k_i$ . The condition  $\mathbf{G}_{hkl} = \mathbf{K}$  is fulfilled each time when the Ewald sphere or the Ewald circle cuts a point of the reciprocal lattice (Fig. 6.26a). The corresponding scattering peak ( $hkl$ ) is observed in the experiment.

From Fig. 6.26a, it is evident that for a fixed experimental geometry, that is, a constant wave vector  $\mathbf{k}_i$  fixed with respect to orientation of the crystal target, only a small number of scattering peaks ( $hkl$ ) will be observed. In order to obtain a larger number of scattering peaks in the experiment, the crystal (lattice) under study is rotated within the incident beam about an axis normal to the beam (*rotating crystal method*). Since the reciprocal lattice is associated in a fixed manner to the real crystal lattice, a rotation of the crystal corresponds to a rotation of the reciprocal lattice through the Ewald sphere (fixed in space by incident beam). One after another the points of the reciprocal lattice pass through the surface of the sphere and give rise to scattering peaks.

**Fig. 6.25** A plane oblique lattice in real space (left) and its corresponding reciprocal lattice (right). The reciprocal vectors  $\mathbf{g}_1$  and  $\mathbf{g}_2$  lie perpendicular to  $\mathbf{b}$  and  $\mathbf{a}$ , respectively





**Fig. 6.26** **a, b** Ewald construction for determining diffraction peaks (reflexes) in X-ray diffraction and elastic scattering of electrons or neutrons on a crystal lattice. **a** Ewald sphere in the reciprocal lattice of a 3D crystal illustrating the scattering condition  $\mathbf{k}_f - \mathbf{k}_i = \mathbf{K} = \mathbf{G} = \mathbf{G}_{hkl}$ . Diffraction beams are produced whenever a reciprocal lattice point coincides with the surface of the sphere. **b** Ewald construction (sphere) for scattering on a 2D surface lattice. The corresponding 2D reciprocal lattice with its lattice points  $(hk)$  is plotted along  $k_x$ . The scattering condition (6.239) is illustrated for the reciprocal lattice point  $(hk) = (2\ 0)$ . Diffracted beams are observed whenever the Ewald sphere is cut by a reciprocal lattice rod (missing 3rd Laue condition)

Another technique to enhance the number of observable scattering peaks is based on the use of a rather broad spectrum of (photon) energies of the incident particles ( $\lambda_{\min} < \lambda < \lambda_{\max}$ ). In this case of the so-called *Bragg method* the Ewald sphere can be imagined as a sphere with varying radius  $k_i$  between  $k_{\max} (\propto \lambda_{\max}^{-1})$  and  $k_{\min} (\propto \lambda_{\min}^{-1})$ . One, then, observes all scattering peaks whose reciprocal lattice points lie between the Ewald spheres corresponding to the minimum and maximum  $k_i$  values of the incident beam.

From (6.228), it is obvious that the intensities of the scattering peaks  $(hkl)$  are modulated with the absolute square  $|f_A|^2$  of the atomic scattering factor, that is, according to (6.229) their intensity decreases with increasing scattering angle  $\vartheta$ . This decrease is the faster the more extended the atomic scattering potential  $v(\mathbf{r}')$  is in space.

For the sake of completeness, it should be mentioned that in case of a non-primitive lattice with more than one atom in the unit cell the intensity of the scattering peaks is furthermore modulated with the so-called *structure factor*. This factor describes the interference effect of radiation scattered from the different atoms at different positions in the unit cell. For further details, one should refer to text books for solid state physics [9].

We now consider scattering from a crystalline surface. Corresponding experiments are performed with particles or radiation which do not penetrate into the crystal or touch the crystal only within a tiny depth below the surface. In Sect. 2.3, we have discussed the scattering of low energy electrons (LEED) and of He atoms on metal

surfaces (Fig. 2.6) as examples. In the limiting case, we assume scattering of the incident particles only on the topmost atomic layer of the crystal. Deeper lying atoms are no more reached by the incident primary particles due to strong interaction with the atomic scattering potential (strong absorption). Scattering, then, occurs on the 2-dimensional (2D) periodic lattice of the surface atoms with the basis vectors  $\mathbf{a}$  and  $\mathbf{b}$ . In this case in (6.228) the sum runs over  $m$  and  $n$  up to very high numbers  $M$  and  $N$ . The sum over  $p$  contains only one single element corresponding to the uppermost atomic layer. Consequently, only the first two Laue equations (6.233) in  $h$  and  $k$  survive.

The 2D lattice of the surface atoms in real space has in correspondence a 2D reciprocal lattice with the points  $(h, k)$  as Miller indices (Fig. 6.26b). The 3rd dimension normal to the  $(h, k)$  plane is not restricted by any condition causing discrete lattice points. We, therefore, attribute so-called lattice rods (Fig. 6.26b) to this direction. This accounts for the free choice of the perpendicular wave vector transfer  $\mathbf{K}_\perp$  upon scattering (missing 3rd Laue condition). With  $\mathbf{K} = (\mathbf{K}_{\parallel}, \mathbf{K}_\perp)$  as the total wave vector transfer the component  $\mathbf{K}_{\parallel}$  parallel to the scattering surface must obey the condition

$$\mathbf{K}_{\parallel} = \mathbf{k}_{f\parallel} - \mathbf{k}_{i\parallel} = \mathbf{G}_{hk} \quad (6.239)$$

for the occurrence of a scattering peak [analogy to (6.238)]. The 2D Ewald construction (Fig. 6.26b) analogous to the 3D case (Fig. 6.26a) shows that the the Ewald sphere cuts all reciprocal lattice rods within reach of the sphere. The corresponding  $(h, k)$  lattice points (Miller indices) give rise to scattering peaks (diffraction spots). Surface scattering on 2D structures does not require further experimental means to visualize enough diffraction spots.

It is worth mentioning that a penetration of the primary beam into the crystal over some few atomic layers restores the 3rd Laue condition (Miller index  $l$ ) to some extent. The rods of the 2D reciprocal lattice can, then, be imagined as modulated in thickness (weighted) along their length, with maxima at those points in  $k_z$  direction where a reciprocal lattice point would be found in case of an infinitely extended lattice. If one varies the energy, that is, also the wave number  $k_i$  of the incident particles, the radius of the Ewald sphere is changed and the sphere runs through “thicker” and “thinner” parts of the lattice rods. The intensity of the scattering peak  $(h, k)$  runs through maxima and minima as a function of the energy of the incident particles. From such an intensity-particle energy plot information about the penetration depth of the primary beam can be extracted.

We now reduce the dimensions of the scattering target even more. We consider scattering of a particle beam (wave) on crystallites with dimensions in the 10 to 100 nm range. Scattering on a nano-crystal differes from scattering on a macroscopic crystal by the size of the numbers  $M, N, P$  in (6.228) respectively (6.232). In contrast to the very large numbers for a macroscopic target  $M, N, P$  assume values between 10 and 50. Consequently there are only a few zeros of the scattering intensity between the main maxima, i.e. the scattering (Bragg) peaks (Fig. 6.24). As a function of the scattering angle, that is, also of  $\mathbf{K} \cdot \mathbf{a}$ , the peaks are relatively broad. We can estimate the angle width of a scattering peak (Bragg reflex) by the distance of the first intensity

zero from the position of the main intensity maximum (Fig. 6.24). From (6.232), we obtain

$$\begin{aligned}\frac{1}{2}(M+1)\mathbf{K} \cdot \mathbf{a} &= \pi, \\ (M+1)\mathbf{K} \cdot \mathbf{a} &= 2\pi.\end{aligned}\quad (6.240)$$

$(M+1)a$  is the diameter of the crystal in  $x$  direction. According to (6.209a), (6.209b) the width  $\Delta K$  of the scattering (Bragg) peak is directly related to the scattering angle width  $\Delta\vartheta$ . This experimentally determined quantity, thus, allows an estimation of the diameter  $d$  of the crystallites in the target:

$$\Delta Kd \approx 2\pi. \quad (6.241)$$

For crystallites not being little spheres but rather little cuboids, the angle width of the scattering peaks has to be measured for all three space directions in order to estimate the dimensions in  $x$ ,  $y$  and  $z$  direction. For randomly oriented crystallites in a target, only their average dimension can be obtained.

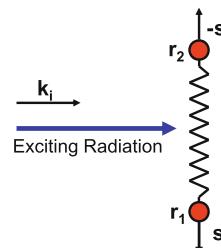
### 6.6.5 Inelastic Scattering on a Molecule

As an example for inelastic scattering the collision of particles, for example, electrons with a two-center molecule is treated.

The molecule contains two identical atoms which are covalently bonded to each other. The chemical bond is described by a potential which depends in good approximation quadratically on the distance between the atomic nuclei. Within this potential the two nuclei, respectively, atoms vibrate against each other; the energy of this vibration can be exchanged with other particles. Scattering of an electron on the molecule, thus contains, beside the elastic part, an inelastic contribution with energy exchange between target and scattering particle.

With  $\mathbf{r}_1$  and  $\mathbf{r}_2$  as the time-independent positions of the two nuclei, the time-dependent vibration coordinates are described as

**Fig. 6.27** Inelastic scattering of a particle beam with incident wave vector  $\mathbf{k}_i$  on a two-center molecule with identical atoms at  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . A molecular vibration is described by the atomic dislocations  $\pm\mathbf{s}$



$$\rho_1 = \mathbf{r}_1 + s e^{-i\omega t}, \quad (6.242a)$$

$$\rho_2 = \mathbf{r}_2 - s e^{-i\omega t}. \quad (6.242b)$$

$\omega$  is the eigenfrequency of the vibration of the nuclei which is determined by the restoring force of the harmonic binding potential (spring-like). The elongations of the two nuclei have opposite sign, but their absolute amount  $|s|$  is equal because of conservation of the center of gravity (equal atomic masses) (Fig. 6.27).

In the following, the inelastic scattering of electrons on the molecule is again described in Born approximation, that is, with neglection of multiple scattering. According to Sect. 6.6.2, the scattering amplitude  $f(\vartheta, \varphi)$  (6.204) can then also be expressed as a transition amplitude  $a_{fi}(t)$  [in Sect. 6.4 denoted as  $a_f(t)$ ] from an initial state  $|i\rangle$ , that of the incident primary electron, into the scattered final state  $|f\rangle$ . This is in the sense of time-dependent perturbation theory (6.104), to which Born approximation can be reduced (Sect. 6.6.2).

When the incident primary electron interacts with the scattering molecule during a time interval  $\tau$  the transition amplitude is expressed according to (6.104) and (6.204) as

$$a_{fi}(\vartheta, \varphi, \tau) = -\frac{i}{\hbar} \int_0^\tau dt \int d^3 r' \langle f | V(\mathbf{r}', t) | i \rangle e^{i(E_f - E_i)t/\hbar}. \quad (6.243)$$

The volume integral encloses the scattering target, that is, the molecule. Scattering on the underlying surface is excluded from the calculation. The scattering potential  $V(\mathbf{r}', t)$  exhibits the time dependence of the harmonic oscillation of the two atomic nuclei:

$$V(\mathbf{r}', t) = v[\mathbf{r}' - \rho_1(t)] + v[\mathbf{r}' - \rho_2(t)]. \quad (6.244)$$

With  $\rho_n = \rho_1, \rho_2$  we express the total scattering amplitude (6.243) as

$$\begin{aligned} a_{fi} &= -\frac{i}{\hbar} \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} \int d^3 r' e^{-i\mathbf{k}_f \cdot \mathbf{r}'} V(\mathbf{r}', t)' e^{i\mathbf{k}_i \cdot \mathbf{r}'} \\ &= -\frac{i}{\hbar} \sum_{n=1}^2 \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} \int d^3 r' e^{-i(\mathbf{k}_f - \mathbf{k}_i) \cdot \mathbf{r}'} v[\mathbf{r}' - \rho_n(t)]. \end{aligned} \quad (6.245)$$

We substitute  $\mathbf{r}' - \rho_n(t) = \xi$  and  $d^3 r = d^3 \xi$  and obtain

$$a_{fi} = -\frac{i}{\hbar} \sum_{n=1}^2 \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} e^{-i(\mathbf{k}_f - \mathbf{k}_i) \cdot \rho_n(t)} \int d^3 \xi v(\xi) e^{-i(\mathbf{k}_f - \mathbf{k}_i) \cdot \xi}. \quad (6.246)$$

The last integral running over  $\xi$  is identical for both atoms of the molecule. It describes scattering on a single atom with the scattering potential  $v(\xi)$ . With  $\mathbf{K} = \mathbf{k}_f - \mathbf{k}_i$  as scattering vector (wave vector transfer), it represents the atomic scattering factor  $f_A$  as it was introduced in (6.227) in connection with elastic scattering on a crystal lattice.

Equation (6.246) then yields

$$a_{fi} = -\frac{i}{\hbar} \sum_{n=1}^2 \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} e^{-i\mathbf{K} \cdot \rho_n(t)} f_A(\mathbf{K}). \quad (6.247)$$

By means of the momentary nuclear coordinates (6.242a), (6.242b), the exponential functions with  $\rho_n$  in the exponent can be represented as

$$e^{-i\mathbf{K} \cdot \rho_1} = e^{-i\mathbf{K} \cdot (\mathbf{r}_1 + \mathbf{s} \exp\{-i\omega t\})} \approx e^{-i\mathbf{K} \cdot \mathbf{r}_1} (1 - i\mathbf{K} \cdot \mathbf{s} e^{-i\omega t}), \quad (6.248a)$$

$$e^{-i\mathbf{K} \cdot \rho_2} \approx e^{-i\mathbf{K} \cdot \mathbf{r}_2} (1 + i\mathbf{K} \cdot \mathbf{s} e^{-i\omega t}). \quad (6.248b)$$

In the approximate expression, the oscillation amplitude  $\mathbf{s}$  of the nuclei is assumed to be small in comparison with the nuclear coordinates  $\mathbf{r}_1$  and  $\mathbf{r}_2$  ( $\ll 10\%$ ).

The transition amplitude is, thus, obtained as

$$\begin{aligned} a_{fi} \approx & -\frac{i}{\hbar} f_A(\mathbf{K}) \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} \\ & \times [(e^{-i\mathbf{K} \cdot \mathbf{r}_1} + e^{-i\mathbf{K} \cdot \mathbf{r}_2}) + i\mathbf{K} \cdot \mathbf{s} e^{-i\omega t} (e^{-i\mathbf{K} \cdot \mathbf{r}_2} - e^{-i\mathbf{K} \cdot \mathbf{r}_1})]. \end{aligned} \quad (6.249)$$

For the calculation of the scattering rate  $R_{fi}$  (Sect. 6.4.1), which is observed in the experiment, we have to take into account that the transition probability  $w_{fi}$  is proportional to the interaction time  $\tau$ . The rate  $R_{fi}$ , therefore, requires a division of  $w_{fi}$  by  $\tau$ :

$$R_{fi} = w_{fi}/\tau = |a_{fi}|^2/\tau. \quad (6.250)$$

Furthermore, squaring the transition amplitude  $|a_{fi}|^2$  causes an interference term due to the two summands in the squared bracket in (6.249). If one measures, however, the elastically and inelastically scattered particles separately (as is usually done), this interference term is not of interest. We, therefore, discuss the elastic part of the amplitude (6.249)

$$a_{fi}^{\text{el}} \approx -\frac{i}{\hbar} f_A(\mathbf{K}) (e^{-i\mathbf{K} \cdot \mathbf{r}_1} + e^{-i\mathbf{K} \cdot \mathbf{r}_2}) \int_0^\tau dt e^{i(E_f - E_i)t/\hbar} \quad (6.251a)$$

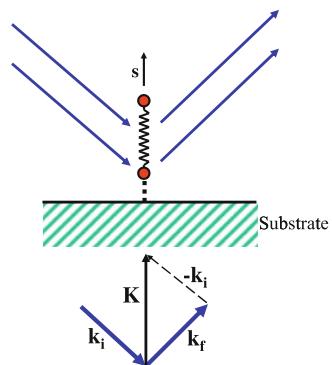
and the inelastic part

$$a_{fi}^{\text{inel}} \approx -\frac{i}{\hbar} f_A(\mathbf{K}) (\mathbf{K} \cdot \mathbf{s}) (e^{-i\mathbf{K} \cdot \mathbf{r}_1} + e^{-i\mathbf{K} \cdot \mathbf{r}_2}) \int_0^\tau dt e^{i[(E_f - E_i)/\hbar - \omega]t} \quad (6.251b)$$

separately.

We use the relations ((6.106a), (6.106b)) to ((6.111a), (6.111b)) of Sect. 6.4.1 and derive the elastic scattering rate from (6.250) and (6.251a) as

**Fig. 6.28** Inelastic scattering of a particle beam on the two-center molecule of Fig. 6.27, but now adsorbed on a surface. During scattering the particle beam is reflected on the surface (*above*). For the considered scattering geometry the wave vector transfer  $\mathbf{k}_f - \mathbf{k}_i = \mathbf{K}$  has a component in the direction of the molecular vibration



$$\begin{aligned} R_{fi}^{\text{el}} &= \frac{2\pi}{\hbar^2} |f_A|^2 |(e^{-i\mathbf{K}\cdot\mathbf{r}_1} + e^{-i\mathbf{K}\cdot\mathbf{r}_2})|^2 \delta(E_f - E_i) \\ &= \frac{2\pi}{\hbar^2} |f_A|^2 2[1 + \cos \mathbf{K} \cdot (\mathbf{r}_2 - \mathbf{r}_1)] \delta(E_f - E_i). \end{aligned} \quad (6.252)$$

The  $\delta$  function expresses elastic scattering, that is, the identity of final and initial particle energy. The squared bracket describes the intensity modulation of scattering (diffraction) on a double slit. The two atoms represent two neighbouring scattering centers for an incident plane wave of particles as does a double slit in the analogous optical experiment. Compared with the previous representation of the double slit interference (3.5), the wave vector difference  $\mathbf{K} = \mathbf{k}_f - \mathbf{k}_i$  appears in (6.252) rather than  $\mathbf{k}$  in (3.5). In (3.5), the initial wave vector  $\mathbf{k}_i$  was not considered. Both interfering waves started at the double slit.

We now consider the inelastic scattering amplitude  $a_{fi}^{\text{inel}}$  (6.251b) and calculate the scattering rate:

$$R_{fi}^{\text{inel}} = \frac{2\pi}{\hbar^2} |f_A|^2 (\mathbf{K} \cdot \mathbf{s})^2 [1 - \cos \mathbf{K} \cdot (\mathbf{r}_2 - \mathbf{r}_1)] \delta(E_f - E_i \pm \hbar\omega). \quad (6.253)$$

The  $\delta$  function describes the energy conservation during the inelastic collision, that is, the energy exchange between molecule and scattered particle. The energy transfer in both directions (+ for energy emission from molecule, - for energy absorption by molecule) is identical with the oscillation quantum  $\hbar\omega$  of the molecular vibration. The scattered intensity is again modulated as a function of  $\mathbf{K}$  (also scattering direction), as in the case of inelastic scattering from the two-center system (double slit interference). Atomistic details of the particle interaction are contained in the atomic scattering factor  $f_A$ . The factor  $(\mathbf{K} \cdot \mathbf{s})^2$  is an important ingredient; inelastic scattering only occurs, if the scattering vector (wave vector transfer)  $\mathbf{K} = \mathbf{k}_f - \mathbf{k}_i$  has a vector component in the direction  $\mathbf{s}$  of the molecular vibration. This situation is ideally given for a molecule adsorbed perpendicular on a solid surface and slow electrons with a primary energy between 10 and 100 eV are incident under an oblique angle to the substrate surface (Fig. 6.28). The observation of electrons scattered into

a Bragg diffraction peak under the reflection angle guarantees a wave vector transfer component in  $\mathbf{s}$  direction.

## References

1. H.A. Kramers, Z. Physik **39**, 828 (1926)
2. S. Fölsch, J. Martínez-Blanco, J. Yang, K. Kanisawa, S.C. Erwin, Nat. Nanotechnol. **9**, 505 (2014)
3. J. Yang, C. Nacci, J. Martínez-Blanco, K. Kanisawa, S. Fölsch, J. Phys.: Condens. Matter **24**, 354008 (2012)
4. H. Lüth, *Solid Surfaces, Interfaces and Thin Films*, 5<sup>th</sup> edition (Springer, Berlin, Heidelberg, New York, 2010), p. 253
5. J. Stark: Verhandlungen der Deutschen Physikalischen Gesellschaft, XVI(7), Braunschweig (1914)
6. G. Wentzel, Z. Physik **43**, 524 (1927)
7. E. Fermi, Rev. Mod. Phys. **4**, 87 (1932)
8. E.R. Andrew, in NMR in Physiology and Biomedicine, ed. by R.J. Gillies (Academic Press, San Diego, 1994), pp. 1–55
9. H. Ibach, H. Lüth, *Solid State Phys.*, 4th edn. (Springer, Berlin, 2009)
10. L.C. West, S.J. Egash, Appl. Phys. Lett. **46**, 1156 (1985)
11. A.F.J. Levi, *Appl. Quantum Mech.* (Cambridge University Press, Cambridge, 2003)
12. A.-M. Oros-Peusquens, *Institut für Medizin* (Forschungszentrum Jülich; private Mitteilung, MR-Gruppe, 2008)

# Chapter 7

## Superposition, Entanglement and Other Oddities

Quantum mechanical behavior of atomic and subatomic particles is different from our familiar perception of the macroscopic world around us. The wave-particle duality as it appears in the double slit experiment with electrons (Sect. 2.4) is the deep underlying reason for the strange, counterintuitive phenomena in the quantum world. In this chapter, we will encounter some more oddities of quantum physics. Mathematically these counterintuitive phenomena are directly derived from the fact that the fundamental equation, the Schrödinger equation, only yields a probability (wave) amplitude rather than a probability for the observation of a certain experimental result. Squaring the complex probability amplitude only leads to observable quantities. As in the double slit experiment, wave amplitudes are summed up and the sum is subsequently squared to yield observable probabilities or average measurement results. This is different from classical statistics, where the probabilities for different events are added for the calculation of the total probability.

The quantum mechanical calculation of the total probability generates additional interference terms which make quantum physics different from classical physics. Note the interference pattern for the detection of electrons having passed the double slit arrangement. An important question in this context concerns the disappearance of the interference pattern by observation of the exact path (through which of the two slits?) of the electron (Sect. 2.4.2). “Which way” information destroys the quantum mechanical interference. It restores classical behavior of the particles. We approach the important question how classical behavior of macroscopic bodies is compatible with the strange phenomena of quantum particles which constitute the macroscopic world. Why are quantum interferences not observed in everyday life?

It might be noted in this context that these kind of questions are of enormous importance for the modern field of quantum information (Sect. 7.5).

## 7.1 Superposition of Quantum States

Superposition of states is an important characteristic property of quantum mechanics. The Schrödinger equation is a linear differential equation. If two solutions  $\psi_1$  and  $\psi_2$  to the equation exist, the linear superposition  $\psi = c_1\psi_1 + c_2\psi_2$  also represents a solution. The same, of course, is true for the corresponding Hilbert vectors  $|\psi_1\rangle$ ,  $|\psi_2\rangle$  and  $|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle$ .

This superposition principle in general holds for all kinds of eigensolutions: With  $|n\rangle$  being the eigenstates of the time-independent Schrödinger equation  $\hat{H}|n\rangle = E_n|n\rangle$  the most general solution is the linear superposition of all eigenstates:

$$|\psi\rangle = \sum_n c_n |n\rangle. \quad (7.1)$$

An energy measurement (operator  $\hat{H}$ ) on the quantum state  $|\psi\rangle$  forces the system into one of the eigenstates  $|n\rangle$ . The probability to prepare (to find) the state  $|n\rangle$  by the measurement is given by  $|c_n|^2$ . The energy measurement destroys the superposition state (7.1) and filters out one single energy eigenstate (collapse of the wave function). Superposition states exist as long as no other measurements (incommensurable) are performed on them. They continue to exist as long as they are not exposed to an interaction with a measurement probe or, more generally, with any kind of external environment.

We encountered this property already in the discussion of the double slit experiment with particles (Sect. 2.4.2). The electron can pass through slit (1) with an amplitude  $\psi_1$  or through slit (2) with the amplitude  $\psi_2$ . As long as we do not perform any additional measurement to find out through which slit the electron has propagated (which way information), the superposition state (wave function)  $\psi_1 + \psi_2$  persists and the detection probability (intensity) on the observation screen is given by  $|\psi_1 + \psi_2|^2$ , that is, the interference pattern representing the particle-wave duality. The interference pattern after the detection of a huge number of particles indicates the existence of the superposition wave function (state)  $\psi_1 + \psi_2$  before the detection event on the screen.

The essential issue in the double slit experiment is the following: There are two possibilities for an electron to reach the detection screen (via slit 1 or via slit 2) between which the observer does not distinguish by an experiment. In that case the correct description of the experiment, that is, the observed result, is obtained by superposition of the amplitudes  $\psi_1$  and  $\psi_2$  (rather than the probabilities  $|\psi_1|^2$  and  $|\psi_2|^2$ ) and subsequent calculation of the detection probability by the absolute square  $|\psi_1 + \psi_2|^2$ .

This *superposition principle* of quantum mechanics is formulated in the most general way as follows: If a physical event, for example, the detection of a particle on a screen, can be realized in several different ways, between which it is not distinguished by experiment, the different probability amplitudes (wave functions) must be superimposed (added) for the calculation of the total probability. This probability

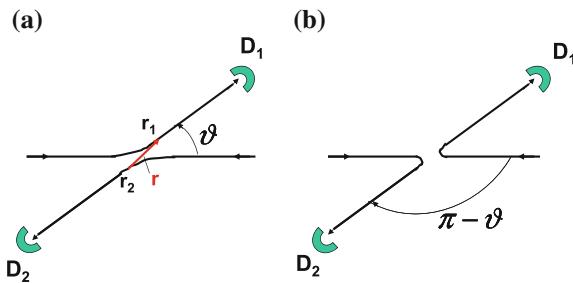
is obtained as the squared absolute amount of the superposition wave function. If an additional experiment, however, distinguishes between the different realization alternatives, the probabilities (squared amplitudes) for the different realizations must be added rather than the amplitudes. Then, the typical quantum mechanical behavior, the occurrence of interferences, disappears and the phenomena obey again classical laws.

Before we will elucidate the deeper reason for this behavior, some more examples for the application of the superposition principle:

- The hydrogen ion  $H_2^+$ : We discussed this topic already in detail within the framework of approximation techniques in quantum mechanics (Sect. 6.2.3). In order to find an approximate ansatz for a trial wave function in the variational method (Sect. 6.2) the two alternatives, electron close to proton  $A$  and electron close to proton  $B$ , are superimposed. We can justify this ansatz for the wave function even well by means of the superposition principle: Both realizations of the  $H_2^+$  ion, electron close to  $A$  and electron close to  $B$ , are equivalent and indistinguishable by an additional experiment. Therefore, the corresponding probability amplitudes for the realization of the two limiting cases,  $\psi_1$  and  $\psi_2$ , must be superimposed to obtain the most general solution to the problem. The superposition of the two wave amplitudes causes splitting of an atomic quantum state (energy level) into a bonding and antibonding state in the molecule. This is the basis of the covalent bond which links the two protons in the molecule.
- In elementary particle physics (Sect. 5.6.4) quarks, which constitute hadrons (mesons and baryons), possess the color quality. Color appears in three different types which are described by the quantum “numbers” Red ( $R$ ), Green ( $G$ ) and Blue ( $B$ ). Without an additional experiment for distinguishing between these colors, the most general meson state  $|q\bar{q}\rangle$  being composed of a quark  $q$  and the antiquark  $\bar{q}$  is a linear (additive) superposition of the three color states  $|R\bar{R}\rangle$ ,  $|G\bar{G}\rangle$  and  $|B\bar{B}\rangle$ . The linear superposition (5.177) guarantees a symmetrical wave function of the boson, that is, its bosonic character.

The analogue situation is given with baryons which consist of three quarks as e.g. the proton  $|p\rangle = |uud\rangle$  or the neutron  $|n\rangle = |udd\rangle$  (Table 5.2). For the baryon  $|uuu\rangle$ , which consists of three identical “up” quarks, its fermionic character expressed by an antisymmetric wave function (upon exchange of two quarks) is only warranted by the existence of quarks with three different colors. The most general state  $|uuu\rangle$  is a linear superposition of triple quark states with  $u$  quarks carrying different color (5.178). The choice of the signs in (5.178) guarantees the fermionic character of  $|uuu\rangle$ . The color superposition state (5.178), of course, persists only as long as no additional experiment for distinguishing between the different colors is performed.

- Scattering of two identical particles on each other is a further interesting example to demonstrate the importance of superposition states. Figure 7.1 schematically shows the scattering process of two electrons on each other. Due to Coulomb repulsion the electrons avoid hitting each other. When they approach each other, they might be deflected from their path by an angle  $\vartheta$  (Fig. 7.1a) or they might be



**Fig. 7.1** Schematic representation of a scattering process of two identical particles on each other, e.g., of two electrons being scattered by Coulomb repulsion. Since the particles are undistinguishable, the detectors  $D_1$  and  $D_2$  can not distinguish between the two processes (a) and (b), forward and backward scattering, respectively. The plotted trajectories can only classically be understood; they are irrelevant in quantum mechanics. Processes a and b participating equally to the scattering process must be superimposed in the quantum mechanical description.  $\vartheta$ , respectively  $\pi - \vartheta$  are the scattering angles;  $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$  is the momentary particle distance

backscattered with a scattering angle  $\pi - \vartheta$  (Fig. 7.1b). Since electrons in identical quantum states are indistinguishable, detection of one electron in detector  $D_1$  and detection of the other one in detector  $D_2$  does not allow a distinction between these two cases (a) and (b). In quantum mechanics, the electron trajectories in Fig. 7.1a, b do not have a real meaning. The position of the particles as a function of time can not be given precisely (uncertainty principle), only the detection events in  $D_1$  and  $D_2$  are well defined.

The most general scattering state of the two particles is, therefore, a superposition state of the two scattering events (a) and (b).

The two electrons are indistinguishable as long as we do not perform an additional experiment to distinguish between forward and backscattering in Fig. 7.1. As will be shown, such an experiment is possible by measuring additionally the spin states of the electrons.

In the next section, we will discuss the consequences of spin preparation and detection in the two electron scattering process.

### 7.1.1 Scattering of Two Identical Particles: A Special Superposition State

Scattering of two particles on each other is described by means of an interaction potential  $V(\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2)$  which depends on the relative coordinate  $\mathbf{r}$ , the distance  $\mathbf{r}_1 - \mathbf{r}_2$  between the two particles (Fig. 7.1). In the present case of two electrons,  $V(\mathbf{r})$  is the Coulomb repulsion potential.

In the formal treatment of the collision process, the initial state of the two electrons, long before the scattering event, is described by the two-particle wave function

$$\psi_{\text{in}} \propto e^{ik_1 \cdot r_1} e^{ik_2 \cdot r_2}. \quad (7.2a)$$

This two-particle wave function can be rewritten as

$$\begin{aligned} \psi_{\text{in}} &\propto \exp\left[i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \frac{\mathbf{r}_1 + \mathbf{r}_2}{2}\right] \exp\left[i\frac{\mathbf{k}_1 - \mathbf{k}_2}{2} \cdot (\mathbf{r}_1 - \mathbf{r}_2)\right] \\ &= \psi_{\text{in}}^{\text{CM}}(\mathbf{r}_{\text{CM}}) \psi_{\text{in}}^{\text{rel}}(\mathbf{r}). \end{aligned} \quad (7.2b)$$

Hereby,  $\mathbf{r}_{\text{CM}} = (\mathbf{r}_1 - \mathbf{r}_2)/2$  is the center-of-mass coordinate of the two particles. Thus,  $\psi_{\text{in}}^{\text{CM}}$  describes the motion of the center of mass, while  $\psi_{\text{in}}^{\text{rel}}(\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2)$  represents the wave function of the relative motion of the two particles against each other. The center-of-mass (CM) coordinate system is defined by  $\mathbf{k}_1 - \mathbf{k}_2 = \mathbf{0}$ . Consequently in the CM system we have  $\psi_{\text{in}}^{\text{CM}} = 1$  and the incident wave function gets a simple representation containing only the relative motion of the particles:

$$\psi_{\text{in}} \propto e^{ik \cdot (\mathbf{r}_1 - \mathbf{r}_2)} = e^{ik \cdot \mathbf{r}}. \quad (7.3)$$

Expressed in the relative coordinate this two-electron wave function in the CM coordinate system formally equals a simple incident plane wave as in (6.181) or (6.190). Since the scattering potential  $V(\mathbf{r})$  depends only on the relative coordinate  $\mathbf{r}$ , we can formally transfer all relations from Sects. 6.6.1 and 6.6.2. The following two-particle scattering wave function with  $\psi_{\text{sc}}$  as scattering solution is, thus, obtained in the center-of-mass coordinate system:

$$\psi(\mathbf{r}_1 - \mathbf{r}_2) \propto [e^{ikz} + \psi_{\text{sc}}(\mathbf{r}_1 - \mathbf{r}_2)] = e^{ikz} + f(\vartheta, \varphi) \frac{e^{ikr}}{r}. \quad (7.4)$$

As in Sect. 6.6.1,  $f(\vartheta, \varphi)$  describes the scattering amplitude. In (7.4), the  $z$ -axis is identical with the trajectories of the incident particles and  $\vartheta$  as scattering angle is defined as in Fig. 7.1. The scattering angle  $\varphi$  runs around the direction of incidence ( $z$  axis).

The two scattering geometries in Fig. 7.1 are distinguished by the different scattering amplitudes  $f(\vartheta, \varphi)$  for forward scattering and  $f(\pi - \vartheta, \varphi)$  for backscattering, respectively after neglecting  $\varphi$  by  $f(\vartheta)$  and  $f(\pi - \vartheta)$ . In case of electron–electron scattering by Coulomb repulsion  $f(\vartheta)$ , of course, assumes the value (6.213), where  $r$  is the distance between the two electrons.

As discussed before, we can not distinguish between the two scattering geometries (a) and (b) in Fig. 7.1. The detectors  $D_1$  and  $D_2$  do not distinguish between particles originating from forward or backscattering events. Both scenarios are indistinguishable alternatives of the two-particle scattering process. The superposition principle, thus, requires the most general scattering solution to be a superposition of both alternatives (a) and (b) in Fig. 7.1. In the general solution the scattering amplitudes,  $f(\vartheta)$  and  $f(\pi - \vartheta)$  must be linearly superimposed.

At this point the bosonic or fermionic character of the two scattering particles decisively enters the problem.

Exchange of the two particles in Fig. 7.1 transforms  $\mathbf{r}$  into  $-\mathbf{r}$  and  $\vartheta$  into  $\pi - \vartheta$ . Scattering of two bosons on each other, for example, of two He atoms, requires a symmetric wave function upon particle exchange, that is, also a symmetric scattering amplitude:

$$f_{\text{bos}} = f(\vartheta) + f(\pi - \vartheta). \quad (7.5a)$$

In case of fermion scattering, for example, electron-electron scattering, the wave function is antisymmetric and consequently also the total scattering amplitude:

$$f_{\text{fermi}} = f(\vartheta) - f(\pi - \vartheta). \quad (7.5b)$$

It is immediately evident that the differential scattering cross section  $d\sigma/d\Omega = |f_{\text{bos}}|^2 \neq |f_{\text{fermi}}|^2$  for boson and fermion scattering differs by the interference term. Because of the plus sign in front of the interference term in the absolute square in (7.5a) the bosonic scattering cross section significantly exceeds that for fermions with a minus sign in (7.5b).

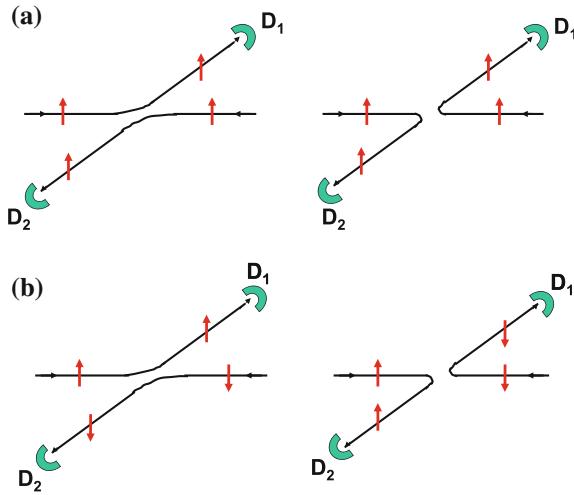
Even more interesting is the two-particle scattering experiment with two identical fermions, in which the spin of the incident particles is prepared in a fixed direction and then the spin orientation is measured after scattering. For sufficiently small spin orbit interaction, the spin orientation is conserved during the scattering process. The detectors  $D_1$  and  $D_2$  are sensitive with respect to spin orientation; they can distinguish between the spin-up  $|\uparrow\rangle$  and the spin-down state  $|\downarrow\rangle$  of the arriving fermion (electron).

For the mathematical description of the experiment, the two-particle wave function (position representation of state)  $\psi_{\text{sc}} = \langle \mathbf{r} | 12 \rangle$  (7.4) must be complemented by the spin part  $|s_1 s_2\rangle$  of the two-particle state. According to Fig. 7.2, spin sensitive detection of the scattered particles in  $D_1$  and  $D_2$ , now, allows a distinction between forward and backscattering, if the spin of the two incident particles was initially prepared with opposite orientation. In analogy with the double slit experiment, we expect different measurement results for scattering with and without distinction between back and forward scattering (which way information).

We, thus, consider the following fermionic two-particle solution [only the scattering part  $\psi_{\text{sc}}$  in (7.4)]:

$$\begin{aligned} \psi_{\text{sc}}(r)|s_1 s_2\rangle &\propto f(\vartheta) \frac{e^{ikr}}{r} |s_1 s_2\rangle - f(\pi - \vartheta) \frac{e^{ikr}}{r} |s_2 s_1\rangle \\ &= f(\vartheta) \psi_{12}(r) |s_1 s_2\rangle - f(\pi - \vartheta) \psi_{12}(r) |s_2 s_1\rangle. \end{aligned} \quad (7.6)$$

We have used the invariance of  $\psi_{12}(r) = e^{ikr}/r$  upon exchange of the two particles. Furthermore, the two-particle spin state  $|s_1 s_2\rangle$  can vary between the states  $|\uparrow\downarrow\rangle$ ,  $|\downarrow\uparrow\rangle$ ,  $|\uparrow\uparrow\rangle$ ,  $|\downarrow\downarrow\rangle$ . Different states are orthogonal with respect to each other and identical states yield a unity modulus, for example,  $\langle \uparrow\uparrow | \uparrow\uparrow \rangle = 1$ , respectively  $\langle \uparrow\downarrow | \downarrow\uparrow \rangle = 0$ .



**Fig. 7.2** **a, b** Schematic representation of the two-particle scattering process of Fig. 7.1, but now with initial preparation of particle spin states (red arrows) and spin resolving detection in the detectors  $D_1$  and  $D_2$ . **a** For equal spin orientation before the scattering process the detectors  $D_1$  and  $D_2$  detect equal spin direction both in forward and backward scattering geometry. Both types of scattering are undistinguishable. **b** For opposite spin direction before scattering the detectors  $D_1$  and  $D_2$  can distinguish between forward and backward scattering

The spatial part of the scattering solution  $\psi_{\text{sc}}(\mathbf{r}_1 - \mathbf{r}_2)$  (7.4) is described as in (7.6) by  $f(\vartheta, \varphi)e^{ikr}/r = \psi_{12}(r)f(\vartheta)$ , respectively after exchange of the particles by  $\psi_{12}(r)f(\pi - \vartheta)$ .

Firstly, we consider the case of the two particles being incident with identical spin orientation. In that situation, the detectors can not distinguish between back and forward scattering (Fig. 7.2a). Both detectors (1) and (2) measure equal spin directions  $|\uparrow\uparrow\rangle = |\uparrow\rangle_1|\uparrow\rangle_2$  and the total solution is expressed as

$$\psi_{\text{sc}}(r)|s_1s_2\rangle = f(\vartheta)\psi_{12}(r)|\uparrow\uparrow\rangle - f(\pi - \vartheta)\psi_{12}(r)|\uparrow\uparrow\rangle. \quad (7.7a)$$

The spin states, here, indicate the results of the detector measurements. For the evaluation of the scattering probability, the absolute square of the amplitude (7.7a) is calculated and because of  $|\psi_{12}|^2 = 1$  we obtain

$$\begin{aligned} |\psi_{\text{sc}}|^2 &= |f(\vartheta)|^2\langle\uparrow\uparrow|\uparrow\uparrow\rangle + |f(\pi - \vartheta)|^2\langle\uparrow\uparrow|\uparrow\uparrow\rangle \\ &\quad - [f^*(\vartheta)f(\pi - \vartheta) + f(\vartheta)f^*(\pi - \vartheta)]\langle\uparrow\uparrow|\uparrow\uparrow\rangle \\ &= |f(\vartheta)|^2 + |f(\pi - \vartheta)|^2 - [f^*(\vartheta)f(\pi - \vartheta) + f(\vartheta)f^*(\pi - \vartheta)]. \end{aligned} \quad (7.7b)$$

Apart from the probabilities  $|f(\vartheta)|^2$  and  $|f(\pi - \vartheta)|^2$  for forward and backscattering the interferences between the two processes appear.

We now consider the case where both electrons are incident with opposite spin orientation (Fig. 7.2b). The detectors, then, can distinguish between forward and backscattering. Analogously to (7.7a) the scattering solution is represented as

$$\psi_{\text{sc}}(r)|s_1s_2\rangle = f(\vartheta)\psi_{12}(r)|\uparrow\downarrow\rangle - f(\pi - \vartheta)\psi_{12}(r)|\downarrow\uparrow\rangle. \quad (7.8a)$$

Again, the spin states are those being detected in  $D_1$  and  $D_2$ . The scattering probability is obtained as

$$\begin{aligned} |\psi_{\text{sc}}|^2 &= |f(\vartheta)|^2\langle\uparrow\downarrow|\uparrow\downarrow\rangle + |f(\pi - \vartheta)|^2\langle\downarrow\uparrow|\downarrow\uparrow\rangle \\ &\quad - f^*(\vartheta)f(\pi - \vartheta)\langle\uparrow\downarrow|\downarrow\uparrow\rangle - f(\vartheta)f^*(\pi - \vartheta)\langle\downarrow\uparrow|\uparrow\downarrow\rangle \\ &= |f(\vartheta)|^2 + |f(\pi - \vartheta)|^2. \end{aligned} \quad (7.8b)$$

Due to the orthogonality of spin states with opposite spin orientation the interference terms between back and forward scattering as appearing in (7.7b) are canceled.

By measurement of the spin directions, we have obtained a “which way” information, that is, the distinction between back and forward scattering (Fig. 7.1). The additional information about the detailed particle trajectory destroys the interference terms in the scattering probability. The total scattering probability for scattering of the two particles on each other, now, is the sum of the two single probabilities  $|f(\vartheta)|^2$  and  $|f(\pi - \vartheta)|^2$  for forward and backscattering, just as it is expected in classical physics for two “either-or” events.

The additional “which way” information transforms quantum mechanical into classic behavior. The typically quantum mechanical interferences in the probability disappear by regarding the spin degree of freedom which enables the registration of detailed path information.

## 7.2 Entanglement

In Sect. 7.1.1, the additional measurement of the particle spin orientation causes the disappearance of quantum interferences in the two-particle scattering probabilities. This issue shall be analysed in the following more in detail by considering the properties of the two-particle spin states with opposite spin orientation (7.8a), (7.8b), those states which provide the distinction between back and forward scattering, i.e. the “which way” information. In our gedanken experiment of Sect. 7.1.1, we have neglected interactions between the two spins of the electrons. The two-particle spin states  $|s_1s_2\rangle$  (7.5a), (7.5b) can, therefore, also be written as a product of two single particle spin states (Sect. 5.6.1):

$$|s_1s_2\rangle = |s_1\rangle|s_2\rangle. \quad (7.9)$$

The numeration of the particles might also be given by indices at the kets, for example, as  $|s_1 s_2\rangle = |\uparrow\rangle_1 |\downarrow\rangle_2$  for electron (1) with spin up and electron (2) with spin down. The two-particle state for electrons incident with opposite spin orientation (7.8a) is, then, be expressed as

$$\psi_{sc}|s_1 s_2\rangle = \psi_{12}(r)[f(\vartheta)|\uparrow\rangle_1 |\downarrow\rangle_2 - f(\pi - \vartheta)|\downarrow\rangle_1 |\uparrow\rangle_2]. \quad (7.10)$$

Note that the spin directions of the two particles are those measured in the detectors  $D_1$  and  $D_2$  (Fig. 7.2b).

A peculiarity of the two-particle state (7.10) is the fact that it can not be written as a product of two single-particle spin states. This can easily be proven. If we assume a product representation of the two-particle spin state  $|s_1 s_2\rangle$  in terms of two single-particle states  $|\alpha\rangle$  and  $|\beta\rangle$ , the spin state in (7.10) must be written as

$$|s_1 s_2\rangle = |\alpha\rangle_1 |\beta\rangle_2. \quad (7.11)$$

We now decompose the single-particle states  $|\alpha\rangle_1$  and  $|\beta\rangle_2$  into the two basis states “up” and “down” of the spin Hilbert space:

$$|\alpha\rangle_1 = a|\uparrow\rangle_1 + b|\downarrow\rangle_1, \quad (7.12a)$$

$$|\beta\rangle_2 = a'|\uparrow\rangle_2 + b'|\downarrow\rangle_2 \quad (7.12b)$$

and calculate the product state (7.11):

$$|\alpha\rangle_1 |\beta\rangle_2 = aa'|\uparrow\rangle_1 |\uparrow\rangle_2 + bb'|\downarrow\rangle_1 |\downarrow\rangle_2 + ab'|\uparrow\rangle_1 |\downarrow\rangle_2 + a'b|\downarrow\rangle_1 |\uparrow\rangle_2. \quad (7.13)$$

If we require that the two-particle scattering state (7.10) shall be represented as a product state of two single-particle spin states as in (7.13), the following conditions have to be fulfilled:

$$aa' = 0, \quad bb' = 0, \quad (7.14a)$$

$$ab' = f(\vartheta), \quad a'b = -f(\pi - \vartheta). \quad (7.14b)$$

These equations lead to a contradiction. Using (7.14a), we conclude:

$$-f(\vartheta)f(\pi - \vartheta) = ab'a'b = 0. \quad (7.15)$$

A consistent solution of (7.14a), (7.14b) is only possible for vanishing scattering amplitudes  $f(\vartheta)$  and  $f(\pi - \vartheta)$ , respectively. Consequently, the scattering problem does not have a solution. The two-particle scattering solution (7.10) can not be represented as a product of two single-particle states.

In the two-particle scattering experiment with spin detection, the two electrons can not be described as two independent, uncoupled particles. Via their spin degree of freedom the two electrons are intimately tied to each other. This typically quantum

mechanical correlation was first considered by Schrödinger in 1935 in a famous publication [1–3]. He called the phenomenon *entanglement* or originally in German “*Verschränkung*”.

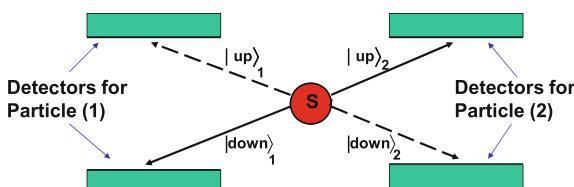
Entanglement of particles is a frequent quantum mechanical phenomenon. It appears in many experiments. Two further examples shall be considered.

A particle source ( $S$ ), for example, an unstable atom or nucleus decays under emission of two identical particles. Because of momentum conservation, the wave vectors of the two particles must obey the relation  $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{0}$ , if the source is at rest. The particles are therefore emitted in opposite directions. The particles are detected by four large area detectors spatially arranged at opposite positions (Fig. 7.3). The detectors can distinguish if particle (1), emitted to the left, appears in the upper (up) or lower (down) detector as well as if particle (2), emitted to the right, reaches the upper or lower detector. Since the emission probability of the source ( $S$ ) is isotropic, the detection event particle (1) in the upper detector  $|up\rangle_1$  and particle (2) in the lower detector  $|down\rangle_2$  appears with the same probability as the event particle (1) below  $|down\rangle_1$  and particle (2) above  $|up\rangle_2$ . Consequently, the amplitudes of both detection events are equal and the most general state is a linear superposition of both amplitudes:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|up\rangle_1|down\rangle_2 \pm |down\rangle_1|up\rangle_2). \quad (7.16)$$

Plus or minus sign indicate the fermionic (−) or bosonic (+) character of the emitted particles. A comparison of (7.16) and (7.10) shows the identical formal structure of both two-particle states. The state vector (7.16) can not be factorized into two single-particle states of the particles (1) and (2). Both particles are entangled with each other.

Another example of an entangled two-particle state is particularly interesting. It is the basis of the famous Einstein, Rosen, Podolsky (EPR) paradox (1935) [4], which is frequently used, even until today, to reveal the oddities and counter-intuitive aspects of the quantum world (Sect. 7.2.1). The gedanken experiment in its simplified version after Bohm [5] is based on an unstable source (atom or nucleus) which randomly emits two identical fermions (or bosons) in opposite directions. If the source particle (at rest) has zero spin (singlet state), the two emitted decay products must have opposite spin orientations because of spin conservation (Fig. 7.4).



**Fig. 7.3** Schematic plot of an experiment to entangle two particles which are emitted in opposite direction from a source ( $S$ ), e.g., a decaying atomic nucleus. Since the emission probability of the source ( $S$ ) is isotropic, the emission processes plotted in *solid* and in *broken line* (particle trajectories) must be superimposed

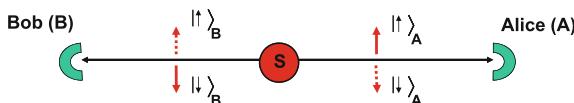
The spins of the emitted particles are anti-correlated, but the spin orientation in a single emission process is fully random, it changes statistically from one emission event to the next one (according to the indeterminacy in quantum physics).

The emitted particles are detected by Alice ( $A$ ) and Bob ( $B$ ) in two detectors positioned in opposite directions (Fig. 7.4). The names Alice (right detector) and Bob (left detector) are frequently used in books about quantum information in connection with this and similar gedanken experiments. Both detection events  $|\uparrow\rangle_A|\downarrow\rangle_B$  and  $|\downarrow\rangle_A|\uparrow\rangle_B$  belonging to opposite spin orientations appear with the same probability. Furthermore, we assume that both emitted particles are quite localized entities, that is, they are described by spatially restricted wave functions (wave packets)  $\psi(\mathbf{r}_A)$  and  $\psi(\mathbf{r}_B)$ . The two-particle state describing the simultaneous emission of two particles with anti-correlated spin orientation, then, results as a linear superposition of the two equally probable states with opposite spin directions:

$$\psi_{AB}|s_A s_B\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A|\downarrow\rangle_B - |\downarrow\rangle_A|\uparrow\rangle_B)\psi(\mathbf{r}_A)\psi(\mathbf{r}_B). \quad (7.17)$$

The minus sign is valid for fermions as decay products. Bosons require a plus sign. This two-particle state is again an entangled state, in analogy to (7.10). It can not be decomposed into two separate single-particle states, one related to the particle ( $A$ ) detected by Alice and the other one ( $B$ ) detected by Bob. Even if we consider the state (7.17) after a sufficiently long time, when the two spatially restricted wave packets have removed from each other far away (no overlap) and an interaction between the particles is not possible anymore, the anti-correlation between the spins persists. If Alice, for example, measures a spin value  $+\hbar/2$  in  $z$  direction (arbitrarily fixed) on her particle  $\psi(\mathbf{r}_A)$ , then Bob definitely finds a spin  $-\hbar/2$  in  $z$  direction on his particle  $\psi(\mathbf{r}_B)$ . If Alice incidentally measures a positive spin in  $x$  direction, Bob detects a negative spin in the same direction.

According to the spin (angular momentum) commutation rules (Sect. 5.2) the spin in  $x$  and  $z$  direction can not simultaneously be measured. Since Alice can perform measurements for both spin orientations without affecting Bob's particle ( $B$ ) the spin of particle ( $B$ ) should be well defined in reality. The same issue should hold vice versa also for particle ( $A$ ). This is at least the classical conclusion which was drawn also by EPR using the argument of classical local reality. EPR use the following def-



**Fig. 7.4** Experiment to entangle two fermions via their spin degree of freedom. The unstable source ( $S$ ), an atom or an atomic nucleus, in a singlet state emits two identical fermions  $A$  and  $B$  with anti-correlated spin orientation in opposite direction. These particles are detected by detectors being run by the two persons *Alice* ( $A$ ) and *Bob* ( $B$ ). The two-particle state of the two simultaneously emitted particles must be represented as a superposition of the two states (*spins as solid and broken red line arrows*) with opposite spin direction

inition of local reality: When particle ( $A$ ) and particle ( $B$ ) are two spatially restricted entities which interacted in the past but are now separated from each other over a long distance, the state of particle ( $A$ ) does not depend on a momentary action on particle ( $B$ ).

From this discussion in connection with the entangled two-particle state (7.17), EPR draw the conclusion that quantum mechanics is incomplete in the existing form. There should be unknown *hidden local variables* which govern quantum mechanical behavior below the formalism of wave function and bra and ket states. These variables should allow for an exchange of information between the particles before they separated and flew away with anti-correlated spins. It must already be mentioned at this point, that hidden variables have never been detected in quantum physics so far.

The counter-intuitive aspect of quantum mechanical behavior as it is expressed in (7.17) shall be further clarified by a paradigmatic classical scenario. Imagine, there are two balls with different color in a box, a blue one and a yellow one. Without looking at it, one takes one of the balls, puts it into his pocket, goes to work and takes the ball out of the pocket there. One realizes, it is a blue ball; then, it is instantly clear, that the yellow ball was left behind and is still in the box at home. This is classical correlation with the implication of strict locality.

Now imagine the scenario happens to be in the quantum world where entanglement and those odd things are encountered. One “quantum ball” is put into the pocket at home without looking at it. Then, at work one looks at it (measurement of a quantum phenomenon). In contrast to classical behavior there is now an equal 50 % probability that one has the blue or the yellow ball in his hand. The outcome of the quantum mechanical check (measurement) is only probabilistic rather than deterministic. But nevertheless, irrespective of the outcome of the check at work, be it the yellow or the blue ball, it is clear that the blue ball is in the box at home, in case one took the yellow ball out of the pocket at work or vice versa. In this type of quantum correlation (entanglement), no one of the two balls has a definite color before one makes the check (measurement) at work. After having found one particular color at work, the color of the ball at large distance at home is also defined. This is non-local quantum behavior.

In conclusion, we approach the philosophically interesting question if the two particles ( $A$ ) and ( $B$ ) in (7.17) have local reality or if the entangled state (7.17) is kind of a pending possibility for the results of Alice's and Bob's spin direction measurements, of course, under the condition of strict anti-correlation of the spin orientations. Is there a contradiction between local reality and the laws of quantum mechanics? As will be shown, this question can be answered experimentally and the answer is that quantum physical reality is non-local.

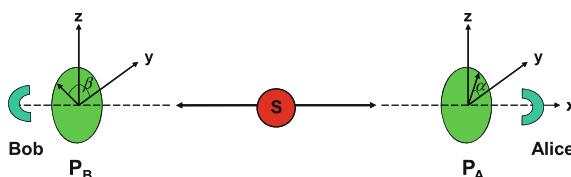
### 7.2.1 Bell's Inequality and Its Experimental Check

In 1964, John Bell succeeded in establishing inequality relations [6, 7] which are based on the assumption of local reality and which can simultaneously be checked experimentally.

The inequality is based on the discussed gedanken experiment (Sect. 7.2) in which an unstable source ( $S$ ) emits anti-correlated spin-1/2 particles into opposite directions (Fig. 7.5). At large distances from the source polarization analyzers  $P_A$  and  $P_B$  are positioned in front of the detectors of Alice (A) and Bob (B). Only particles can pass the analyzers which have a spin orientation  $+\hbar/2$  in the momentary analyzer direction  $\alpha$  in front of Alice's detector, respectively  $\beta$  in front of Bob's detector. Because of the properties of the source and the formal structure of the entangled state of the emitted particles (7.17), it is evident that a detection event at the detector A under the polarizer angle  $\alpha$  ensures the emission of a particle with spin  $-\hbar/2$  under  $\alpha$  direction along the path to the detector B. With the polarizer  $P_B$  in  $\beta$  direction, then, no particle is detected by Bob (detector B).

In order to measure correlations between detection events at the detectors A and B, each adjustment of the analyzers  $P_A$  and  $P_B$  must be random and an action of one analyzer on the other one must be excluded. For this purpose analyzer  $P_B$  must be adjusted within such a short time period after the adjustment of  $P_A$  that a light signal emitted from  $P_A$  can not reach  $P_B$  and vice versa. In a relevant correlation experiment, many detection events must be measured by Alice and Bob for ever new randomly adjusted polarizer orientations  $\alpha$  (near Alice's detector) and  $\beta$  (near Bob's detector).

Before we discuss the experimental realization of the difficult experiment, we want to analyze the theoretical background a little bit further: With common sense we assume the standpoint of local realism: Each particle has a well defined, fixed spin value in each given direction. This value is not changed by some action or modification of the apparatus far away from the considered particle. Alice and Bob perform many measurements of the spin orientation using always different analyzer angles  $\alpha$  and  $\beta$ . They register a detection event only if the particle arriving at their detector has a spin value  $+\hbar/2$  in the randomly adjusted analyzer direction  $\alpha$  or  $\beta$ .



**Fig. 7.5** Experiment for investigating Bell's inequality relation [6, 7] according to Bohm [5]. As in Fig. 7.4, two detectors at Alice and Bob detect two particles with anti-correlated spins emitted from the source ( $S$ ) in opposite direction. The spin direction of the emitted particles can be determined by the polarizers  $P_A$  and  $P_B$ . The polarizer orientation can be randomly adjusted in order to measure correlations between the spin orientations at A and B

We assume  $N(\alpha/\beta)$  to be the number of detection events where Alice ( $A$ ) finds positive spin under the analyzer angle  $\alpha$  and Bob ( $B$ ) simultaneously registers positive spin under the angle  $\beta$ .  $N(\alpha/\beta)$  can, however, also be interpreted as the number of events where  $A$  detects positive spin under  $\alpha$ , but under the angle  $\beta$  nothing is found in the detector  $A$ . The reason is that  $B$  detects positive spin in  $\beta$  direction [see definition of  $N(\alpha/\beta)$ ], which requires the detection of a particle with negative spin under  $\beta$  direction by the detector  $A$ . The negative spin direction, however, can not pass the analyzer and the particle is not detected in  $A$ .

We now consider an additional measurement with the polarizer in  $\gamma$  orientation. The number  $N(\alpha/\beta)$  can, then, be divided into two groups of particles, those which can pass with positive spin  $+\hbar/2$  under the analyzer angles  $\alpha$  and  $\gamma$ ,  $N(\alpha\gamma/\beta)$ , and those which can pass with positive spin under  $\alpha$  direction only, but not under the directions  $\beta$  and  $\gamma$ ,  $N(\alpha/\beta\gamma)$ .

Both groups of particles, those which can pass under the angle  $\gamma$  and those which cannot pass under  $\gamma$ , must add up to the total number  $N(\alpha/\beta)$ , i.e. we must require:

$$N(\alpha/\beta) = N(\alpha\gamma/\beta) + N(\alpha/\beta\gamma). \quad (7.18)$$

$N(\alpha\gamma/\beta)$  is a subclass of the number  $N(\gamma/\beta)$  since in addition to the requirement of passing the analyzer with positive spin direction under the angle  $\gamma$  the particles within  $N(\alpha\gamma/\beta)$  must also pass with spins in  $\alpha$  direction (more stringent condition). An analogous condition is derived from the comparison of  $N(\alpha/\beta\gamma)$  with  $N(\alpha/\gamma)$ . We thus obtain the following relations:

$$N(\alpha\gamma/\beta) \leq N(\gamma/\beta), \quad (7.19a)$$

$$N(\alpha/\beta\gamma) \leq N(\alpha/\gamma). \quad (7.19b)$$

From (7.18), (7.19a), (7.19b), Bell's inequality is derived as

$$N(\alpha/\beta) \leq N(\alpha/\gamma) + N(\gamma/\beta). \quad (7.20a)$$

In this form, the inequality can be checked experimentally with the conditions that the spin analyzers  $P_A$  and  $P_B$  are adjusted each time randomly and that any kind of interaction between  $P_A$  and  $P_B$  is excluded. It is emphasized again that (7.20a), (7.20b) are derived under the condition of local reality, that is, the two  $\hbar/2$  spin particles emitted in opposite directions with anti-correlated spin orientation possess their spins as real, well-defined qualities on the whole way from the source to the detector. The spin orientations are determined by the history of the particles within the unstable source. Note that this predetermined, well defined reality of the spin orientations is not contained in the representation (7.17) of the entangled two-particle state.

What is the answer of quantum mechanics to Bell's inequality (7.20a) which is based on the assumption of local reality? To this purpose, we must calculate correlations between detection events for randomly adjusted spin analyzers on the

basis of the entangled two-particle state (7.17). In quantum physics, probabilities are calculated. We, therefore, transfer (7.20a) into a relation for probabilities  $w$  ( $0 \leq w \leq 1$ ) by dividing the number  $N$  by the total number of all detection events  $N_0$ :

$$w(\alpha/\beta) \leq w(\alpha/\gamma) + w(\gamma/\beta). \quad (7.20b)$$

For the quantum mechanical calculation of correlations, we need an operator  $\hat{O}_A(\alpha)$  which describes the probability for a detection event at Alice's detector ( $A$ ), that is, the probability for an incident particle with positive  $(+\hbar/2)$  spin orientation in  $\alpha$  direction passing the polarizer  $P_A$  (adjusted at an angle  $\alpha$ ). For an incident particle with spin  $-\hbar/2$  in  $\alpha$  direction the average (expectation) value  $\langle s_A | \hat{O}_A(\alpha) | s_A \rangle$  must vanish. In other words, we need an operator  $\hat{O}_A(\alpha)$  which fulfills the conditions  $_A \langle \uparrow | \hat{O}_A(\alpha) | \uparrow \rangle_A = 1$  and  $_A \langle \downarrow | \hat{O}_A(\alpha) | \downarrow \rangle_A = 0$ . One can easily show that these conditions are fulfilled by the following operator:

$$\hat{O}_A(\alpha) = \frac{1}{2}(1 + \hat{\sigma}_A \cdot \mathbf{n}_\alpha). \quad (7.21)$$

Hereby,  $\mathbf{n}_\alpha$  is the unity vector in  $\alpha$  direction within the  $y-z$  plane of the analyzer  $P_A$  at Alice's place and  $\hat{\sigma}_A$  is the spin operator belonging to the particle incident in Alice's direction. For detection in  $z$  direction, that is,  $\alpha = 90^\circ$ , we find  $\hat{\sigma}_A \cdot \mathbf{n}_{90} |\uparrow\rangle = \hat{\sigma}_z |\uparrow\rangle = |\uparrow\rangle$  and  $\hat{\sigma}_A \cdot \mathbf{n}_{90} |\downarrow\rangle = -|\downarrow\rangle$ . The operator (7.21) filters out the value one for positive spin direction and the value zero for negative spin with the same analyzer orientation, here  $90^\circ$ . This is exactly the property we need for the evaluation of the probability that Alice ( $A$ ) detects a particle with positive spin orientation under the angle  $\alpha$ . An analogous consideration for Bob's detector yields the following detection operators:

$$\hat{O}_A(\alpha) = \frac{1}{2}(1 + \hat{\sigma}_A \cdot \mathbf{n}_\alpha), \quad (7.22a)$$

$$\hat{O}_B(\beta) = \frac{1}{2}(1 + \hat{\sigma}_B \cdot \mathbf{n}_\beta). \quad (7.22b)$$

By means of these operators we can calculate, within the framework of quantum mechanics and given the entangled two-particle state (7.17), the probability that Alice ( $A$ ) detects a particle with positive spin  $+\hbar/2$  under the angle  $\alpha$  and Bob a particle with positive spin under the angle  $\beta$ . We consider only the spin part of (7.17) and calculate the correlation probability for the two detection events as follows:

$$\begin{aligned} & \langle s_A s_B | \frac{1}{2}(1 + \hat{\sigma}_A \cdot \mathbf{n}_\alpha) \frac{1}{2}(1 + \hat{\sigma}_B \cdot \mathbf{n}_\beta) | s_A s_B \rangle \\ &= \frac{1}{4} \langle s_A s_B | 1 + \hat{\sigma}_A \cdot \mathbf{n}_\alpha + \hat{\sigma}_B \cdot \mathbf{n}_\beta + (\hat{\sigma}_A \cdot \mathbf{n}_\alpha)(\hat{\sigma}_B \cdot \mathbf{n}_\beta) | s_A s_B \rangle. \end{aligned} \quad (7.23)$$

In the further calculation, the coordinate system of Fig. 7.5 with the  $y-z$  plane for the description of the polarizer (analyzer) angles and  $x$  as the propagation direction of the emitted particles is used. We, then, obtain:

$$\hat{\sigma}_A \cdot \mathbf{n}_\alpha = (\cos \alpha) \hat{\sigma}_y^A + (\sin \alpha) \hat{\sigma}_z^A, \quad (7.24a)$$

$$\hat{\sigma}_B \cdot \mathbf{n}_\beta = (\cos \beta) \hat{\sigma}_y^B + (\sin \beta) \hat{\sigma}_z^B. \quad (7.24b)$$

Using the representation of the Pauli spin matrices (5.116) and the 2D vector representation of spin states  $|\uparrow\rangle = (1, 0)$  and  $|\downarrow\rangle = (0, 1)$ , we derive the following relations:

$$\hat{\sigma}_x |\uparrow\rangle = |\downarrow\rangle, \quad \hat{\sigma}_x |\downarrow\rangle = |\uparrow\rangle, \quad (7.25a)$$

$$\hat{\sigma}_y |\uparrow\rangle = i |\downarrow\rangle, \quad \hat{\sigma}_y |\downarrow\rangle = -i |\uparrow\rangle, \quad (7.25b)$$

$$\hat{\sigma}_z |\uparrow\rangle = |\uparrow\rangle, \quad \hat{\sigma}_z |\downarrow\rangle = -|\downarrow\rangle. \quad (7.25c)$$

By use of these relations the  $\hat{\sigma}_A \cdot \mathbf{n}_\alpha$  term in (7.23) follows as:

$$\begin{aligned} & \langle s_{ASB} | \hat{\sigma}_A \cdot \mathbf{n}_\alpha | s_{ASB} \rangle \\ &= \frac{1}{\sqrt{2}} \left( {}_A \langle \uparrow | {}_B \langle \downarrow | - {}_A \langle \downarrow | {}_B \langle \uparrow | \right) \\ &\quad \cdot \frac{1}{\sqrt{2}} \left\{ i \cos \alpha (|\downarrow\rangle_A |\downarrow\rangle_B + |\uparrow\rangle_A |\uparrow\rangle_B) + \sin \alpha (|\uparrow\rangle_A |\downarrow\rangle_B + |\downarrow\rangle_A |\uparrow\rangle_B) \right\} \\ &= \frac{1}{2} \left( {}_A \langle \uparrow | \uparrow \rangle_{AB} \langle \downarrow | \downarrow \rangle_B - {}_A \langle \downarrow | \downarrow \rangle_{AB} \langle \uparrow | \uparrow \rangle_B \right) \sin \alpha = 0. \end{aligned} \quad (7.26)$$

Analogously the  $\hat{\sigma}_B \cdot \mathbf{n}_\beta$  term in (7.23) vanishes. This is nothing to wonder about since the state (7.17) has singlet character with spin compensation.

We now consider the term in (7.23) which is quadratic in the spin operators and use (7.24a), (7.24b) for the evaluation:

$$\begin{aligned} & \langle s_{ASB} | (\hat{\sigma}_A \cdot \mathbf{n}_\alpha)(\hat{\sigma}_B \cdot \mathbf{n}_\beta) | s_{ASB} \rangle \\ &= \langle s_{ASB} | (\cos \alpha \cos \beta) \hat{\sigma}_y^A \hat{\sigma}_y^B + (\cos \alpha \sin \beta) \hat{\sigma}_y^A \hat{\sigma}_z^B + (\sin \alpha \cos \beta) \hat{\sigma}_z^A \hat{\sigma}_y^B \\ &\quad + (\sin \alpha \sin \beta) \hat{\sigma}_z^A \hat{\sigma}_z^B | \frac{1}{\sqrt{2}} (|\uparrow\rangle_A |\downarrow\rangle_B - |\downarrow\rangle_A |\uparrow\rangle_B) \rangle \\ &= \frac{1}{2} \left( {}_A \langle \uparrow | {}_B \langle \downarrow | - {}_A \langle \downarrow | {}_B \langle \uparrow | \right) \{ \cos \alpha \cos \beta (|\downarrow\rangle_A |\uparrow\rangle_B - |\uparrow\rangle_A |\downarrow\rangle_B) \} \\ &\quad - i \cos \alpha \sin \beta (|\downarrow\rangle_A |\downarrow\rangle_B - |\uparrow\rangle_A |\uparrow\rangle_B) \\ &\quad - i \sin \alpha \cos \beta (|\uparrow\rangle_A |\uparrow\rangle_B - |\downarrow\rangle_A |\downarrow\rangle_B) \\ &\quad - i \sin \alpha \sin \beta (|\uparrow\rangle_A |\downarrow\rangle_B - |\downarrow\rangle_A |\uparrow\rangle_B) \\ &= -\langle s_{ASB} | \cos \alpha \cos \beta - \sin \alpha \sin \beta | s_{ASB} \rangle = -\cos(\alpha - \beta). \end{aligned} \quad (7.27)$$

We, thus, obtain the probability (7.23), that simultaneously Alice ( $A$ ) detects a particle with spin  $+\hbar/2$  under  $\alpha$  direction and Bob ( $B$ ) a particle with the same spin under  $\beta$  direction, i.e. the quantum mechanical correlation  $w_Q(\alpha/\beta)$  of an  $AB$  detection event with positive spin under  $\alpha$  and  $\beta$  direction as

$$\begin{aligned} w_Q(\alpha/\beta) &= \langle s_A s_B | \frac{1}{2}(1 + \hat{\sigma}_A \cdot \mathbf{n}_\alpha) \frac{1}{2}(1 + \hat{\sigma}_B \cdot \mathbf{n}_\beta) | s_A s_B \rangle \\ &= \frac{1}{4}[1 - \cos(\alpha - \beta)] = \frac{1}{2} \sin^2 \frac{\beta - \alpha}{2}. \end{aligned} \quad (7.28)$$

By simply plugging in different detection angles (polarizer orientations)  $\alpha, \beta, \gamma$  we can check if the quantum mechanically calculated result obeys Bell's inequality (7.20b). With the assumptions

$$\alpha = 0^\circ, \quad \beta = 90^\circ, \quad \gamma = 45^\circ. \quad (7.29a)$$

Equation (7.28) yields

$$w_Q(\alpha/\beta) = 0.5, \quad w_Q(\alpha/\gamma) = 0.146, \quad w_Q(\gamma/\beta) = 0.146. \quad (7.29b)$$

The quantum mechanical correlation, thus, obeys

$$w_Q(\alpha/\beta) > w_Q(\alpha/\gamma) + w_Q(\gamma/\beta). \quad (7.29c)$$

This is in contradiction to Bell's inequality (7.20b) which has been derived under the condition of local reality for the two emitted particles.

A further example with polarizer angle adjustments  $\alpha = 0^\circ, \beta = 60^\circ, \gamma = 30^\circ$  shows a similar contradiction (7.29c) to Bell's inequality (7.20b):

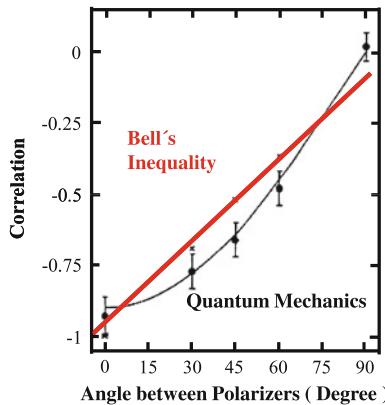
$$\begin{aligned} w_Q(\alpha/\beta) &= \frac{1}{2} \sin^2 30^\circ > \frac{1}{2} \sin^2 15^\circ + \frac{1}{2} \sin^2 15^\circ \\ &= w_Q(\alpha/\gamma) + w_Q(\gamma/\beta). \end{aligned} \quad (7.30)$$

Quantum mechanical reality is obviously in contradiction to common sense thinking based on local reality as it is expressed by Bell's inequality. The contradiction might be illuminated again by phrasing it in other words: In common sense thinking the two emitted particles possess local reality concerning their spin orientation. Due to their history, they have a well defined spin orientation, which is revealed as anti-correlated by Alice's and Bob's measurements. In quantum mechanical thinking expressed by the entangled state (7.17), neither of the two emitted particles has a well defined spin state. Only after measurement of the spin direction in a particular basis (angle adjustment of the polarizers) the particle in Alice's detector incidentally assumes one of the two possible spin directions and instantaneously Bob's particle has an orthogonal (anti-correlated) spin state.

Note that this non-local phenomenon is not in contradiction to the theory of relativity: The instantaneous anti-correlation of the spin orientations can only be verified by Alice's and Bob's measurements, an information exchange between the detectors with super-light-velocity is not possible.

So far we have considered merely theoretical concepts. The interesting question is, what answer gives the experiment. Is the answer in favor of Bell's inequality or is quantum mechanics valid?

In 1976, Lamehi-Rachti and Mittig [8] published a paper about an experimental check of Bell's inequality. By irradiation of a hydrogen target (poly-ethylene foil) with protons from the Saclay accelerator (Paris) they produced two spin anti-correlated protons being emitted in opposite directions. Correlations between their spin orientations were investigated by two spin-resolving detectors. Statistically chosen polarizer adjustments defined particular spin directions in planes perpendicular to the opposite flight directions of the two protons and correlations between detection events in the two channels ( $A$  and  $B$  detection) were measured. The experiment did not fulfill the condition of local separation of the  $A$  and  $B$  detection events, that is, the distance between the detectors  $A$  and  $B$  was not large enough that a mutual influence of the detectors could be excluded (because of finite light velocity). Figure 7.6 shows the outcome of the experiment. For angular differences of 30, 45, 60° between the spin directions at the  $A$  and  $B$  detector the data points, that is, the correlations between the two detection events, clearly lie below the straight line predicted by Bell's inequality. The quantum mechanical prediction well matches the experimental data points.



**Fig. 7.6** Correlation between the spin orientation of two protons emitted from a hydrogen target in opposite directions [8]. The data points have been experimentally determined, while the *curved solid line* describes the quantum mechanical prediction. In contrast, the *straight red line* shows the classically expected correlation according to Bell's inequality relation [6, 7]

As already emphasized, the described experiment with anti-correlated protons [8] did not fulfill the condition of spatial separation of the detection events in the two detectors, since it was performed statically with fixed spin polarizers.

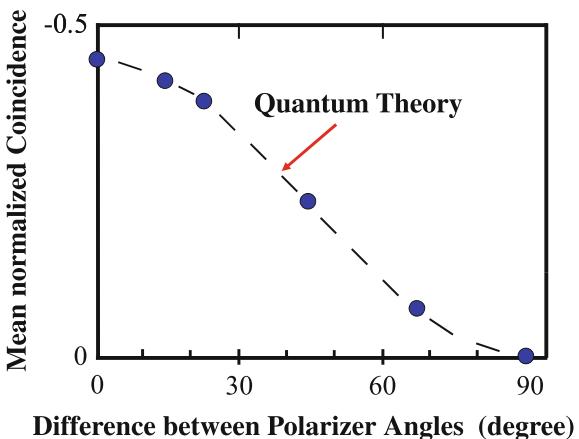
Analogous experiments can also be performed with photons, where fast optical tools allow spatial separation of the detection events. Photons are bosons with spin  $\pm\hbar$ . The two spin orientations correspond to right and left hand circularly polarized light which can be realized experimentally by two linearly polarized waves with mutually perpendicular polarization. The analogy with the spin two-level system is evident. Linearly polarized photons can be analyzed and switched by very fast operating optical polarizers or switches operating in connection with fixed polarizers. By means of these tools, the angular adjustments of the photon spins at the detectors A and B can be made randomly in the two channels and sufficiently fast, that spatial separation between the detection events is guaranteed. The random angular adjustments are made separately while the two photons are on their way to the detectors. Thus, Bob's measurement can not influence that of Alice and vice versa.

At the beginning of the eighties of the 20th century, Alain Aspect and coworkers [9] performed such measurements at the Paris University. The spin anti-correlated photons originate from optically excited Cs-atoms. The measurement results are shown in Fig. 7.7. The coincidences between measurements with different polarizer angles at the detectors A and B fit extremely well with the prediction of the quantum mechanical calculation ( $\pm 1$  standard deviation).

Also further experiments about the correlation or anti-correlation of fermion, respectively, boson spins in an entangled state of the type (7.17) have yielded results in contradiction to Bell's inequality. Quantum theory in the present form describes all the experimental results extremely well.

We have to accept that quantum physical reality is non-local. Evidently, there are no so-called hidden variables which give local reality to the entangled particles in (7.17) and describe their fate in time in a deterministic way below the level of probabilistic wave functions.

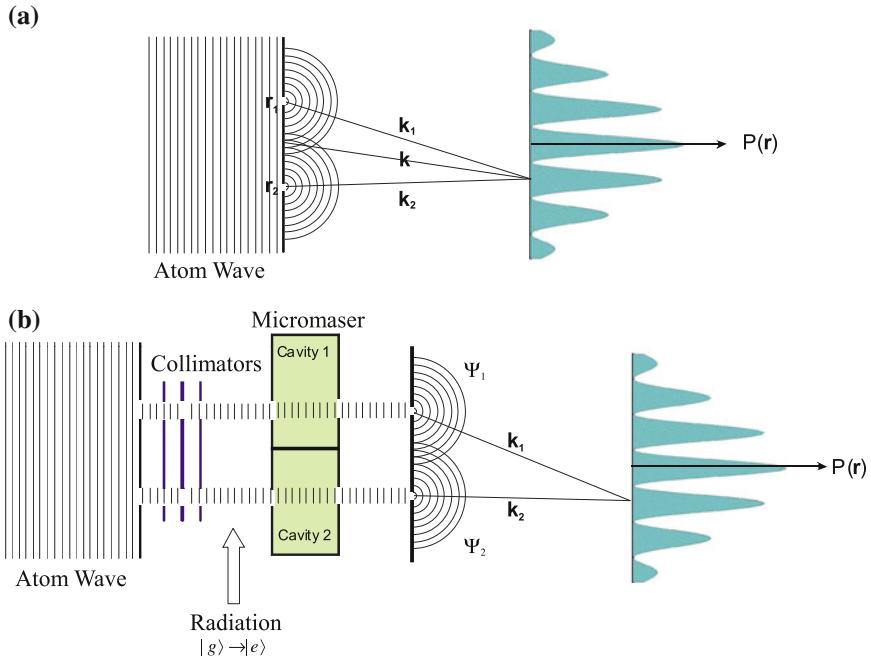
**Fig. 7.7** Measured detection coincidence of Alice and Bob (Fig. 7.5) for spin anti-correlated photons emitted from excited Cs atoms as a function of the randomly adjusted polarizer angles  $\alpha$  and  $\beta$  ( $P_A$  and  $P_B$  in Fig. 7.5). The data points are the experimental results. The curve in broken line is the quantum mechanical prediction [9]



### 7.2.2 “Which Way” Information and Entanglement: A Gedanken Experiment

In the two-particle scattering experiment with fermions (Sect. 7.1.1), entanglement of the spin degree of freedom with the spatial part of the state vector (wave function) made the interference terms of the scattering amplitudes for back and forward scattering disappear in case, that anti-parallel spin orientations allow the distinction between both scattering processes. The distinction between back and forward scattering is equivalent to a “Which Way” information about details of the particle trajectory (Fig. 7.1). An experimental probe providing this information—here opposite spin orientations of the incident particles—washes out the interference terms. One immediately remembers the double slit interference experiment (Sect. 2.4.2) where the additional observation of the particle path, that is, the information about passing this or that slit, destroys the interference pattern. We might ask if again entanglement between two subsystems, the observed particle and the “Which Way” detector, causes the double slit interference to disappear. In 1991, Scully et al. published a gedanken experiment [10] which suggests this conclusion. The experimental proof was finally achieved not earlier than 1998 by Dürr et al. [11] in a quite complex experimental set-up which is discussed in Sect. 8.2.4. Nevertheless, it is worth discussing the early gedanken experiment [10], because it brings the essential issues very clearly to the point.

In Scully’s et al. experiment [10], the double slit interference of a beam of rubidium (Rb) atoms is considered. In contrast to “simple” particles as electrons composed particles as atoms possess internal degrees of freedom (spin, electronic excitations) which provide possibilities for “Which Way” detection. The Rb atoms generating the interference (Fig. 7.8a) exhibit long living excited electronic states, so-called Rydberg states, of the outer electronic shell. An electron excited into these states by a spectrally sharp (laser) light wave of adequate photon energy returns into its ground state by deexcitation through a sequence of other electronic states. When the excited atom, on its way to the double slit, passes an electromagnetic resonator cavity being tuned exactly to the deexcitation frequency of the Rb atom, the atom strongly couples to the cavity and spontaneously emits radiation (Fig. 7.8b). Upon passing the resonator cavity the atom is deexcited and the cavity gains the corresponding energy as a light quantum (photon). Then, a highly precise measurement of the radiation power in the cavity allows to determine, whether the Rb atom has passed the resonator or not. An electronic transition of the Rb atom being suitable for the experiment is in the low energy microwave regime; its frequency amounts to 21 GHz. At room temperature the resonator cavity contains already a high number of these microwave photons, that is, many standing electromagnetic waves with this frequency are excited; an additional photon emitted from the passing Rb atom is not detectable. The resonator must be on extremely low temperature, such that its thermal energy is not sufficient for the excitation of microwave photons. Only on the background of a negligible microwave excitation level an additional photon providing the “Which Way” information can be detected. Having in mind these conditions, the double slit experiment with Rb atoms



**Fig. 7.8** **a, b** Gedanken experiment to elucidate the relation between entanglement and “Which Way” information in the double slit diffraction of a Rb atomic beam [10]. **a** Double slit experiment with incident atomic beam described by a plane wave. Partial waves originating from the slits at  $r_1$  and  $r_2$  interfere with each other. **b** Similar experimental set-up as in (a), but additional equipment between incident atomic beam (particle wave) and double slit allows the determination of the detailed path way (slit 1 or slit 2) of a Rb atom: Collimator apertures prepare two spatially restricted partial atomic beams from the incident beam, which separately pass two distinct cavities of micro-masers. The cavity resonators are tuned to a characteristic interatomic transition  $|g\rangle \leftrightarrow |e\rangle$  of the Rb atoms at around 21 GHz. For recording the “Which Way” information the atoms are excited by suited electromagnetic radiation before entering the micro-masers from the ground state  $|g\rangle$  into the excited state  $|e\rangle$ . Depending on the way the atom moves an RF photon is detected in cavity 1 or in cavity 2 due to de-excitation of the atom from  $|e\rangle$  to  $|g\rangle$

as in Fig. 7.8 can be performed as follows: Rb atoms are evaporated from a furnace and two well defined atom beams are formed by two apertures and a subsequent collimator system (Fig. 7.8b). The two beams described by two spatially restricted wave packets travel through two separate so-called maser microwave resonators (1) and (2), essentially two electro-magnetic resonator cavities, and pass the double slit aperture (1) and (2), where the interfering particle waves  $\psi_1$  and  $\psi_2$  are formed. These waves propagate into open space where they interfere with each other. The interference intensity pattern is registered at long distance by a large area detector screen consisting of many laterally arranged particle detectors. Before the Rb atoms enter the cavity resonators, they can be excited by a laser beam whose monochromatic photons have the energy of the described electronic transitions. Without laser irradi-

ation, the atoms are in their ground state and can not transfer a micro-wave quantum (photon) to the resonator cavity. “Which Way” information can not be obtained from the experiment. After switching on the laser light, the Rb atoms are excited and can emit a photon into resonator (1) or resonator (2) depending on the path the atom moves.

In the bra-ket notation, we can express the states of the resonators as follows:  $|1\ 0\rangle$  describes the creation of a microwave photon ( $\approx 21$  GHz) in resonator (1), that is, the excited Rb atom has passed just this resonator.  $|0\ 1\rangle$  means emission of a photon into resonator (2), but no photon in resonator (1). The Rb atom has passed resonator (2). The exact meaning of the creation of a photon in resonator (1) or resonator (2) will be explained more in detail in Chap. 8. The internal states of the Rb atom simplified as a 2-level system are  $|g\rangle$  and  $|e\rangle$  for the ground and for the excited states, respectively. Creation of a microwave photon in one of the two resonator cavities, thus, is due to de-excitation of the passing excited Rb atom from  $|e\rangle$  to  $|g\rangle$  (Sect. 8.2.2).

The two spherical waves of Rb atoms starting from the two holes (double slit) in the screen behind the two maser cavities (Fig. 7.8b) are described at large distance (detector screen) by two plane wave amplitudes  $\psi_j = C \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_j) - i\omega t]$  with  $j = 1, 2$ , as in Sect. 3.1. For large distances between the detector screen and the double slit arrangement, we can introduce a common relative coordinate for the two particle waves by  $\mathbf{R} \approx \mathbf{r} - \mathbf{r}_1 \approx \mathbf{r} - \mathbf{r}_2$  (Fig. 7.8b). When the laser beam in front of the cavities is not switched on, the outer electron of the Rb atoms is in its ground state  $|g\rangle$  and no photons can be created in the cavities. “Which Way” information can not be obtained. The electronic state of the Rb atoms at the detector, then, must be expressed as

$$|\psi(\mathbf{R})\rangle = \frac{1}{\sqrt{2}} [|\psi_1(\mathbf{R})\rangle + |\psi_2(\mathbf{R})\rangle] |g\rangle \quad (7.31)$$

or in space representation by wave functions:

$$\psi(\mathbf{R}) = \langle r | \psi \rangle \frac{1}{\sqrt{2}} [|\psi_1(\mathbf{R})\rangle + |\psi_2(\mathbf{R})\rangle] |g\rangle. \quad (7.32)$$

Consequently, the intensity of the particle waves on the detector screen follows as

$$I = \psi^* \psi = \frac{1}{2} [|\psi_1|^2 + |\psi_2|^2 + \psi_1^* \psi_2 + \psi_2^* \psi_1] \langle g | g \rangle. \quad (7.33)$$

Because of  $\langle g | g \rangle = 1$  and because of

$$\psi_1^* \psi_2 + \psi_2^* \psi_1 = 2C^2 \cos \mathbf{k} \cdot (\mathbf{r}_2 - \mathbf{r}_1) \quad (7.34)$$

we obtain the familiar double slit interference pattern which is always observed, when we do not have any information about the particle path.

Now, we switch on the laser radiation such that the Rb atoms are excited into the state  $|e\rangle$  before they pass the maser resonators. Depending on the detailed path

an atom can create a microwave photon whether in cavity (1) or cavity (2) by de-excitation. Cavity (1) assumes the state  $|1\ 0\rangle$  or cavity (2) the state  $|0\ 1\rangle$ . Measurement of the resonator cavity states, that is, detection of a photon here or there, yields the “Which Way” information. Taking into account the different cavity states for the two possible particle trajectories the quantum state of the particles at the detector is expressed as:

$$|\psi\rangle = \frac{1}{\sqrt{2}}[|\psi_1\rangle|1\ 0\rangle + |\psi_2\rangle|0\ 1\rangle]|g\rangle. \quad (7.35)$$

The factor  $|g\rangle$  represents the fact that the particles are again in their ground state after passage of the resonators.

In space representation, we obtain the following wave amplitude from (7.35):

$$\psi(\mathbf{R}) = \frac{1}{\sqrt{2}}[\psi_1|1\ 0\rangle + \psi_2|0\ 1\rangle]|g\rangle \quad (7.36)$$

and finally the probability density, that is, the particle wave intensity  $I$  on the detector screen:

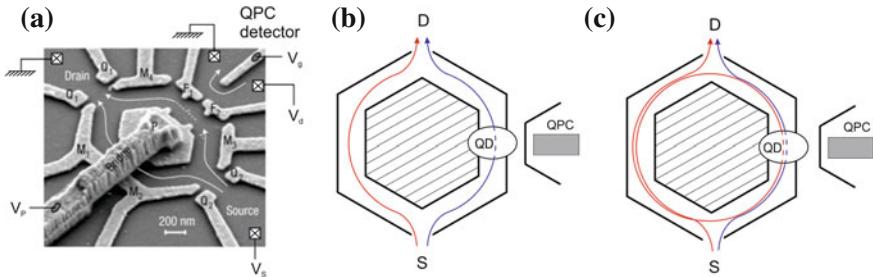
$$I = \psi^*\psi = \frac{1}{2}(|\psi_1|^2 + |\psi_2|^2 + \psi_1^*\psi_2\langle 1\ 0|0\ 1\rangle + \psi_2^*\psi_1\langle 0\ 1|1\ 0\rangle)\langle g|g\rangle. \quad (7.37)$$

The two states  $|0\ 1\rangle$  and  $|1\ 0\rangle$  of the cavity resonators (1) and (2), each time with one additional photon in one resonator and none in the other one, are, of course, orthogonal with respect to each other. In (7.37), therefore, the interference terms with  $\langle 1\ 0|0\ 1\rangle$  and  $\langle 0\ 1|1\ 0\rangle$  vanish. The double slit interference has disappeared due to probing the detailed path of the atom, be it through hole (1) or hole (2) in the double slit arrangement. This is exactly the phenomenon which we have observed in an analogue experiment in Sect. 2.4.2. In the present gedanken experiment, we recognize the underlying reason for the destruction of the double slit interference. When we switch on the experimental tool for obtaining the “Which Way” information, in the present case laser radiation and microwave resonator cavities, the particle states  $|\psi_j\rangle$  in (7.35) become entangled with the states  $|1\ 0\rangle$  and  $|0\ 1\rangle$  of the probe set-up, the resonator cavities. The detected particles and the measurement apparatus can no longer be considered any more as separate, distinct systems. They obey the typically quantum mechanical correlation of entanglement. Non-locality and entanglement explain the influence of the measurement apparatus on the experimental observation in the double slit experiment. We will see that this issue is also relevant for the so-called collapse of the wave function (Sect. 7.4) upon a quantum mechanical measurement.

### 7.2.3 “Which Way” Probing in an Aharonov-Bohm Interference Experiment

Meanwhile some experiments on lithographically prepared nano-rings based on semiconductor heterostructures (Appendix B) have been performed with the aim to study the effect of “Which Way” detection on Aharonov-Bohm (AB) interferences of electrons [12–14]. This is in principle the question for the path an electron travels in a double beam (slit) interference experiment, similarly as in the gedanken experiment of Sect. 7.2.2. In an AB experiment two electron partial waves propagating through the two halves of a ring structure between opposite source and drain contacts undergo a relative phase shift due to a magnetic field which penetrates the ring conductor (Fig. 5.8a). At sufficiently low temperatures and small dimensions the electron transport is ballistic or diffusive without inelastic scattering and in the drain region interference occurs which shows up as magneto-conductance oscillations when the magnetic flux through the conducting loop is varied (Sect. 5.4.4). The most common way to prepare AB structures is based on high mobility 2D electron gases (2DEGs) at the interface of a modulation doped AlGaAs/GaAs heterostructure and using split gate technology for the definition of the various isolating and conducting areas (Appendices A and B). As in Sect. 5.4.4 described also mesa-type semiconductor ring structures are used for the studies. For getting “Which Way” information a quantum dot (QD) is embedded within one path of the ring interferometer. An electron travelling from source to drain through this path can be trapped for some dwelling time in the QD. The increased charge in the dot can be detected by a QPC arranged in close vicinity (Sect. 3.8). The QPC charge detector thus yields the “Which Way” information in the double beam interference experiment.

In Fig. 7.9a a scanning electron micrograph of the nanoscaled AB device of the experiment of Chang et al. [12] is shown. The open Aharonov-Bohm ring (radius about 550 nm) is defined by a number of hexagonally arranged metal gates  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$  and a central island gate  $P$  connected to a metal bridge on top of an AlGaAs/GaAs heterostructure with a 2DEG about 80 nm below the surface (Appendix B). Electron depletion zones below the metal gates (often enhanced by applied negative voltages) pinch the 2DEG and thus separate the different well conducting areas of the 2DEG from each other. The two gate electrodes  $F_1$  and  $F_2$  as well as the two little nipples at the central island gate opposite  $F_1$  and  $F_2$  define the QD in the right path of the AB interferometer and separate it from the QPC charge detector. The island gate  $P$  is electrically controlled through a metal bridge. Measurements at a temperature of 140 mK are made in lock-in technique by applying a  $10\ \mu\text{V}$  ac excitation voltage to the source and monitoring the output current at the drain contact. Charge measurements on the QD to get “Which Way” information were made by probing the conductance of the QPC detector. For this purpose the QPC conductance had been adjusted by the gate voltage  $V_g$  to a particular value where the conductance jumps between two plateaus (Fig. 3.22) and highest sensitivity to potential variations in the surrounding is given. Details of this charge measurement are described in Sect. 3.8. The correct function of the coupling between QD and QPC detector is checked



**Fig. 7.9** Aharonov Bohm (AB) interference experiment with “Which Way” probing [12]. **a** Scanning electron micrograph of the nanostructured Aharonov Bohm ring device prepared on a GaAs/AlGaAs semiconductor heterostructure with a 2D electron gas (2DEG) at the interface about 80 nm below the surface. The ring-like electron channel between source (S) and drain (D) contacts is defined in split-gate technology by the hexagonally arranged metal gates  $Q_i$ ,  $M_i$ , and the central island P being electrically controlled by an air bridge contact. Electron depletion zones below the metal gates isolate uncovered areas of the 2DEG from each other. In the right interferometer arm a quantum dot (QD) is defined by the gates  $F_1$ ,  $F_2$  and the two opposite nipples at the central island. The contacts  $F_1$  and  $F_2$  together with the gate contact ( $V_g$ ) define the quantum point contact (QPC) used as a charge detector. **b** Electron trajectories (red and blue) between S and D regions produce the 1st harmonic of the AB interference oscillations with the period of the flux quantum  $h/e$ . “Which Way” information can be obtained by measuring the presence of an additional electron in the QD by means of the QPC detector. **c** Electron trajectories causing the 2nd harmonic of the AB oscillations with the period of half the flux quantum  $h/2e$ . Both trajectories (red and blue) pass the QD

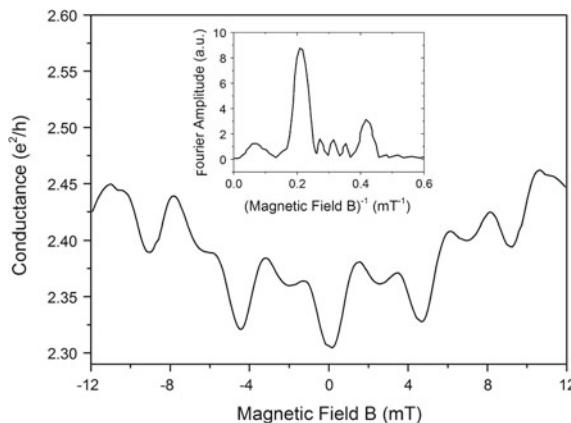
in advance by pinching off the left path of the interferometer loop by large negative voltages on the side gates  $M_1$  and  $M_2$  and measuring simultaneously the QPC and the QD conductance. In case of correct bias adjustments on the QD and the QPC correlated resistance and conductance curves as in Fig. 3.23 are obtained.

For the investigation of AB interferences and the effect of “Which Way” probing the gate potentials are adjusted for an open loop and a magnetic field variable between zero and 14 T is applied normal to the loop plain. In this configuration the ring conductance measured between source and drain contacts as a function of the magnetic flux through the loop ( $\propto$  magnetic field) exhibits two series of magneto-conductance oscillations with periods of one flux quantum  $h/e$  (5.93) and of half the flux quantum  $h/2e$  (2nd harmonic) (Fig. 7.10). The interpretation of the 1st harmonic is in terms of Aharonov-Bohm interferences of the two electron partial waves which travel along the left and the right half circle of the AB ring structure (in red and blue in Fig. 7.9b). According to Sect. 5.4.4 electrons travelling left and right handed around the enclosed magnetic field  $\mathbf{B} = \text{curl} \mathbf{A}$ , which penetrates the loop, gain opposite phase increments, which appear in exponential prefactors in front of the wave functions of the interfering electrons (5.88). The interference is described by a linear superposition of the two partial waves in the drain region and the intensity (absolute square of the amplitudes), respectively the electronic transmission (proportional to intensity) of the loop, contains an interference term which oscillates as a function of  $\Phi/\Phi_0$

with  $\Phi$  as the magnetic flux which penetrates the loop.  $\Phi_0 = h/e$  is the magnetic flux quantum (5.93) and the calculation is given in detail in Sect. 5.4.4 (5.87–5.93).

The 2nd harmonic with the oscillation period of half the flux quantum  $\Phi_0/2 = h/2e$  (Fig. 7.10) originates from interference between an electron partial wave travelling through the QD on the right path and a partial wave starting at the left interferometer arm but encircling the interferometer loop one times before it interferes in the drain region with the wave having travelled through the right arm (Fig. 7.9c). The partial wave starting on the left is exposed to the magnetic field (flux) for one additional circulation around the loop as compared to the partial wave in Fig. 7.9b (1st harmonic). Correspondingly it collects double the phase shift and the interference pattern repeats already at half the period of the 1st harmonic. In general even higher harmonics with multiple circulations around the loop are possible, but their intensity decays rapidly and they are not observed anymore.

For the 1st harmonic (Fig. 9.7b) only one electron partial wave, that through the right interferometer path, passes the QD and can be detected by the coupled QPC. From the QPC conductance a decision is possible about the way the electron has taken (“Which Way” information), through the left or the right interferometer path. This is different for the 2nd harmonic. Both electron partial waves pass the QD one times, of course, at different times. Consequently, without time dependent measurements a decision about the electron path, be it through the right or the left interferometer arm, is not possible. “Which Way” information can not be obtained and we expect a different dephasing behaviour for the 1st and the 2nd harmonic of the AB magneto-conductance oscillations. This is indeed found in the experiment. In Fig. 7.11 the intensities of the Fourier transformed AB magneto-conductance oscillations of the



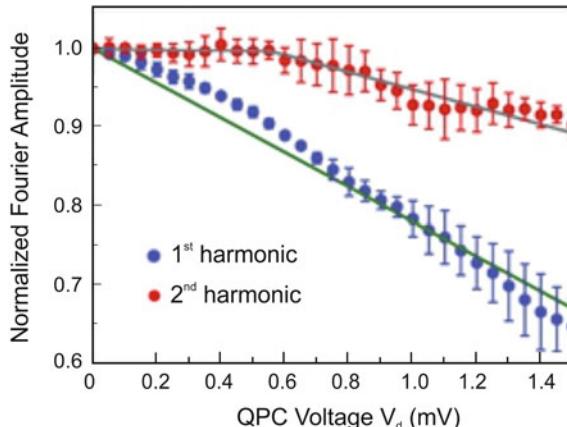
**Fig. 7.10** Aharonov Bohm magneto-conductance oscillations in units of the conductance quantum  $e^2/h$  measured between the S and D regions at a gate voltage  $V_d = 0$  V at the QPC. The stronger oscillation has flux quantum ( $h/e$ ) periodicity (1st harmonic). The 2nd harmonic with periodicity  $h/2e$  is also observed. *Inset* Fourier transform of the magneto-conductance oscillations with 1st and 2nd harmonic peaks [12]

1st and the 2nd harmonic (see inset of Fig. 7.10) are plotted versus the bias  $V_d$  at the QPC charge detector (Fig. 7.9a). An increase of the detector bias  $V_d$  enhances the number of electrons which pass the QPC channel during the dwelling time of an electron in the QD. A higher  $V_d$  value thus amplifies the detection mechanism and enhances dephasing of the interfering electron partial waves. In Fig. 7.11 the normalized 1st harmonic intensity decreases from 1 to about 0.65 for  $V_d$  varying between zero and 1.4 mV, while the 2nd harmonic changes only marginally within this bias range.

The formal description of the experiment can be made in a similar way as for the gedanken experiment in Sect. 7.2.2. Electronic partial waves on their way through the left ( $l$ ) and the right ( $r$ ) half (with imbedded QD) of the interferometer loop are described by the quantum states  $|\psi_l\rangle$  and  $|\psi_r\rangle$ , respectively. A distinction between an electron being in state  $|\psi_l\rangle$  or in state  $|\psi_r\rangle$ , at least for the 1st AB harmonic, is possible by reading out the QPC conductance. Two states of the QPC within the whole variety of possible states are of interest here and will be further considered: one state  $|QPC+\rangle$  which is related to an additional electron in the QD and which indicates the presence of state  $|\psi_r\rangle$  of the electron in the AB loop. The other state  $|QPC-\rangle$  of the QPC charge detector indicates that the electron in the loop has taken the way through the left half circle (state  $|\psi_l\rangle$ ) being related to no additional charge in the QD.

As in Sect. 7.2.2 we can, thus, write the complete superposition wave function of the two interfering partial waves coupled to their linked states of the QPC charge detector as follows:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\psi_l\rangle|QPC-\rangle + |\psi_r\rangle|QPC+\rangle). \quad (7.38a)$$



**Fig. 7.11** Normalized Fourier amplitudes from Fig. 7.10 measured as a function of the gate voltage  $V_d$  at the QPC detector. While the 2nd harmonic amplitude (red) changes only slightly, the 1st harmonic (blue) is considerably suppressed with increasing gate voltage  $V_d$  [12]

As in (7.35) this is an entangled state between interfering electrons and states of the “Which Way” detection apparatus. In space representation this reads:

$$\langle r | \Psi \rangle = \frac{1}{\sqrt{2}} (\langle \mathbf{r} | \psi_l \rangle |QPC-\rangle + \langle \mathbf{r} | \psi_r \rangle |QPC+\rangle). \quad (7.38b)$$

The transmission of the loop between source and drain  $T_{SD}$  is proportional to the probability of finding an electron in the drain region and, therefore, to the absolute square of (7.38b):

$$\begin{aligned} T_{SD} &\propto |\langle \mathbf{r} | \psi_l \rangle|^2 + |\langle \mathbf{r} | \psi_r \rangle|^2 + \langle \mathbf{r} | \psi_l \rangle \langle \psi_r | \mathbf{r} \rangle \langle QPC- | QPC+ \rangle \\ &+ \langle \mathbf{r} | \psi_r \rangle \langle \psi_l | \mathbf{r} \rangle \langle QPC+ | QPC- \rangle \\ &= |\psi_l|^2 + |\psi_r|^2 + \psi_r^* \psi_l \langle QPC- | QPC+ \rangle + \psi_l^* \psi_r \langle QPC+ | QPC- \rangle. \end{aligned} \quad (7.39)$$

In analogy to (7.37) the two terms containing  $\langle QPC- | QPC+ \rangle$  and  $\langle QPC+ | QPC- \rangle$  are the interference terms which contain in the products  $\psi_r^* \psi_l$  and  $\psi_l^* \psi_r$  the magneto-conductance oscillations of the first harmonic with the periodicity of the single flux quantum  $\Phi_0 = h/e$ . This is obvious, since the wave functions  $\Psi_l$  and  $\Psi_r$  contain exponential prefactors which describe the opposite phase increments of the electrons encircling the magnetic flux  $\Phi$  left and right handed. In the calculation in Sect. 5.4.4 the path integrals along the right and the left half of the loop add up to a loop integral around the closed loop which yields the magnetic flux  $\Phi$  through the loop and finally a magneto-conductance oscillation period of  $\Phi_0 = h/e$ .

The interference terms in (7.39) are cancelled out if the QPC detector states  $|QPC+\rangle$  and  $|QPC-\rangle$  are mutually perpendicular. The AB oscillations would not be observed if a unequivocal read out of the “Which Way” information would be possible in the experiment. In the other limiting case of no “Which Way” probing the QPC detection mechanism is skipped, the detector states  $|QPC+\rangle$  and  $|QPC-\rangle$  do not exist or they are mutually parallel. In this case the interference term keeps its full intensity in (7.39) and the AB magneto-conductance oscillations are observed in the experiment. In the described experiment we have a situation in between. The AB interferences can obviously not be destroyed completely by increasing the QPC bias  $V_d$ . The detection mechanism for “Which Way” information can not be made such unequivocal within the possible experimental parameter range that the two indicating QPC states are fully orthogonal. A number of reasons might be discussed; in particular, there might be some interaction also between an electron passing the left arm of the AB ring and the QPC detector. Nevertheless, the outcome of the experiment can be explained in terms of entanglement between the interfering electron waves and the probing QPC detector, in a similar fashion as was discussed in Sect. 7.2.2 for the double slit gedanken experiment.

### 7.3 Pure and Mixed States: The Density Matrix

So far we have seen that superposition states as in the double slit interference or in a two-particle scattering experiment loose their quantum mechanical interference terms in the probability expressions (absolute square of the wave amplitudes) by entanglement of the particle wave function with additional degrees of freedom as spin or other internal excitation states. This transition from quantum mechanical behavior to classical behavior is of central importance for a deeper understanding of the oddities of quantum mechanics. In other words: classical behavior means that we have to add up probabilities in a superposition of different events rather than wave amplitudes in quantum mechanics. For a more extended discussion of these issues, the introduction of a new operator, the so-called *density matrix* will be helpful. The density matrix is often called *statistical operator*, a name which might be more appropriate, as the operator appears as a matrix only in special Hilbert space representations.

Before we proceed into this direction, it is useful to elucidate a little bit more in detail the physical reason for the loss of quantum interferences. For this purpose a further discussion of the Stern–Gerlach experiment (Sect. 5.5.1) is helpful.

#### 7.3.1 Quantum Mechanical and Classical Probabilities

In the Stern–Gerlach experiment a beam of silver atoms—also other atoms can be used—is produced by evaporation from an oven. The atom beam passes a strongly inhomogeneous (in  $z$  direction) magnetic field and is split into two beams corresponding to the two possible spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  with respect to the  $z$  direction of the magnetic field. The atom beam has been separated into the two spin eigenstates, which are then given as two well defined (prepared) quantum states in the two beams. A further subsequent measurement of the spin direction in one of the two beams will always yield only a spin-up  $|\uparrow\rangle$  and a spin-down  $|\downarrow\rangle$  result.

Now we modify the spin orientation in one of the two beams. Atoms in the beam with only spin-up orientation  $|\uparrow\rangle$  in the static magnetic field  $B_0$  in  $z$  direction are exposed to an oscillating magnetic field  $B_{\approx}$  directed in  $x$  direction (normal to  $z$ ). The exposure to this alternating magnetic field should have the duration time of a  $\pi/2$  pulse (Sect. 6.5.2). This  $\pi/2$  pulse flips the spin from the  $z$ -direction into the  $x$  direction and, thus, creates the superposition state

$$|s\rangle_{\text{sup}} = \frac{1}{\sqrt{2}}(|\uparrow\rangle + e^{i\alpha}|\downarrow\rangle). \quad (7.40)$$

In this state, both spin orientations in  $z$  direction are represented with equal weight. The spin rotation axis is directed perpendicular to the magnetic field  $B_0$  and

rotates around the  $B_0$  direction with the frequency  $(E_\uparrow - E_\downarrow)/\hbar = \omega_0$  (Sect. 6.5.2). For reasons of generalization, a phase factor  $\exp(i\alpha)$  has been added to the second term in (7.40). The phase angle  $\alpha$  is determined by experimental details.

For the superposition state (7.40), we can calculate the expectation values  $\langle \hat{s}_x \rangle$ ,  $\langle \hat{s}_y \rangle$ ,  $\langle \hat{s}_z \rangle$ , that is, the average values of the different spin components which are obtained in a measurement of the spin orientations in the three space directions  $x$ ,  $y$ ,  $z$ . For this purpose, many measurements on the ensemble of particles are performed after the beam has passed the experimental set-up. The calculation yields the following results:

$$\begin{aligned}\langle \hat{s}_z \rangle &= \frac{1}{2} \hbar \langle \hat{\sigma}_z \rangle = \frac{1}{2} \hbar \frac{1}{2} (\langle |\uparrow\rangle + e^{-i\alpha} \langle |\downarrow\rangle) \hat{\sigma}_z (\langle |\uparrow\rangle + e^{i\alpha} \langle |\downarrow\rangle) \\ &= \frac{1}{4} \hbar (\langle |\uparrow|\uparrow\rangle - e^{i\alpha} \langle |\uparrow|\downarrow\rangle + e^{-i\alpha} \langle |\downarrow|\uparrow\rangle - \langle |\downarrow|\downarrow\rangle) = 0\end{aligned}\quad (7.41a)$$

$$\begin{aligned}\langle \hat{s}_x \rangle &= \frac{1}{2} \hbar \langle \hat{\sigma}_x \rangle = \frac{1}{4} \hbar (\langle |\uparrow\rangle + e^{-i\alpha} \langle |\downarrow\rangle) \hat{\sigma}_x (\langle |\uparrow\rangle + e^{i\alpha} \langle |\downarrow\rangle) \\ &= \frac{1}{4} \hbar (\langle |\uparrow\rangle + e^{-i\alpha} \langle |\downarrow\rangle) (\langle |\downarrow\rangle + e^{i\alpha} \langle |\uparrow\rangle) \\ &= \frac{1}{4} \hbar (\langle |\uparrow|\downarrow\rangle + \langle |\uparrow|\uparrow\rangle e^{i\alpha} + e^{-i\alpha} \langle |\downarrow|\downarrow\rangle + \langle |\downarrow|\uparrow\rangle) \\ &= \frac{1}{2} \hbar \cos \alpha\end{aligned}\quad (7.41b)$$

$$\begin{aligned}\langle \hat{s}_y \rangle &= \frac{1}{2} \hbar \langle \hat{\sigma}_y \rangle = \frac{1}{4} \hbar (\langle |\uparrow\rangle + e^{-i\alpha} \langle |\downarrow\rangle) \hat{\sigma}_y (\langle |\uparrow\rangle + e^{i\alpha} \langle |\downarrow\rangle) \\ &= \frac{1}{4} \hbar (\langle |\uparrow\rangle + e^{-i\alpha} \langle |\downarrow\rangle) (i \langle |\downarrow\rangle + i e^{i\alpha} \langle |\uparrow\rangle) \\ &= -\frac{1}{2} \hbar \sin \alpha.\end{aligned}\quad (7.41c)$$

For the derivation, the relations (7.25a)–(7.25c) for spin operators have been used. The expectation values (7.41a)–(7.41c) are quantum mechanical averages based on the quantum state (7.40), which is itself defined in probabilistic terms as usual in quantum mechanics (see wave function). The average values contain the maximum knowledge we can have about the system. Since nature is inherently random on the quantum mechanical level [expressed by (7.40)] measurement results are average values obtained from many identical measurements repeatedly performed on one and the same system. No more information is possible about the system than that given by the quantum state (7.40).

Now, we consider a different experiment, where the atomic beam does not pass the Stern–Gerlach set-up which prepares the well defined spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  in the two split beams. A spin resolving detector measures the spin orientation in  $z$  direction immediately after the beam has been produced by evaporation from the

oven. Since both spin orientations, up and down, that is, the states  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , are emitted with equal probability from the source, we will detect the two states each with a probability of 50 %. The same result would be obtained if we mix the two atom beams containing the spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$  behind the Stern–Gerlach set-up in a proportion 1:1. The spin average value in  $z$  direction obtained from repeatedly performed measurements on the mixed beam, then, follows as:

$$\langle\bar{s}_z\rangle = \frac{1}{2}\langle\uparrow|\hat{s}_z|\uparrow\rangle + \frac{1}{2}\langle\downarrow|\hat{s}_z|\downarrow\rangle = \sum_{i=1}^2 p_i \langle s_i |\hat{s}_z | s_i \rangle. \quad (7.42)$$

This average characterized by the upper bar contains the inherently quantum mechanical average (expectation) values  $\langle\uparrow|\hat{s}_z|\uparrow\rangle$  and  $\langle\downarrow|\hat{s}_z|\downarrow\rangle$ , but additionally the average over the classical probabilities  $p_i$  (here:  $p_i = p_\uparrow = p_\downarrow$ ) takes into account that the lack of spin preparation by a Stern–Gerlach set-up prevents information about the initial quantum state. This classical average is necessary because of the lack of knowledge about the system which could in principle be avoided by an additional measurement (spin preparation). The classical probabilities  $p_i$  in (7.42) are based on avoidable or wanted lack of knowledge as in classical thermodynamic statistics. There, we are not interested in the path of single molecules when variables as temperature or pressure of a macroscopic system are considered.

Corresponding to (7.42) the statistical average spin values  $\langle\bar{s}_x\rangle$  and  $\langle\bar{s}_y\rangle$  measured on a 1:1 mixed electron beam with up and down spin orientations or on a non-polarized beam can be calculated. We obtain:

$$\begin{aligned} \langle\bar{s}_z\rangle &= \sum_{i=1}^2 p_i \langle s_i |\hat{s}_z | s_i \rangle = \frac{1}{2}\langle\uparrow|\hat{s}_z|\uparrow\rangle + \frac{1}{2}\langle\downarrow|\hat{s}_z|\downarrow\rangle \\ &= \frac{\hbar}{2}\left(\frac{1}{2}\langle\uparrow|\uparrow\rangle - \frac{1}{2}\langle\downarrow|\downarrow\rangle\right) = 0, \end{aligned} \quad (7.43a)$$

$$\begin{aligned} \langle\bar{s}_x\rangle &= \frac{1}{2}\langle\uparrow|\hat{s}_x|\uparrow\rangle + \frac{1}{2}\langle\downarrow|\hat{s}_x|\downarrow\rangle = \frac{\hbar}{4}\left(\langle\uparrow|\hat{\sigma}_x|\uparrow\rangle + \langle\downarrow|\hat{\sigma}_x|\downarrow\rangle\right) \\ &= \frac{\hbar}{4}\left(\langle\uparrow|\downarrow\rangle + \langle\downarrow|\uparrow\rangle\right) = 0 \end{aligned} \quad (7.43b)$$

and in analogy:

$$\langle\bar{s}_y\rangle = 0. \quad (7.43c)$$

In the mixed or non-polarized beam all spin directions occur with the same (classical) probability as their opposite directions; the average values  $\langle\bar{s}_z\rangle$ ,  $\langle\bar{s}_x\rangle$ ,  $\langle\bar{s}_y\rangle$  vanish. This is in contrast to quantum mechanical average (expectation) values  $\langle\hat{s}_x\rangle$  and  $\langle\hat{s}_y\rangle$  (7.41b), (7.41c) which are measured on the well prepared quantum state where all possible information is given.

We, thus, must strictly distinguish between systems and their quantum states which are determined according to their maximum quantum mechanical information and those in which pure quantum mechanical states (prepared by a previous measurement) are mixed. The first states which are identical with all quantum states considered so far are called *pure states*. The latter states, for which a principally possible information is lacking but is only given in terms of classical probabilities, are called *mixed states*. A frequently used example for a mixed state is an ensemble in which a first series of measurements of one particular observable has prepared all possible eigenstates of the corresponding measurement operator, but where these eigenstates have not been registered afterwards. Only classical probabilities for their occurrence can, then, be given.

For further clarification, we assume an observable  $A$  described by the operator  $\hat{A}$ . Furthermore, a quantum mechanical system should be in a pure state  $|\psi\rangle$ , which has been prepared by a corresponding measurement set-up. An  $A$  measurement on that pure state, then, yields the quantum mechanical average value:

$$\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle. \quad (7.44)$$

To be more precise, an ensemble is in the pure state  $|\psi\rangle$  and we perform an  $A$  measurement (operator  $\hat{A}$ ). Since the outcome of the measurement is only random for  $|\psi\rangle$  not being an eigenstate of the operator  $\hat{A}$ , we must repeat the same measurement many times on the identical system or we have to perform the  $A$  measurement on many identical systems simultaneously. Note that quantum mechanical randomness requires quite difficult measurement procedures. Each measurement yields one of the many eigenvalues  $A_n$  of the operator  $\hat{A}$ . After these  $A$  measurements the eigenstates  $|n\rangle$  of the operator  $\hat{A}$  are prepared according to

$$\hat{A}|n\rangle = A_n|n\rangle \quad \text{and} \quad |\psi\rangle = \sum_n a_n|n\rangle. \quad (7.45)$$

Hereby,  $|a_n|^2$  is the probability to find the eigenstate  $|n\rangle$  in the mixture of all other eigenstates which are also prepared by the sequence of measurements on the ensemble.  $p_n = |a_n|^2$  is the classical probability for the occurrence of  $|n\rangle$  in the mixture of states.

If we now perform a  $B$  measurement (operator  $\hat{B}$  not commuting with  $\hat{A}$ ) on this mixture of states which are not eigenstates of  $\hat{B}$ , the measurements can only yield quantum mechanical expectation values  $\langle n | \hat{B} | n \rangle$ . But because of the presence of a mixture of  $|n\rangle$  states, each found with a probability  $p_n = |a_n|^2$  the final average over all measurement results amounts to

$$\langle \bar{B} \rangle = \sum_n p_n \langle n | \hat{B} | n \rangle. \quad (7.46)$$

This average value contains the quantum mechanical averages expressed by the matrix elements  $\langle n | \hat{B} | n \rangle$  (necessary since the system was not in eigenstates of  $\hat{B}$ )

and the classical average over the probability distribution  $p_n$  which describes the mixture of pure quantum mechanical eigenstates of the operator  $\hat{A}$  (resulting from previous measurement).

To summarize, so far we have calculated quantum mechanical average (expectation) values as  $\langle \psi | \hat{A} | \psi \rangle$  on an ensemble of  $N$  systems which were all in the same (pure) quantum state  $|\psi\rangle$ . In a mixed state, on the other hand, there is an ensemble of  $N$  systems of which  $n < N$  are in pure states  $|n\rangle$ . Consequently for large numbers,  $p_n = n/N$  is the probability of finding the state  $|n\rangle$  in the mixed ensemble. Therefore, the average describing the result of a  $B$  measurement must be calculated according to (7.46).

Is there now a way of representing an average value of a measurement result disregarding the type of the given system ensemble, be it pure or mixed? We are going to approach this question in the next section.

### 7.3.2 The Density Matrix

To find a common representation of average (expectation) values of an observable both for mixed and pure ensembles, we start with the average value of an observable  $B$  in a mixed ensemble (7.46) and apply the completeness condition (4.69a) for an arbitrary orthonormal eigenfunction system  $|i\rangle$ :

$$\langle \bar{B} \rangle = \sum_n p_n \langle n | \hat{B} | n \rangle = \sum_{ni} p_n \langle n | i \rangle \langle i | \hat{B} | n \rangle = \sum_{ni} p_i \langle n | i \rangle \langle i | \hat{B} | n \rangle. \quad (7.47)$$

We have replaced  $p_n$  by  $p_i$  because of  $\langle n | i \rangle = \delta_{ni}$ . From (7.47), the definition of the *density matrix*, or better the *statistical operator* ( $\hat{\rho}$  is a matrix only in special representations), is derived as follows:

$$\hat{\rho} = \sum_i p_i |i\rangle \langle i|. \quad (7.48)$$

By use of (7.47), the average value (7.47) can be written as

$$\langle \bar{B} \rangle = \sum_n \langle n | \hat{\rho} \hat{B} | n \rangle. \quad (7.49)$$

$\hat{\rho} \hat{B}$  is a product of operators where first  $\hat{B}$  is applied to a ket and subsequently  $\hat{\rho}$ . For a discrete orthogonal basis system  $|n\rangle$ , the expressions  $\langle n | \hat{\rho} \hat{B} | n \rangle$  are the diagonal elements of the matrix  $\hat{\rho} \hat{B}$  (Sect. 4.3.1). The expression (7.49) being the sum of the diagonal elements of a matrix is called *trace* (Tr) of the matrix. A special name is appropriate for this quantity since the trace of Hermitian matrices (operators) does not depend on the particular basis system  $|n\rangle$  in which it is calculated. This is evident

because such matrices can be transformed into diagonal form where the diagonal elements are the eigenvalues of the matrix (Sect. 4.3.1). The eigenvalues, on the other hand, are unequivocally attributed to the matrix and their sum (trace) is thus characteristic for the matrix. The average value (7.49) over a mixed ensemble is therefore written independently on any particular basis set as

$$\langle \bar{B} \rangle = \text{Tr}(\hat{\rho} \hat{B}). \quad (7.50)$$

The expression (7.50) evidently contains the average value of a measurement on a pure state (ensemble)  $A$ , too. In case of a pure state there is only one single state  $|\psi\rangle$  in the ensemble, that is, apart from one special  $p_o = 1$  all other probabilities  $p_n$  vanish. The density matrix of the pure state  $|\psi\rangle$  is obtained as

$$\hat{\rho}_{\text{pure}} = |\psi\rangle\langle\psi|. \quad (7.51)$$

According to (7.50) the average value is calculated in any arbitrary orthogonal basis system  $|n\rangle$  as

$$\begin{aligned} \langle \hat{A} \rangle &= \text{Tr}(\hat{\rho}_{\text{pure}} \hat{A}) = \sum_n \langle n | \hat{\rho}_{\text{pure}} \hat{A} | n \rangle = \sum_n \langle n | \psi \rangle \langle \psi | \hat{A} | n \rangle \\ &= \sum_n \langle \psi | \hat{A} | n \rangle \langle n | \psi \rangle = \langle \psi | \hat{A} | \psi \rangle. \end{aligned} \quad (7.52)$$

As expected the trace of  $\hat{\rho}_{\text{pure}} \hat{A}$  yields the quantum mechanical average of the observable  $A$  in the pure state  $|\psi\rangle$ . The definition of the density matrix (statistical operator) (7.50), respectively (7.51) allows the calculation of average (expectation) values both for mixed and pure system ensembles by means of the same procedure.

The density matrix obviously contains all possible information about a quantum state as does the wave function  $\psi$  or the state vector  $|\psi\rangle$ . The probability to find a particle at a position  $\mathbf{r}$ ,  $P(\mathbf{r}) = \psi^* \psi$ , can be obtained in terms of  $\hat{\rho}$  by means of its diagonal elements in position representation:

$$\langle \mathbf{r} | \hat{\rho} | \mathbf{r} \rangle = \langle \mathbf{r} | \psi \rangle \langle \psi | \mathbf{r} \rangle = \psi^*(\mathbf{r}) \psi(\mathbf{r}). \quad (7.53)$$

Using the projection operator  $\hat{P}_{\mathbf{r}} = |\mathbf{r}\rangle\langle\mathbf{r}|$  which projects on the position state  $|\mathbf{r}\rangle$ , we can again calculate the probability  $P(\mathbf{r})$  by tracing a matrix:

$$\begin{aligned} P(\mathbf{r}) &= \psi^* \psi = \text{Tr}(\hat{P}_{\mathbf{r}} \hat{\rho}) \\ &= \sum_n \langle n | \mathbf{r} \rangle \langle \mathbf{r} | \psi \rangle \langle \psi | n \rangle \\ &= \sum_n \langle \mathbf{r} | \psi \rangle \langle \psi | n \rangle \langle n | \mathbf{r} \rangle = \langle \mathbf{r} | \psi \rangle \langle \psi | \mathbf{r} \rangle; \end{aligned} \quad (7.54a)$$

that is, we can also write:

$$\psi^* \psi = \text{Tr}(\hat{P}_\mathbf{r} \hat{\rho}). \quad (7.54b)$$

By means of the density matrix, we can furthermore derive a criterion for distinguishing between pure and mixed states.

For pure states, the following relation holds:

$$\hat{\rho}_{\text{pure}} = |\psi\rangle\langle\psi|, \quad (7.55a)$$

$$\hat{\rho}_{\text{pure}}^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = \hat{\rho}_{\text{pure}}. \quad (7.55b)$$

The density matrix of a pure ensemble  $\hat{\rho}_{\text{pure}}$  represented in any arbitrary orthogonal basis system and traced yields:

$$\text{Tr } \hat{\rho}_{\text{pure}} = \sum_n \langle n | \psi \rangle \langle \psi | n \rangle = \langle \psi | \psi \rangle = 1, \quad (7.55c)$$

$$\text{Tr } \hat{\rho}_{\text{pure}}^2 = \text{Tr } \hat{\rho}_{\text{pure}} = 1. \quad (7.55d)$$

For mixed ensembles with the density matrix

$$\hat{\rho} = \sum_n p_n |n\rangle\langle n|, \quad (7.56a)$$

where the states  $|n\rangle$  are found with a (classical) probability  $p_n$  the following relations are valid:

$$\begin{aligned} \text{Tr } \hat{\rho} &= \sum_{in} p_n \langle i | n \rangle \langle n | i \rangle = \sum_{in} p_n \langle n | i \rangle \langle i | n \rangle \\ &= \sum_n p_n \langle n | n \rangle = \sum_n p_n = 1, \end{aligned} \quad (7.56b)$$

$$\begin{aligned} \text{Tr } \hat{\rho}^2 &= \text{Tr} \left( \sum_{nm} p_n p_m |n\rangle\langle n|m\rangle\langle m| \right) \\ &= \text{Tr} \left( \sum_{nm} p_n p_m |n\rangle\langle m| \delta_{nm} \right) = \sum_{in} p_n^2 \langle i | n \rangle \langle n | i \rangle \\ \text{Tr } \hat{\rho}^2 &= \sum_n p_n^2 < 1. \end{aligned} \quad (7.56c)$$

While for pure quantum states (ensembles) the traces of both  $\hat{\rho}$  and  $\hat{\rho}^2$  are equal 1 this is true only for  $\text{Tr } \hat{\rho}$  in case of a mixed ensemble. Since for classical probabilities  $p_n < 1$ , we obtain  $\text{Tr } \hat{\rho}^2 < 1$  for the mixed state (7.56c).

Because of the simpleness of the spin system with its two eigenstates  $|\uparrow\rangle$  and  $|\downarrow\rangle$  only the density matrix formalism is easily manageable here. Just for exercise, we assume a general spin state

$$|s\rangle = a|\uparrow\rangle + b|\downarrow\rangle = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}. \quad (7.57)$$

The vector representation has been chosen in basis eigenvectors of the  $\hat{\sigma}_z$  component. In this representation, the density matrix of the pure state (7.57) is written as

$$|s\rangle\langle s| = \begin{pmatrix} a \\ b \end{pmatrix} (a^*b^*) = \begin{pmatrix} aa^* & ab^* \\ ba^* & bb^* \end{pmatrix}. \quad (7.58)$$

Using the rules of matrix multiplication (Sect. 4.3.1), we immediately show that the butterfly symbol  $|s\rangle\langle s|$  indeed represents a matrix. Note the difference between  $|s\rangle\langle s|$  and the scalar product of the two states:

$$\langle s|s\rangle = (a^*b^*) \begin{pmatrix} a \\ b \end{pmatrix} = a^*a + b^*b. \quad (7.59)$$

While expressions of the type  $\langle\psi|\psi\rangle$  always represent scalar numbers, the butterfly symbols  $|\psi\rangle\langle\psi|$ , first a ket and then a bra, denote operators or matrices.

As an exercise we consider the density matrices of the pure spin superposition state (7.40) and that of the mixed state where atom beams with spin-up  $|\uparrow\rangle$  and spin-down  $|\downarrow\rangle$  polarization are mixed in a 50 % ratio. Using the rules for spin state representation and multiplication, we obtain the density matrix of the pure state as

$$\begin{aligned} \hat{\rho}_{\text{pure}} &= \frac{1}{2}(|\uparrow\rangle + e^{i\alpha}|\downarrow\rangle)(\langle\uparrow| + e^{-i\alpha}\langle\downarrow|) \\ &= \frac{1}{2}(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|) + \frac{1}{2}(e^{-i\alpha}|\uparrow\rangle\langle\downarrow| + e^{i\alpha}|\downarrow\rangle\langle\uparrow|) \\ &= \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) + \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) + \frac{1}{2}e^{-i\alpha} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}e^{i\alpha} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & e^{-i\alpha} \\ e^{i\alpha} & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & e^{-i\alpha} \\ e^{i\alpha} & 1 \end{pmatrix}. \end{aligned} \quad (7.60)$$

The density matrix of the mixed state, on the other hand, results as

$$\hat{\rho} = p_{\uparrow}|\uparrow\rangle\langle\uparrow| + p_{\downarrow}|\downarrow\rangle\langle\downarrow| = \frac{1}{2}(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|) = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7.61)$$

While in the pure state (7.60) the non-diagonal elements  $\exp(\pm i\alpha)$  in the density matrix contain the phase information of the quantum interference, these non-diagonal interference terms are lost in the density matrix of the mixed state (7.61). The density matrix of a quantum state clearly reflects the difference between pure and mixed ensembles. In a mixed ensemble, there are no interference terms which are characteristic for coherent quantum states. Mixed systems behave as classical systems.

The transition from a pure superposition state into a mixed (classical) state is therefore connected with the loss of non-diagonal elements in the density matrix. In

a mixed state probabilities are added, while in a pure state probability amplitudes are superimposed; in the absolute square of these amplitudes interference terms appear as non-diagonal elements in the density matrix.

Because of

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\pm i\alpha} d\alpha = 0 \quad (7.62)$$

we can describe this behavior also by saying that the transition from a pure into a mixed quantum state is related to averaging the phases of the interference terms in the density matrix:

$$\hat{\rho} = \frac{1}{2\pi} \int_0^{2\pi} d\alpha \hat{\rho}_{\text{pure}}. \quad (7.63)$$

## 7.4 Quantum Environment, Measurement Process and Entanglement

In Sect. 7.2.1, we have learnt that quantum physical reality is non-local. Everything is linked with everything, or more precisely entangled. In physical experiments, on the other hand, we want to study a particular distinct subsystem of the total reality. When we are interested in the excited electronic states of an atom, we want to exclude from our consideration interactions with an eventual solid state matrix or with electromagnetic fields in an ion trap. It is essential in physical sciences that we separate systems under study from their surrounding and from the measurement set-up for reasons of simpler abstract mathematical description. This is the heart of physical abstraction, which has led to the incomparable advances in physical thinking. Even in the physical measurement process, which is described several times in this book as a collapse of the wave packet into an eigenstate of the measurement operator, the interaction, even the existence, of the measurement set-up is excluded from the description.

In the following, we are going to approach the question, what consequences the exclusion of the physical environment (also the measurement apparatus) has on the subsystem under study. This subsystem embedded in its physical environment is linked by entanglement with this environment because of the non-locality of the quantum world. The exclusion of the environment upon the measurement process, thus, concerns a subtle question about the basis of perception of nature.

### 7.4.1 Subsystem and Environment

To elucidate the basis of physical abstraction, that is, the restriction on a mathematically treatable subsystem of nature, we consider a system which is composed of two subsystems (1) and (2).

A quantum state of the total system is described by a state vector

$$|\psi\rangle = \sum_{nm} c_{nm} |n\rangle_1 |m\rangle_2. \quad (7.64)$$

$|n\rangle_1$  and  $|n\rangle_2$  are orthonormal eigenvector systems of the two subsystems which are mutually linked by the matrix  $\{c_{nm}\}$ . This matrix causes entanglement of the two subsystems with each other, because  $|\psi\rangle$  cannot be represented as a product of state vectors of the two systems (1) and (2). Only a representation  $c_{nm} = \alpha_n \beta_m$  would allow such a separation of the two subsystems, written as

$$|\psi\rangle = \sum_{nm} \alpha_n \beta_m |n\rangle_1 |m\rangle_2 = \left( \sum_{nm} \alpha_n |n\rangle_1 \right) \left( \sum_{nm} \beta_m |m\rangle_2 \right). \quad (7.65)$$

Equation (7.65) describes two decoupled systems (Sect. 5.6.1). We return to the coupled total system (7.64) and calculate the density matrix of this pure ensemble:

$$\hat{\rho} = |\psi\rangle\langle\psi| = \sum_{\substack{nm \\ n'm'}} c_{nm} c_{n'm'}^* |n\rangle_1 |m\rangle_2 \langle n'|_2 \langle m'|. \quad (7.66)$$

We now consider system (2) as the environment to be neglected, into which the subsystem (1) is embedded. Subsystem (1) is the system under study on which an observable  $A$  shall be measured. According to the rules in Sect. 7.3.2 we calculate the expectation value  $\langle A \rangle$  by tracing the operator  $\hat{\rho}\hat{A}$ . Because of the limitation on system (1) (only this system is concerned by the measurement), the trace calculation with  $\hat{A}$  is performed only in system (1):

$$\langle A \rangle = \text{Tr}_1 \text{Tr}_2 (\hat{\rho}\hat{A}) = \text{Tr}_1 [(\text{Tr}_2 \hat{\rho}) \hat{A}]. \quad (7.67)$$

$\text{Tr}_2 \hat{\rho}$  is calculated by taking into account eigenstates of subsystem (2) only. By this procedure, the total density matrix  $\hat{\rho}$  is reduced to a so-called *reduced density matrix*  $\hat{\rho}_{\text{red}}$  which is relevant for the calculation of the expectation value  $\langle A \rangle$ . More details about the motivation of the reduced density matrix are given in Appendix C.

The trace calculation in subsystem (2) yields the reduced density matrix:

$$\begin{aligned} \hat{\rho}_{\text{red}} &= \text{Tr}_2 \hat{\rho} = \sum_{\substack{inm \\ n'm'}} c_{nm} c_{n'm'}^* |n\rangle_1 \langle n'|_2 \langle m'|_2 \langle i|m\rangle_2 \\ &= \sum_{\substack{inm \\ n'm'}} c_{nm} c_{n'm'}^* |n\rangle_1 \langle n'|_2 \langle m'|_2 \langle i|m\rangle_2 \\ &= \sum_{nn'm} c_{nm} c_{n'm}^* |n\rangle_1 \langle n'|. \end{aligned} \quad (7.68a)$$

With

$$p_{nn'} = \sum_m c_{nm} c_{n'm}^*, \quad (7.68b)$$

the reduced density matrix is written as

$$\hat{\rho}_{\text{red}} = \sum_{nn'} p_{nn'} |n\rangle_1 \langle n'|. \quad (7.68c)$$

The comparison with the density matrix of a mixed state (7.48), respectively (7.56a), reveals a similar structure. Entanglement of the system under study (1) with subsystem (2), the environment, obviously leads to an expectation value (average)  $\langle A \rangle$  for a measurement on system (1), where the relevant density matrix  $\hat{\rho}_{\text{red}}$  for the calculation of  $\langle A \rangle$  contains the more general ket-bra combination  $|n\rangle\langle n'|$  rather than  $|n\rangle\langle n|$ . In analogy we argue that the matrix elements  $p_{nn'}$  are classical probabilities which are due to neglecting information about subsystem (2), the environment.

To check this assumption, we calculate the average value  $\langle A \rangle$  (7.67), that is, the expectation value of an  $A$  measurement, under the condition that the environment [subsystem (2)] is excluded from the consideration:

$$\begin{aligned} \langle A \rangle &= \text{Tr}_1(\hat{\rho}_{\text{red}} \hat{A}) = \sum_{\substack{i \\ inn' \\ m}} c_{nm} c_{n'm}^* \langle i | n \rangle_1 \langle n' | \hat{A} | i \rangle \\ &= \sum_{\substack{i \\ inn' \\ n'}} c_{nm} c_{n'm}^* \langle n' | \hat{A} | i \rangle \langle i | n \rangle_1 \\ &= \sum_{nn'm} c_{nm} c_{n'm}^* \langle n' | \hat{A} | n \rangle_1 = \sum_{nn'} p_{nn'} \langle n' | \hat{A} | n \rangle_1. \end{aligned} \quad (7.69)$$

$\langle A \rangle$  is indeed a sum of quantum mechanical matrix elements  $\langle n' | \hat{A} | n \rangle_1$  which contains also non-diagonal elements in contrast to an average in a pure ensemble.  $p_{nn'}$  are, thus, probabilities for the occurrence of  $\langle n' | \hat{A} | n \rangle_1$  as results of an  $A$  measurement.

This result (7.69) looks like an average value in a mixed ensemble, where the mixed ensemble originates from the exclusion of the environment in the  $A$  measurement.

To prove this issue, we calculate  $\text{Tr}_1 \hat{\rho}_{\text{red}}^2$  because the criterion  $\text{Tr}_1 \hat{\rho}_{\text{red}}^2 < 1$  doubtlessly shows the presence of a mixed ensemble.

$$\begin{aligned} \text{Tr}_1 \hat{\rho}_{\text{red}}^2 &= \text{Tr}_1 \sum_{\substack{nn' \\ vv'}} p_{nn'} p_{vv'} |n\rangle_1 \langle n'| \langle v' | v \rangle_1 \\ &= \sum_{\substack{inn' \\ vv'}} p_{nn'} p_{vv'} \langle i | n \rangle_1 \langle n' | v \rangle_1 \langle v' | i \rangle \\ &= \sum_{\substack{inn' \\ vv'}} p_{nn'} p_{vv'} \langle v' | i \rangle \langle i | n \rangle_1 \langle n' | v \rangle_1 \\ &= \sum_{\substack{nn' \\ vv'}} p_{nn'} p_{vv'} \delta_{v'n} \delta_{n'v} = \sum_{nn'} p_{nn'} p_{n'n} < 1. \end{aligned} \quad (7.70)$$

The last conclusion in (7.70) is based on the interpretation of  $p_{nn'}$  as classical probabilities, that is, the validity of  $\sum_{nn'} p_{nn'} = 1$ . From this relation we infer  $p_{nn'} < 1$  and therefore also  $\sum_{nn'} p_{nn'} p_{n'n} < 1$ .

From (7.70) it is evident that the neglect of the environment (subsystem 2) upon the  $A$  measurement on the subsystem (1) forces the system into a mixed state. In real experiments, the described situation is mostly given. Commonly we abstract the environment from the special physical system on which a measurement is performed.

In literature, a trendy term is used for this issue. A quantum system which is coupled to further degrees of freedom in its environment (not considered in the special context) is called an *open quantum system*. The total system, the one considered in the special experiment and the environment together, is called a *closed quantum system*.

In most cases, quantum systems studied in an experiment are coupled to an environment which is excluded from the consideration. These systems under study are therefore open systems and mixed quantum states are commonly encountered in experiments.

### 7.4.2 Open Quantum Systems, Decoherence and Measurement Process

According to what we have learnt about open and closed quantum systems, we expect that in an open system coherent superposition states can exist only for a finite limited time. They will decay by coupling (entanglement) to the environment which is mostly excluded from the theoretical description. This *decoherence* process shall be considered a little bit more in detail in the following. For this purpose, we assume a 2-level atom with the two electronic states  $|g\rangle$  and  $|e\rangle$  as in Sect. 7.2. This 2-level atom shall be embedded into an environment  $E$ , for example, an electromagnetic field. The environment is globally described by a complex quantum state  $|E\rangle$ , which might contain a high number of degrees of freedom. The time evolution of this state is described by a unitary transformation (Sect. 4.3.5).

After some time  $t$  the time evolution of the states, then, can be written as:

$$|g\rangle|E\rangle \xrightarrow{\hat{U}(t)} |g\rangle|E_0(t)\rangle, \quad (7.71a)$$

$$|e\rangle|E\rangle \xrightarrow{\hat{U}(t)} |e\rangle|E_1(t)\rangle. \quad (7.71b)$$

We have assumed that at the beginning the environment is in the state  $|E\rangle$  irrespective of the state of the 2-level atom. Depending on coupling to the ground state  $|g\rangle$  or to the excited state  $|e\rangle$  the unitary operator  $\hat{U}(t)$  transforms the environment state into  $|E_0(t)\rangle$  or  $|E_1(t)\rangle$ . In this simple treatment, the atomic states  $|g\rangle$  and  $|e\rangle$  as eigenstates of the considered open quantum system shall not change during the time evolution interval.

We now assume that the atom is in a superposition state of  $|g\rangle$  and  $|e\rangle$  at the beginning. This superposition state might be generated by irradiation of a  $\pi/2$  pulse of electromagnetic radiation having a frequency  $(E_e - E_g)/\hbar = \omega$  (Sect. 6.5.1). The time evolution of the superposition state  $\alpha|g\rangle + \beta|e\rangle$  ( $\alpha = \beta$  for exact  $\pi/2$  pulse) is then obtained as:

$$(\alpha|g\rangle + \beta|e\rangle)|E\rangle \xrightarrow{\hat{U}(t)} \alpha|g\rangle|E_0\rangle + \beta|e\rangle|E_1\rangle. \quad (7.72)$$

The resulting state after time  $t$  is again a superposition state, but now the states  $|g\rangle$  and  $|e\rangle$  of the open quantum system (2-level atom) are entangled with the environment states  $|E_0\rangle$  and  $|E_1\rangle$  (Sect. 7.2). The atomic states can no longer be considered as decoupled from the environment states.

In the following, this issue shall be considered by means of the density matrix formalism. The density matrix of the entangled state (7.72) of the closed system (2-level atom plus environment) reads as

$$\begin{aligned} \hat{\rho} &= (\alpha|g\rangle|E_0\rangle + \beta|e\rangle|E_1\rangle)(\alpha^*|g\rangle\langle E_0| + \beta^*|e\rangle\langle E_1|) \\ &= \alpha\alpha^*|g\rangle\langle g||E_0\rangle\langle E_0| + \beta\beta^*|e\rangle\langle e||E_1\rangle\langle E_1| \\ &\quad + \alpha\beta^*|g\rangle\langle e||E_0\rangle\langle E_1| + \beta\alpha^*|e\rangle\langle g||E_1\rangle\langle E_0|. \end{aligned} \quad (7.73)$$

We are not interested in the environment but rather in the 2-level atom which is an open quantum system in this context. The relevant density matrix is thus the reduced density matrix  $\hat{\rho}_{\text{red}}$  which is obtained from (7.73) by tracing over the environment states (Sect. 7.4.1). For the calculation of the trace, we represent the environment states  $|E_0\rangle$  and  $|E_1\rangle$  in terms of an arbitrary orthonormal basis system  $|i\rangle$ :

$$|E_0\rangle = \sum_i c_{0i}|i\rangle, \quad (7.74a)$$

$$|E_1\rangle = \sum_i c_{1i}|i\rangle. \quad (7.74b)$$

A short remark concerning the environment states: In general  $|E_0\rangle$  and  $|E_1\rangle$  are not mutually orthogonal. At least for small time intervals after preparation of the superposition state (7.72) they might be similar with much overlap. The calculation of  $\hat{\rho}_{\text{red}}$  is performed by tracing the environment states in the basis system  $|i\rangle$ :

$$\begin{aligned}
\hat{\rho}_{\text{red}} &= \text{Tr}_E \hat{\rho} \\
&= \alpha\alpha^*|g\rangle\langle g| \sum_i \langle i|E_0\rangle\langle E_0|i\rangle \\
&\quad + \beta\beta^*|e\rangle\langle e| \sum_i \langle i|E_1\rangle\langle E_1|i\rangle \\
&\quad + \alpha\beta^*|g\rangle\langle e| \sum_i \langle i|E_0\rangle\langle E_1|i\rangle \\
&\quad + \beta\alpha^*|e\rangle\langle g| \sum_i \langle i|E_1\rangle\langle E_0|i\rangle.
\end{aligned} \tag{7.75}$$

For the system  $|i\rangle$ , the completeness condition is used as

$$\sum_i \langle E_0|i\rangle\langle i|E_1\rangle = \langle E_0|E_1\rangle \tag{7.76}$$

and we obtain from (7.75):

$$\begin{aligned}
\hat{\rho}_{\text{red}} &= \text{Tr}_E \hat{\rho} \\
&= \alpha\alpha^*|g\rangle\langle g| + \beta\beta^*|e\rangle\langle e| \\
&\quad + \alpha\beta^*|g\rangle\langle e|\langle E_1|E_0\rangle \\
&\quad + \beta\alpha^*|e\rangle\langle g|\langle E_0|E_1\rangle.
\end{aligned} \tag{7.77}$$

Now the Hilbert vectors  $|g\rangle$  and  $|e\rangle$  are expressed as spinors in the common 2D vector space:

$$|g\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |e\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7.78}$$

The butterfly operators in (7.77), then, read as

$$|g\rangle\langle g| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.79a}$$

$$|g\rangle\langle e| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (0 \ 1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{7.79b}$$

Similar expressions are obtained for  $|e\rangle\langle e|$  and for  $|e\rangle\langle g|$ , 2D matrices with one on the lower diagonal and one left on the bottom (non-diagonal), respectively.

The reduced density matrix  $\hat{\rho}_{\text{red}}$  of the open 2-level system in matrix representation is finally obtained as

$$\hat{\rho}_{\text{red}} = \text{Tr}_E \hat{\rho} = \begin{pmatrix} |\alpha|^2 & \alpha\beta^*\langle E_1|E_0\rangle \\ \beta\alpha^*\langle E_0|E_1\rangle & |\beta|^2 \end{pmatrix}. \tag{7.80}$$

As long as the environment states  $|E_0\rangle$  and  $|E_1\rangle$  are not mutually orthogonal the density matrix of the open system of the 2-level atom has non-diagonal elements. There is still a coherent superposition state containing the states  $|g\rangle$  and  $|e\rangle$ . We remind the spin density matrices (7.60) and (7.61) where the pure state is characterized by a density matrix with phase factors  $\exp(\pm i\alpha)$  on the non-diagonal (7.60) while the mixed state is represented by a density matrix without any non-diagonal elements (7.61). In the spin case the diagonal elements, both equal  $1/2$ , are the probabilities for finding the two spin orientations in the mixed state.

In analogy, the diagonal elements  $|\alpha|^2$  and  $|\beta|^2$  in (7.80) represent the probabilities for finding the states  $|g\rangle$  and  $|e\rangle$  if a mixed state develops by loosing the non-diagonal elements in (7.80) (orthogonalization of environment states). On the other hand, the non-diagonal elements  $\alpha\beta^*\langle E_1|E_0\rangle$  and  $\beta\alpha^*\langle E_0|E_1\rangle$  in (7.80) describe phase relations between the two parts in the coherent superposition state (7.72).

While at the beginning after preparation of the superposition state (7.72) one and the same environment state  $|E\rangle$  is given, this state splits and develops more and more into two different states  $|E_0\rangle$  and  $|E_1\rangle$  depending on the interaction of the environment with the ground state  $|g\rangle$  or the excited state  $|e\rangle$  of the 2-level atom. During the time evolution, the states  $|E_0\rangle$  and  $|E_1\rangle$  gradually become more and more dissimilar with less overlap. Finally, the overlap between the two environment states is lost and they are mutually orthogonal.

We have to assume a time evolution of the type

$$\langle E|E\rangle = 1 \xrightarrow{U(t)} \langle E_0|E_1\rangle = 0 \quad (7.81)$$

which might approximately be described by an exponential decay of the scalar product of the two environment states:

$$\langle E_0(t)|E_1(t)\rangle = e^{-\gamma t}. \quad (7.82)$$

In a model calculation, Palme et al. [15] and Unruh [16] have coupled a 2-level system to an environment ensemble of harmonic oscillators and have indeed found an exponential decay of the phase factors (7.82) in the density matrix of the open 2-level system.

This process where a pure superposition state changes into a mixed state by entanglement with the environment is called *dephasing*. Dephasing results in decoherence and the characteristic time  $\tau = 1/\gamma$  in (7.82) during which coherence is destroyed is called *dephasing time*. According to our present knowledge, this time is very short in most cases, for example, in the order of  $10^{-12}$  s for electron hole excitations in semiconductors [17]. In other cases of weak coupling to the environment, as for example, nuclear spins to the electronic shell of paramagnetic atoms (Sect. 6.5.3), the dephasing time can reach values up to  $10^{-4}$  s.

One must strictly distinguish the dephasing time from the time constant during which the diagonal matrix elements of the density matrix (7.80) decay in particular processes. This time is relevant for de-excitation of thermally excited systems into

their ground state. We assume, for example, a 2-level system being excited into its state  $|e\rangle$ . Due to coupling to a thermal bath, it returns into its ground state  $|g\rangle$  within this time interval. Spin systems build up a magnetic spin polarization in an external magnetic field, which decays into thermal equilibrium within some time interval after switching off the field. The underlying reason is thermal coupling to the environment. For this kind of decay processes being related to thermal energy transfer to the environment, the relevant decay time is frequently called  $T_1$ , while the dephasing time is usually called  $T_2$ . This time  $T_2$ , the same as the time  $\tau$  defined above, is related to the decay of a coherent superposition state, that is, the typically quantum mechanical interferences (non-diagonal elements of the density matrix). For certain spin systems, Golovach et al. [18] have estimated the dephasing time  $T_2$  to be double as high as the decay time  $T_1$  for thermal de-excitation.

The 2-level system (atom) coupled to its environment, which is discussed so far, might also be considered as a paradigm of the quantum mechanical measurement process. We assume that a measurement is performed for finding out the energy eigenstate of a 2-level system which is in the superposition state of  $|g\rangle$  and  $|e\rangle$ . For this purpose, a macroscopic measurement apparatus is used which finally indicates the measurement result, via complex intermediate mechanisms, by a pointer orientation. Corresponding to the presence of  $|g\rangle$  or  $|e\rangle$  the pointer should point upwards or downwards. We can, then, identify the whole measurement set-up with the environment considered so far. The two environment states  $|E_0\rangle$  and  $|E_1\rangle$  are attributed in this case to the two states of the measurement set-up which result in the two pointer orientations, upwards and downwards. In the measurement process we abstract the 2-level system from its environment, the measurement apparatus. We are only interested in properties of the 2-level system, whether it is in state  $|g\rangle$  or state  $|e\rangle$ , rather than in properties of the measurement set-up. More precisely: we are interested in the probabilities of finding  $|g\rangle$  or  $|e\rangle$  in the measurement.

In analogy to the above discussion, we can describe the measurement on the general superposition state  $|\psi\rangle = \alpha|g\rangle + \beta|e\rangle$  of the 2-level system in terms of entanglement of the two states  $|g\rangle$  and  $|e\rangle$  with the two states of the measurement set-up  $|E_0\rangle$  and  $|E_1\rangle$  (7.72). Within the dephasing time  $\tau (= T_2)$  the phase factors (non-diagonal elements) in the reduced density matrix  $\hat{\rho}_{\text{red}}$  (7.80) decay to zero. The result of the measurement is a mixed state in which the two states  $|g\rangle$  and  $|e\rangle$  of the 2-level system are registered with the probabilities  $|\alpha|^2$  and  $|\beta|^2$ . The disappearance of the two scalar products  $\langle E_1 | E_0 \rangle$ , respectively  $\langle E_0 | E_1 \rangle$  after dephasing guarantees the unequivocal read-out of the measurement result by means of the two pointer orientations, downwards or upwards. These two states are mutually orthogonal.

This is the solution to the puzzle of the collapse of the wave packet in early quantum mechanics (Sect. 4.1). The measurement process destroys a previously prepared quantum state  $|\psi\rangle$  to force the system under study into one of the eigenstates of the operator of the measurement observable. As we have seen here, the underlying reason is the entanglement of the given quantum state with states of the measurement set-up.

### 7.4.3 Schrödinger's Cat

In the early days of quantum mechanics (1935) Schrödinger, one of the founders of this theory, has described the totally counter-intuitive aspects of a quantum mechanical measurement in terms of a gedanken experiment [2, 3]. The apparent paradox in this gedanken experiment, well known under the name “Schrödinger's Cat” is even today frequently used to demonstrate the odd and counter-intuitive property of entanglement along with the transition of a quantum system to macroscopic behavior.

In Schrödinger's own words, the cat paradox [2, 3] reads as follows (translation from the German according to [19]):

A cat is penned up in a steel chamber, along with the following diabolical device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, so small, that perhaps in the course of one hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives if meanwhile no atom has decayed. The first atomic decay would have poisoned it. The  $\psi$ -function of the entire system would express this by having in it the living and the dead cat mixed or smeared out in equal parts.

If there is no observer, we would be confronted with the odd situation of a superposition state of a living and a dead cat, an idea which contradicts every perception of reality. By means of the cat paradigm Schrödinger wanted to demonstrate the apparent inconsistency of the quantum mechanical measurement process. We can easily see how entanglement solves the apparent contradiction in the observation of the cat being simultaneously dead and alive.

For the solution to the problem we must analyze the entire (closed) system cat plus killing device, i.e. unstable atom, relay, bottle etc. rather than only the open system cat. The atomic state of the unstable atom which initiates the cat killing relay is described by  $|1\rangle$  in case of atom decay and by  $|0\rangle$  for the non-decaying atom. The most general state of the atom, then, is the superposition state  $(|1\rangle + |0\rangle)/\sqrt{2}$ . This state interacts via the killing device with the cat. The simplest description of the entire system, therefore, must at least take into account entanglement of the atomic state with that of the cat:

$$|1\rangle|\text{cat-dead}\rangle + |0\rangle|\text{cat-alive}\rangle. \quad (7.83)$$

This representation of the entire system is analogous to that of a 2-level system embedded in a macroscopic environment (7.72). In the present case, all macroscopic objects as the cat and the killing device with their enormous number of degrees of freedom are transformed by their interaction with the unstable atom into final states with completely different wave functions without any overlap. A glimpse at the reduced density matrix (7.80) and the process of dephasing (7.82), that is, the establishment of decoherence, reveals the destruction of the coherent superposition state (7.83) after the dephasing time  $\tau$ . There is no superposition of a dead and a living cat anymore, the cat is alive or dead.

By means of the concepts of entanglement and dephasing of superposition states, the old puzzle of Schrödinger's cat is solved.

## 7.5 Superposition States for Quantum-Bits and Quantum Computing

In Sect. 7.4.2, we have studied the phenomenon of decoherence of a quantum mechanical superposition state. As an example, a 2-level atom with the states  $|g\rangle$  and  $|e\rangle$  coupled to its environment has been considered. By irradiation of an electromagnetic high-frequency  $\pi/2$  pulse with the photon energy  $\hbar\omega = E_e - E_g$ , the superposition state

$$|\psi\rangle = \alpha|g\rangle + \beta|e\rangle \quad (7.84)$$

is prepared (Sect. 6.5.1). This is the most general state of the 2-level system as long as no measurement has been performed. By means of the infinite variety of probability amplitudes  $\alpha$  and  $\beta$  with  $|\alpha|^2 + |\beta|^2 = 1$ , this superposition state can carry an infinite amount of information in terms of the two complex numbers  $\alpha$  and  $\beta$ . A unequivocal operation without dephasing, that is, a unitary transformation of the state (7.84) which results in a new superposition state, thus, allows an extremely parallel information processing. Ideas being based on this concept are at the basis of the currently highly attractive field of quantum information and quantum computing [19, 20].

In a classical computer, digital information processing is based on operations with two well defined, distinct electrical states (voltage, current, resistance, ...), which represent the values 1 and 0 in the dual number system. This smallest unit of information is called a *bit*. In analogy the superposition state (7.84) of a quantum mechanical 2-level system is called a *quantum bit* (Q-bit).

For processing the highly parallel quantum information contained in the superposition state of a Q-bit operations are necessary, which transform a Q-bit into another one without loss of information. The new transformed probability amplitudes must unequivocally be linked to the previous ones. The operations must not cause dephasing of the superposition state (Q-bit), because the phase of the wave function ( $\alpha$  and  $\beta$  are complex) is an essential part of the information content. This requirement is only fulfilled by unitary transformations (Sect. 4.3.5).

In analogy to classical computer gates, which transform one bit into another one, we use the term *quantum gate* for a device, which transforms one Q-bit into another one obeying the rules of a unitary transformation. Although a quantum gate deterministically transforms one Q-bit into another one, the information contained in the probability amplitudes  $\alpha$  and  $\beta$  (7.84) is probabilistic. This is important to keep in mind, when we consider the input and the read-out of a quantum computer. In contrast, a classical computer (Turing machine) operates fully deterministically, both as far as the read-in and the read-out as well as the processing of information are concerned.

As was discussed in Sect. 7.4.2, Q-bits physically realized by 2-level systems always interact with their environment, that is, dephasing of the Q-bit superposition state can not fully be suppressed, but only diminished to a certain extent. For a quantum computer, one must require that the 2-level systems for realization of the Q-bits guarantee dephasing times small in comparison with the computing time for a quantum algorithm [20].

Experimental research in the field of quantum computing at this early stage predominantly aims at the realization of 2-level systems with sufficiently long dephasing times of superposition states which realize the Q-bits. As far as solid state research is concerned in this field, the crucial point is the interaction of the particular 2-level system (spins, supercurrents etc.) with the solid state matrix by means of lattice vibrations, thermal fluctuations, electromagnetic coupling etc.

### 7.5.1 Coupled Quantum Dots as Quantum-Bits

Quantum-bits and quantum computing are physically realized by 2-level systems. In a quantum computer, of course, many 2-level systems must be coupled with each other to enable the running of complex quantum algorithms. All possible 2-level systems are suited if they exhibit sufficiently long dephasing times. Because of their long dephasing times nuclear spin based systems and coherent superpositions of two supercurrents with opposite direction in a current loop are particularly interesting in the solid state physics field.

With respect to the realization of complex quantum computers with many coupled 2-level systems preferentially those systems are of interest which can be integrated as nanostructures on a solid state chip analogously to present semiconductor nano(micro)-electronics. In this respect, Q-bits realized by coupled quantum dots (Sects. 6.2.3 and 6.5.1) are of particular interest. For quantum dots an established and sophisticated integration technique has been developed in semiconductor nano-electronics (Appendix B).

For the following discussion of coupled quantum dots as Q-bits, we refer to the results of Sect. 6.2.3, where two quantum dots, left and right, with the ground states  $|L\rangle$  and  $|R\rangle$  were considered. For identical potential wells left and right, the energies of the two ground states are equal:  $E_L = E_R = E_0$ . For sufficiently close quantum dots, the ground states  $|L\rangle$  and  $|R\rangle$  overlap. The interaction between the two states, or in other words electronic tunneling between the states, creates a 2-level system with the energies  $E_+ = E_0 + t_{LR}$  and  $E_- = E_0 - t_{LR}$  with  $t_{LR} = \langle L | \hat{V}_L | R \rangle = \langle R | \hat{V}_R | L \rangle$  as the interaction matrix element. It contains the effect of the potential parts  $\hat{V}_L$ , respectively  $\hat{V}_R$  of the left and the right quantum dot on the ground states of the respectively neighbouring quantum dot.

This 2-level system described in Sect. 6.2.3 with its energies  $E_+$  and  $E_-$  (identical with  $E_e$  and  $E_g$ , respectively  $|e\rangle$  and  $|g\rangle$ ) as the corresponding states in Sect. 6.2.3 exists in the described form as long as the two potential wells are equivalent. If we apply an electric field  $\mathcal{E}$  parallel to the connection axis of the two dots, an electron

becomes more localized in the left or in the right dot depending on the field direction. At complete localization we prepare the state  $|L\rangle$  or  $|R\rangle$  (Fig. 7.12). The superposition states (6.57a), (6.57b)

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|L\rangle \pm |R\rangle), \quad (7.85)$$

characteristic for a delocalized electron in two equivalent dot potentials do no longer exist. A sufficiently strong electric field causes an asymmetric two-dot potential and destroys the 2-level system of the Q-bit. By means of an electric field we, thus, can switch off and on a Q-bit, an effect which is of interest for the experimental study of such systems (Sect. 7.5.2).

For the calculation of the energy eigenvalues of the Schrödinger equation, we describe the action of the electric field  $\mathcal{E}$  by adding an electric dipole contribution  $\pm\mu\mathcal{E}$  resulting from the electron localization to the ground state energy  $E_0$  of the unperturbed 2-level system (Fig. 7.12b, d).

As in Sect. 6.2.3, we insert the general superposition state of the 2-level system  $|\psi\rangle = c_L|L\rangle + c_R|R\rangle$  into the Schrödinger equation of the coupled quantum dots

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = \hat{H}|\psi\rangle. \quad (7.86)$$

Apart from the unperturbed Hamilton operator  $H_0$  of the two separate quantum dots the Hamiltonian  $\hat{H}$  contains an interaction term  $\hat{h}$  which describes the influence of the neighbouring dot potentials on the ground states  $|L\rangle$  and  $|R\rangle$ . According to Sect. 6.2.3, this small interaction potential induces the matrix elements  $H_{LR} = H_{RL} \approx t_{LR}$ , which describe the tunneling amplitude of an electron between the two dots through the potential barrier in between (Fig. 7.12a).

Without an external electric field (7.86), then, reads:

$$i\hbar\frac{\partial}{\partial t}(c_L|L\rangle + c_R|R\rangle) = (\hat{H}_0 + \hat{h})(c_L|L\rangle + c_R|R\rangle). \quad (7.87)$$

Using the approximation  $S \approx 0$ , which was already made in Sect. 6.2.3 (6.58a), (6.58b), we can write:

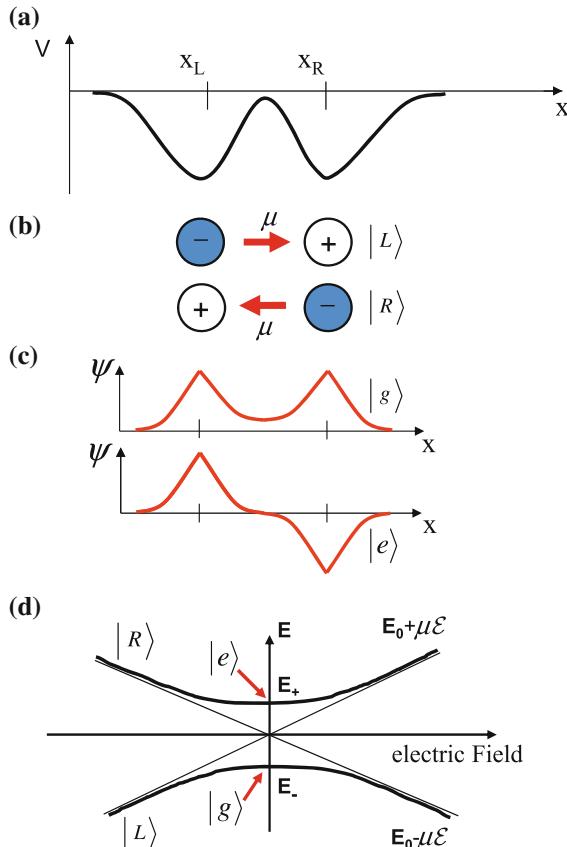
$$\langle L|\hat{H}_0 + \hat{h}|L\rangle = \langle R|\hat{H}_0 + \hat{h}|R\rangle \approx \langle L|\hat{H}_0|L\rangle = \langle R|\hat{H}_0|R\rangle = E_0, \quad (7.88a)$$

$$\begin{aligned} \langle L|\hat{H}_0 + \hat{h}|R\rangle &= \langle R|\hat{H}_0 + \hat{h}|L\rangle = E_0\langle L|R\rangle + \langle L|\hat{h}|R\rangle \\ &\approx H_{LR} = H_{RL} = t_{LR}. \end{aligned} \quad (7.88b)$$

By multiplying from left with the bras  $\langle L|$  and  $\langle R|$ , respectively, we obtain:

$$i\hbar\frac{\partial}{\partial t}c_L = E_0c_L + t_{LR}c_R, \quad (7.89a)$$

$$i\hbar\frac{\partial}{\partial t}c_R = t_{LR}c_L + E_0c_R. \quad (7.89b)$$



**Fig. 7.12 a–d** Manipulation of a quantum bit (Q-bit) which is realized by an electron in two coupled quantum dots. **a** Two binding potentials (qualitatively) of quantum dots at  $X_L$  and  $X_R$ . **b** By an applied electric field  $\mathcal{E}$  the electron can be localized in the left quantum dot  $|L\rangle$  or in the right dot  $|R\rangle$ . This localization generates an electric dipole  $\mu$ , the direction of which depends on the electron position. **c** For symmetrical quantum dots without electric field the ground state energy levels of the quantum dots split into a bonding and an anti-bonding energy level  $E_-$  and  $E_+$  belonging to the states  $|g\rangle$  and  $|e\rangle$ . This 2-level system represents the Q-bit (see also Fig. 6.4). **d** Electron energy in the two coupled quantum dots as function of an externally applied electric field  $\mathcal{E}$ . *Straight thin lines* correspond to the case of two uncoupled quantum dots, while the *curves in thick solid line* describe the situation of the coupled dots. Without electric field the Q-bit is realized by the quantum states  $|g\rangle$  and  $|e\rangle$  with the energies  $E_-$  and  $E_+$ . The electron is delocalized between  $|g\rangle$  and  $|e\rangle$ . For increasing electrical field strength this superposition state of  $|g\rangle$  and  $|e\rangle$  (Q-bit) decays and the electron becomes localized whether in  $|L\rangle$  or in  $|R\rangle$ .

Taking into account the harmonic time dependences  $c_L \propto \exp(-i\omega t)$  and  $c_R \propto \exp(-i\omega t)$  of energy eigenstates (7.89a), (7.89b) finally yields

$$\hbar\omega \begin{pmatrix} c_L \\ c_R \end{pmatrix} = \begin{pmatrix} E_0 & t_{LR} \\ t_{LR} & E_0 \end{pmatrix} \begin{pmatrix} c_L \\ c_R \end{pmatrix}. \quad (7.90)$$

This equation is solved for  $\hbar\omega$  by setting the matrix determinant to zero. This yields the two eigenvalues

$$\hbar\omega = E_0 \pm t_{LR} = E_{\mp} \quad (7.91)$$

which represent the energies of the bonding and antibonding eigenstates  $|g\rangle$  and  $|e\rangle$  of the 2-level system of the two coupled quantum dots. Since  $t_{LR}$  has a negative sign,  $E_-$  belongs to the bonding ground state  $|g\rangle$  and  $E_+$  to the antibonding excited state  $|e\rangle$ . As in Sect. 6.2.3 (6.58c), the energetic distance between the two eigenstates of the coupled quantum dots amounts to  $2|t_{LR}|$ .

We keep in mind that the states  $|g\rangle$  and  $|e\rangle$  are eigenstates of the coupled quantum dots with the quantum energies  $E_{\mp}$  (7.89a), (7.89b). In contrast, the states  $|L\rangle$  and  $|R\rangle$  are not eigenstates. The occupation probability amplitudes  $c_L$  and  $c_R$  of these states oscillate in time; the electron might be imagined as tunneling back and forth between the two dots. This picture is obvious from solving (7.89a), (7.89b) by an alternative method: One adds and subtracts (7.88a) and (7.88b) from each other and obtains:

$$i\hbar \frac{\partial}{\partial t} (c_L + c_R) = (E_0 + t_{LR})(c_L + c_R), \quad (7.92)$$

$$i\hbar \frac{\partial}{\partial t} (c_L - c_R) = (E_0 - t_{LR})(c_L - c_R). \quad (7.93)$$

These equations are solved by

$$c_L + c_R = \alpha e^{-i(E_0+t_{LR})t/\hbar}, \quad c_L - c_R = \beta e^{-i(E_0-t_{LR})t/\hbar}. \quad (7.94)$$

By adding and subtracting these solutions, we get

$$c_L = \frac{\alpha}{2} e^{-i(E_0+t_{LR})t/\hbar} + \frac{\beta}{2} e^{-i(E_0-t_{LR})t/\hbar}, \quad (7.95)$$

$$c_R = \frac{\alpha}{2} e^{-i(E_0+t_{LR})t/\hbar} - \frac{\beta}{2} e^{-i(E_0-t_{LR})t/\hbar}. \quad (7.96)$$

Symmetry requires  $\alpha = \beta = 1$  which yields

$$c_L(t) = e^{-iE_0t/\hbar} \cos(t_{LR}t/\hbar), \quad (7.97a)$$

$$c_R(t) = e^{-iE_0t/\hbar} [-i \sin(t_{LR}t/\hbar)]. \quad (7.97b)$$

This result, indeed, shows that  $|L\rangle$  and  $|R\rangle$  are not stationary states and that the electron oscillates between the left and the right quantum well. The probability to find the electron in the right dot, that is, in the quantum state  $|R\rangle$ , follows from (7.97b) as

$$|c_R(t)|^2 = \sin^2(t_{LR}t/\hbar). \quad (7.98)$$

Both representations, in terms of stationary solutions  $|g\rangle$ ,  $|e\rangle$  and of non-stationary ones,  $|L\rangle$ ,  $|R\rangle$  are equivalent and can be transformed into each other by a unitary matrix:

$$\begin{pmatrix} |g\rangle \\ |e\rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} |L\rangle \\ |R\rangle \end{pmatrix}. \quad (7.99)$$

What is the consequence of an electric field applied along the connection axis between the dots? The electron will be localized left or right as already discussed above. For the mathematical description, the ground state energy  $E_0$  in (7.89a), (7.89b) must be augmented with an electrical dipole contribution  $\mu\mathcal{E}$ , whose sign depends on the field direction, and from (7.90) we obtain:

$$\hbar\omega \begin{pmatrix} c_L \\ c_R \end{pmatrix} = \begin{pmatrix} E_0 + \mu\mathcal{E} & t_{LR} \\ t_{LR} & E_0 - \mu\mathcal{E} \end{pmatrix} \begin{pmatrix} c_L \\ c_R \end{pmatrix}. \quad (7.100)$$

In analogy to (7.91), there are two stationary solutions with the energies

$$E_{\mp} = E_0 \pm \sqrt{t_{LR}^2 + \mu^2\mathcal{E}^2}. \quad (7.101)$$

For vanishing electric field, the stationary states with the energies  $E_{\mp}$  are obtained. Their superposition represents the Q-bit. For strong electric fields ( $\mu^2\mathcal{E}^2 \gg t_{LR}^2$ ) the electron is localized in  $|L\rangle$  or  $|R\rangle$  depending on the field direction. Both states are then stationary states of the system. This is immediately obvious from the state amplitudes (7.95), (7.96):  $E_0$  must be replaced by  $E_0 \pm \mu\mathcal{E}$  and for a sufficiently large electric field  $\mu\mathcal{E}$  becomes the leading term, also in comparison with  $t_{LR}$ . The time dependencies of  $c_L$  and  $c_R$  are then obtained as  $\exp(\pm i\mu\mathcal{E}t/\hbar)$ . Due to the localization of the electron in  $|L\rangle$  or  $|R\rangle$ , the Q-bit has been switched off (Fig. 7.12d).

### 7.5.2 Experimental Realization of a Quantum-Bit by Quantum Dots

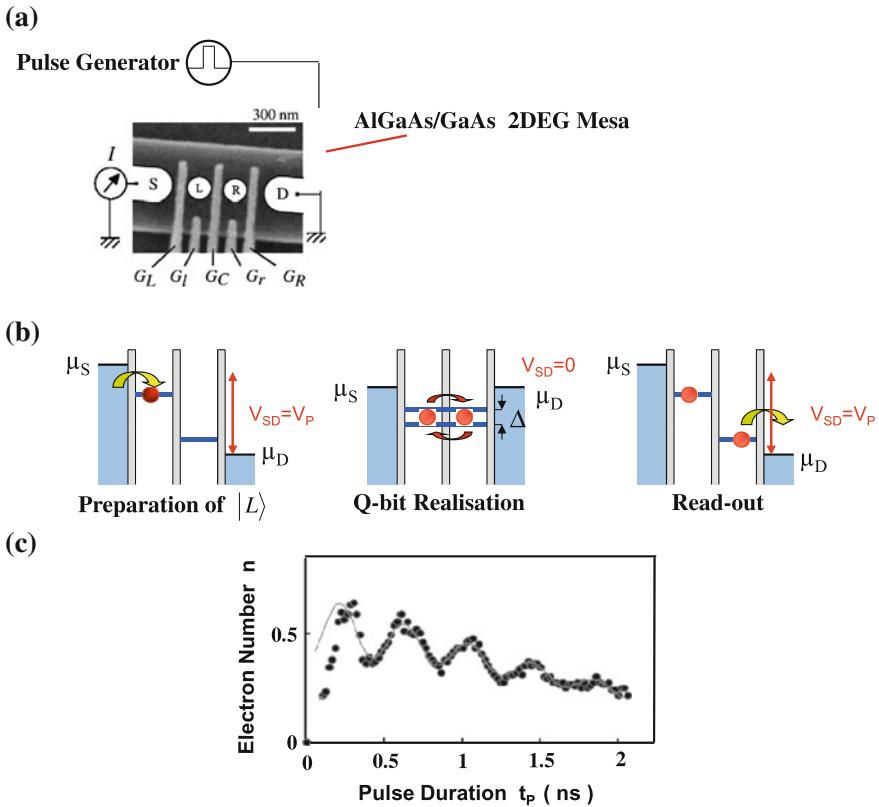
The representation of a Q-bit described in Sect. 7.5.1 has been realized experimentally by two coupled semiconductor quantum dots. Hayashi et al. [21] prepared the coupled quantum dots in split-gate technology (Sect. 5.7.2, Appendix B) on a GaAs/AlGaAs heterostructure which contains a 2D electron gas (2DEG) at its interface (Appendix A). Three parallel metal stripes were lithographically deposited as

electrical gate contacts  $G_L, G_C, G_R$  (applied voltages  $V_L, V_C, V_R$ ) on the heterostructure. Under these metal contacts, the 2DEG is depleted from electrons because of the existing depletion space charge layer (Schottky contact, Appendix A) (Fig. 7.13). The depleted areas under the metal act as energetic barriers for the 2DEG electrons whose height can be varied by the applied voltages  $V_L, V_C, V_R$ . The two laterally restricted areas in between occupied by electrons of the 2DEG define two quantum dots with locally lowered potential which are separated by  $G_C$  (applied voltage  $V_C$ ). The potential of the quantum dots can be separately controlled with respect to the source region by means of the gates  $G_l$  and  $G_r$  with the applied voltages  $V_l$  and  $V_r$ . The voltages  $V_L, V_C, V_R$  are adjusted in advance in control experiments to obtain suited tunnel barriers between the quantum dots ( $V_C$ ) as well as between source region and the left quantum dot ( $V_L$ ) respectively, drain region and the right quantum dot ( $V_R$ ). The potentials at source and drain as well as the voltages  $V_l$  and  $V_r$  on the other hand, allow a variation of the dot potentials with respect to each other and with respect to source and drain in the experiment itself (Fig. 7.13b).

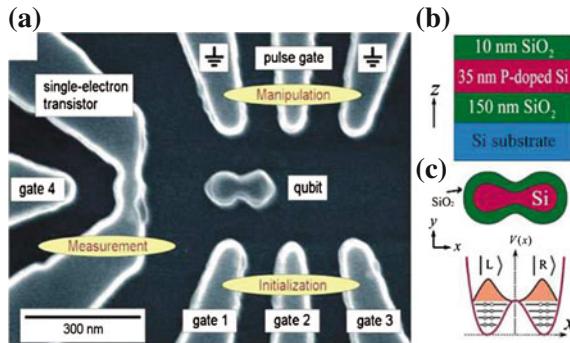
Depending on the potential adjustments an electron can tunnel through the left barrier ( $G_L$ ) from source into the left quantum dot where it occupies the corresponding ground state of the confined electronic levels, or with equal potentials in the two quantum dots it can occupy a superposition state covering both dots. This superposition state, which represents a Q-bit, might be interpreted as electron tunneling back and forth through the central barrier ( $G_C$ ). By lowering the drain potential relative to that of the right quantum dot the electron can tunnel through the right barrier ( $G_R$ ) and is detected as a current at the drain contact. This current measurement, of course, destroys the Q-bit.

We describe the experiment for proving the realization of a Q-bit in the two coupled quantum dots now in detail: For the preparation of the Q-bit at first a voltage  $V_{SD}$  of about  $600 \mu\text{V}$  is applied between source and drain. The two ground states  $|L\rangle$  and  $|R\rangle$  are then not in resonance ( $\mu_S > E_L > E_R > \mu_D = \mu_S - eV_{SD}$ ). If the outer tunneling barriers ( $G_L, G_R$ ) exhibit a higher tunneling probability than the central one ( $G_C$ ), an electron tunnels into the left quantum dot and the state  $|L\rangle$  is prepared (Fig. 7.13b) due to a relatively strong electric field between the quantum dots (Sect. 7.5.1).

The Q-bit is prepared by changing the potentials within 100 ps, such that the source and drain regions assume equal potential. Consequently the two ground state energies of the quantum dots are equal ( $E_L = E_R = E_0$ ) and lower in energy than the Fermi energies  $\mu_S$  and  $\mu_D$  of source and drain (Fig. 7.13b). Electrons can neither tunnel into the source nor into the drain region; the double quantum dot, that is, the Q-bit is electrically isolated from its environment, the two current electrodes. The initially prepared state  $|L\rangle$  of the electron has changed into the superposition state (7.85), where the two ground states  $E_L$  and  $E_R$  are split into the states  $E_{\pm}$  (Sect. 7.5.1). The electron might be imagined as tunneling back and forth between the two quantum dots. In the experiment, the corresponding potentials are adjusted with vanishing source-drain voltage  $V_{SD} = 0$  during pulse duration times  $t_p$  between 80 and 2000 ps. Within these time intervals the Q-bit exists as a superposition state.



**Fig. 7.13** **a–c** Experimental realization of a quantum bit (Q-bit) by two coupled GaAs quantum dots [21]. **a** Scanning electron micrograph of the lithographically prepared double quantum dot structure with schematically drawn measurement circuit (Appendix B). The quantum dots  $L$  and  $R$  as well as the source ( $S$ ) and drain ( $D$ ) electrodes are formed in a 2DEG at the interface of an AlGaAs/GaAs heterostructure. For this purpose metallic electrode fingers  $G_L, G_I, G_C, G_r$ , and  $G_R$  were deposited on the heterostructure (prepared as mesa). Below these metallic electrodes the 2DEG is depleted from electrons and the resulting Schottky barriers separate the different conducting areas  $S, L, R, D$  in the 2DEG. **b** Manipulation of the Q-bit shown schematically in the electronic band scheme of the double dot, source, drain structure: By means of a suited voltage pulse  $V_P = V_{SD}$  between source and drain the ground state energy levels of the two quantum dots are shifted with respect to each other such that an electron can tunnel from the source region into the left quantum dot: preparation of the state  $|L\rangle$ . By means of a subsequent voltage pulse source and drain potential are made equal ( $V_{SD} = 0$ ). The coupled quantum dots are isolated from source and drain. Because of the Fermi level position  $E_F = \mu_S = \mu_D$  the electronic occupation of the two new levels  $E_-$  (bonding) and  $E_+$  (anti-bonding) generated by the dot coupling and energetically separated by  $\Delta$  is equally probable. Thus, the superposition state of the Q-bit is realized. After varying time intervals (pulse duration) the voltage condition  $V_{SD} = 0$  for Q-bit realization is switched off and the system returns in its initial state (preparation of  $|L\rangle$ ). Depending on the pulse duration for Q-bit realization the electron is localized whether in  $|L\rangle$  or in  $|R\rangle$  at the end. If the state  $|R\rangle$  is realized (electron localized in right Q-dot), a current pulse is measured between  $S$  and  $D$  (recording). Localization of the electron in  $|L\rangle$  does not cause a current pulse. **c** Measured current signal, i.e., essentially electron number  $n$  as function of the pulse duration time  $t_p$ , during which the Q-bit was realized



**Fig. 7.14** a–c Realization of a quantum bit (Q-bit, qubit) by two silicon (Si) quantum dots prepared lithographically (Appendix B) from a thin Si layer (35 nm thick) on an oxidized Si wafer (SOI = silicon on insulator) [22]. **a** Scanning electron micrograph of the nanostructures prepared on the SOI wafer with explanations. The *bright areas* are the conducting Si structures which remained after the lithographical removal of the areas in between. These structures are covered with a 10 nm thick SiO<sub>2</sub> film for passivation. **b** Layer structure of the bright areas in (a). **c** Scheme of the coupled Si quantum dots which realize the Q-bit (qubit). At the bottom the potential  $V(x)$ , the energy levels and the wave functions of the states  $|L\rangle$  and  $|R\rangle$  in the quantum dots are indicated

The electron undergoes Rabi oscillations between the two quantum states of the coupled quantum dots (Sect. 6.5.1).

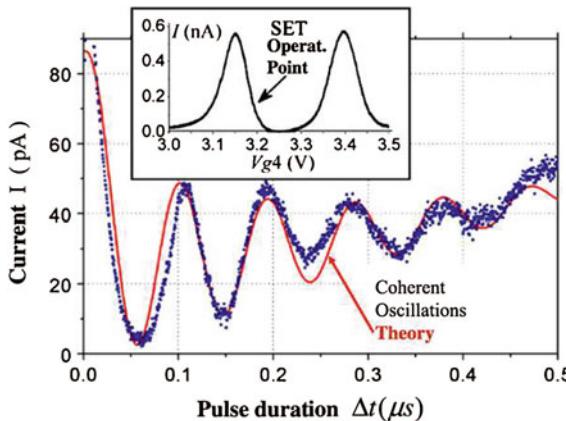
As the Q-bit is, however, coupled to its environment in manifold ways (phonons, leak currents, capacities etc.) the superposition state dephases within the pulse duration time  $t_p$ , i.e. the amplitude of the Rabi oscillations decreases. This effect is measured by readjusting the voltage drop  $V_{SD} \approx 600 \mu\text{V}$  between source and drain within 100 ps after the pulse duration  $t_p$ . The ground states with energies  $E_L$  and  $E_R$  are not in resonance anymore. If in this moment the electron occupies the right quantum dot, that is, the state  $|R\rangle$ , it can tunnel through the right barrier ( $G_R$ ) into the drain region and is detected as a current. An electron being just at that moment in the left potential well (state  $|L\rangle$ ) during its Rabi oscillation can not contribute to the current measured at the drain region. A measurement of the source-drain current as a function of pulse duration time  $t_p$ , thus, yields oscillations which demonstrate the presence of the superposition state (before each measurement), that is, the alternating occurrence of the states  $|R\rangle$  with a current maximum and of  $|L\rangle$  with a current minimum (Fig. 7.13c). The corresponding experimental curve, that is, the average number of pulse induced tunneling electrons  $\langle n_p \rangle$  as a function of pulse duration, shows the Q-bit Rabi oscillations decaying in amplitude due to dephasing (Fig. 7.13c). The numerical analysis of the data yields an oscillation frequency  $\Omega/2\pi \approx 2.3 \text{ GHz}$  and a dephasing time (decay time of the oscillation amplitude) of about 1 ns.

An even more elegant experiment on two coupled Si quantum dots was performed by Gorman et al. [22]. The two dots were lithographically prepared from a thin Si layer on an insulating silicon dioxide (SiO<sub>2</sub>) substrate (SOI = silicon on insulator) (Fig. 7.14). The basic material was a commercial SOI wafer with a phosphorus doped

35 nm thick Si layer. The double quantum dot is formed by two circular Si segments which are connected by a 20 nm wide constriction. Because of the depletion space charge layers at the edges of this constriction (Appendix A) the potential is elevated there in comparison with the potential wells in the two circular quantum dots. The constriction, thus, represents a tunnel barrier for an electron moving back and forth between the dots. Six further contact fingers (2 ground contacts, pulse gate, gates 1, 2, 3) were prepared lithographically in the conducting Si epilayer. Voltages applied to the gates 1, 2, 3 produce electric fields directed essentially parallel to the axis of the double dot structure; they are used to localize the electron whether in the left  $|L\rangle$  or in the right quantum dot  $|R\rangle$ . Manipulation in the time domain, i.e. initialization, manipulation and measurement of the Q-bit superposition state is performed by a voltage pulse of varying duration time at the pulse gate (Fig. 7.14). During the pulse duration the two ground states of the quantum dots are in resonance and the Q-bit exists as a superposition state. After switching off the pulse the electron is localized whether in the left or in the right quantum dot. This situation corresponds to the measurement: In particular, localization of the electron in the left quantum dot (state  $|L\rangle$ ) is detected by a single-electron transistor (SET) which has also been prepared lithographically in the thin Si epilayer. The SET consists of a quantum dot which is separated from two source and drain contacts by two lateral constrictions (tunneling barriers). Single electron tunneling through these barriers is controlled by the gate voltage at gate 4 (Sect. 3.7). A suited bias at gate 4 establishes Coulomb blockade in the transistor (Sect. 3.7). Localization of an electron in the left quantum dot, then, changes the SET potential and releases the Coulomb blockade; consequently a current signal is detected in the SET. The SET allows an extremely sensitive detection of electron localization in the left quantum dot, that is, the occupation of the quantum state  $|L\rangle$ .

A current measurement through the SET, thus, yields the information, if after switching off the Q-bit superposition state the state  $|L\rangle$  or  $|R\rangle$  is occupied by the electron. As a function of the pulse duration time  $\Delta t$  (time during which the two dot ground states are in resonance) the SET current shows the characteristic Rabi oscillations which indicate the existence of the superposition state (Q-bit) during the pulse duration. The decay of the oscillation amplitude with time (pulse duration) indicates the dephasing process of the Q-bit. The data points in Fig. 7.15 can be fitted to a damped sine function. Its decay time, i.e. the dephasing time of the Q-bit, is obtained as about 200 ns. It is longer by a factor of about 200 than that in the experiment of Hayashi et al. [21] described above.

The explanation of this much longer existence time of the Q-bit is straight forward: In case of the SOI quantum dots the coupling of the dots to their direct environment, the measurement and the control electrodes is merely by capacities, without any direct conductive coupling. On the other hand, in the split-gate experiment of Hayashi et al. [21] the Q-bit is coupled to its environment by tunnel contacts carrying electrical currents. This stronger coupling to the environment causes much faster dephasing of the Q-bit.



**Fig. 7.15** Electron current measured in the single electron transistor (SET) of Fig. 7.14 as function of the pulse duration time  $\Delta t$  during which the potential adjustment at the electrodes in Fig. 7.14 generates a Q-bit in the Si double quantum dot. Maxima of the current oscillations correspond to lift of the Coulomb blockade in the SET, i.e., to a current flow through the transistor. These current maxima indicate localization of the electron in the left quantum dot  $|L\rangle$ . *Inset* Coulomb Blockade oscillations of the single-electron transistor (SET in Fig. 7.14a) used for measurement of electron localization. The operating point for the measurement of the state  $|L\rangle$  in the double quantum dot is indicated [22]

From the comparison of the two experiments, we learn that the manipulation of Q-bits in eventual future quantum computers is only possible by means of weak coupling to the read-in and read-out as well as the manipulation hardware.

## References

1. E. Schrödinger, Proc. Camb. Phil. Soc. **31**, 555 (1935)
2. E. Schrödinger, *Abhandlungen zur Wellenmechanik* (J.A. Barth, Leipzig, 1927)
3. E. Schrödinger, *Briefe zur Wellenmechanik* (Springer, Wien, 1963)
4. A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. **47**, 777 (1935)
5. D. Bohm, Phys. Rev. **85**, 166 (1952)
6. J.S. Bell, Physics **1**, 195 (1964)
7. J.S. Bell, Rev. Mod. Phys. **38**, 447 (1966)
8. M. Lamehi-Rachti, W. Mittig, Phys. Rev. **14**, 2543 (1976)
9. A. Aspect, P. Grangier, G. Roger, Phys. Rev. Lett. **49**, 91 (1982)
10. M.O. Scully, B.-G. Englert, H. Walther, Nature **351**, 111 (1991)
11. S. Dürr, T. Nonn, G. Rempe, Nature **395**, 33 (1998)
12. D.-I. Chang, G.L. Khym, K. Kang, Y. Chung, H.-J. Lee, M. Seo, M. Heiblum, D. Mahalu, V. Umansky, Nat. Phys. **4**, 205 (2008)
13. E. Buks, R. Schuster, M. Heiblum, D. Mahalu, V. Umansky, Nature **391**, 871 (1998)
14. S. Gustavsson, R. Leturcq, M. Stnder, T. Ihn, K. Ensslin, D. C. Driscoll, A. C. Gossard, NanoLetters **8**, 2547 (2008)
15. G.M. Palma, K.A. Suominen, A.K. Eckert, Proc. R. Soc. Lond. A **452**, 567 (1996)

16. W. Unruh, Phys. Rev. A **51**, 992 (1995)
17. D. Di Vincenzo, Phys. Rev. A **50**, 1015 (1995)
18. V.N. Golovach, A. Khaetskii, D. Loss, Phys. Rev. Lett. **93**, 016601–1 (2004)
19. H.-K. Lo, S. Popescu and T. Spiller (eds.), Introduction to Quantum Computation and Information, World Scientific, Singapore (1998)
20. D. Bouwmeester, A. Ekert, A. Zeilinger (eds.), *The Physics of Quantum Information* (Springer, Berlin, 2000)
21. T. Hayashi, T. Fujisawa, H.D. Cheong, Y.H. Jeong, Y. Hirayama, Phys. Rev. Lett. **91**, 226804–1 (2003)
22. J. Gorman, D.G. Hasko, D.A. Williams, Phys. Rev. Lett. **95**, 090502–1 (2005)

# Chapter 8

## Fields and Quanta

The introduction into quantum mechanics has been essentially based on the experimentally well founded assumption that all matter has simultaneously wave and particle character. We have subsequently found a suited mathematical formalism for this concept. The classical concept of a particle propagating along a well defined trajectory had to be given up and operators must be introduced for the description of observables. Essential ingredients of this description are the wave function, its interpretation as a probability amplitude and the commutativity or non-commutativity of operators. The latter properties are at the basis of quantum physics, as non-commutativity of operators stands for the impossibility of simultaneously measuring two observables, for example, position and momentum. This guarantees the statistical character of the results of subsequent position and momentum measurements.

The described concept has been applied so far only to the dynamics of one single particle.

Even in cases of complex systems with several or many electrons, as atoms, molecules or solids the so-called single-particle approximation has been used. Hereby, the action of all other electrons except the one under consideration is put into an effective potential which determines the dynamics of the one electron (Sects. 8.3.2 and 8.3.4).

By no means, the formalism of single particle dynamics used so far allows the description of Einstein's fundamental light quantum hypothesis where the electromagnetic field extended over large areas of space (described by field equations) is composed of light particles, the photons with a well-defined quantum energy  $E = \hbar\omega$ . At this phenomenon, we obviously encounter the simultaneous nature of a continuous field, that is, electromagnetic waves, and many particles. By means of modern highly sensitive semiconductor detectors, the detection of single separate photons in a weak light field is possible. Photons as particles constitute the light field. Depending on the measurement, the field character described by Maxwell's equations is more apparent; in other cases, particularly in weak fields the particle character is evident.

A similar problem occurs when, instead of only one single electron, a large ensemble of many interacting electrons must be described adequately. Then, the many-body wave function  $\psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N)$  can not be factorized any more into single-electron wave functions. As the electromagnetic field the wave function exists as a field function continuously in the whole space. Nevertheless, we can detect a single localized electron at a certain time and a certain position within this field by means of an appropriate electron detector. It, therefore, seems more adequate to attribute a field function to the many-body system which extends over large space areas, similarly as the electromagnetic field. Using an appropriate detector, the fields must show the presence of a particle, depending on the field, the presence of an electron or of a photon. A suited formalism would yield a unified description of many-particle fields and single detectable quanta (particles) in these fields. The distinction between fields, e.g. the light field, and particles as electrons and photons would be eliminated. A unified field-quantum description [1, 4] indeed shows the complementarity of waves (fields) and particles which is the basis of reality both with respect to classical fields as the electromagnetic field and particles being considered in classical physics.

Somewhat exaggerating the picture one might imagine the field-quantum system as an ocean (the field) which creates at certain positions and at particular moments a fountain like burst, the creation of a particle. This particle carries a well defined energy, the quantum energy  $\hbar\omega$ , which is transferred from the field to the detector during the measurement.

## 8.1 Ingredients of a Quantum Field Theory

The formalism described above as being appropriate to the field-particle duality is called quantum field theory [1–5]. Like Schrödinger's wave mechanics quantum field theory has been “invented” [6, 7], not incidentally, but by deep reasoning about what one has learned from establishing single particle quantum mechanics. Extending single-particle quantum mechanics assumptions (hypotheses) are made which must be falsified or verified by quantitative measurement results. In establishing these hypotheses some requirements had to be fulfilled: The new theory must contain the single-particle quantum mechanics (Schrödinger equation) in case that only one particle is present (Sect. 8.3.2). For the light field, large field amplitudes must obey classical Maxwell's theory. In both cases field differential equations determine the behavior of continuous many-particle fields, the Schrödinger equation for electrons and Maxwell's equations for photons. How can the particle character, the typically quantum property, introduced into the continuous field concept? As in single-particle Schrödinger quantum mechanics position and momentum, there must be field specific variables which can not be measured simultaneously with infinite accuracy. This reflects the stochastic character of the field observables: The measurement of the one field observable allows only a statistically predictable result for the measurement of the other observable. The observables are incommensurable. In this case, the two field observables are canonically conjugated (Sect. 3.4) as are position and momentum in

single particle quantum mechanics. Similarly a description in terms of operators which do not commute is obvious.

In order to make classical field theories compatible with quantum physics, therefore, continuous field variables as the electric or the magnetic field strength must be replaced by operators which are defined in the whole space region where the fields exist. Furthermore, these field operators must obey commutator relations as do position and momentum operators in single particle quantum mechanics. As a recipe, we use the transition from classical Hamilton mechanics to Schrödinger quantum mechanics (Sect. 3.4). In this concept, a decisive role is played by the Hamilton function  $H(p, x)$ , the total energy as a function of the independent variables position  $x$  and momentum  $p = m\dot{x}$  (Sect. 3.4). With the kinetic energy of a particle  $T = p^2/2m$ , the Hamilton function is written as

$$H = T + V(x) = \frac{p^2}{2m} + V(x) = \frac{(m\dot{x})^2}{2m} + V(x). \quad (8.1)$$

From the two Hamilton equations,

$$\frac{\partial H}{\partial p} = \dot{x}, \quad (8.2a)$$

$$-\frac{\partial H}{\partial x} = \frac{\partial V}{\partial x} = K(x) = \dot{p} \quad (8.2b)$$

we derive Newton's classical dynamic equation with the force  $K(x)$  as cause of the momentum change  $\dot{p}$ . Simultaneously the Hamilton formalism (8.2a), (8.2b) makes clear what observables, namely  $p$  and  $x$ , are incommensurable in quantum mechanics, that is, must obey a commutation relation.

The procedure for establishing a quantum field theory, therefore, should start with expressing the total energy of the corresponding field in terms of a Hamilton function. This Hamilton function must yield the dynamic field equations (e.g., Maxwell or Schrödinger equation) by application of the Hamilton equations (8.2a), (8.2b). From this formalism, we derive the two independent canonical field variables which are replaced by field operators obeying a commutation relation, analogously to  $[p, x]$ .

A further tool for the quantization of fields was introduced when dealing with the harmonic oscillator (Sect. 4.4.2). By clever factorization of the Schrödinger equation, we could define step operators  $\hat{b}$  and  $\hat{b}^+$  which lead from one oscillator eigenstate to the next lower ( $\hat{b}$ ) or the next higher one ( $\hat{b}^+$ ). Multiple application of these operators creates a ladder of energetically equidistant energy eigenvalues  $E_n = (n + \frac{1}{2})\hbar\omega$ . Each time one quantum of energy  $\hbar\omega$  is added to or subtracted from the initial energy. This behavior is just what is needed for the description of field quanta which constitute a field. One type of particles, for example, photons of energy  $\hbar\omega$ , are added or removed from the electromagnetic field when the field intensity changes. For this purpose, it would be desirable to have operators as  $\hat{b}$  and  $\hat{b}^+$  available which destroy or create a photon of that type.

Along the described way indeed a quantum field theory is established which describes the behavior of particle fields consistently and in excellent agreement with all experiments performed so far. Of particular importance for application is the interaction of the light field with atoms and solids being the basis of the laser mechanism.

## 8.2 Quantization of the Electromagnetic Field

The electromagnetic field described by its electric field  $\mathcal{E}$  and its magnetic field component  $\mathbf{H}$ , respectively  $\mathbf{B}$ , is defined by six field values at each position  $\mathbf{r}$  in space at time  $t$ . The successful classical electromagnetic field theory of Maxwell is not compatible with quantum theory as is evident from following considerations: How do we measure the electric and the magnetic field? For the measurement of the electric field  $\mathcal{E}$ , we observe the movement of a point charge. The direction of the charge movement indicates the direction of the field, while the amount of the displacement is a measure of the field strength. To determine the strength of the magnetic field  $\mathbf{B}$ , the Lorentz force  $\mathbf{K} = e\mathbf{v} \times \mathbf{B}$  acting on the moving charge might be determined. But in addition a measurement of the charge velocity  $\mathbf{v}$  is necessary. For an electric field orientation normal to  $\mathbf{B}$  the velocity vector of the charge is oriented in the direction of the spatial displacement. According to Heisenberg's uncertainty principle (Sect. 3.3) velocity and position can not be determined simultaneously with infinite accuracy. Consequently, also the electric and the magnetic field components of the electromagnetic field can not be measured simultaneously with infinite precision. There must be an uncertainty principle for the  $\mathcal{E}$  and the  $\mathbf{B}$  field. The electromagnetic field has to be quantized, that is, we must replace the classical field variables by operators and must establish commutation relations between these field operators. The corresponding procedure is suggested in Sect. 8.1.

We start the derivation with the classical Maxwell equations for electric and magnetic fields in vacuum, where space charges and currents do not exist, that is:

$$\operatorname{div} \mathbf{D} = \operatorname{div} \varepsilon_0 \mathcal{E} = 0, \quad (8.3a)$$

$$\operatorname{div} \mathbf{B} = \operatorname{div} \mu_0 \mathbf{H} = 0, \quad (8.3b)$$

and

$$\operatorname{curl} \mathcal{E} = -\dot{\mathbf{B}}, \quad (8.4a)$$

$$\operatorname{curl} \mathbf{H} = \dot{\mathbf{D}}. \quad (8.4b)$$

Using the operator relation  $\operatorname{curl} \operatorname{curl} = \operatorname{grad} \operatorname{div} -\Delta$  one easily derives the wave differential equation for the  $\mathcal{E}$  and  $\mathbf{B}$  fields of the transverse electromagnetic wave ( $\mathcal{E}$  and  $\mathbf{B}$  normal to each other) from (8.4a), (8.4b). Hereby, the vacuum light velocity follows as  $c = 1/\sqrt{\varepsilon_0 \mu_0}$ .

As in Sect. 5.4.4, the vector potential  $\mathbf{A}$  appears as the fundamental variable, from which the magnetic and the electric field are derived. As in Sect. 5.4.2,  $\mathbf{A}$  is introduced by the relations

$$\mathbf{B} = \operatorname{curl} \mathbf{A}, \quad \operatorname{div} \mathbf{A} = 0. \quad (8.5)$$

By application of the curl operation on (8.5), one obtains from (8.4a), (8.4b):

$$\operatorname{curl} \operatorname{curl} \mathbf{A} = \operatorname{grad} \operatorname{div} \mathbf{A} - \Delta \mathbf{A} = \mu_0 \varepsilon_0 \dot{\mathbf{E}}, \quad (8.6a)$$

and together with (8.4a), i.e.  $\mathcal{E} = -\dot{\mathbf{A}}$ , finally:

$$\Delta \mathbf{A} - \frac{1}{c^2} \ddot{\mathbf{A}} = 0. \quad (8.6b)$$

This is a wave differential equation. Similarly as for the  $\mathcal{E}$  and the  $\mathbf{H}$  field this equation is solved by a transversal vector potential ( $\mathbf{A}$ ) wave with two  $\mathbf{A}$  components normal to the wave vector  $\mathbf{q}$  (propagation direction). Their frequencies are identical  $\omega_{\mathbf{q}}$ . The solution to (8.6b) can, thus, be expressed as a Fourier expansion:

$$\mathbf{A}(\mathbf{r}, t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{q}, \lambda} A_{\mathbf{q}}(t) \mathbf{s}_{\mathbf{q}, \lambda} e^{i\mathbf{q} \cdot \mathbf{r}}. \quad (8.7)$$

In this chapter, we will denote the wave vector of photons by  $\mathbf{q}$  to distinguish it from that of electrons ( $\mathbf{k}$ ).  $s_{\mathbf{q}, \lambda}$  denotes the two unity vectors ( $\perp \mathbf{q}$ ) of the mutually orthogonal oscillation directions (polarization vectors) of the  $\mathbf{A}$  field.  $\lambda = 1, 2$  numerates the two components. The requirement that  $\mathbf{A}(\mathbf{r}, t)$  is a real function is guaranteed by

$$\mathbf{A}_{\mathbf{q}\lambda} = \mathbf{A}_{-\mathbf{q}\lambda}^* \quad (8.8)$$

with the assumption  $s_{\mathbf{q}, \lambda} = s_{-\mathbf{q}, \lambda}$ .

Similarly as in Sect. 3.6.1 for electron waves, we have assumed a large box of volume  $V = L^3$  as the spatial region in which the electromagnetic field is defined. Consequently, the wave vectors (numbers)  $\mathbf{q}$  of the orthonormal expansion eigenfunctions (waves) in (8.7) are quantized, however quasi-continuously with distances  $\Delta q = 2\pi/L$ . The transition to a continuous description ( $L \rightarrow \infty$ ) is easily made by replacing sums by integrals (4.74). By inserting (8.7) into (8.6a), (8.6b) and using the light dispersion relation,

$$\omega_q = c|q| \quad (8.9)$$

one immediately obtains:

$$\ddot{A}_{\mathbf{q}\lambda} - \omega_{\mathbf{q}}^2 A_{\mathbf{q}\lambda} = 0. \quad (8.10)$$

Note that both light polarizations  $\lambda = 1, 2$  oscillate with the same frequency  $\omega_{\mathbf{q}}$ . The oscillator differential equation (8.10) yields amplitudes  $A_{\mathbf{q}\lambda}$  which oscillate in time as  $\exp(-i\omega_{\mathbf{q}}t)$ .

Starting point for the field quantization is the Hamilton function  $H$  which represents the total energy of the field:

$$H = E = \frac{1}{2} \int d^3r (\mathbf{E} \cdot \mathbf{D} + \mathbf{H} \cdot \mathbf{B}) = \int d^3r |\mathcal{E}|^2 = \int d^3r |H|^2. \quad (8.11)$$

We use the relation  $\mathcal{E} = -\dot{\mathbf{A}}$  and express the magnetic field by means of (8.7):

$$\mathbf{B} = \text{curl } \mathbf{A} = \frac{1}{\sqrt{V}} \sum_{\mathbf{q}, \lambda} A_{\mathbf{q}\lambda}(t) [\mathbf{s}_{\mathbf{q}, \lambda} \times \mathbf{q}] e^{i\mathbf{q} \cdot \mathbf{r}}. \quad (8.12)$$

By means of the identity

$$(\mathbf{s}_{\mathbf{q}\lambda} \times \mathbf{q}) \cdot (\mathbf{s}_{\mathbf{q}'\lambda'} \times \mathbf{q}') = (\mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{s}_{\mathbf{q}'\lambda'}) (\mathbf{q} \cdot \mathbf{q}') - (\mathbf{q} \cdot \mathbf{s}_{\mathbf{q}'\lambda'}) (\mathbf{s}_{\mathbf{q}'\lambda'} \cdot \mathbf{q}) \quad (8.13)$$

and using (8.8) the total energy (8.11) is calculated as

$$H = E = \frac{1}{2} \varepsilon_0 \sum_{\mathbf{q}\lambda} (\dot{A}_{\mathbf{q}\lambda}^* \dot{A}_{\mathbf{q}\lambda} + \omega_{\mathbf{q}}^2 A_{\mathbf{q}\lambda}^* A_{\mathbf{q}\lambda}). \quad (8.14)$$

Beside  $(\mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{q}) = 0$  the following representation of the  $\delta$ -function has been used:

$$\frac{1}{V} \int d^3r e^{i(\mathbf{q}-\mathbf{q}') \cdot \mathbf{r}} = \delta(\mathbf{q} - \mathbf{q}'). \quad (8.15)$$

In analogy to the Hamiltonian of the harmonic oscillator (4.109), we try to factorize the Hamilton function (8.14), similarly to (4.110) and (4.112a), (4.112b), by the following self-evident ansatz:

$$a_{\mathbf{q}\lambda} = \frac{1}{2} \left( A_{\mathbf{q}\lambda} + \frac{i}{\omega_{\mathbf{q}}} \dot{A}_{\mathbf{q}\lambda} \right), \quad a_{\mathbf{q}\lambda}^* = \frac{1}{2} \left( A_{\mathbf{q}\lambda} - \frac{i}{\omega_{\mathbf{q}}} \dot{A}_{\mathbf{q}\lambda} \right). \quad (8.16)$$

This ansatz with the prefactor 1/2 has been chosen to fulfill the condition (8.8) being relevant for the electromagnetic field and also to yield correct dimensions by the factor  $i/\omega_{\mathbf{q}}$  when calculating the time differentiation  $\dot{A}_{\mathbf{q}\lambda}$ . The important condition (8.8) is easily seen to be fulfilled by calculating the inverted relations to (8.16):

$$A_{\mathbf{q}\lambda} = a_{\mathbf{q}\lambda} + a_{-\mathbf{q}\lambda}^*, \quad \dot{A}_{\mathbf{q}\lambda} = \frac{\omega_{\mathbf{q}}}{i} (a_{\mathbf{q}\lambda} - a_{-\mathbf{q}\lambda}^*), \quad (8.17a)$$

$$A_{\mathbf{q}\lambda}^* = a_{\mathbf{q}\lambda}^* + a_{-\mathbf{q}\lambda}, \quad \dot{A}_{\mathbf{q}\lambda}^* = \frac{\omega_{\mathbf{q}}}{-i} (a_{\mathbf{q}\lambda}^* - a_{-\mathbf{q}\lambda}). \quad (8.17b)$$

Inserting (8.17a), (8.17b) into (8.14) yields:

$$H = \frac{1}{2} \varepsilon_0 \sum_{\mathbf{q}\lambda} \omega_{\mathbf{q}\lambda}^2 (2a_{\mathbf{q}\lambda}^* a_{\mathbf{q}\lambda} + 2a_{-\mathbf{q}\lambda}^* a_{-\mathbf{q}\lambda}). \quad (8.18a)$$

By replacing  $-\mathbf{q}$  by  $\mathbf{q}$  in the second sum, we obtain the favored product representation as for the harmonic oscillator:

$$H = \varepsilon_0 \sum_{\mathbf{p}\lambda} 2\omega_{\mathbf{q}\lambda}^2 (a_{\mathbf{q}\lambda}^* a_{\mathbf{q}\lambda}). \quad (8.18b)$$

Hereby, we have interchanged  $a_{\mathbf{q}\lambda}$  and  $a_{\mathbf{q}\lambda}^*$ , an obviously allowed procedure for normal numbers.

A comparison of the product representation (8.18b) and the oscillator Hamiltonian (4.115) shows that  $a_{\mathbf{q}\lambda}$  and  $a_{\mathbf{q}\lambda}^*$  are the canonical variables which are relevant for quantization. For the quantization of the electromagnetic field they must be replaced by non-commuting field operators. Non-commutability of the operators is expressed as  $[\hat{a}_{\mathbf{q}\lambda}, \hat{a}_{\mathbf{q}\lambda}^+] \neq 0$ .

To obtain the energy in the field Hamiltonian in units of  $\hbar\omega_{\mathbf{q}}$ , we introduce the operators  $\hat{b}_{\mathbf{q}\lambda}$  and  $\hat{b}_{\mathbf{q}\lambda}^+$  instead of  $\hat{a}_{\mathbf{q}\lambda}$  and  $\hat{a}_{\mathbf{q}\lambda}^+$  by means of the relation

$$\hat{a}_{\mathbf{q}\lambda} = \sqrt{\frac{\hbar}{2\varepsilon_0\omega_{\mathbf{q}}}} \hat{b}_{\mathbf{q}\lambda}. \quad (8.19)$$

In analogy to the oscillator (Sect. 4.4.2) the commutation relations, then, are written as

$$[\hat{b}_{\mathbf{q}'\lambda'}, \hat{b}_{\mathbf{q}\lambda}^+] = \hat{b}_{\mathbf{q}'\lambda'} \hat{b}_{\mathbf{q}\lambda}^+ - \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}'\lambda'} = \delta_{\mathbf{q}\lambda, \mathbf{q}'\lambda'}, \quad (8.20a)$$

$$[\hat{b}_{\mathbf{q}'\lambda'}^+, \hat{b}_{\mathbf{q}\lambda}^+] = [\hat{b}_{\mathbf{q}'\lambda'}, \hat{b}_{\mathbf{q}\lambda}] = 0. \quad (8.20b)$$

As the operators  $\hat{a}_{\mathbf{q}\lambda}$ ,  $\hat{a}_{\mathbf{q}\lambda}^+$ , respectively  $\hat{b}_{\mathbf{q}\lambda}$ ,  $\hat{b}_{\mathbf{q}\lambda}^+$  do not commute, the simple calculation step from (8.18a) to (8.18b) can not be performed. According to the commutation relations (8.20a), (8.20b), the sequence of the operators  $\hat{a}_{\mathbf{q}\lambda}$ ,  $\hat{a}_{\mathbf{q}\lambda}^+$ , respectively  $\hat{b}_{\mathbf{q}\lambda}$ ,  $\hat{b}_{\mathbf{q}\lambda}^+$  is well defined and must be obeyed in this calculation. Using (8.20a), (8.20b) we, thus, obtain:

$$\hat{H} = \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} (\hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + \hat{b}_{-\mathbf{q}\lambda}^+ \hat{b}_{-\mathbf{q}\lambda} + 1). \quad (8.21)$$

By replacing  $-\mathbf{q}$  by  $\mathbf{q}$  in the second sum the Hamilton operator of the light field is obtained as

$$\hat{H} = \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + \frac{1}{2} \right). \quad (8.22)$$

For a particular  $(\mathbf{q}, \lambda)$  pair the sum elements in (8.22) are identical with the Hamiltonian of the harmonic oscillator (4.115). The light field obviously consists of a high

number of oscillators, at least formally. We will soon identify these oscillators with photons.

We denote a general multi-photon state of the electromagnetic field by  $|\Phi\rangle$ . As in single particle quantum mechanics the total energy  $E$  attributed to this field is obtained as the eigenvalue of the field Hamiltonian (8.22):

$$\hat{H}|\phi\rangle = E|\phi\rangle. \quad (8.23)$$

As expected we find properties of the  $\hat{b}_{\mathbf{q}\lambda}$ ,  $\hat{b}_{\mathbf{q}\lambda}^+$  operators which are analogous to those of the step operators in Sect. 4.4.2. Application of  $\hat{b}_{\mathbf{q}'\lambda'}$  from the left side on (8.23) yields:

$$\sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( \hat{b}_{\mathbf{q}'\lambda'} \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + \frac{1}{2} \hat{b}_{\mathbf{q}'\lambda'} \right) |\phi\rangle = E \hat{b}_{\mathbf{q}'\lambda'} |\phi\rangle. \quad (8.24a)$$

The first summand on the left side is simplified by use of the commutation relations (8.20a), (8.20b). Hereby  $\hat{b}$  operators with differing  $\mathbf{q}\lambda$  commute, while those with equal  $\mathbf{q}\lambda$  do not commute. We thus obtain:

$$\begin{aligned} \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( \hat{b}_{\mathbf{q}\lambda}^+ b_{\mathbf{q}\lambda} + \frac{1}{2} \right) (\hat{b}_{\mathbf{q}'\lambda'} |\phi\rangle) &= (E - \hbar\omega_{\mathbf{q}'}) (\hat{b}_{\mathbf{q}'\lambda'} |\phi\rangle), \\ \hat{H} (\hat{b}_{\mathbf{q}'\lambda'} |\phi\rangle) &= (E - \hbar\omega_{\mathbf{q}'}) (\hat{b}_{\mathbf{q}'\lambda'} |\phi\rangle). \end{aligned} \quad (8.24b)$$

According to (8.24b) the field state  $(\hat{b}_{\mathbf{q}'\lambda'} |\Phi\rangle)$  is an energy eigenstate with a total field energy eigenvalue being diminished by  $\hbar\omega_{\mathbf{q}'}$ . The operator  $\hat{b}_{\mathbf{q}'\lambda'}$  destroys an energy quantum  $\hbar\omega_{\mathbf{q}'}$  in the electromagnetic field. A particle of this energy, a photon, has been removed from the field. In this context,  $\hat{b}_{\mathbf{q}\lambda}$  represents a photon (particle) *destruction* (or *annihilation*) operator.

By an analogous calculation, we can demonstrate that the operator  $\hat{b}_{\mathbf{q}\lambda}^+$  creates a new energy quantum (particle) of energy  $\hbar\omega_{\mathbf{q}}$  in the field. Within the frame of quantum field theory  $\hat{b}_{\mathbf{q}\lambda}^+$  and  $\hat{b}_{\mathbf{q}\lambda}$  are particle creation and destruction operators, respectively. In quantum field theory particles, in the present case photons, are excitation states of the field. A general field state, thus, exists in manifold excitation states, it contains  $n_{\mathbf{q}\lambda}, n_{\mathbf{q}'\lambda'}, n_{\mathbf{q}''\lambda''}, \dots$  photons, each type with quantum energies  $\hbar\omega_{\mathbf{q}}, \hbar\omega_{\mathbf{q}'}, \hbar\omega_{\mathbf{q}''}, \dots$  Such a multi-photon state is often denoted as

$$|\phi\rangle = |\dots, n_{\mathbf{q}\lambda}, n_{\mathbf{q}'\lambda'}, n_{\mathbf{q}''\lambda''}, \dots\rangle = |\dots, n_{\mathbf{q}\lambda}, \dots\rangle. \quad (8.25)$$

Application of the field Hamilton operator (8.22) on this state yields in analogy to the harmonic oscillator (Sect. 4.4.2):

$$\begin{aligned}\hat{H}|\phi\rangle &= \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( \hat{b}_{\mathbf{q}\lambda}^+ b_{\mathbf{q}\lambda} + \frac{1}{2} \right) |\dots, n_{\mathbf{q}\lambda}, \dots\rangle \\ &= \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( n_{\mathbf{q}\lambda} + \frac{1}{2} \right) |\dots, n_{\mathbf{q}\lambda}, \dots\rangle,\end{aligned}\quad (8.26a)$$

The energy eigenvalues of the field are thus obtained as

$$E = \sum_{\mathbf{q}\lambda} n_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} + \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}}. \quad (8.26b)$$

For obvious reasons, the operator  $\hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda}$  is called number operator. Its eigenvalues  $n_{\mathbf{q}\lambda}$  represent the number of photons  $\hbar\omega_{\mathbf{q}}$  in the field state  $|\Phi\rangle$ .

We assume a field  $|\Phi\rangle$  which contains only  $n_{\mathbf{q}\lambda}$  photons of energy  $\hbar\omega_{\mathbf{q}}$  (photons of type  $\mathbf{q}, \lambda$ ). Now we apply the destruction operator  $\hat{b}_{\mathbf{q}\lambda}$  on this field state  $n_{\mathbf{q}\lambda}$  times:

$$(\hat{b}_{\mathbf{q}\lambda})^{n_{\mathbf{q}\lambda}} |\phi\rangle = |0\rangle. \quad (8.27)$$

The field is then empty of photons. The ground state of the field, the so-called *vacuum state*  $|0\rangle$ , is obtained. According to (8.26b) this vacuum state nevertheless possesses an energy

$$E_0 = \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}}. \quad (8.28)$$

This is in principle an infinite amount of energy since  $\mathbf{q}$  is not limited in the sum. For many applications, this energy of the vacuum state does not play a role since we are free in the choice of the energy scale. We can put the energy zero point just on the vacuum energy level  $E_0$ . Nevertheless there are experimental issues (Casimir force, Sect. 8.2.5) which can only be understood on the basis of the vacuum state energy of the electromagnetic field.

The normalization factors of the multi-photon states are calculated in analogy to the states of the harmonic oscillator (Sect. 4.4.2). From (8.24b), the operators  $\hat{b}_{\mathbf{q}\lambda}^+$  and  $\hat{b}_{\mathbf{q}\lambda}$  are known to create, respectively to destroy one single photon of energy  $\hbar\omega_{\mathbf{q}}$  in the field, that is, for the photon creation we have

$$\hat{b}_{\mathbf{q}\lambda}^+ |\dots, n_{\mathbf{q}\lambda}, \dots\rangle = C |\dots, (n_{\mathbf{q}\lambda} + 1), \dots\rangle. \quad (8.29)$$

Now the factor  $C$  must be determined such that both states on the left and on the right side of the equation are normalized. For this purpose, we multiply (8.29) from

the left with the bra  $\langle \dots, n_{\mathbf{q}\lambda}, \dots | \hat{b}_{\mathbf{q}\lambda}$  and take into account that  $\hat{b}_{\mathbf{q}\lambda}$  is the adjoint operator to  $\hat{b}_{\mathbf{q}\lambda}^+$ , i.e.

$$(\hat{b}_{\mathbf{q}\lambda}^+)^+ = \hat{b}_{\mathbf{q}\lambda}. \quad (8.30)$$

We, thus, conclude

$$\begin{aligned} & \langle \dots, n_{\mathbf{q}\lambda}, \dots | \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} | \dots, n_{\mathbf{q}\lambda}, \dots \rangle \\ &= C^2 \langle \dots, (n_{\mathbf{q}\lambda} + 1), \dots | \dots, (n_{\mathbf{q}\lambda} + 1), \dots \rangle. \end{aligned} \quad (8.31)$$

Using the commutation relation  $\hat{b}_{\mathbf{q}\lambda} \hat{b}_{\mathbf{q}\lambda}^+ = \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + 1$  one concludes from (8.31):

$$\begin{aligned} & \langle \dots, n_{\mathbf{q}\lambda}, \dots | \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} | \dots, n_{\mathbf{q}\lambda}, \dots \rangle + \langle \dots, n_{\mathbf{q}\lambda}, \dots | \dots, n_{\mathbf{q}\lambda}, \dots \rangle \\ &= C^2 \langle \dots, (n_{\mathbf{q}\lambda} + 1), \dots | \dots, (n_{\mathbf{q}\lambda} + 1), \dots \rangle, \end{aligned} \quad (8.32a)$$

and with the precondition that the many-body states are already normalized:

$$n_{\mathbf{q}\lambda} + 1 = C^2. \quad (8.32b)$$

A calculation analogous to (8.29)–(8.32b) for the destruction operators  $\hat{b}_{\mathbf{q}\lambda}$  finally yields the relations

$$\boxed{\hat{b}_{\mathbf{q}\lambda}^+ | \dots, n_{\mathbf{q}\lambda}, \dots \rangle = \sqrt{n_{\mathbf{q}\lambda} + 1} | \dots, (n_{\mathbf{q}\lambda} + 1), \dots \rangle}, \quad (8.33a)$$

$$\boxed{\hat{b}_{\mathbf{q}\lambda} | \dots, n_{\mathbf{q}\lambda}, \dots \rangle = \sqrt{n_{\mathbf{q}\lambda}} | \dots, (n_{\mathbf{q}\lambda} - 1), \dots \rangle}. \quad (8.33b)$$

Multiple applications of the creation operators  $\hat{b}_{\mathbf{q}\lambda}^+$  on the (vacuum) ground state of the field  $|0\rangle$  allow the following representation of a general normalized state of the electromagnetic field:

$$| \dots, n_{\mathbf{q}\lambda}, \dots \rangle = \prod_{\mathbf{q}\lambda} \frac{1}{\sqrt{n_{\mathbf{q}\lambda}!}} (\hat{b}_{\mathbf{q}\lambda}^+)^{n_{\mathbf{q}\lambda}} | 0 \rangle. \quad (8.34)$$

In conclusion, it must be emphasized that the described quantization of the electromagnetic field is compatible with the theory of special relativity. This is in contrast to single-particle quantum mechanics, where the Schrödinger equation is a non-relativistic approximation. The reason for the relativistic compatibility of the present field quantization is due to the fact that Maxwell's theory of the electromagnetic field is consistent with special relativity (Lorentz invariant). Without knowing it Maxwell has created a relativistic theory of the electromagnetic field. Classical (Hamilton) mechanics being the starting point for Schrödinger's theory, on the other hand is a non-relativistic approximation.

Quantization of the electromagnetic field within the formal frame of the 4-dimensional space-time of special relativity does not yield results different from those presented in this section.

Note furthermore, that reality of the vector potential  $\mathbf{A}$  as required in (8.8) is essential for the quantized light field. Complex field amplitudes would not be compatible with Maxwell's theory.

### 8.2.1 What Are Photons?

In Chap. 2, we have discussed Einstein's light quanta hypothesis, where the light field is assumed to be composed of photon particles having an energy  $\hbar\omega$  and a momentum  $\hbar k$  (in this chapter:  $\hbar q$ ). The term photon as a light particle was used to explain the experimental findings of the photo-effect and of Compton scattering. This particle could by no means be understood within the framework of single-particle quantum mechanics. Only on the basis of the light field quantization a deeper understanding of the nature of the photon has been achieved. According to Sect. 8.2, photons are excitation states of the electromagnetic field. These excitations are characterized by a wave vector  $\mathbf{q}$  and a polarization ( $\perp \mathbf{q}$ ) having two mutually perpendicular orientations  $\lambda = 1, 2$ . The excitation states exhibit a photon energy  $\hbar\omega_{\mathbf{q}}$ . In a field state  $| \dots, n_{\mathbf{q}\lambda}, \dots \rangle$  with  $n_{\mathbf{q}\lambda}$  excitations, that is, photons of the type  $\hbar\omega_{\mathbf{q}}(\lambda)$ , additional photons can be created or others destroyed by the action of the creation, respectively destruction operators  $\hat{b}_{\mathbf{q}\lambda}^+$  and  $\hat{b}_{\mathbf{q}\lambda}$ . Thus, photons are not particles in the familiar classical sense, they are not necessarily localized in space. Nevertheless a single photon can be detected locally in a detector by a single "click". We will soon see, that this particle picture as an excitation state of a field is also valid for other particles, for example, electrons.

What's now about the momentum  $\hbar\mathbf{q}$  of a photon? Is this quantity also linked to classical field variables of Maxwell's theory? We must figure out what a momentum or a momentum density means for a classical electromagnetic field.

In the electromagnetic field  $(\mathcal{E}, \mathbf{B})$ , a charge  $q$  moving with velocity  $\mathbf{v}$  is subject to the Lorentz force (Sect. 5.4.1):

$$\mathbf{K} = q(\mathcal{E} + \mathbf{v} \times \mathbf{B}). \quad (8.35)$$

For many particles in the unit volume the charges add up to the charge density  $\rho$  and the quantity  $q\mathbf{v}$  to the current density  $\mathbf{j}$ . The total force as the time derivative of the mechanical momentum for a volume  $V$  is, thus, obtained as:

$$\frac{d}{dt} \mathbf{P}_{\text{mech}} = \int_V d^3r (\rho \mathcal{E} + \mathbf{j} \times \mathbf{B}). \quad (8.36)$$

Using the Maxwell relations

$$\rho = \operatorname{div}(\varepsilon_0 \mathcal{E}), \quad \mathbf{j} = \frac{1}{\mu_0} \operatorname{curl} \mathbf{B} - \varepsilon_0 \dot{\mathcal{E}} \quad (8.37)$$

we obtain the integrand in (8.36) as

$$\rho \mathcal{E} + \mathbf{j} \times \mathbf{B} = \varepsilon_0 (\mathcal{E} \operatorname{div} \mathcal{E} + \mathbf{B} \times \dot{\mathcal{E}} - c^2 \mathbf{B} \times \operatorname{curl} \mathbf{B}). \quad (8.38)$$

According to the chain rule of differentiation one concludes:

$$\mathbf{B} \times \dot{\mathcal{E}} = -\frac{\partial}{\partial t} (\mathcal{E} \times \mathbf{B}) + \mathcal{E} \times \dot{\mathbf{B}}. \quad (8.39)$$

By means of (8.39) equation (8.36) yields:

$$\frac{d}{dt} \mathbf{P}_{\text{mech}} = \varepsilon_0 \int d^3r \left[ \mathcal{E} \operatorname{div} \mathcal{E} - \frac{\partial}{\partial t} (\mathbf{B} \times \mathcal{E}) + \mathcal{E} \times \dot{\mathbf{B}} - c^2 \mathbf{B} \times \operatorname{curl} \mathbf{B} \right], \quad (8.40a)$$

$$\begin{aligned} \frac{d}{dt} \mathbf{P}_{\text{mech}} + \frac{d}{dt} \int d^3r \varepsilon_0 (\mathcal{E} \times \mathbf{B}) \\ = \varepsilon_0 \int d^3r (\mathcal{E} \operatorname{div} \mathcal{E} + \mathcal{E} \times \dot{\mathbf{B}} - c^2 \mathbf{B} \times \operatorname{curl} \mathbf{B}). \end{aligned} \quad (8.40b)$$

As is shown in textbooks of electrodynamics, the integral on the right side of (8.40b) can be rewritten into an integral over the surface area of the volume  $V$ . The integral, then, describes a momentum flux through this area. We do not give the detailed derivation at this point, since even without this derivation (8.40b) is recognized as an equation for momentum conservation in the volume  $V$ : The time derivative of the total momentum in  $V$  (both integral terms on the left) are compensated by momentum flux into and out of the volume  $V$ .

The momentum change consists of two contributions, one related to total mechanical momentum  $\mathbf{P}_{\text{mech}}$  of all moving particles and another field related quantity:

$$\mathbf{P}_{\text{field}} = \frac{1}{c^2} \int d^3r (\mathcal{E} \times \mathbf{H}). \quad (8.41)$$

This momentum is attributed to the electromagnetic field as a whole. We can define the following momentum density for the electromagnetic field:

$$\mathbf{p} = \frac{1}{c^2} (\mathcal{E} \times \mathbf{H}). \quad (8.42)$$

Apart from the factor  $1/c^2$  this expression is identical with that of the Pointing vector of the energy current density.

By replacing the field amplitudes  $\mathbf{E}$  and  $\mathbf{H}$  by their corresponding field operators one can calculate the quantum mechanical momentum of photons. For this purpose, we use the relations  $\mathbf{E} = -\dot{\mathbf{A}}$  and  $\mathbf{H} = \mu_0^{-1} \operatorname{curl} \mathbf{A}$  to rewrite the fields in (8.41) into the vector potential  $\mathbf{A}$ .

By use of the Fourier representations (8.7) and (8.12), the total field momentum ( $\mathbf{P}$  instead of  $\mathbf{P}_{\text{field}}$ ) in the volume  $V$  is obtained as:

$$\mathbf{P} = \varepsilon_0 \frac{1}{V} \sum_{\mathbf{q}, \mathbf{q}', \lambda, \lambda'} \int d^3r e^{i(\mathbf{q}+\mathbf{q}')} (-i) \dot{A}_{\mathbf{q}\lambda} A_{\mathbf{q}'\lambda'} (\mathbf{s}_{\mathbf{q}\lambda} [\mathbf{q}' \times \mathbf{s}_{\mathbf{q}'\lambda'}]). \quad (8.43)$$

For a sufficiently large volume, the integral transforms into a  $\delta$ -function and  $\mathbf{q}'$  changes into  $-\mathbf{q}$ , i.e.:

$$\mathbf{P} = \varepsilon_0 \sum_{\mathbf{q}\lambda} i \dot{A}_{\mathbf{q}\lambda} A_{-\mathbf{q}\lambda} (\mathbf{s}_{\mathbf{q}\lambda} \times [\mathbf{q} \times \mathbf{s}_{-\mathbf{q}\lambda}]). \quad (8.44)$$

The double cross product can be expressed as

$$\mathbf{s}_{\mathbf{q}\lambda} \times [\mathbf{q} \times \mathbf{s}_{-\mathbf{q}\lambda}] = \mathbf{q} (\mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{s}_{-\mathbf{q}\lambda}) - \mathbf{s}_{-\mathbf{q}\lambda} (\mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{q}). \quad (8.45)$$

Using furthermore the relation  $s_{\mathbf{q}\lambda} = s_{-\mathbf{q}\lambda}$  and  $s_{\mathbf{q}\lambda} \perp \mathbf{q}$ , we obtain the following expression for the field momentum:

$$\mathbf{P} = \varepsilon_0 \sum_{\mathbf{q}\lambda} i \mathbf{q} \dot{A}_{\mathbf{q}\lambda} A_{-\mathbf{q}\lambda} = i \varepsilon_0 \sum_{\mathbf{q}\lambda} \mathbf{q} \dot{A}_{\mathbf{q}\lambda} A_{\mathbf{q}\lambda}^*. \quad (8.46a)$$

The last expression in (8.46a) is a direct consequence of (8.17a), (8.17b). By means of (8.17a), (8.17b), one can express (8.46a) also as

$$\mathbf{P} = \varepsilon_0 \sum_{\mathbf{q}\lambda} \mathbf{q} \omega_{\mathbf{q}} (a_{\mathbf{q}\lambda} - a_{-\mathbf{q}\lambda}^*) (a_{\mathbf{q}\lambda}^* + a_{-\mathbf{q}\lambda}). \quad (8.46b)$$

At this point, we introduce the quantized field description by replacing the scalar variables  $a_{\mathbf{q}\lambda}$ ,  $a_{-\mathbf{q}\lambda}^*$ , etc. by non-commuting operators  $\hat{a}_{\mathbf{q}\lambda}$ ,  $\hat{a}_{-\mathbf{q}\lambda}^+$ , etc. By means of (8.19), we also introduce the creation and destruction operators  $\hat{b}_{\mathbf{q}\lambda}^+$  and  $\hat{b}_{\mathbf{q}\lambda}$ , which yields the following expression for the field momentum operator:

$$\begin{aligned} \hat{\mathbf{P}} &= \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} (\hat{b}_{\mathbf{q}\lambda} - \hat{b}_{-\mathbf{q}\lambda}^+) (\hat{b}_{\mathbf{q}\lambda}^+ + \hat{b}_{-\mathbf{q}\lambda}) \\ &= \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} (\hat{b}_{\mathbf{q}\lambda} \hat{b}_{-\mathbf{q}\lambda} - \hat{b}_{-\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda}^+ + \hat{b}_{\mathbf{q}\lambda} \hat{b}_{\mathbf{q}\lambda}^+ - \hat{b}_{-\mathbf{q}\lambda}^+ \hat{b}_{-\mathbf{q}\lambda}). \end{aligned} \quad (8.47)$$

The relation  $\hat{b}_{-\mathbf{q}\lambda} = \hat{b}_{\mathbf{q}\lambda}^+$ , a consequence of (8.17a), (8.17b), is used to rewrite the first two summands and in the last sum the summation index is chosen as  $\mathbf{q}$  instead of  $-\mathbf{q}$ . The result is

$$\hat{\mathbf{P}} = \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} (\hat{b}_{\mathbf{q}\lambda} \hat{b}_{\mathbf{q}\lambda}^+ - \hat{b}_{\mathbf{q}\lambda} \hat{b}_{\mathbf{q}\lambda}^+) + \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} (\hat{b}_{\mathbf{q}\lambda} \hat{b}_{\mathbf{q}\lambda}^+ + \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda}). \quad (8.48)$$

By means of the commutation relation, (8.20a), (8.20b) applied to the second sum we obtain

$$\hat{\mathbf{P}} = \frac{1}{2} \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} (2\hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + 1). \quad (8.49a)$$

Adding up all  $\hbar \mathbf{q}$  values yields equal numbers of positive and negative contributions ( $\pm \mathbf{q}$ ); they cancel each other and the momentum operator of the electromagnetic field is finally obtained as:

$$\hat{\mathbf{P}} = \sum_{\mathbf{q}\lambda} \hbar \mathbf{q} \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda}. \quad (8.49b)$$

To determine the momentum of a photon with wave vector  $\mathbf{q}'$  (in Chap. 2,  $\mathbf{k}$  was used instead of  $\mathbf{q}$ ) we calculate the eigenvalue of  $\hat{\mathbf{P}}$  in a field quantum state  $|\phi\rangle$  containing only one single photon of the type  $(\hbar\omega_{\mathbf{q}'}, \mathbf{q}')$ ; there is no other excited photon in the state, that is,

$$|\phi\rangle = |0, 0, \dots, \hbar\mathbf{q}', \dots, 0, 0, \dots\rangle = |\hbar\mathbf{q}'\rangle. \quad (8.50)$$

The eigenvalue equation reads:

$$\hat{\mathbf{P}} |\hbar\mathbf{q}'\rangle = \hbar\mathbf{q}' |\hbar\mathbf{q}'\rangle. \quad (8.51)$$

The  $\hat{b}_{\mathbf{q}\lambda}$  operators in the sum of (8.49b) generate a zero for all  $\mathbf{q} \neq \mathbf{q}'$ . Only for  $\mathbf{q} = \mathbf{q}'$  the eigenvalue  $\hbar\mathbf{q}'$  is obtained in (8.51).

A single photon excited in the field has the energy  $\hbar\omega_{\mathbf{q}'}$  and the momentum  $\hbar\mathbf{q}'$  just as was required in Chap. 2 for the interpretation of the basic experiments with light quanta. Using the classical expressions for field energy and momentum of Maxwell's theory quantization of the field yields the corresponding expressions for the photon, the field quantum of the electromagnetic field.

Also the bosonic character (integer spin) of photons is immediately concluded from the quantization, that is, the commutation relations of the creation and destruction operators of the field. For this purpose, we consider a two-photon state (photons  $\mathbf{q}\lambda$  and  $\mathbf{q}'\lambda'$ ) which is generated by the operation of the two corresponding creation operators on the vacuum ground state of the field:

$$|\mathbf{q}\lambda, \mathbf{q}'\lambda'\rangle = \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}'\lambda'}^+ |0\rangle. \quad (8.52a)$$

The two photons are interchanged and we obtain

$$|\mathbf{q}'\lambda', \mathbf{q}\lambda\rangle = \hat{b}_{\mathbf{q}'\lambda'}^+ \hat{b}_{\mathbf{q}\lambda}^+ |0\rangle. \quad (8.52b)$$

The creation and destruction operators are derived from the canonical variables of classical mechanics, respectively single particle quantum mechanics. Creation operators for two different particles mutually commute as do the momentum operators in single particle quantum mechanics, that is,

$$[\hat{b}_{\mathbf{q}\lambda}^+, \hat{b}_{\mathbf{q}'\lambda'}^+] = 0. \quad (8.53)$$

Subtraction of the two equations (8.52a), (8.52b), then, yields:

$$|\mathbf{q}\lambda, \mathbf{q}'\lambda'\rangle = |\mathbf{q}'\lambda', \mathbf{q}\lambda\rangle. \quad (8.54)$$

The two-photon field state keeps its sign upon interchange of the two photons. According to Sect. 5.6.2, this is the criterion for bosons: photons are bosons with integer spin.

In analogy to the momentum, also the spin of a photon can be evaluated by starting with classical Maxwell's theory, that is, by using the expression for the angular momentum density of the electromagnetic field  $\mathbf{l} = c^{-2} \mathbf{r} \times (\mathcal{E} \times \mathbf{H})$ . In this expression the continuous field variables must be replaced by operators which must obey the common commutation relations. We do not present the detailed calculation but rather report the major result: There is an angular momentum contribution parallel to the photon wave vector  $\mathbf{k}$ , which can assume the two values  $\pm\hbar$ . Since the angular momentum values are inherent to a single photon, the interpretation in term of a spin is obvious. The integer values  $\pm\hbar$  are consistent with the bosonic character of photons. The two spin orientations correspond to the two circular polarization states of the light wave, left and right handed around its wave vector  $\mathbf{q}$ .

The photon spin of  $\pm\hbar$  has already been concluded from the discussion of the transition matrix elements between electronic states of atoms (Sect. 6.4.4).

To summarize, photons are quantum mechanical excitations (particles) of the electromagnetic field: they have an energy  $\hbar\omega_{\mathbf{q}}$ , a momentum  $\hbar\mathbf{q}$  and two possible spin values  $\pm\hbar$  corresponding to the two circular polarizations around the wave vector  $\mathbf{q}$  of the electromagnetic wave.

### 8.2.2 2-Level Atom in the Light Field: Spontaneous Emission

As an application example for the quantized electromagnetic field we discuss the interaction of a 2-level system (atom) with the light field. In contrast to Sect. 6.5.1, we describe the 2-level atom in terms of operator algebra and the light field by the quantized formalism introduced in Sect. 8.2. In this context, we will encounter a new

phenomenon, the spontaneous emission of photons. This effect is inherently related to the quantization of the light field.

As in Sect. 6.5.1, the atom is simplified by a system with two electronic states  $|g\rangle$  and  $|e\rangle$  having energies  $E_g$  and  $E_e$ . Only these two states are relevant for the coupling to the light field. The transition energy between the two levels amounts to  $\hbar\omega = E_e - E_g$ . The atom is, thus, described in a 2-dimensional Hilbert space with basis vectors  $|g\rangle$  and  $|e\rangle$ . Electronic transitions between these states, that is, excitation of an electron from  $|g\rangle$  to  $|e\rangle$  or de-excitation from  $|e\rangle$  to  $|g\rangle$ , are formally identical with a spin flip process. The spin is also a 2-level system. Operators, which describe a spin flip, that is, the transition from one spin orientation into the other one, have already been introduced in Sect. 5.5.2 (5.113a), (5.113b) as combinations of spin matrices ( $\hat{\sigma}_x \pm i\hat{\sigma}_y$ ). In analogy to the spin we can, therefore, define step operators for the atom, which transform  $|g\rangle$  into  $|e\rangle$  and vice versa by

$$\hat{\sigma}^+ = \frac{1}{2}(\hat{\sigma}_x + i\hat{\sigma}_y), \quad (8.55a)$$

$$\hat{\sigma} = \frac{1}{2}(\hat{\sigma}_x - i\hat{\sigma}_y). \quad (8.55b)$$

A more general definition of the step operators  $\hat{\sigma}^+$  and  $\hat{\sigma}$  might be given in terms of “butterfly operators” formed from the state vectors  $|g\rangle$  and  $|e\rangle$ :

$$\hat{\sigma}^+ = |e\rangle\langle g|, \quad (8.56a)$$

$$\hat{\sigma} = |g\rangle\langle e|. \quad (8.56b)$$

From the following relations, it is obvious that the two states of the atom are linked by these operators:

$$\hat{\sigma}^+|g\rangle = |e\rangle\langle g|g\rangle = |e\rangle, \quad \hat{\sigma}^+|e\rangle = |e\rangle\langle g|e\rangle = 0, \quad (8.57a)$$

$$\hat{\sigma}|g\rangle = |g\rangle\langle e|g\rangle = 0, \quad \hat{\sigma}|e\rangle = |g\rangle\langle e|e\rangle = |g\rangle. \quad (8.57b)$$

The two states are, of course, supposed to be orthogonal, that is,  $\langle e|g\rangle = \langle g|e\rangle = 0$ .

Supposing the 2-level atom is not moving, that is, having zero kinetic energy, the Hamilton operator of the atom (4.72) can be expressed by its energy eigenvalues as follows:

$$\hat{H} = E_g|g\rangle\langle g| + E_e|e\rangle\langle e|. \quad (8.58)$$

This is easily shown by multiplying  $\hat{H}$  from the left and from the right by the unity operator  $\hat{1} = |g\rangle\langle g| + |e\rangle\langle e|$ . Frequently, the energy zero point is set at the energy  $E_g = 0$  of the ground state and with  $\hbar\omega = E_e$  as excitation energy into the state  $|e\rangle$  the atom Hamiltonian reads as

$$\hat{H} = E_e|e\rangle\langle e| = \hbar\omega\hat{\sigma}^+\sigma. \quad (8.59)$$

We now turn to the interaction with the electromagnetic field described by its vector potential  $\mathbf{A}$ . According to Sect. 5.4.2, the interaction Hamilton operator between an electron (charge  $-e$ ) and the  $\hat{\mathbf{A}}$  field for sufficiently weak electromagnetic fields (5.67) is given by

$$\hat{W} = -\frac{e}{m} \hat{\mathbf{A}} \cdot \hat{\mathbf{p}}. \quad (8.60)$$

According to Sect. 6.4.2 (6.119), the electron momentum operator  $\mathbf{p}$  of the atom can be replaced by the position operator:

$$\hat{\mathbf{p}} = i\omega \hat{\mathbf{r}}. \quad (8.61)$$

From the Fourier representation of the vector potential  $\mathbf{A}$  of the light field, we obtain by the use of (8.17a):

$$\mathbf{A} = \frac{1}{\sqrt{V}} \sum_{\mathbf{q}\lambda} a_{\mathbf{q}\lambda} \mathbf{s}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + \frac{1}{\sqrt{V}} \sum_{\mathbf{q}\lambda} a_{-\mathbf{q}\lambda}^* \mathbf{s}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}}. \quad (8.62a)$$

Replacing the summation over  $\mathbf{q}$  by  $-\mathbf{q}$  in the second sum yields

$$\mathbf{A} = \frac{1}{\sqrt{V}} \sum_{\mathbf{q}\lambda} \mathbf{s}_{\mathbf{q}\lambda} (a_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + a_{\mathbf{q}\lambda}^* e^{-i\mathbf{q}\cdot\mathbf{r}}). \quad (8.62b)$$

We replace the scalar variables  $a_{\mathbf{q}\lambda}$ ,  $a_{\mathbf{q}\lambda}^*$  by the corresponding quantized (non-commuting) operators and use the representation (8.19) in terms of photon creation and destruction operators. By means of these operators, the vector potential operator of the light field is obtained as:

$$\hat{\mathbf{A}} = \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\mathbf{q}\lambda} \frac{1}{\sqrt{\omega_{\mathbf{q}}}} \mathbf{s}_{\mathbf{q}\lambda} (\hat{b}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + \hat{b}_{\mathbf{q}\lambda}^+ e^{-i\mathbf{q}\cdot\mathbf{r}}). \quad (8.63)$$

Equation (8.63) inserted into the interaction operator (8.60) yields

$$\hat{W} = -\frac{e}{m} \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\mathbf{q}\lambda} \frac{\mathbf{s}_{\mathbf{q}\lambda} \cdot \hat{\mathbf{p}}}{\sqrt{\omega_{\mathbf{q}}}} (\hat{b}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + \hat{b}_{\mathbf{q}\lambda}^+ e^{-i\mathbf{q}\cdot\mathbf{r}}), \quad (8.64a)$$

and finally rewritten in terms of the position operator by means of (8.61):

$$\hat{W} = -ie\omega \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\mathbf{q}\lambda} \frac{\mathbf{s}_{\mathbf{q}\lambda} \cdot \hat{\mathbf{r}}}{\sqrt{\omega_{\mathbf{q}}}} (\hat{b}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + \hat{b}_{\mathbf{q}\lambda}^+ e^{-i\mathbf{q}\cdot\mathbf{r}}). \quad (8.64b)$$

The product  $(-e\mathbf{r})$  in this expression represents the electrical dipole moment  $\hat{\mathbf{d}}$  of the 2-level atom. As discussed in Sect. 6.4.2 this dipole moment related to electronic transitions mediates the coupling of the atom to the electromagnetic field. According to (4.72), any operator can be represented by its matrix elements in an orthonormal system of eigenstates. The Hilbert space of the 2-level system is spanned by the states  $|g\rangle$  and  $|e\rangle$ . Consequently, we obtain the representation of  $\hat{\mathbf{d}}$  by multiplying the operator from the left and from the right by the unity operator (4.69a):

$$\begin{aligned}\hat{\mathbf{d}} &= (|e\rangle\langle e| + |g\rangle\langle g|)\hat{\mathbf{d}}(|e\rangle\langle e| + |g\rangle\langle g|) \\ &= |e\rangle\langle e|\hat{\mathbf{d}}|e\rangle\langle e| + |e\rangle\langle e|\hat{\mathbf{d}}|g\rangle\langle g| \\ &\quad + |g\rangle\langle g|\hat{\mathbf{d}}|e\rangle\langle e| + |g\rangle\langle g|\hat{\mathbf{d}}|g\rangle\langle g|.\end{aligned}\tag{8.65a}$$

The matrix elements  $\langle e|\hat{\mathbf{d}}|e\rangle$  and  $\langle g|\hat{\mathbf{d}}|g\rangle$  vanish, as the integrals over positive and negative integration regions cancel each other for identical final and initial states (because of  $\mathbf{r}$  in the integrand). Equation (8.65a) thus simplifies to

$$\hat{\mathbf{d}} = |e\rangle\langle e|\hat{\mathbf{d}}|g\rangle\langle g| + |g\rangle\langle g|\hat{\mathbf{d}}|e\rangle\langle e|.\tag{8.65b}$$

Note that the matrix elements obey the relations  $\langle e|\hat{\mathbf{d}}|g\rangle = \mathbf{d}$  and  $\langle g|\hat{\mathbf{d}}|e\rangle = \mathbf{d}^*$ . Using the representation of the step operators of the 2-level system yields:

$$\hat{\mathbf{d}} = \mathbf{d}\hat{\sigma}^+ + \mathbf{d}^*\hat{\sigma}.\tag{8.66}$$

The interaction operator (8.64a), (8.64b) can then be written as

$$\hat{W} = i\omega \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\mathbf{q}\lambda} \frac{1}{\sqrt{\omega_{\mathbf{q}}}} (\mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{d}\hat{\sigma}^+ + \mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{d}^*\hat{\sigma}) (\hat{b}_{\mathbf{q}\lambda} e^{i\mathbf{q}\cdot\mathbf{r}} + \hat{b}_{\mathbf{q}\lambda}^+ e^{-i\mathbf{q}\cdot\mathbf{r}}).\tag{8.67}$$

The total Hamiltonian of the 2-level atom imbedded in the radiation field (Jaynes–Cummings Hamiltonian [21]) is composed of contributions of the atom  $\hat{H}_{\text{at}}$  (8.59), of the radiation field (light:  $L$ )  $\hat{H}_L$  (8.22) and of the interaction operator  $\hat{W}$  (8.67), i.e.

$$\hat{H} = \hat{H}_{\text{at}} + \hat{H}_L + \hat{W} = \hbar\omega\hat{\sigma}^+\hat{\sigma} + \sum_{\mathbf{q}\lambda} \hbar\omega_{\mathbf{q}} \left( \hat{b}_{\mathbf{q}\lambda}^+ \hat{b}_{\mathbf{q}\lambda} + \frac{1}{2} \right) + \hat{W}.\tag{8.68}$$

The quantum states of this composed system contain information about the state of the atom ( $v = g, e$ ) and about the state of the electromagnetic field, i.e. the information how many photons of each type ( $\hbar\omega_{\mathbf{q}\lambda}$ ,  $\mathbf{q}\lambda$ ) are excited. The composed quantum state of the atom plus field might, thus, be written as

$$|\phi\rangle = |v, \dots, n_{\mathbf{q}\lambda}, \dots\rangle.\tag{8.69}$$

Now we consider the action of the interaction operator  $\hat{W}$  (8.67) on this state. The action of the operator product  $\hat{\sigma}^+ \hat{b}_{\mathbf{q}\lambda}^+$  means simultaneous creation of a photon and excitation of the atom from  $|g\rangle$  to  $|e\rangle$ . The latter process would need a second photon. Since the photon creation and destruction operators exhibit a time dependence  $\exp(\pm i\omega_{\mathbf{q}} t)$  this process requires the consideration of the double photon frequency, as does the process belonging to the operator combination  $\hat{\sigma} \hat{b}_{\mathbf{q}\lambda}$ . In the sense of the rotating wave approximation (Sect. 6.5.1), terms containing two photons are neglected and by multiplication of the remaining one-photon terms in (8.67) the interaction operator is obtained as

$$\hat{W} = i\omega \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\mathbf{q}\lambda} \frac{1}{\sqrt{\omega_{\mathbf{q}}}} (e^{i\mathbf{q}\cdot\mathbf{r}} \mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{d} \hat{\sigma}^+ \hat{b}_{\mathbf{q}\lambda} + e^{-i\mathbf{q}\cdot\mathbf{r}} \mathbf{s}_{\mathbf{q}\lambda} \cdot \mathbf{d}^* \hat{\sigma} \hat{b}_{\mathbf{q}\lambda}^+). \quad (8.70)$$

In the reasonable limit that the atom dimensions are small in comparison to the light wavelength the calculation of the transition matrix elements (6.111a), (6.111b), that is, the integration over the atom volume, can be performed with the assumption  $\exp(\pm i\mathbf{q} \cdot \mathbf{r}) \approx 1$ .

For simplicity reason, we consider a one-dimensional problem in the following and only one single sharp spectral line with the frequency  $\omega_{\mathbf{q}} = \omega_L$ . The realization might be a laser beam in a resonator. In this case, the interaction operator is written as

$$\hat{W} = D \hat{\sigma}^+ \hat{b}_L + D^* \hat{\sigma} \hat{b}_L^+. \quad (8.71)$$

Setting

$$D = i \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \frac{\omega}{\sqrt{\omega_L}} d \quad (8.72)$$

we have introduced a generalized dipole moment. The photon indices  $\mathbf{q}\lambda$  have been replaced by the subscript  $L$  (for light) as only one single type of photons, those with  $\omega_L$ , are used. The combined states of the total system atom plus light field (8.69), then, read  $|v, n_L\rangle$  with  $v = g, e$ .

For the calculation of transition rates between states of the total system atom plus light field Fermi's golden rule (6.111a), (6.111b) is used as it was discussed in Sect. 6.4.1. In the transition matrix elements  $\langle f | \hat{W} | i \rangle$ , now, the initial ( $i$ ) and final ( $f$ ) states of the combined system atom plus field appear. Application of the operator  $\hat{W}$  on a state  $|v, n_L\rangle$  causes either de-excitation of the atom by an energy  $\hbar\omega$  with simultaneous increase of the phonon number  $n_L$  by one [2nd term in (8.71)] or excitation of the atom from  $|g\rangle$  to  $|e\rangle$  with simultaneous destruction of one photon  $\hbar\omega_L$  [1st term in (8.71)]. Consequently, only two types of matrix elements  $\langle f | \hat{W} | i \rangle$  are relevant for the calculation of the transition rates and by use of (8.33a), (8.33b) they are obtained as

$$\begin{aligned}\langle f | \hat{W} | i \rangle &= \langle g, n_L + 1 | D^* \hat{\sigma} \hat{b}_L^+ | e, n_L \rangle \\ &= D^* \langle g, n_L + 1 | g, n_L + 1 \rangle \sqrt{n_L + 1},\end{aligned}\quad (8.73a)$$

$$\begin{aligned}\langle f | \hat{W} | i \rangle &= \langle e, n_L - 1 | D \hat{\sigma}^+ \hat{b}_L | g, n_L \rangle \\ &= D \langle e, n_L - 1 | e, n_L - 1 \rangle \sqrt{n_L}.\end{aligned}\quad (8.73b)$$

Equation (8.73a) describes the emission of a photon  $\hbar\omega_L$  with simultaneous de-excitation of the atom from  $|e\rangle$  to  $|g\rangle$ , while (8.73b) is the expression for photon absorption from the light field by the atom.

According to Fermi's golden rule the transition rates (6.111b) are, thus, obtained as

$$R_{\text{em}} = \frac{2\pi}{\hbar} |D|^2 (n_L + 1) \delta(E_f^{\text{tot}} - E_i^{\text{tot}}) \quad (8.74a)$$

for emission of a photon from the excited atomic state  $|e\rangle$  and

$$R_{\text{abs}} = \frac{2\pi}{\hbar} |D|^2 n_L \delta(E_f^{\text{tot}} - E_i^{\text{tot}}) \quad (8.74b)$$

for absorption of a photon by the atom in its ground state  $|g\rangle$ .

For the final and initial state energies  $E_f^{\text{tot}}$  and  $E_i^{\text{tot}}$  of the combined system, atom and light field taken together, the following balances are obtained:

$$\begin{aligned}\text{Emission: } E_f^{\text{tot}} - E_i^{\text{tot}} &= E_g + (n + 1)\hbar\omega_L - (E_e + n\hbar\omega_L) \\ &= E_g - E_e + \hbar\omega_L\end{aligned}\quad (8.75a)$$

$$\begin{aligned}\text{Absorption: } E_f^{\text{tot}} - E_i^{\text{tot}} &= E_e + (n - 1)\hbar\omega_L - (E_g + n\hbar\omega_L) \\ &= E_e - E_g - \hbar\omega_L\end{aligned}\quad (8.75b)$$

According to the discussion in Sect. 6.4.1 the  $\delta$ -functions in (8.74a), (8.74b) represent the energy conservation  $E_f^{\text{tot}} - E_i^{\text{tot}} = 0$  during the emission and the absorption process for the combined total system. The energy conservation is given with the spectral sharpness of the  $\delta$ -Function, in resonance with  $\omega = \omega_L$ . By comparing (8.74a) with (8.74b), we realize that the following relation is valid for part of the transition rates:

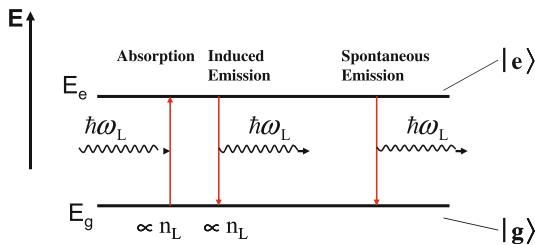
$$R_{\text{em}}^{\text{stim}} = R_{\text{abs}}^{\text{stim}} \propto \frac{2\pi}{\hbar} |D|^2 n_L \quad (8.76a)$$

In the emission process, an additional term exists:

$$R_{\text{em}}^{\text{spon}} \propto \frac{2\pi}{\hbar} |D|^2 \quad (8.76b)$$

The three different processes are depicted in Fig. 8.1.

Since the photon number  $n_L$  is proportional to the intensity of the electromagnetic field, we regain the results of Sect. 6.4.2 in (8.76a), namely that the absorption and



**Fig. 8.1** Schematic representation of possible interactions between a 2-level system with quantum states  $|g\rangle$ ,  $|e\rangle$  and an electromagnetic field being in resonance with the electronic transition  $|g\rangle \leftrightarrow |e\rangle$ . In the photon absorption process the electron is excited from the ground state energy  $E_g$  to the energy  $E_e$  of the excited state; transition intensity (probability proportional to photon density  $n_L$ ). In stimulated (induced) emission a transition from  $E_e$  to  $E_g$  is induced and a photon of the energy  $\hbar\omega_L$  is emitted; transition intensity proportional to  $n_L$ . In the spontaneous emission process an electron is de-excited from  $E_e$  to  $E_g$  without stimulation by an external light field. Simultaneously a photon of energy  $\hbar\omega_L$  is emitted

the stimulated emission process have the same transition rate which is proportional to the light intensity (6.116a), (6.116b). Both processes are triggered by photons in the light field.

Quantization of the light field has generated a new term in the transition rates for emission (8.76a), (8.76b). This term describes the so-called *spontaneous emission*, a process which is missing in the semi-quantum mechanical description neglecting field quantization (Sect. 6.4.2). Spontaneous emission of photons does not depend on the photon number in the electromagnetic field. This process also occurs for  $n_L = 0$ , i.e. without any excited photons and the light field being in its ground state  $|0\rangle$ .

Spontaneous emission of photons is inherently related to the quantization of the electromagnetic field; it might be traced back to the (random) interaction of an excited atom with the vacuum state of the electromagnetic field. One might also consider spontaneous emission as an emission process which is stimulated by random vacuum fluctuations of the electromagnetic field.

### 8.2.3 Atom Diffraction by Light Waves

In Sect. 2.4.2, we have discussed an experiment having essential importance for quantum behavior, namely the diffraction of a Rb atom beam on a high intensity standing laser light wave. By the diffraction process, the atomic beam is split into a transmitted and a diffracted (1st order) beam [8]. Both beams interfere with each other after a second diffraction process and the experiment allows the observation of the impact of Which-Way-Information on the double slit interference of atoms. Details of the diffraction process could not be explained in this earlier context.

In the following, we will see that diffraction of atoms by a high-intensity standing light wave is a direct manifestation of the quantum character of the electromagnetic

field. As in Sect. 8.2.2, we consider the interaction of a 2-level system (atom) with the electromagnetic (light) field described by the Hamilton operator (8.68). The zero point of the energy scale shall be assumed such that the vacuum state energy of the field  $\sum_{\mathbf{q}\lambda}(1/2)\hbar\omega_{\mathbf{q}}$  vanishes. As in Sect. 8.2.2, only one single spectrally sharp light wave with  $\omega_L, n_L$  is considered in a one-dimensional problem. Using the expression (8.71) for the interaction Hamiltonian  $\hat{W}$  the total Hamilton operator is obtained as:

$$\hat{H} = \hbar\omega\hat{\sigma}^+\hat{\sigma} + \hbar\omega_L\hat{b}_L^+\hat{b}_L + (D\hat{\sigma}^+\hat{b}_L + D^*\hat{\sigma}\hat{b}_L^+). \quad (8.77a)$$

Allowing a detuning  $\Delta = \omega_L - \omega$  between the laser frequency  $\omega_L$  and the angular frequency  $\omega = (E_e - E_g)/\hbar$  of the electronic transition of the 2-level atom (8.77a) is written as

$$\hat{H} = \hbar(\omega_L - \Delta)\hat{\sigma}^+\hat{\sigma} + \hbar\omega_L\hat{b}_L^+\hat{b}_L + (D\hat{\sigma}^+\hat{b}_L + D^*\hat{\sigma}\hat{b}_L^+). \quad (8.77b)$$

Unlike in Sect. 8.2.2, we can add the phase factors  $\exp(\pm iqx)$  neglected in (8.70) to the effective dipole moments  $D$  and  $D^*$  (8.72). These factors, however, would not modify the following calculation, since the effective dipole moments only enter as  $|D|^2$ .

We consider the action of  $\hat{H}$  (8.77b) on a general state  $|v, n_L\rangle$  ( $v = e, g$ ) of the combined system atom plus light field. The interaction term  $\hat{W}$  decreases the photon number  $n_L$  with simultaneous excitation of the atom or vice versa. Therefore, only the following two states of the system are of interest:

$$|g, n_L + 1\rangle \leftrightarrow |e, n_L\rangle. \quad (8.78)$$

The most general state of the system to be considered is then:

$$|\phi\rangle = c_g|g, n_L + 1\rangle + c_e|e, n_L\rangle. \quad (8.79)$$

In this two-dimensional basis, we must solve the stationary Schrödinger equation using the Hamilton operator (8.77b):

$$\hat{H}|\phi\rangle = E|\phi\rangle. \quad (8.80)$$

For the purpose of using the matrix formalism in the calculation, we insert (8.79) into (8.80) and multiply from the left with  $\langle g, n_L + 1|$ , respectively  $\langle e, n_L|$ . The calculation yields:

$$\begin{aligned} \langle g, n_L + 1|\hat{H}|g, n_L + 1\rangle c_g + \langle g, n_L + 1|\hat{H}|e, n_L\rangle c_e \\ = E\langle g, n_L + 1|g, n_L + 1\rangle c_g, \end{aligned} \quad (8.81a)$$

$$\begin{aligned} \langle e, n_L|\hat{H}|g, n_L + 1\rangle c_g + \langle e, n_L|\hat{H}|e, n_L\rangle c_e \\ = E\langle e, n_L|e, n_L\rangle c_e. \end{aligned} \quad (8.81b)$$

Taking into account the action of the operators  $\hat{b}_L^+, \hat{b}_L$  on the light field (8.33a), (8.33b) and that of  $\hat{\sigma}^+, \hat{\sigma}$  on the states of the 2-level atom (8.57a), (8.57b) one obtains the following relations:

$$\langle g, n_L + 1 | \hat{H} | g, n_L + 1 \rangle = \hbar\omega_L(n_L + 1), \quad (8.82a)$$

$$\langle g, n_L + 1 | \hat{H} | e, n_L \rangle = D^* \sqrt{n_L + 1}, \quad (8.82b)$$

$$\langle e, n_L | \hat{H} | e, n_L \rangle = \hbar(\omega_L - \Delta) + \hbar\omega_L n_L, \quad (8.82c)$$

$$\langle e, n_L | \hat{H} | g, n_L + 1 \rangle = D \sqrt{n_L + 1}. \quad (8.82d)$$

By use of (8.22), the relations (8.81a), (8.81b) can be written as a matrix equation representing the Schrödinger equation:

$$\begin{pmatrix} \hbar\omega_L(n_L + 1) & D^* \sqrt{n_L + 1} \\ D \sqrt{n_L + 1} & \hbar\omega_L(n_L + 1) - \hbar\Delta \end{pmatrix} \begin{pmatrix} c_g \\ c_e \end{pmatrix} = E \begin{pmatrix} c_g \\ c_e \end{pmatrix}. \quad (8.83a)$$

For symmetry reasons, we rearrange the different matrix elements as follows:

$$\begin{bmatrix} \left( (n_L + 1)\hbar\omega_L - \frac{1}{2}\hbar\Delta \right) & 0 \\ 0 & \left( (n_L + 1)\hbar\omega_L - \frac{1}{2}\hbar\Delta \right) \\ + \left( \begin{array}{cc} \frac{1}{2}\hbar\Delta & D^* \sqrt{n_L + 1} \\ D \sqrt{n_L + 1} & -\frac{1}{2}\hbar\Delta \end{array} \right) \end{bmatrix} \begin{pmatrix} c_g \\ c_e \end{pmatrix} = E \begin{pmatrix} c_g \\ c_e \end{pmatrix}. \quad (8.83b)$$

The first matrix on the left contains only the total energy of the combined system, light field plus atom (conserved quantity), and the detuning  $\Delta$ . For sufficiently small detuning spontaneous emission might be neglected (Sect. 8.2.2) and the matrix elements can be assumed as constant. We shift the energy zero point into the value of these constant matrix elements and obtain the simplified Schrödinger matrix equation

$$\begin{pmatrix} \frac{1}{2}\hbar\Delta - E & D^* \sqrt{n_L + 1} \\ D \sqrt{n_L + 1} & -\frac{1}{2}\hbar\Delta - E \end{pmatrix} \begin{pmatrix} c_g \\ c_e \end{pmatrix} = 0. \quad (8.84)$$

The two energy eigenvalues  $E$  of the 2-dimensional problem follow from the vanishing determinant:

$$\det \begin{vmatrix} \frac{1}{2}\hbar\Delta - E & D^* \sqrt{n_L + 1} \\ D \sqrt{n_L + 1} & -\frac{1}{2}\hbar\Delta - E \end{vmatrix} = 0. \quad (8.85)$$

This equation is solved by the eigenvalues

$$E_{\pm} = \pm \frac{1}{2}\hbar \sqrt{\Delta^2 + \frac{4}{\hbar^2} |D|^2 (n_L + 1)}. \quad (8.86)$$

The number of photons  $n_L$  in the laser beam is proportional to the intensity (energy density) of the electromagnetic field. In a standing wave between two mirrors with locally fixed nodes and bellies, therefore,  $n_L(x)$  is a position dependent function. For large photon numbers the quantity  $[n_L(x) + 1]$  in (8.86) can be expressed in terms of a spatially periodic electric field  $\mathcal{E}(x)$  by using the expression for the energy density of the field:

$$\frac{1}{2}\varepsilon_0\mathcal{E}^2 = \frac{1}{V}(n_L + 1)\hbar\omega_L. \quad (8.87)$$

For the generalized dipole moment, we use the definition (8.72) and the detuning is assumed to be approximately zero, that is,  $\omega_L \approx \omega$ .  $\Delta$  is certainly small as compared to  $\omega_L$  and  $\omega$ . The energy eigenvalues (8.86) can then be written as

$$E_{\pm} = \pm \frac{1}{2}\hbar\sqrt{\Delta^2 + \frac{1}{\hbar^2}d^2\mathcal{E}^2(x)}. \quad (8.88)$$

Here,  $d = \langle e|ex|g \rangle$  is the dipole moment between ground and excited state of the 2-level atom and  $\mathcal{E}(x)$  is the oscillating electric field of the standing laser light wave with locally fixed nodes.

A comparison of (8.88) with (6.154) reveals the second term in the square root to be the squared Rabi frequency. If the light field is in resonance with the excitation energy of the atom ( $\Delta = 0$ ), the eigenvalues (8.88) represent the positive and negative Rabi frequencies (6.154). For the case of finite detuning ( $\Delta \neq 0$ ), a generalized Rabi frequency can be defined by:

$$\Omega_{\text{gen}} = \sqrt{\Delta^2 + \frac{1}{\hbar^2}d^2\mathcal{E}^2}. \quad (8.89)$$

This frequency describes how fast the 2-level atom oscillates between the states  $|g, n_L + 1\rangle$  and  $|e, n_L\rangle$  (Rabi flopping) in case that the photon energy  $\hbar\omega_L$  of the light field is detuned by  $\Delta$  with respect to the excitation energy of the atom. This phenomenon of Rabi-flopping was already described in the approximation of the semi-classical non-quantized electromagnetic field in Sect. 6.5.1 (Fig. 6.12).

Corresponding to the eigenvalues  $E_{\pm}$  (8.86), (8.88) there are two eigenvectors  $(c_g^{\pm}, c_e^{\pm})$  as solutions to (8.84). They are calculated by inserting  $E_{\pm}$  (8.86) into (8.84), solving for  $c_g$ , respectively  $c_e$  and normalizing the obtained vectors. Without presenting the detailed calculation, the obtained eigenvector components are given as [9]:

$$c_g^{\pm} = \sqrt{1 \pm \Delta/\Omega_{\text{gen}}}, \quad c_e^{\pm} = \sqrt{1 \mp \Delta/\Omega_{\text{gen}}}. \quad (8.90)$$

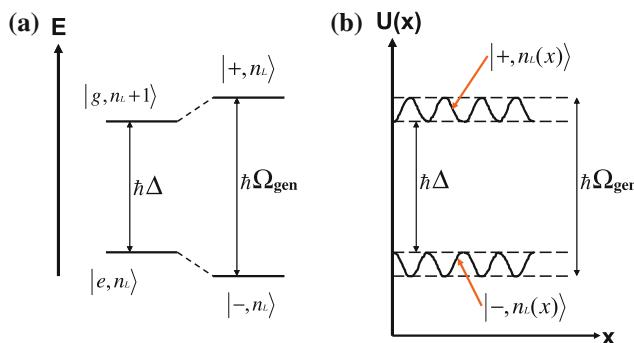
Similarly as in Sect. 6.5.1 linear superposition allows the generation of two new eigenstates:

$$|\pm, n_L\rangle = (c_g^{\pm}|g, n_L + 1\rangle \pm c_e^{\pm}|e, n_L\rangle) \frac{1}{\sqrt{2}}. \quad (8.91)$$

The eigenstates  $|\pm, n_L\rangle$  are called *dressed states*, since they are used for the description of the atom-light interaction. The energy eigenvalues calculated on their basis contain an energy amount due to atom-light interaction. In contrast, the states  $|g, n_L + 1\rangle$  and  $|e, n_L\rangle$  are states which merely describe the combined system atom plus light field, without taking notice of their interaction; they are called *bare states*.

The bare states belong to energy eigenvalues, which are obtained from (8.86) by neglecting the second interaction term under the square root. The energetic distance between the bare state levels, thus, amounts to  $E_+ - E_- = \hbar\Delta$ , while for the dressed states the energy level distance is  $\hbar\Omega_{\text{gen}}$  according to (8.86), (8.88), (8.89). Note that in the resonance case, where the photon energy of the light field matches the transition energy between ground and excited state of the atom, the energies of the two bare states  $|g, n_L + 1\rangle$  and  $|e, n_L\rangle$  are identical. This is not astonishing because states of the total, combined system are concerned. Atomic transition energies are balanced by excitation energies of the electromagnetic field ( $|n_L\rangle \leftrightarrow |n_L + 1\rangle$ ). Transitions between the dressed states described by the transition energy  $\hbar\Omega_{\text{gen}}$  belong to higher energies, since they contain the interaction energy between atom and light field. The situation is qualitatively shown in Fig. 8.2a.

In the Schrödinger representation (Sects. 3.5, 4.3.5), the time dependence of states is contained in the states themselves, that is, in the probability amplitudes  $c_g^\pm, c_e^\pm$ . According to Sect. 3.5, this time dependence is determined by the energy eigenvalues (8.88) of the time-independent problem [solution of the time-independent Schrödinger equation (8.80)]:



**Fig. 8.2** **a, b** Bare and dressed electronic states of a 2-level system (atom) in a quantized monochromatic electromagnetic field. The photon energy of the light field  $\hbar\omega_L$  is detuned by an amount  $\hbar\Delta$  with respect to the transition energy  $\hbar\omega$  between the states  $|g\rangle$  and  $|e\rangle$  of the 2-level system. **a** Energy levels of the bare (left) and of the dressed (right) states of the total system 2-level atom plus light field. The dressed state levels are shifted with respect to the bare levels by the interaction between atom and light field.  $\Omega_{\text{gen}}$  is the generalized Rabi frequency which is determined by the detuning  $\Delta$ . **b** Energies of the dressed states  $|+, n_L(x)\rangle$  and  $|-, n_L(x)\rangle$  as function of position  $x$  in a standing light wave with frequency  $\omega_L$ . The oscillating energy levels act on the 2-level atom like a potential  $U(x)$

$$c_g^\pm \propto \exp\left(\pm \frac{i}{2} \Omega_{\text{gen}} t\right). \quad (8.92)$$

Because of (8.91), this is also the time dependence of the dressed states. Note that according to (8.88), (8.89) for a standing laser light wave in a resonator (locally fixed nodes)  $\Omega_{\text{gen}}(x)$  depends on the spatial position (Fig. 8.2b). Assume that the laser light acts on the 2-level atom only during a short pulse time  $\tau$ , but with a sufficiently large detuning  $\Delta$ . In that case, we can have

$$\Omega_{\text{gen}} \tau \approx \Delta \tau > 1, \quad \text{or} \quad \Delta > 1/\tau \quad (8.93)$$

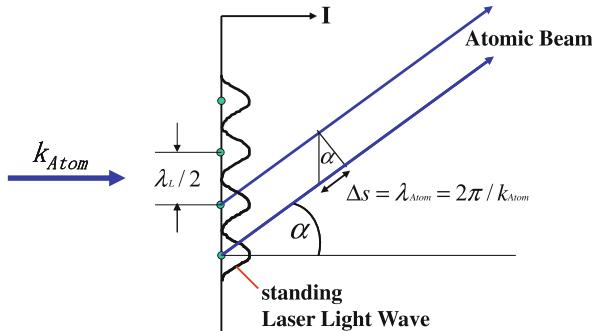
and according to (8.92) switching between the two atomic levels is quite rapid.

The bare initial state  $|g, n_L + 1\rangle$  adiabatically switches into one of the two dressed states (8.91), i.e. nearly instantaneously in comparison with the time change of the energy eigenvalues (8.88). The eigenvalues can be considered as a light-atom potential  $U(x)$  for the atom dynamics [9]. According to (8.88) and to Fig. 8.2b this potential is spatially periodic in  $x$  because of the spatial structure of the standing wave. On the 2-level atoms passing the light wave it acts like a diffraction grating.

A standing laser light wave with wave length  $\lambda_L$  has nodes at a distance  $\lambda_L/2$  (Fig. 8.3). This is the periodicity distance of the diffraction grating. According to common rules for diffraction on a grating (Fig. 8.3) the path difference of the atom beam diffracted into the 1st maximum amounts to  $\Delta s = \lambda_{\text{atom}} = 2\pi/k_{\text{atom}}$ . Observation of this maximum under a diffraction angle  $\alpha$  requires:

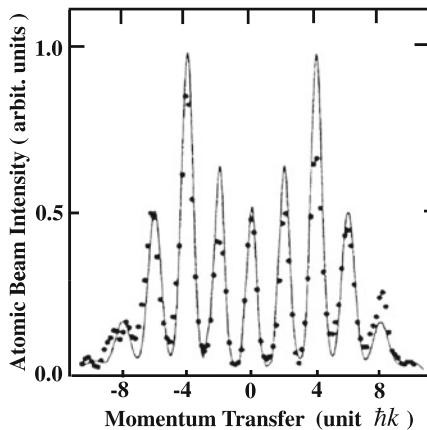
$$\lambda_{\text{atom}} \sin \alpha = \lambda_L/2. \quad (8.94)$$

Diffraction of the atom beam into the 1st diffraction maximum is, thus, related to a momentum transfer to the atoms ( $q_L$  wave vector of laser light), perpendicular to the incident beam direction of



**Fig. 8.3** Schematic representation of the diffraction of a 2-level atomic beam (wave vector  $\mathbf{k}_{\text{atom}}$ ) on the periodic grating of a standing light wave (wave length  $\lambda_L$ ), the frequency of which is detuned with respect to the energy of the electronic transition  $|g\rangle \leftrightarrow |e\rangle$  of the 2-level atom (see Fig. 8.2b)

**Fig. 8.4** Atom beam interference pattern of He atoms which are diffracted by a standing light wave (see Figs. 8.2b and 8.3). Generalized Rabi frequency  $\Omega_{\text{gen}} = 150 \text{ MHz}$ ; Detuning  $\Delta = 120 \text{ MHz}$  [9]



$$\delta p = \hbar k_{\text{atom}} \sin \alpha = 2\hbar q_L. \quad (8.95)$$

In Fig. 8.4, the experimentally measured diffraction pattern of He atoms diffracted from a standing laser light wave ( $\lambda_L = 1083 \text{ nm}$ ) is shown as an example [9]. The diffraction pattern is plotted as a function of the perpendicular momentum transfer (8.95) in units of  $\hbar k$ . The experimental parameters Rabi frequency of 150 MHz, detuning  $\Delta$  between atomic resonance and laser frequency of 120 MHz and interaction time of atom beam with standing laser wave of  $\tau = 20 \text{ ns}$  fulfill the condition (8.93) for diffraction on the light grating.

#### 8.2.4 Once Again: “Which Way” Information and Entanglement

After we have learnt more details about quantized light fields and diffraction of a 2-level atom on standing light waves, the fundamental experiment of Dürr et al. [8, 10] about the destruction of double slit interferences by observation of the detailed particle path (Sect. 2.4.2) can be understood more thoroughly.

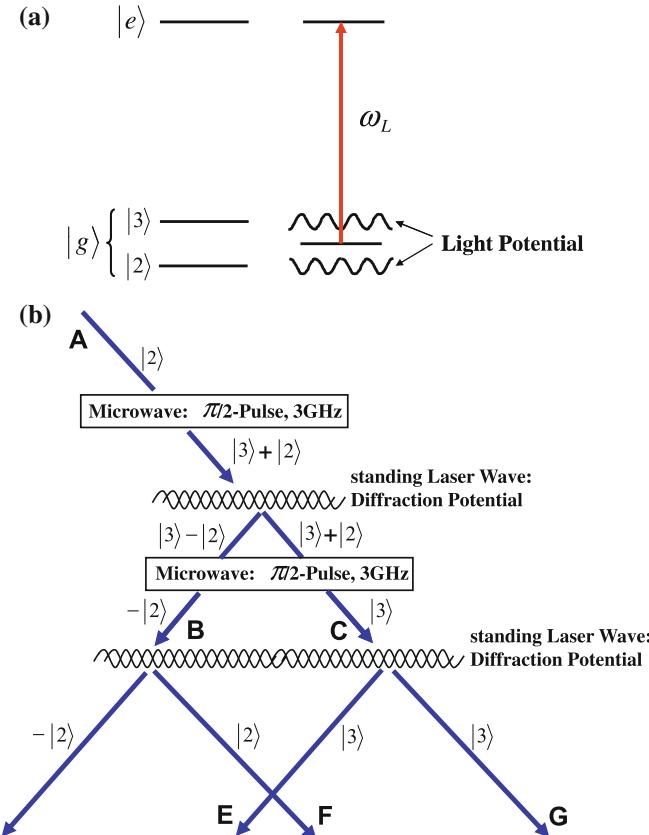
According to Sect. 8.2.3, it is obvious that the standing laser light wave in Fig. 2.10a and in Fig. 8.5b splits the incident Rb atom beam into a transmitted beam ( $C$ ) and into a beam diffracted into 1st order ( $B$ ). The necessary condition is the matching of the light frequency  $\omega_L$  to the transition frequency  $\omega$  between two significant energy levels of the atoms (2-level system). Apart from the Rabi frequency, that is, essentially the atomic transition dipole moment multiplied with the electric field strength of the light field, the detuning  $\Delta = \omega_L - \omega$  between light frequency and atomic transition frequency determines the periodic potential of the diffraction grating. After the atoms have passed the standing light wave two times splitting into the two partial beams  $B$  and  $C$  (Figs. 2.10a and 8.5b) and subsequently (2nd grating)

into the beams  $D$  and  $F$ , respectively  $E$  and  $G$  is achieved (Fig. 2.10a). The experimental parameters, in particular the light intensity, are adjusted that the beam splitter produce a 50 % partitioning of the atomic beams. The atoms in the beam propagate with such a low velocity of 2 m/s that repeated switch on and off of the laser beam (switched on during the interaction time between atoms and light) allows the use of only one single standing wave for the two subsequent diffraction processes. Between two light pulses realizing the 1st and 2nd diffraction grating (splitting into  $B$  and  $C$  and subsequently into  $D$  and  $F$ , respectively into  $E$  and  $G$ ) the atoms move freely.

The special trick in the experiment is the use of Rb atoms whose energy level scheme can be simplified as is shown in Fig. 8.5a. Without presenting details [10], we note that the Rb 2-level atom should more accurately be represented by a 3-level system, where the ground state  $|g\rangle$  is split into two tightly separated electronic states  $|2\rangle$  and  $|3\rangle$ . The splitting is due to nuclear spin interaction and the states are called according to their total angular momentum quantum numbers 2 and 3. While the transition frequency between  $|2\rangle$  and  $|3\rangle$  amounts to about 3 GHz in the microwave frequency range, the electronic transition between  $|e\rangle$  and  $|g\rangle$ , be it from  $|2\rangle$  or from  $|3\rangle$ , requires photon energies of the visible light spectrum. In the experiment, laser light with a wavelength  $\lambda$  of 780 nm is used. This wavelength corresponds to a photon energy or frequency  $\omega_L$  which fits exactly to the average of the transition frequencies  $|3\rangle \rightarrow |e\rangle$  and  $|2\rangle \rightarrow |e\rangle$  (Fig. 8.5a). Consequently, the detuning  $\Delta_2$  and  $\Delta_3$  for the two relevant electronic transitions producing the corresponding diffraction potentials have the same amount in (8.88) but their sign (phase) is different:  $\Delta_2 < 0$ ,  $\Delta_3 > 0$ . Independent whether the Rb atoms initially occupy the state  $|2\rangle$  or  $|3\rangle$ , according to (8.88) they are diffracted equally. The interference between the partial beams  $D$  and  $E$ , respectively between  $F$  and  $G$  in Fig. 2.10a occurs equally both for atoms in states  $|2\rangle$  and  $|3\rangle$ .

By additional manipulation of the occupation of the initial states  $|2\rangle$  or  $|3\rangle$ , i.e. the internal degree of freedom of the Rb atoms, however, we can stamp the atoms with information about their path. A distinction between atoms originating from the partial beams  $B$  and  $C$  becomes possible. Irradiation of the transition frequency between  $|2\rangle$  and  $|3\rangle$ , i.e. of microwave radiation of 3 GHz frequency, induces a total reshuffle of the states  $|2\rangle$  and  $|3\rangle$  by a  $\pi$ -pulse or a superposition state of  $|2\rangle$  and  $|3\rangle$  after a  $\pi/2$ -pulse (Sect. 6.5.1).

For the read-out of the “Which Way” information, i.e. the determination of the origin of the interfering atom beams, partial beam  $B$  or  $C$  after the first diffraction process (Fig. 2.10a), the sign (phase) of the diffracting light wave potential is essential. It is important whether the detuning  $\Delta$  is positive (initial state  $|3\rangle$ ) or negative (initial state  $|2\rangle$ ). The absolute diffraction probability of an atom does not depend on the sign of the potential. In contrast, however, the phase of the diffracted atom wave does depend on the sign of the diffracting potential. Depending on the initial states  $|2\rangle$  or  $|3\rangle$  of the Rb atoms their wave phase differs after diffraction. Atom waves behave exactly as light waves. Already in Sect. 3.6.3, where electrons running against a potential step were treated, or in connection with the WKB approximation in Sect. 6.1 the electronic wave function exhibits a phase containing the term  $\sqrt{2m(E - V)}$ .  $E$  is the electron energy and  $V$  the potential the electron has passed.



**Fig. 8.5 a, b** Scheme of a Rb atomic beam diffraction experiment for the study of the effect of “Which Way” information on the two-beam interference of the atoms [8, 10]. **a** Simplified energy level scheme of the Rb atoms: The excited state  $|e\rangle$  is represented by the state  $5^2 P_{3/2}$  in atomic nomenclature. The ground state  $|g\rangle$  (in atomic nomenclature  $5^2 S_{1/2}$ ) is split into two energetically close states  $|3\rangle$  and  $|2\rangle$  due to hyperfine interaction via spins. These states have the angular momenta  $3\hbar$  and  $2\hbar$ . On the right side the effect of a standing light wave with frequency  $\omega_L$  is shown: For the states  $|3\rangle$  and  $|2\rangle$  the amount of the detuning are identical but its sign is opposite. Scheme of the experimental set-up of the Rb atomic beam diffraction experiment [8, 10]. **b** Plot of the elements of the interference experiment: The incident Rb atom beam  $A$  is split into two partial beams  $B$  and  $C$  on a 1st standing laser light wave. A 2nd standing wave splits these beams into the beams  $D$  and  $F$ , respectively  $E$  and  $G$ . The beams  $D$  and  $E$ , respectively  $F$  and  $G$ , propagate in parallel and interfere with each other in double beam interference (see Fig. 2.10a, b). Information about the detailed path of single Rb atoms is obtained by additional irradiation of 3 GHz microwave pulses ( $\pi/2$ ). The 1st pulse produces a superposition state  $|3\rangle + |2\rangle$  which is split into the states  $|3\rangle - |2\rangle$  and  $|3\rangle + |2\rangle$  by diffraction on the standing light wave. The 2nd  $\pi/2$  pulse prepares the two states  $-|2\rangle$  and  $|3\rangle$  in the beams  $B$  and  $C$ . After diffraction on the 2nd standing light wave these internal states of the Rb atom allow a distinction between the situations that the Rb atom originates from  $B$  or  $C$  (through what slit). Both situations differ by interferences of the internal states  $-|2\rangle$  and  $|3\rangle$ , respectively  $|2\rangle$  and  $|3\rangle$ .

Depending on the value of  $E$  in comparison to  $V$  this phase factor might be real or imaginary. The latter case involves a phase change of the wave function.

Without going into mathematical details [10], we use the classical optical analogue for light: A light wave travelling in an optically thin medium is diffracted at the interface to an optically denser medium with a phase change  $\pi$ . On the other hand, diffraction on an optically thin medium or transmission into a thin or dense medium does not change the phase of the light wave. A negative diffraction potential causes an imaginary phase term in the diffracted particle wave, that is, a phase change. We, therefore, conclude:

- Rb atom in  $|2\rangle \Rightarrow$  potential negative  $\Rightarrow$  optically dense  $\Rightarrow$  phase change  $\pi$
- Rb atom in  $|3\rangle \Rightarrow$  potential positive  $\Rightarrow$  optically thin  $\Rightarrow$  no phase change

By means of these conclusions, we can understand the sequence of internal quantum states of the Rb atoms in Fig. 8.5b when they have passed the diffracting laser light waves and have undergone the twofold application of  $\pi/2$  pulses of 3 GHz radiation.

We assume the Rb atoms to be initially before being exposed to the 3 GHz  $\pi/2$  pulse in their ground state  $|2\rangle$ . This is the energetically lowest state which is preferentially occupied in a cold atomic beam. For obtaining the “Which Way” information, a 1st  $\pi/2$  microwave pulse transfers the atoms into the superposition state  $(|3\rangle + |2\rangle)/\sqrt{2}$  (Sect. 6.5.1). After the subsequent splitting of the atom beam into the two partial beams  $B$  and  $C$  the diffracted beam  $B$  has experienced a  $\pi$  phase shift (sign change) of its  $|2\rangle$  component and the atoms are in the superposition state  $(|3\rangle - |2\rangle)/\sqrt{2}$ . The component  $|3\rangle$  remains unchanged, since it is exposed to a positive diffraction potential, that is, it is diffracted into an optically thinner medium. The transmitted beam  $C$  does not change its phase, analogously to the optical case. After passing the 1st diffraction potential (laser light wave), the atoms of the beams are in superposition states in which the external positional state (wave functions  $\psi_B$  and  $\psi_C$ ) is entangled with the internal degree of freedom of the atoms. The total entangled state of the atom beams is obtained as

$$|\psi\rangle \propto |\psi_B\rangle(|3\rangle - |2\rangle) + |\psi_C\rangle(|3\rangle + |2\rangle). \quad (8.96)$$

This entanglement is the basis for the storage of the “Which-Way” information, as we will see.

The 2nd  $\pi/2$  microwave pulse after passage of the 1st beam splitter transfers the internal superposition states  $(|3\rangle - |2\rangle)/\sqrt{2}$  and  $(|3\rangle + |2\rangle)/\sqrt{2}$  again into states where only  $|2\rangle$  and  $|3\rangle$  are occupied (Sect. 6.5.1). The 2nd  $\pi/2$  pulse restores the states  $|2\rangle$  and  $|3\rangle$  of the 2-level system again. The effect is easily clarified by comparing the present 2-level atom with the spin 2-level system (Sect. 6.5.2, Fig. 6.14). The superposition state  $(|3\rangle - |2\rangle)/\sqrt{2}$  corresponds to a spin state with spin orientation in the  $x$ ,  $y$  plane  $\perp z$ , but in negative direction because of the negative sign. Correspondingly a  $\pi/2$  pulse flips the spin into the  $-z$  direction. This corresponds to the  $-|2\rangle$  state of the 2-level atom.

Thus, the state of the two partial beams  $B$  and  $C$  after having been exposed to the  $\pi/2$  pulse and before they reach the 2nd diffraction grating (laser light wave) is represented as:

$$|\psi\rangle \propto -|\psi_B\rangle|2\rangle + |\psi_C\rangle|3\rangle. \quad (8.97)$$

The 2nd diffraction grating (laser light wave) splits the beams  $B$  and  $C$  further into the beams  $D$  and  $E$ , respectively  $F$  and  $G$  which finally produce the double slit interference by superposition. The internal states of the transmitted beams  $D$  and  $G$  do not change, while the atoms in the diffracted beam  $F$  undergo a  $\pi$  phase shift, that is, their internal state changes from  $-|2\rangle$  to  $|2\rangle$ , since the atoms pass a negative potential and are, thus, diffracted into an optically more dense medium. Atoms in the state  $|3\rangle$  in beam  $E$  experience a positive diffraction potential and, thus, keep their internal state  $|3\rangle$ .

The total quantum state  $|\psi\rangle$  of the four resulting partial beams  $D, E, F, G$  is consequently represented as

$$|\psi\rangle \propto -|\psi_D\rangle|2\rangle + |\psi_E\rangle|3\rangle + |\psi_F\rangle|2\rangle + |\psi_G\rangle|3\rangle. \quad (8.98)$$

This is an entangled state in which the position wave functions are entangled with the internal states  $|2\rangle$  and  $|3\rangle$  of the atoms. In the far field, the two pairs of parallel beams  $D, E$  and  $F, G$  propagate into spatially distinct regions. Pair wise, by superposition, they produce the measured interference patterns shown in Fig. 2.10b, on the left side originating from the beams  $D$  and  $E$ , on the right side from the interference of  $F$  and  $G$ . As is described in Sect. 2.4.2, the interference pattern is only observed as long as no “Which Way” information is stored by application of microwave pulses which enable a distinction between the internal atomic states  $|2\rangle$  and  $|3\rangle$  and thus a determination of the detailed particle trajectory. Read-out of the “Which Way” information requires the experimental determination of the internal state of the interfering atoms, whether they are in state  $|2\rangle$  or in state  $|3\rangle$  (Fig. 8.5b).

To elucidate the underlying reason for the destruction of the interference pattern by gaining “Which Way” information, we calculate the interference intensity originating from the two partial beams  $D$  and  $E$ , that is, the probability  $P = \langle\psi_{DE}|\psi_{DE}\rangle$  for detecting an electron at distance  $z$  from the 2nd diffraction grating (laser wave):

$$\begin{aligned} P(z) &\propto (\langle\psi_D|\langle 2| + \langle\psi_E|\langle 3|)(-|\psi_D\rangle|2\rangle + |\psi_E\rangle|3\rangle) \\ &= |\psi_D|^2\langle 2|2\rangle + |\psi_E|^2\langle 3|3\rangle - \langle\psi_D|\psi_E\rangle\langle 2|3\rangle - \langle\psi_E|\psi_D\rangle\langle 3|2\rangle \\ &= |\psi_D|^2 + |\psi_E|^2 - \psi_D^*\psi_E\langle 2|3\rangle - \psi_E^*\psi_D\langle 3|2\rangle. \end{aligned} \quad (8.99)$$

The two partial beams are described by atomic wave functions  $\psi_D$  and  $\psi_E$  of freely propagating atoms (3.4) as in the double slit experiment in Sect. 3.1. Without taking notice of the internal states  $|2\rangle$  and  $|3\rangle$ , (8.99) is identical with the expression for the interference intensity of two electron beams in the double slit experiment (3.5). The mixed terms  $\psi_D^*\psi_E$  and  $\psi_E^*\psi_D$  in (8.99) produce the interference pattern described by  $\cos \mathbf{k} \cdot (\mathbf{r}_2 - \mathbf{r}_1)$  in (3.5). Without storage of the “Which Way” information, that

is, by switching off the microwave radiation probe, the double slit interference of the atomic beams  $D$  and  $E$  originating from the sources  $B$  and  $C$  are observed.

Because of the orthogonality of the internal states  $|2\rangle$  and  $|3\rangle$  we require in (8.99):

$$\langle 2|3\rangle = \langle 3|2\rangle = 0. \quad (8.100)$$

The entanglement of the position wave functions  $\psi_D$  and  $\psi_E$  with the internal degrees of freedom of the atom used for storage of the “Which Way” information causes destruction of the interference pattern (Sect. 2.4.2).

From the discussed experiment, it is obvious that the destruction of the interference pattern by switching on the probe for “Which Way” information (microwave pulse) is caused by entanglement rather than by a momentum transfer from the measurement probe (3 GHz photons) to the interfering atoms, as was frequently assumed previously. The Rb atoms of the atomic beams flying with a velocity of 2 m/s have a mass of about  $85m_{\text{Proton}}$  and, thus, carry a mechanical momentum of about  $3 \times 10^{-25}$  kg m/s. In contrast, the radiation momentum  $\hbar k$  of 3 GHz photons is merely  $6 \times 10^{-33}$  kg m/s. This momentum is comparatively much too small to perturb the propagating Rb atoms and, thus, to destroy their interference pattern.

Therefore, the experiment undoubtedly proves that the quantum mechanical correlation due to entanglement between measurement probe and observed particle destroys the interference pattern. The entanglement occurs fully independently whether the “Which Way” information is read out and registered by a human observer. Merely switching on the measurement probe is decisive. This is a clear rejection of the idealistic idea that the human observer might influence quantum mechanical behavior.

### 8.2.5 The Casimir Effect

In 1948, the Dutch physicists Hendrik B.G. Casimir and Dirk Polder theoretically predicted an effect, which can only be understood on the basis of the zero-point energy (8.28) of the electromagnetic field [11].

To remind some results of Sect. 8.2: The infinitely large ground state vacuum energy of the electromagnetic field  $E_0 = (\sum_{q\lambda} \hbar\omega_{q\lambda})/2$  results from the zero-point energies of single oscillators (Sect. 4.4.2) which describe photons of energy  $\hbar\omega_{q\lambda}$ . These are typically quantum mechanical zero-point fluctuations of the vacuum, which carries an internal energy even without photons being present. According to (8.28) these fluctuations are vibrational modes of the vacuum with frequency,  $\omega_{q\lambda}$ .

In accord with Casimir’s considerations we, now, assume two large area metallic (electrically conductive) plates mutually parallel at a short distance from each other. The vibrational modes of the electromagnetic field, also those of the vacuum fluctuations, must fulfill the boundary conditions of the electromagnetic field on the metallic surfaces. The electric field must exhibit nodes on the conductive surfaces. Consequently, between the plates only standing wave modes, as in a resonator, can

exist. This is in contrast to the vacuum state outside of the two plates. There, all kinds of modes do exist. The variety of modes between the plates is limited in comparison to those outside the plates. Therefore, the energy per volume between the plates is different from that in the external space, that is, a force on the plates results.

Using (8.28), we calculate the energy of the vacuum state between the two parallel plates by assuming plates with an edge length  $L$  at a distance  $x$  from each other. The wave vectors (numbers)  $q = 2\pi/\lambda$  of the standing waves between the plates must then fulfill the condition

$$x = v \frac{\lambda}{2} = v \frac{\pi}{q}, \quad v = 1, 2, 3, \dots \quad (8.101)$$

As in Sect. 3.6.1, the field modes are, thus, quantized in wave vector space by intervals  $\Delta q = \pi/x$  for the direction normal to the plates and by  $\Delta q = \pi/L$  parallel to the plates. Consequently in wave vector space the state of a vibration mode occupies the volume

$$(\Delta q)^3 = \frac{\pi^3}{x L^2}. \quad (8.102)$$

The sum of the zero-point modes in (8.18a), (8.18b) can be transformed into an integral and the ground state vacuum energy of the electromagnetic field between the plates is obtained as

$$U \simeq \frac{L^2 x}{\pi^3} \int \hbar \omega_{\mathbf{q}} d^3 q. \quad (8.103a)$$

Using the 2D plate geometry with  $d^3 q = \pi q^2 dq$  the 3D integration is transformed into an integration over only one coordinate  $q$ :

$$U \simeq \frac{L^2 x}{\pi^3} \int_{\pi/x}^{\pi/a_0} \hbar c q^3 dq = \frac{L^2 x}{\pi^2} \hbar \frac{q^4}{4} \Big|_{\pi/x}^{\pi/a_0}. \quad (8.103b)$$

The integral covers a  $q$  interval from the lower wave vector limit  $\pi/x$  determined by the plate distance  $x$  up to the upper limit  $\pi/a_0$ , where  $a_0$  describes a finite atomic length. This length is in some meaningful way related to the boundary conditions of the electric field of the standing wave mode at the metal plate surface. By assuming such a finite length  $a_0$ , we avoid the problem of extending the integral to infinitely large wave vectors  $q$  upon approaching the plate ( $a_0^{-1} \rightarrow \infty$ ). Because of the atomic dimensions of surface details of the metallic plates and the resulting boundary conditions of the electric field a reasonable assumption for  $a_0$  is that of the Bohr radius. With this approximation the energy of the vacuum field between the plates is obtained as:

$$U \simeq \frac{1}{4} \frac{L^2 \hbar c}{\pi^2} x \left( \frac{\pi^4}{a_0^4} - \frac{\pi^4}{x^4} \right) = \frac{1}{4} \pi^2 L^2 \hbar c \left( \frac{x}{a_0^4} - \frac{1}{x^3} \right), \quad (8.103c)$$

with  $a_0$  as Bohr radius.

From (8.103c), the attractive force between the two plates per unit area, the *Casimir force*, is calculated as

$$F_c = -\frac{1}{L^2} \frac{dU}{dx} \simeq \frac{-1}{4} \pi^2 \hbar c \left( \frac{1}{a_0^4} + \frac{3}{x^4} \right), \quad (8.104a)$$

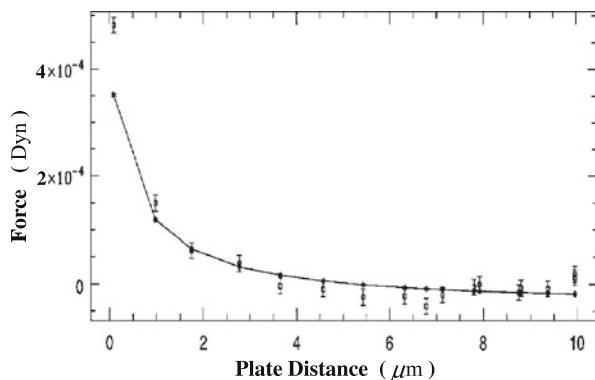
$$F_c = K_c \frac{1}{x^4} \quad \text{with } K_c = \frac{\pi \hbar c}{480} = 1.3 \cdot 10^{-27} \text{ N m}^2. \quad (8.104b)$$

From the derivation, it is obvious that the Casimir force is not limited to two opposite parallel plates, but the absolute amount and the distance dependence of the force depend on the shape of the two opposing metallic objects. The force between a plane plate and a spherical surface has a distance dependence  $F_c \propto x^{-3}$  [12]. For this case, Lamoreaux has performed a precise measurement of the distance dependence for distances between plate and sphere ranging between 0.6 and 10  $\mu\text{m}$  [12]. According to the author, the results shown in Fig. 8.6 confirm the theoretical prediction within deviations in the 5 % range.

The great experimental challenge in these measurements is the exclusion or corrective compensation of superimposed disturbing effects, such as a residual electrostatic interaction. Even more precise measurements than in Fig. 8.6 have been performed by Bressi et al. [13] on a system of two parallel plates. One of these had been realized by a silicon cantilever. The Casimir force has been determined from the vibrational frequency of the cantilever arm. Even the absolute value of the constant  $K_c$  in (8.104b) has been determined with an accuracy of about 15 % to be  $K_c = (1.22 \pm 0.18) \times 10^{-27} \text{ N m}^2$ . The agreement with the theoretical value is astonishingly good. These kinds of measurements of the Casimir force are, of course, far away from achieving the precision of spectroscopic data to confirm quantum electrodynamics (Sect. 5.5.1).

However, it is interesting to see that one encounters effects of the quantized light field even in the field of micro- and nano-mechanics. Only the precision of the measurement must be sufficiently high.

**Fig. 8.6** Casimir force as function of distance between an electrically conducting plane surface and a spherical surface (gold evaporated on quartz). The measured data points are connected with an optimum fit curve (minimum quadratic deviation) [12]



### 8.3 The Quantized Schrödinger Field of Massive Particles

Particles without mass as photons obey the formalism of the quantized electromagnetic field (Sect. 8.2). Electrons carrying a mass are described by the Schrödinger equation, both in case of a single particle as well as of a system of many particles. In the non-relativistic limit of particle velocities far below light speed Schrödinger equation yields an adequate description of the particle-wave dynamics. This is true for fermionic electrons but also for bosons such as  $\text{He}^4$  atomic nuclei. Both fermions and bosons obey the Schrödinger equation for sufficiently small particle velocities. In the relativistic range of high velocities reaching light speed the Schrödinger equation must be modified. This modification results in two different new fundamental dynamic equations, the *Dirac equation* for fermions and the *Klein–Gordon equation* for bosons. We will not deal with these relativistic equations in this book, since our interest is focused on non-relativistic phenomena in solids and nanostructures.

After we have learnt much about the quantized field of photons, particles without mass, in Sect. 8.2 we will now approach the problem of quantization of a field of massive particles. The most familiar example is that of a many-electron system.

As was already pointed out at the beginning of Chap. 8 and in Sect. 5.6.1, electrons in a many-body wave function do not have an identity anymore. When the single-particle wave functions spatially overlap in an ensemble of particles, we can not distinguish to what particle a particular wave function must be attributed. Only one single global many-electron wave function is defined over the whole space of the many-body system. An adequate measurement at a position  $\mathbf{r}$  at a time  $t$  yields a certain probability (amplitude) to detect a particle there. The situation is fully analogous to what we have learnt in Sect. 8.2 about the electromagnetic field and its constituents, the photons. In the present case of many electrons the relevant field function is, of course, the many-particle wave function  $\psi$ , which is calculated as the solution to a many-particle Schrödinger equation. In contrast to the electromagnetic field with the vector potential  $\mathbf{A}$  being a vector field the electron wave function  $\psi$  is a scalar field as long as the spin is not taken notice of. In analogy to the electromagnetic field we, thus, expect that the quantization of the many-particle  $\psi$  field yields electrons as excitations of this field. As the wave function  $\psi$  already originates from a quantum theory developed for single particle dynamics, the quantization of the many-particle wave function is frequently called *second quantization*, a more historical term.

So far the Schrödinger field  $\psi(\mathbf{r})$  was used to describe the dynamics of electrons which are fermions because of their half integer spin. For one single particle it is not important in the wave function formalism whether this particle is a fermion or a boson. Only for two and more particles the fermionic or bosonic character of the wave function becomes decisive (Sect. 5.6.2). An interchange of particles defining a bosonic or fermionic wave function (symmetric or anti-symmetric) is only possible for more than one particle.

In the following, we will quantize the many-particle Schrödinger field without taking notice whether the massive particles are bosons or fermions. Using the com-

mon commutation rules of Sect. 8.2 for photons, we expect to end up with a boson field as for  $\text{He}^4$  nuclei.

In analogy to the many-photon electromagnetic field, the many-particle Schrödinger field is described by a many-body wave function  $\psi(\mathbf{r}, t)$ . But in contrast to the real valued photon field the Schrödinger field is complex, its amplitude has a real and an imaginary value at each position  $\mathbf{r}$  at time  $t$ . Rather than describing the field by its real and imaginary part of the amplitude we can use just as well the two independent functions  $\psi(\mathbf{r}, t)$  and  $\psi^*(\mathbf{r}, t)$ . These two functions appear in the expression for the total energy  $H$  of the many-particle field:

$$H = \int d^3r \psi^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \psi(\mathbf{r}). \quad (8.105)$$

This expression is easily understood as a straight forward extension of the energy calculated for many particles at positions  $\mathbf{r}$ . While in the earlier treatment (Sect. 5.6.2), the particles were found at discrete coordinates  $\mathbf{r}_i$ , the continuous particle coordinate  $\mathbf{r}$  (simultaneously field coordinate) requires an integral for the summation over all the particles. The expression in brackets is nothing else than the one-particle Hamilton operator of Schrödinger wave mechanics for a particle at position  $\mathbf{r}$ . Considering  $\psi(\mathbf{r}, t)$  and  $\psi^*(\mathbf{r}, t)$  as field functions (as in Schrödinger's theory) rather than operators (8.105) represents the energy expectation value of the total number of particles in the field. This energy  $H$  corresponds to the total energy of the electromagnetic field (8.11) in Sect. 8.2, respectively to the Hamilton function (8.1) in classical mechanics.

The energy expression (8.105) being the analogue to the classical Hamilton function with  $\psi(\mathbf{r}, t)$  and  $\psi^*(\mathbf{r}, t)$  as scalar field functions is, therefore, the starting point for the quantization of the many-particle Schrödinger field. Consequently, the energy functional  $H$  can be used, as in Sect. 8.1, to find the canonically conjugated variables which must obey the commutation relations analogously to  $[p, x] = \hbar/i$ . We proceed as in (8.1), (8.2a), (8.2b) and consider  $\psi^*(\mathbf{r}, t)$  to be linearly independent of  $\psi(\mathbf{r}, t)$ . According to the Hamilton formalism, the derivation of the Hamilton function  $H$  with respect to a momentum-like variable ( $p$ ) must yield the time derivative (8.2a) of a position-like variable ( $\dot{x}$ ), while the negative derivative of  $H$  with respect to the position-like variable ( $x$ ) must result in the dynamic equation of the problem, that is, the Schrödinger equation in the present case [see (8.2b) for the classical case]. The structure of the Hamilton function (8.105) suggests as a trial derivative  $\partial/\partial\psi^*$ . We might eventually expect the Schrödinger equation as the result. Some remarks in advance: The operation  $\partial/\partial\psi^*$  is a generalization of a derivative; in this derivative with respect to a function the function  $\psi^*(\mathbf{r})$  is treated as a variable. But the derivation is performed at a particular fixed position  $\mathbf{r}$  of the function, while the integral in (8.105) contains all positions  $\mathbf{r}$  of the considered space volume. We solve this problem formally by extending the integral over all positions  $\mathbf{r}'$  but selecting out the position  $\mathbf{r}$  by inserting the function  $\delta(\mathbf{r} - \mathbf{r}')$  into the integral. This delta function fixes the position of the  $\psi^*$  function for the functional derivative:

$$\begin{aligned}\frac{\partial H}{\partial \psi^*(\mathbf{r})} &= \frac{\partial}{\partial \psi^*(\mathbf{r})} \int d^3 r' \delta(\mathbf{r} - \mathbf{r}') \psi^*(\mathbf{r}') \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}') \right] \psi(\mathbf{r}') \\ &= \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \psi(\mathbf{r}).\end{aligned}\quad (8.106)$$

As expected the derivative  $\partial H / \partial \psi^*$  yields the Schrödinger dynamic equation if (8.106) would be equal to  $i\hbar \dot{\psi}(\mathbf{r})$ . With this assumption, we obtain:

$$\frac{\partial H}{\partial \psi^*} = i\hbar \dot{\psi}, \quad \text{respectively} \quad \frac{\partial H}{\partial (i\hbar \psi^*)} = \dot{\psi}. \quad (8.107)$$

According to (8.107), we regain the classical Hamilton formalism (8.2a), (8.2b) by introducing  $\psi(\mathbf{r})$  as the position-like ( $x$ ) and  $i\hbar \psi^*(\mathbf{r}) = \Pi(\mathbf{r})$  as the momentum-like variable ( $p$ ). From (8.107), we further conclude

$$\frac{\partial}{\partial \Pi} H = \dot{\psi} = \frac{1}{i\hbar} \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \psi. \quad (8.108)$$

The Schrödinger equation is obtained, which is simultaneously one of the two Hamilton equations (8.2a). The second Hamilton equation (8.2b) is represented as

$$\begin{aligned}-\frac{\partial}{\partial \psi} H &= -\frac{\partial}{\partial \psi(\mathbf{r})} \int d^3 r' \delta(\mathbf{r} - \mathbf{r}') \psi^*(\mathbf{r}') \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}') \right] \psi(\mathbf{r}') \\ &= -\frac{\partial}{\partial \psi(\mathbf{r})} \int d^3 r' \delta(\mathbf{r} - \mathbf{r}') \left\{ \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}') \right] \psi^*(\mathbf{r}') \right\} \psi(\mathbf{r}') \\ &= -\left[ \frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \psi^*(\mathbf{r}) = \hat{\Pi} = i\hbar \dot{\psi}^*(\mathbf{r}).\end{aligned}\quad (8.109)$$

For this calculation, the energy operator had to be shifted in front of  $\psi^*$  by using (4.12a), (4.12b). It now operates on  $\psi^*$  rather than on  $\psi$ . Otherwise the derivative with respect to  $\psi$  according to (8.106) could not have been calculated. By setting  $x \rightarrow \psi$  and  $p \rightarrow \Pi = i\hbar \dot{\psi}$  (8.109) represents the second Hamilton equation (8.2b) and simultaneously the Schrödinger equation for the complex conjugate wave function  $\psi^*$ .

The Hamilton formalism has been shown to be applicable and the canonically conjugated variables of the Schrödinger field are identified as

$$x \rightarrow \psi(\mathbf{r}), \quad (8.110a)$$

$$p \rightarrow \Pi(\mathbf{r}) = i\hbar \psi^*(\mathbf{r}). \quad (8.110b)$$

For the quantization of the Schrödinger field, these field functions must be replaced by field operators  $\hat{\psi}$  and  $\hat{\Pi}$ , which obey the familiar commutation relation (4.25). As the common procedure (Chap. 4) requires, the field function  $\psi^*$  is replaced by the

operator  $\hat{\psi}^+$ . Then, the commutation relations for the field operators are written as:

$$[\hat{H}(\mathbf{r}), \hat{\psi}(\mathbf{r}')] = \frac{\hbar}{i} \delta(\mathbf{r} - \mathbf{r}'), \quad \text{respectively} \quad (8.111\text{a})$$

$$[\hat{\psi}^+(\mathbf{r}), \hat{\psi}(\mathbf{r}')] = \delta(\mathbf{r} - \mathbf{r}'), \quad (8.111\text{b})$$

$$[\hat{\psi}(\mathbf{r}), \hat{\psi}^+(\mathbf{r}')] = [\hat{\psi}^+(\mathbf{r}), \hat{\psi}^+(\mathbf{r}')]=0. \quad (8.111\text{c})$$

Since the field operators are continuous operators in the 3-D space, the unity operator in the operator space must be the  $\delta$ -function according to Sect. 4.3.4. Equation (8.111c) expresses the well known fact [see classical commutation relations (4.25)], that both the position-like and the momentum-like field operators mutually commute.

Analogously to the electromagnetic field (Sect. 8.2), there is the many-particle state  $|\phi\rangle$  on which the field operators  $\hat{\psi}$  and  $\hat{\psi}^+$  operate. These field operators  $\hat{\psi}$ ,  $\hat{\psi}^+$  operate at a special position  $\mathbf{r}$  of the field; they correspond to the position dependent operators  $\hat{\mathbf{A}}(\mathbf{r})$  and  $\hat{\mathbf{A}}^+(\mathbf{r})$ , the vector potentials of the electromagnetic field. In Sect. 8.2, these vector potential operators have been expanded in a Fourier series. The corresponding Fourier components represent the eigenstates of the electromagnetic field in free space. They have been used to determine the creation and destruction operators  $\hat{b}_{\mathbf{q}\lambda}^+$  and  $\hat{b}_{\mathbf{q}\lambda}$  for photons. For the Schrödinger field, we proceed analogously. In contrast to the electromagnetic field, we do not want to limit the treatment to plane waves as eigenfunctions for free particles in vacuum. We assume a certain physical problem described by the single particle Schrödinger equation

$$\left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \varphi_i = \varepsilon_i \varphi_i, \quad (8.112)$$

$V(\mathbf{r})$  might be the attractive potential of a positive nucleus. This potential might bind one single electron as described by (8.112). But there might also be a high number of electrons exposed to this potential. In this case, the field theoretical description of the problem is adequate. The single particle eigenfunctions  $\varphi_i(\mathbf{r})$  of the single electron Hamiltonian in (8.112) are then the adequate eigenstates for the series expansion of the field operators  $\hat{\psi}$ ,  $\hat{\psi}^+$ . The most general wavefunction for the problem is

$$\psi(\mathbf{r}) = \sum_i b_i \varphi_i(\mathbf{r}). \quad (8.113\text{a})$$

Consequently the expansion of the field operators adequate to the described physical problem is

$$\hat{\psi}(\mathbf{r}) = \sum_i \hat{b}_i \varphi_i(\mathbf{r}), \quad (8.113\text{b})$$

$$\hat{\psi}^+(\mathbf{r}) = \sum_i \hat{b}_i^+ \varphi_i^*(\mathbf{r}). \quad (8.113\text{c})$$

Here, the Fourier expansion coefficients  $b_i$  in (8.113a) are simply interpreted in (8.113a), (8.113b) as non-commuting operators.

All physical quantities, which can be defined in the many-particle field, such as energy, momentum, angular momentum etc., must now be rewritten as operators  $\hat{\Omega}$ , which act on the field state  $|\phi\rangle$ . Measured results of those observables are calculated as expectation values as in the single-particle Schrödinger formalism:

$$\langle \Omega \rangle = \langle \phi | \hat{\Omega} | \phi \rangle. \quad (8.114)$$

We start with the observable energy of the field. The expression for the energy of the non-quantized field is given by (8.105). We transform this expression into a field operator by replacing the wave functions by the corresponding field operators:

$$\hat{H} = \int d^3r \psi^+(\mathbf{r}') \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \hat{\psi}(\mathbf{r}). \quad (8.115)$$

Now the field operators  $\hat{\psi}$  and  $\hat{\psi}^+$  are expressed by their expansions in terms of single particle wave functions (8.113b), (8.113c) and by means of (8.112) we obtain:

$$\begin{aligned} \hat{H} &= \sum_{ij} \int d^3r \hat{b}_i^+ \varphi_i^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \hat{b}_j \varphi_j(\mathbf{r}) \\ &= \sum_{ij} \int d^3r \hat{b}_i^+ \varphi_i^*(\mathbf{r}) \varepsilon_j \hat{b}_j \varphi_j(\mathbf{r}) = \sum_{ij} \hat{b}_i^+ \hat{b}_j \varepsilon_j \delta_{ij} \\ &= \sum_i \varepsilon_i \hat{b}_i^+ \hat{b}_i. \end{aligned} \quad (8.116)$$

This representation is the analogue to (8.22), the Hamilton operator of the electromagnetic field. The immediate question arises whether the operators  $\hat{b}_i^+$  and  $\hat{b}_i$  are again creation and destruction operators for particles or quanta in the many-body Schrödinger field.

To answer this question, we transfer the commutation relations (8.111a)–(8.111c) to the operators  $\hat{b}_i^+$  and  $\hat{b}_i$ . By inserting (8.113b), (8.113c), the commutation relation (8.111b) yields

$$\begin{aligned} [\hat{\psi}(\mathbf{r}'), \hat{\psi}^+(\mathbf{r})] &= \left[ \sum_i \hat{b}_i \varphi_i(\mathbf{r}'), \sum_j b_j^+ \varphi_j^*(\mathbf{r}) \right] \\ &= \sum_{ij} \hat{b}_i \hat{b}_j^+ \varphi_i(\mathbf{r}') \varphi_j^*(\mathbf{r}) - \sum_{ij} \hat{b}_j^+ \hat{b}_i \varphi_j^*(\mathbf{r}) \varphi_i(\mathbf{r}') \\ &= \delta(\mathbf{r} - \mathbf{r'}). \end{aligned} \quad (8.117)$$

Integration of this relation over both position coordinates  $\mathbf{r}$  and  $\mathbf{r}'$  and the whole system volume yields

$$\begin{aligned} \sum_{ij} \hat{b}_i \hat{b}_j^+ \int d^3r d^3r' \varphi_i(\mathbf{r}') \varphi_j^*(\mathbf{r}) - \sum_{ij} \hat{b}_j^+ \hat{b}_i \int d^3r d^3r' \varphi_j^*(\mathbf{r}) \varphi_i(\mathbf{r}') \\ = \int d^3r d^3r' \delta(\mathbf{r} - \mathbf{r}') = 1. \end{aligned} \quad (8.118)$$

This equation can only be fulfilled for  $\mathbf{r} = \mathbf{r}'$ , that is, with  $\int d^3r \varphi_i(\mathbf{r}) \varphi_j^*(\mathbf{r}) = \delta_{ij}$  we conclude

$$[\hat{b}_i, \hat{b}_j^+] = \delta_{ij}, \quad (8.119a)$$

and analogously

$$[\hat{b}_i, \hat{b}_j] = 0, \quad (8.119b)$$

$$[\hat{b}_i^+, \hat{b}_j^+] = 0. \quad (8.119c)$$

These commutation relations are identical with those of the creation and destruction operators for photons in the light field (8.20a), (8.20b). The total energy of the many-body Schrödinger field (8.116) expressed in terms of these operators is an expression analogous to that of the quantized light field (8.22), apart from the energy of the vacuum state. This clearly demonstrates that the operators  $\hat{b}_i^+$  and  $\hat{b}_i$  are creation and destruction operators for particles (quanta) in the many-particle field. The index  $i$  indicates the possible quantum states, that is, particle excitation states with energy  $\varepsilon_i$ . This is analogous to the light field with its light quanta, the photons of energy  $\hbar\omega_{\mathbf{q}}$ ; as for photons with their quantum numbers  $(\mathbf{q}, \lambda)$  the type of the particular particle excitation in the Schrödinger field is denoted by the quantum number  $i$ . Consequently the many-particle state of the Schrödinger field might be written as

$$|\phi\rangle = |\dots, n_i, n_j, n_k, \dots\rangle = |\dots, n_j, \dots\rangle, \quad (8.120)$$

with  $n_j$  as the number of particles or excitations of type  $j$  in the field. As in Sect. 8.2, the action of the field Hamilton operator (8.116) on a general field state yields the field Schrödinger equation:

$$\hat{H}|\dots, n_i, n_j, \dots\rangle = \sum_i \varepsilon_i \hat{b}_i^+ \hat{b}_i |\dots, n_i, n_j, \dots\rangle = E |\dots, n_j, \dots\rangle, \quad (8.121)$$

$E$  is the total energy of the field, if  $\dots, n_i, n_j, \dots$  particles with the energies  $\dots, \varepsilon_i, \varepsilon_j, \dots$  are excited. By using (8.12) and the commutation relations (8.119a)–(8.119c), the operation of  $\hat{H}$  (8.116) on the many-body state  $\hat{b}_j |\dots, n_i, n_j, \dots\rangle$  yields

$$\begin{aligned}
\hat{H}\hat{b}_j|\dots,n_j,\dots\rangle &= \sum_i \varepsilon_i \hat{b}_i^+ \hat{b}_i \hat{b}_j |\dots,n_j,\dots\rangle = \sum_i \varepsilon_i \hat{b}_i^+ \hat{b}_j \hat{b}_i |\dots,n_j,\dots\rangle \\
&= \sum_i \varepsilon_i (-\delta_{ij} + \hat{b}_j \hat{b}_i^+) b_i |\dots,n_j,\dots\rangle \\
&= (E - \varepsilon_j) \hat{b}_j |\dots,n_j,\dots\rangle.
\end{aligned} \tag{8.122}$$

Therefore,  $\hat{b}_j|\dots,n_i,n_j,\dots\rangle = \hat{b}_j|\phi\rangle$  is a field eigenstate, in which the total energy  $E$  of the field is diminished by a field quantum, a particle  $j$  with energy  $\varepsilon_j$ . One particle of type  $j$  has been destroyed in the field. The complete formalism for photons in Sect. 8.2 can, thus, be transferred to the field of massive particles. As in (8.26b) the total energy of the field can be expressed as

$$E = \sum_i n_i \varepsilon_i. \tag{8.123}$$

By multiple operation of  $\hat{b}_i$  a vacuum or ground state of the field  $|0\rangle$  is reached, in which no particles are found anymore. The application of the operator  $\hat{b}_i^+$  creates a particle of type  $i$  in the field with simultaneous increase of the total field energy by  $\varepsilon_i$ . Since the particles of the considered field are bosons, this increase of the particle numbers can proceed up to infinity.

In analogy to (8.34), the general bosonic field state can be represented as

$$|\dots,n_j,\dots\rangle = \prod_j \frac{1}{\sqrt{n_j!}} (\hat{b}_j^+)^{n_j} |0\rangle. \tag{8.124}$$

As in (8.33a), (8.33b), the action of the creation and destruction operators  $\hat{b}_i^+$  and  $\hat{b}_i$  can be represented by their normalizing factors  $\sqrt{n_i + 1}$  and  $\sqrt{n_i}$ , respectively.

We must stress at this point, that the creation operator  $\hat{b}_j^+$  does not operate within the space of one single many-particle state with a finite particle number  $N$ . It rather operates from a space with  $N$  particles to a space with  $N + 1$  particles. The general many-particle state, thus, contains an infinite number of particles and the adequate most general state is an infinite linear superposition (sum) of all finite many-particle states  $|0\rangle, |n_i\rangle, |n_i, n_j\rangle, |n_i, n_j, n_k\rangle, \dots, \dots$ , each state multiplied by a probability amplitude. These general superposition states of infinitely many many-particle states, each one with a finite number of particles (up to infinitely high numbers), span the so-called *Fock space*. According to (8.124) the creation operators generate the entire Fock space. Within the Fock space the particle number  $N$  is not a simple number, but rather a number operator  $\hat{N}$  (8.138), which has as eigenvalues the possible particle numbers of the quantum field. In general, the number of particles in a field varies and might be concentrated around a certain average value. Only within this field theoretical description in terms of the Fock space the transformation of one particle species into another one as observed in high energy elementary particle physics becomes treatable in mathematical terms. While Einstein's famous formula  $E = mc^2$

predicts the transformation of mass (massive particles) into energy (photon fields) and vice versa on the basis of special relativity, it does not explain, how such a transformation can occur in detail. Within the Fock space of quantum field theory the particle number is not fixed, it can change by interaction between different fields (see Sects. 8.4.2 and 8.4.4). Quantum field theory thus explains for the first time, how particles can transform into each other.

It should be furthermore reminded that in the Schrödinger picture the general many-body state is dependent on time:

$$|\psi_{\{n_j\}}(t)\rangle = |\dots, n_j, \dots\rangle \exp\left(-\frac{i}{\hbar}Et\right), \quad (8.125)$$

This is in contrast to the Heisenberg picture, in which the operators  $\hat{b}_j^+$  and  $\hat{b}_j$  contain the time dependence.

As was emphasized a couple of times already, the formalism described so far is limited to bosons. The creation operator  $\hat{b}_j^+$  can create particles to infinitely high numbers. The question arises, what has to be changed in the formalism to describe fermion fields, such as that of electrons.

### 8.3.1 The Quantized Fermionic Schrödinger Field

Because of the Pauli exclusion principle fermions as for example, electrons can occupy an electronic state only once. The more general formulation of Pauli's principle requires antisymmetry of the many-body wave function upon interchange of two particles. For fermion fields with creation operators  $\hat{a}_i^+$  one must, therefore, require

$$\hat{a}_i^+ \hat{a}_i^+ |0\rangle = 0. \quad (8.126a)$$

The index  $i$  includes the spin quantum number, otherwise two fermions with opposite spin could occupy one single-particle state. Equation (8.126a) is valid not only for the vacuum state  $|0\rangle$  of the field but also for all other possible field states  $|\phi\rangle$ . Therefore the general formulation is

$$\hat{a}_i^+ \hat{a}_i^+ = 0. \quad (8.126b)$$

In contrast to boson fields (Sect. 8.3), the creation and destruction operators are denoted by  $\hat{a}_i^+$  and  $\hat{a}_i$  for fermion fields reminding the antisymmetry property of these fields. For fermion fields one can then formulate a so-called *anti-commutation relation*, which contains (8.126a) as a special case:

$$\hat{a}_i^+ \hat{a}_j^+ + \hat{a}_j^+ \hat{a}_i^+ = [\hat{a}_i^+, \hat{a}_j^+]_+ = 0. \quad (8.126c)$$

This formulation is the analogue to the bosonic commutation relation (8.119c); it merely differs by the plus sign instead of the minus sign in the boson case. The sign inversion gives rise to the denotation anti-commutation relation. The symbol in square brackets is an operator called anti-commutator.

In analogy to (8.119a)–(8.119c) we, thus, require for fermion fields the validity of the following anti-commutation relations as a postulate:

$$\hat{a}_i^+ \hat{a}_j + \hat{a}_j \hat{a}_i^+ = [\hat{a}_i^+, \hat{a}_j]_+ = \delta_{ij}, \quad (8.127a)$$

$$\hat{a}_i \hat{a}_j + \hat{a}_j \hat{a}_i = [\hat{a}_i, \hat{a}_j]_+ = 0, \quad (8.127b)$$

$$\hat{a}_i^+ \hat{a}_j^+ + \hat{a}_j^+ \hat{a}_i^+ = [\hat{a}_i^+, \hat{a}_j^+]_+ = 0. \quad (8.127c)$$

The hypothetical introduction of these anti-commutation rules for fermion fields will be shown to preserve the characteristic quantum properties of the field as in Sect. 8.3; but simultaneously the antisymmetry of the many-body wave function required by the Pauli principle is guaranteed.

Now the anti-commutation relations (8.127a)–(8.127c) can be transferred to the field operators  $\hat{\psi}^+(\mathbf{r})$  and  $\hat{\psi}(\mathbf{r})$ . Like in (8.112), we assume particular single-particle wave functions  $\varphi_i(\mathbf{r})$  as solutions to the corresponding single particle problem (Schrödinger equation). In analogy to (8.113a)–(8.113c) the field operator  $\hat{\psi}(\mathbf{r})$  is expanded in terms of the eigenfunctions  $\varphi_i(\mathbf{r})$ :

$$\hat{\psi}(\mathbf{r}) = \sum_i \hat{a}_i \varphi_i(\mathbf{r}). \quad (8.128)$$

After multiplying (8.128) by  $\varphi_j^*(\mathbf{r})$  and integration over the system volume, we gain an expression for the destruction operators  $\hat{a}_j$  (analogously for the creation operators):

$$\int d^3r \varphi_j^*(\mathbf{r}) \hat{\psi}(\mathbf{r}) = \sum_i \hat{a}_i \int d^3r \varphi_j^*(\mathbf{r}) \varphi_i(\mathbf{r}) = \sum_i \hat{a}_i \delta_{ij} = \hat{a}_j. \quad (8.129)$$

We insert this expression (8.129) for  $\hat{a}_j$  and the analogue for  $\hat{a}_j^+$  into the anti-commutation relation (8.127a)–(8.127c). The calculation is shown for (8.127a):

$$\begin{aligned} [\hat{a}_i^+, \hat{a}_j]_+ &= \int d^3r d^3r' \varphi_i(\mathbf{r}) \varphi_j^*(\mathbf{r}') \hat{\psi}^+(\mathbf{r}) \hat{\psi}(\mathbf{r}') \\ &\quad + \int d^3r d^3r' \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}') \hat{\psi}(\mathbf{r}') \hat{\psi}^+(\mathbf{r}) = \delta_{ij}. \end{aligned} \quad (8.130a)$$

Because of (8.127a) there is the Kronecker symbol on the right side of (8.130a). Therefore, in the integral we can assume  $i = j$  and by using the orthogonality of the eigenfunctions  $\varphi_i(\mathbf{r})$  we obtain

$$[\hat{a}_i^+, \hat{a}_j]_+ = \int d^3r \delta(\mathbf{r} - \mathbf{r}') [\hat{\psi}^+(\mathbf{r}), \hat{\psi}(\mathbf{r}')]_+ = \delta_{ij}. \quad (8.130b)$$

This finally yields

$$[\hat{\psi}^+(\mathbf{r}), \hat{\psi}(\mathbf{r}')]_+ = \delta(\mathbf{r} - \mathbf{r}'), \quad (8.131\text{a})$$

and after an analogous calculation

$$[\hat{\psi}^+(\mathbf{r}), \hat{\psi}^+(\mathbf{r}')]_+ = 0, \quad (8.131\text{b})$$

$$[\hat{\psi}(\mathbf{r}), \hat{\psi}(\mathbf{r}')]_+ = 0. \quad (8.131\text{c})$$

The anti-commutation relations for the fermion field operators (8.31) are fully analogous to the commutation relations for bosons (8.111a)–(8.111c). They only differ by the sign between the two operator products.

For fermion fields, the Hamilton (energy) operator  $\hat{H}$  looks identical with that for boson fields (8.116):

$$\begin{aligned} \hat{H} &= \int d^3r \hat{\psi}^+(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \hat{\psi}(\mathbf{r}) \\ &= \sum_{ij} \int d^3r \hat{a}_i^+ \varphi_i^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \varphi_j(\mathbf{r}) \hat{a}_j \\ &= \sum_{ij} \delta_{ij} \varepsilon_j \hat{a}_i^+ \hat{a}_j = \sum_i \varepsilon_i \hat{a}_i^+ \hat{a}_i. \end{aligned} \quad (8.132)$$

The function of  $\hat{a}_i^+$  as a creation operator is obvious from its action on the vacuum state  $|0\rangle$  and the corresponding field energy:

$$\begin{aligned} \hat{H}(\hat{a}_i^+ |0\rangle) &= \sum_j \varepsilon_j \hat{a}_j^+ \hat{a}_j \hat{a}_i^+ |0\rangle \\ &= \sum_j \varepsilon_j \hat{a}_j^+ (\delta_{ij} - \hat{a}_i^+ \hat{a}_j) |0\rangle = \varepsilon_i (\hat{a}_i^+ |0\rangle). \end{aligned} \quad (8.133)$$

We have used the relation  $\hat{a}_j |0\rangle = 0$ , a direct result from (8.127b), that is, the fact that no more particle can be destroyed in the vacuum state. Equation (8.133) shows that  $\hat{a}_i^+ |0\rangle$  is an eigenstate of the field, in which a particle of energy  $\varepsilon_i$  has been created. A second particle with the same quantum number  $i$  can not be created, because of (8.127c) it is required:

$$\hat{a}_i^+ \hat{a}_i^+ |0\rangle = -\hat{a}_i^+ \hat{a}_i^+ |0\rangle = 0. \quad (8.134)$$

However, for arbitrary particles with  $j, k, \dots, \neq i$  the relation (8.133) is valid. Consequently, the total energy of the field can be written as

$$\hat{H}|\phi\rangle = \sum_j \varepsilon_j \hat{a}_j^+ \hat{a}_j |\phi\rangle = \sum_j n_j \varepsilon_j. \quad (8.135)$$

The particle number  $n_j$  can only assume the values 1 and 0, exactly the property of fermions in a many-particle state. Like in (8.26b) and (8.123) we call

$$\hat{n}_j = \hat{a}_j^+ \hat{a}_j \quad (8.136)$$

the particle number operator for particles of type  $j$ . Its eigenvalues  $n_j$  are the particle numbers of type  $j$  in the fermion field.

In analogy to the bosonic state (8.124), the fermionic many-particle state might be expressed as

$$|\phi\rangle = |\dots, n_i, n_j, \dots\rangle = \prod_i (\hat{a}_i^+)^{n_i} |0\rangle, \quad n_i = 0, 1. \quad (8.137a)$$

Because of

$$(\hat{a}_i^+)^2 = 0 \quad \text{and} \quad (\hat{a}_i^+)^0 = 1, \quad (8.137b)$$

the following simpler representation, however, is possible for fermions:

$$|\phi\rangle = |\dots, n_i, n_j, \dots\rangle = \hat{a}_1^+ \hat{a}_2^+, \dots, \hat{a}_i^+ \hat{a}_j^+, \dots, \hat{a}_N^+ |0\rangle, \quad (8.137c)$$

when the many-body state contains  $N$  particles in total, each of type  $1, 2, \dots, i, j, \dots, N$ .

The operator of the total particle number  $\hat{N}$  is obtained by adding up the number operators (8.136) of all different particle types  $1, 2, \dots, i, j, \dots, N$ :

$$\hat{N} = \sum_j \hat{a}_j^+ \hat{a}_j. \quad (8.138)$$

The action of the creation and destruction operators  $\hat{a}_i^+$  and  $\hat{a}_i$  on a fermion many-body state can also be expressed in terms of occupation numbers, similarly as in (8.29)–(8.33b) for boson fields. In analogy to (8.29), we write:

$$\hat{a}_i^+ |\dots, n_i, \dots\rangle = C |\dots, (n_i + 1), \dots\rangle. \quad (8.139)$$

By multiplying (8.139) from the left with  $\langle \dots, n_i, \dots | \hat{a}_i$  and using (8.30), we obtain

$$\langle \dots, n_i, \dots | \hat{a}_i \hat{a}_i^+ | \dots, n_i, \dots \rangle = C^2 \langle \dots, (n_i + 1), \dots | \dots, (n_i + 1), \dots \rangle. \quad (8.140a)$$

Unlike for boson fields (8.20a), (8.20b) in the present case the fermionic anti-commutation rule (8.127a) is applied:

$$\begin{aligned} & -\langle \dots, n_i, \dots | \hat{a}_i^+ \hat{a}_i | \dots, n_i, \dots \rangle + \langle \dots, n_i, \dots | \dots, n_i, \dots \rangle \\ & = C^2 \langle \dots, (n_i + 1), \dots | \dots, (n_i + 1), \dots \rangle, \end{aligned} \quad (8.140b)$$

that is, with (8.136) we conclude

$$-n_i + 1 = C^2. \quad (8.140c)$$

By means of a similar calculation concerning the operation of  $\hat{a}_i$  on the state  $|\dots, n_i, \dots\rangle$ , we obtain the fermion analogues to (8.33a), (8.33b) as

$$\boxed{\hat{a}_i^+ |\dots, n_i, \dots\rangle = \sqrt{1-n_i} |\dots, (n_i+1), \dots\rangle}, \quad (8.141a)$$

$$\boxed{\hat{a}_i^- |\dots, n_i, \dots\rangle = \sqrt{n_i} |\dots, (n_i-1), \dots\rangle}. \quad (8.141b)$$

For the occupation numbers  $n_i = 0, 1$ , the properties typical for fermions are obtained from (8.141a), (8.141b): A fermion of type (i) can only be created starting from an initial state with  $n_i = 0$ . On the other hand, the action of  $\hat{a}_i$  on  $|\dots, n_i, \dots\rangle$  results in the state  $|\dots, (n_i-1), \dots\rangle$  only for  $n_i = 1$ .

In the following, we will restrict our discussion of the quantized Schrödinger field to fermion fields. In the present context, this is the more important case, since in particular electrons are concerned.

### 8.3.2 Field Operators and Back to the Single Particle Schrödinger Equation

The essential goal of quantum field theory, like any other theory, is the theoretical prediction of measurement results, that is, of expectation values as (8.114). For this purpose, we need a recipe for the construction of field operators  $\hat{\Omega}_{\text{field}}$  operating on the Schrödinger field. The question arises, how we derive the many-body field operator from the known operator of single particle quantum mechanics (Sect. 3.5). The general rule has already been presented in Sect. 8.3 in connection with the representation of the total energy of the field (8.105). The starting point is the many-body field function  $\psi(\mathbf{r})$  defined at each space position  $\mathbf{r}$ , where a particle might be created. For one single particle the single particle operator  $\hat{\Omega}$  operating at the position  $\mathbf{r}$  of that single particle is relevant for the calculation of the corresponding expectation value. In case of many particles in the field the expectation value, thus, is an integral over the volume of the whole system, that is, an integral over all possible particle coordinates  $\mathbf{r}$ :

$$\langle \Omega \rangle = \int d^3r \psi^*(\mathbf{r}) \hat{\Omega}(\mathbf{r}) \psi(\mathbf{r}). \quad (8.142a)$$

In analogy to (8.105) and (8.115), the field operator  $\hat{\Omega}_{\text{field}}$  of the field observable  $\Omega$  is obtained by replacing the field functions  $\psi^*$  and  $\psi$  in (8.142a) by the field operators  $\hat{\psi}^+(\mathbf{r})$  and  $\hat{\psi}(\mathbf{r})$ . In case of fermions these operators, of course, must obey the anti-

commutation relations (8.131a)–(8.131c). The expression for the field operator is written as

$$\hat{\mathcal{O}}_{\text{field}} = \int d^3r \psi^+(\mathbf{r}) \hat{\mathcal{O}}(\mathbf{r}) \hat{\psi}(\mathbf{r}). \quad (8.142b)$$

The action of the operators  $\hat{\psi}^+(\mathbf{r})$  and  $\hat{\psi}(\mathbf{r})$  on a field state  $|\dots, n_i, \dots\rangle$  as well as that of the operators  $\hat{a}_i^+$  and  $\hat{a}_i$  in the series expansion (8.128) is well known according to Sect. 8.3.1. Consequently the action of the field operator  $\hat{\mathcal{O}}_{\text{field}}$  on the field (many-body) state is known, too.

As an example, we calculate the total momentum of a many-electron field. The electrons are free particles described by plane waves with wave vectors  $\mathbf{k}_j$ .

With  $\hat{\mathbf{p}} = (\hbar/i)\nabla$  as single-particle momentum the momentum expectation value for the non-quantized field is

$$\langle \mathbf{p} \rangle = \int d^3r \psi^* \left( \frac{\hbar}{i} \nabla \right) \psi. \quad (8.143a)$$

We quantize by setting  $\psi^* \rightarrow \hat{\psi}^+$  and  $\psi \rightarrow \hat{\psi}$  and obtain the momentum operator of the quantized field as

$$\hat{\mathbf{p}} = \int d^3r \hat{\psi}^+(\mathbf{r}) \left( \frac{\hbar}{i} \nabla \right) \hat{\psi}(\mathbf{r}). \quad (8.143b)$$

Using the series expansion (8.128), the operator might be rewritten in terms of electron creation and destruction operators:

$$\hat{\mathbf{p}} = \sum_{jk} \hat{a}_j^+ \hat{a}_k \frac{\hbar}{i} \int d^3r \varphi_j^*(\mathbf{r}) \nabla \varphi_k(\mathbf{r}). \quad (8.144)$$

According to our initial assumption the electrons are free and described by single-particle wave functions  $\varphi_j(\mathbf{r}) \propto \exp(i\mathbf{k}_j \cdot \mathbf{r})$ . Therefore, the gradient operation on  $\varphi_k(\mathbf{r})$  in (8.144) generates the term  $i\mathbf{k}_k$ , that is,

$$\begin{aligned} \hat{\mathbf{p}} &= \sum_{jk} \hat{a}_j^+ \hat{a}_k \frac{\hbar}{i} \int d^3r (i\mathbf{k}_k) \varphi_j^*(\mathbf{r}) \varphi_k(\mathbf{r}) \\ &= \sum_{jk} \hat{a}_j^+ \hat{a}_k (\hbar \mathbf{k}_k) \delta_{jk} = \sum_j \hbar \mathbf{k}_j \hat{a}_j^+ \hat{a}_j. \end{aligned} \quad (8.145)$$

According to (8.136), the operation  $\hat{a}_j^+ \hat{a}_j$  on a field state  $|\phi\rangle$  yields the particle number  $n_j$ , that is:

$$\langle \phi | \hat{\mathbf{p}} | \phi \rangle = \sum_j \hbar \mathbf{k}_j n_j. \quad (8.146)$$

This is indeed the expected result. The expectation value of the total field momentum is the sum of all single particle momenta  $\hbar\mathbf{k}_j$ .

In order to calculate the total number of particles in the field the probability  $\psi^*(\mathbf{r})\psi(\mathbf{r})$  to find a particle at  $\mathbf{r}$  must be integrated over the system volume. After replacing the wave functions by the corresponding field operators, the operator of the total particle number is

$$\begin{aligned}\hat{N} &= \int d^3r \hat{\psi}^+(\mathbf{r})\hat{\psi}(\mathbf{r}) = \sum_{ij} \hat{a}_i^+ \hat{a}_j \int d^3r \varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r}) \\ &= \sum_{ij} \hat{a}_i^+ \hat{a}_j \delta_{ij} = \sum_j \hat{a}_j^+ \hat{a}_j.\end{aligned}\quad (8.147)$$

This is exactly the representation (8.138), which has already been concluded from a different consideration.

To answer the question for the particle density, that is, what is the probability to find a particle just at the position  $\mathbf{r}'$ , we must select from the integral (8.147) just the position  $\mathbf{r}'$ . This is done by inserting the  $\delta(\mathbf{r}' - \mathbf{r})$  function into the integral. This yields the particle density operator as

$$\hat{\rho}(\mathbf{r}') = \int d^3r \hat{\psi}^+(\mathbf{r})\delta(\mathbf{r}' - \mathbf{r})\hat{\psi}(\mathbf{r}) = \hat{\psi}^+(\mathbf{r}')\hat{\psi}(\mathbf{r}').\quad (8.148)$$

The straight forward way to the operator expression (8.148) would have been to replace the particle density  $\psi^*(\mathbf{r})\psi(\mathbf{r})$  of Schrödinger's theory by the corresponding field operators.

For a deeper understanding of the physical meaning of the field operators, it is interesting to elucidate the action of the particle density operator  $\hat{\rho}(\mathbf{r})$  (8.148) on the field state  $\hat{\psi}^+(\mathbf{r}')|0\rangle$ :

$$\begin{aligned}\hat{\rho}(\mathbf{r})(\hat{\psi}^+(\mathbf{r}')|0\rangle) &= \hat{\psi}^+(\mathbf{r})\hat{\psi}(\mathbf{r})\hat{\psi}^+(\mathbf{r}')|0\rangle \\ &= \hat{\psi}^+(\mathbf{r})[\delta(\mathbf{r}' - \mathbf{r}) - \hat{\psi}^+(\mathbf{r}')\hat{\psi}(\mathbf{r})]|0\rangle \\ &= \delta(\mathbf{r} - \mathbf{r}')(\hat{\psi}^+(\mathbf{r}')|0\rangle).\end{aligned}\quad (8.149)$$

For the calculation, we have used the anti-commutation relation (8.131a) and the fact that the action of the destruction operator  $\hat{\psi}$ , a sum of destruction operators  $\hat{a}_i$  (8.128), on the vacuum state  $|0\rangle$  yields zero.

Equation (8.149) can easily be interpreted: The  $\delta(\mathbf{r} - \mathbf{r}')$  function is the eigenvalue of the particle density operator for the eigenstate  $\hat{\psi}^+(\mathbf{r}')|0\rangle$ . The delta function is the measurement result of a particle density measurement on the field. The particle density is only different from zero near  $\mathbf{r}'$ , otherwise it vanishes.  $\hat{\psi}^+(\mathbf{r}')$  creates a particle at the position  $\mathbf{r}'$ , which is detected by the experimental set-up belonging to a particle density measurement, if this set-up passes along the position  $\mathbf{r}'$ .

We can furthermore conclude from (8.148) that field observables of the Schrödinger theory containing the wave function can directly be transferred into field operators by replacing the wave functions by the corresponding field operators  $\hat{\psi}^+(\mathbf{r})$  and  $\hat{\psi}(\mathbf{r})$ . By this simple rule we can write down the field operator  $\hat{\mathbf{j}}$  of the electrical current density as

$$\hat{\mathbf{j}}(\mathbf{r}) = \frac{e\hbar}{2mi} [\hat{\psi}^+(\mathbf{r}) \nabla \hat{\psi}(\mathbf{r}) - \hat{\psi}(\mathbf{r}) \nabla \hat{\psi}^+(\mathbf{r})]. \quad (8.150)$$

The operator of the total current is obtained as the volume integral over  $\hat{\mathbf{j}}$  (8.150).

To make contact to the single-particle Schrödinger quantum mechanics, we consider the field state  $\hat{a}_i^+ |0\rangle$  which contains only one single electron with wave-vector  $\mathbf{k}_i$  and spin up, i.e. an electron of type  $i = \mathbf{k}_i, \uparrow$ . In the single-particle Schrödinger theory, the probability to find this electron at position  $\mathbf{r}$  is given by  $\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r})$ . To find out what quantized field theory yields for this probability, we calculate the expectation value of the particle density operator  $\hat{\psi}^+\hat{\psi}$  for the single-electron field state  $\hat{a}_i^+ |0\rangle$ :

$$\begin{aligned} \langle \phi | \hat{\rho} | \phi \rangle &= \langle 0 | \hat{a}_i \hat{\psi}^+ \hat{\psi} \hat{a}_i^+ | 0 \rangle \\ &= \sum_{jk} \langle 0 | \hat{a}_i \hat{a}_j^+ \varphi_j^*(\mathbf{r}) \hat{a}_k \varphi_k(\mathbf{r}) \hat{a}_i^+ | 0 \rangle \\ &= \sum_{jk} \varphi_j^* \varphi_k \langle 0 | \hat{a}_i \hat{a}_j^+ \hat{a}_k \hat{a}_i^+ | 0 \rangle \\ &= \sum_{jk} \varphi_j^* \varphi_k \langle 0 | \hat{a}_i \hat{a}_j^+ (\delta_{ki} - \hat{a}_i^+ \hat{a}_k) | 0 \rangle \\ &= \sum_j \varphi_j^* \varphi_i \langle 0 | \hat{a}_i \hat{a}_j^+ | 0 \rangle \\ &= \sum_j \varphi_j^* \varphi_i \langle 0 | \delta_{ij} - \hat{a}_j^+ \hat{a}_i | 0 \rangle \\ &= \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}). \end{aligned} \quad (8.151)$$

For the derivation, we have used the anti-commutation relation (8.127a) two times and furthermore the fact that  $\hat{a}_j$  acting on the vacuum state  $|0\rangle$  yields zero. The field theoretical result is indeed identical with the outcome of the Schrödinger single-particle theory. This is just a requirement for each theory in physics: a higher-ranking theory as that of quantized many-particle fields must contain the limiting case of one single particle which is described by the simpler single-particle Schrödinger theory. Both theories must be consistent, to be “true”.

Therefore we expect that also the single-particle Schrödinger equation can be regained from the field theoretical formalism. For this purpose, we consider the time-independent Schrödinger equation on the quantized fermion field. The energy field operator (8.115), hereby, acts on the single particle field state  $\hat{a}_i^+ |0\rangle$  and the resulting expression must be equal to the energy eigenvalue times the field state

$\hat{a}_i^+ |0\rangle$ . This single-particle field state used so far is a special state, where the particle has a particular  $\mathbf{k}_i$  vector and spin. However, the particle must not be prepared with a well defined  $\mathbf{k}$  vector and spin; the most general single-particle field state is a superposition state containing all possible  $\mathbf{k}$  vectors and spins, that is, a wave packet with a broad  $\mathbf{k}$  distribution and both spin orientations. This general single-particle state, thus, is written as

$$|1\rangle = \sum_i c_i \hat{a}_i^+ |0\rangle. \quad (8.152a)$$

The expansion coefficients  $c_i$  must obey the normalization condition for one particle:

$$\sum_i |c_i|^2 = 1. \quad (8.152b)$$

According to (8.129), the operator  $\hat{a}_i^+$  can also be expressed in terms of the creation operator  $\hat{\psi}^+(\mathbf{r})$  as

$$\hat{a}_i^+ = \int d^3r \varphi_i(\mathbf{r}) \hat{\psi}^+(\mathbf{r}). \quad (8.153)$$

We insert (8.153) into (8.152a) and obtain the general single-particle state as

$$|1\rangle = \int d^3r \sum_i c_i \varphi_i(\mathbf{r}) \hat{\psi}^+(\mathbf{r}) |0\rangle. \quad (8.154a)$$

The sum over the orthonormal eigenfunctions  $\varphi_i(\mathbf{r})$  weighted by the amplitudes  $c_i$  represents a general function  $\chi(\mathbf{r})$ , that is, (8.154a) might also be expressed as

$$|1\rangle = \int d^3r \chi(\mathbf{r}) \hat{\psi}^+(\mathbf{r}) |0\rangle. \quad (8.154b)$$

Now we consider the field theoretical Schrödinger equation with  $\hat{H}$  from (8.115):

$$\hat{H}|1\rangle = E|1\rangle. \quad (8.155)$$

With (8.115) and (8.154b), we calculate the expression on the left side of (8.155):

$$\begin{aligned} \hat{H}|1\rangle &= \int d^3r d^3r' \hat{\psi}^+(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta_{\mathbf{r}} + V(\mathbf{r}) \right] \hat{\psi}(\mathbf{r}) \chi(\mathbf{r}') \hat{\psi}^+(\mathbf{r}') |0\rangle \\ &= \int d^3r d^3r' \left\{ \hat{\psi}^+(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta_{\mathbf{r}} + V(\mathbf{r}) \right] \chi(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') \right\} |0\rangle \\ &= \int d^3r \hat{\psi}^+(\mathbf{r}) |0\rangle \left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \chi(\mathbf{r}). \end{aligned} \quad (8.156)$$

For the calculation, the anti-commutation relation (8.131a) and the symmetry of the  $\delta$  function with respect to interchanging  $\mathbf{r}'$  with  $\mathbf{r}$  [ $\Delta_{\mathbf{r}}\delta(\mathbf{r} - \mathbf{r}') = \Delta_{\mathbf{r}'}\delta(\mathbf{r} - \mathbf{r}')$ ] have been used. According to (8.155), the expression (8.156) must be equal to

$$E|1\rangle = E \int d^3r \chi(\mathbf{r}) \hat{\psi}^+(\mathbf{r}) |0\rangle. \quad (8.157)$$

This relation (8.155) is valid for any position  $\mathbf{r}$  and the states  $\hat{\psi}^+(\mathbf{r}) |0\rangle$  for differing  $\mathbf{r}$  are linearly independent. Consequently the terms being multiplied by the single-particle field states must also be equal on the left and on the right side of (8.155), that is,

$$\left[ -\frac{\hbar^2}{2m} \Delta + V(\mathbf{r}) \right] \chi(\mathbf{r}) = E \chi(\mathbf{r}). \quad (8.158)$$

The probability amplitude of the most general single-particle state

$$\chi(\mathbf{r}) = \sum_i c_i \varphi_i(\mathbf{r}) \quad (8.159)$$

therefore, must obey the familiar single-particle Schrödinger equation. In the many-body field formalism, we have regained the single-particle Schrödinger theory.

In analogy to (8.152a), the general two-particle state  $|2\rangle$  can also be formulated:

$$|2\rangle = \sum_{ij} c_{ij} \hat{a}_i^+ \hat{a}_j^+ |0\rangle. \quad (8.160a)$$

By means of (8.153), one can rewrite (8.160a) in terms of the operators  $\hat{\psi}^+(\mathbf{r})$ :

$$|2\rangle = \int d^3r d^3r' \sum_{ij} c_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}') \hat{\psi}^+(\mathbf{r}) \hat{\psi}^+(\mathbf{r}') |0\rangle. \quad (8.160b)$$

Analogously to (8.159), a two-particle wave function can be identified with the probability amplitude

$$\chi(\mathbf{r}, \mathbf{r}') = \sum_{ij} c_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}'). \quad (8.161)$$

The two-particle state (8.160b) can, thus, be expressed as

$$|2\rangle = \int d^3r d^3r' \chi(\mathbf{r}, \mathbf{r}') \hat{\psi}^+(\mathbf{r}) \hat{\psi}^+(\mathbf{r}') |0\rangle. \quad (8.162)$$

The two-particle wave function  $\chi(\mathbf{r}, \mathbf{r}')$  must be anti-symmetric supposed the formalism correctly describes fermions. This anti-symmetry is directly evident from an interchange of the two particles at  $\mathbf{r}$  and  $\mathbf{r}'$ . A change of the coordinates yields:

$$\begin{aligned} & \int d^3r d^3r' \chi(\mathbf{r}', \mathbf{r}) \hat{\psi}^+(\mathbf{r}') \hat{\psi}^+(\mathbf{r}) |0\rangle \\ &= \int d^3r d^3r' \chi(\mathbf{r}', \mathbf{r}) [-\hat{\psi}^+(\mathbf{r}) \hat{\psi}^+(\mathbf{r}')] |0\rangle = -|2\rangle. \end{aligned} \quad (8.163)$$

By using the anti-commutation relation (8.131b) the interchange of the two particles causes a transformation of  $|2\rangle$  into  $-|2\rangle$ . The Pauli principle is fulfilled, indeed the requirement for fermions. This result demonstrates again that the formalism of the quantized fermionic Schrödinger field contains all those properties which have been presented in Chaps. 4 and 5 in the context of single-particle quantum physics.

### 8.3.3 The Particle Picture in Quantum Field Theory

At this point we must again stress the fundamentally distinct view of a particle in quantum field theory as compared to classical physics or even single particle Schrödinger quantum mechanics [14]. In quantum field theory a particle is a more or less localized excitation of the quantum field, a field quantum, which is created by the field operator  $\hat{\Psi}^+(\mathbf{r}')$  at the position  $\mathbf{r}'$  and which is destroyed by  $\hat{\Psi}(\mathbf{r})$  at  $\mathbf{r}$  [see (8.149)]. The particle itself has no identity. We can not follow its path. Movement of a particle through space can solely be described by sequential annihilation ( $\hat{\Psi}$ ) and creation ( $\hat{\Psi}^+$ ) of field quanta along that path. There is no entity which continuously proceeds along the path. The dynamical laws of the field (energy and momentum conservation) impose the correct link between the annihilated and the created field quantum (particle), such that a continuous path of an identifiable particle is feigned. But note, in quantum field theory one can not talk about one and the same particle at the starting point and at the end of the path..

Since the field theoretical formalism (Sects. 8.3.1 and 8.3.2) can analogously be transferred to relativistic fields (Dirac field for fermions and Klein-Gordon field for bosons), all kind of elementary particles (Sect. 5.6.4) are described as field excitations in their particular field. Similarly as in the present context non-relativistic electrons (as e.g. in solid material) are treated as excitations of the Schrödinger quantum field and relativistic photons as excitations of the quantized electromagnetic field, relativistic electrons in vacuum or electrons in general are described as excitations of the relativistic electronic Dirac field and quarks as excitations of the chromodynamics quantum fields of strong interaction (Sect. 5.6.4). We can thus assume quantum fields as the basic reality rather than single particles. Particles are only excitations of quantum fields. Since quantum fields are extended non-local entities, similarly as light waves, particle-wave duality, as we encounter it in the double slit experiment, appears in a new light, maybe less bizarre.

In the picture of the non-local quantum field, where particles lack their own identity, some counterintuitive aspects of quantum physics become more acceptable to our mind. In the double slit experiment with electrons (Sect. 2.4) the quantum field, which is extended over the whole space, determines, at which positions in space

electrons can be created and annihilated (detected). The shape and the dynamics of the field sensitively depend on boundary conditions, which differ significantly for a double slit or a single slit geometrical arrangement. The structure of the quantum field determines whether we detect electrons at positions restricted by the field interference pattern of the double slit (as in the case of light interference) or just simply in one broad single peak behind the single slit, when no interference pattern is imposed by the boundary conditions. In this view it is not relevant, whether the electrons are detected in large number simultaneously or one after each other in low density without having any mutual contact (Sect. 2.4.1). In both cases the geometrical shape of the quantum field determines, where electrons can be detected.

Also tunnelling of particles through a barrier (Sect. 3.6.4) can be looked at in a somewhat different light. The quantum field exists outside and within the barrier. An electron in front of the barrier is not identical with the electron which is detected after tunnelling behind the barrier. There are only two excitations of the quantum field, in front and behind the barrier, which obey the laws of the non-local coherent field. In this picture one cannot follow one and the same particle on its way through the barrier. Experiments which have been interpreted in the classical way in terms of a tunnelling particle with identity yield tunnelling velocities significantly higher than the speed of light [15]. This contradiction to special relativity is easily removed on the basis of the quantum field theoretical interpretation.

Even the counterintuitive outcome of correlation statistical experiments regarding the Einstein-Podolsky-Rosen (EPR) paradox and Bell's inequality relation (Sect. 7.2) appears acceptable in the view of quantum field theory. In these experiments two particles are produced in a common source (S in Fig. 7.5) and move away in opposite directions. Even at large distances from each other the measurement of the spin direction at one particle with a random result instantaneously imposes opposite spin direction of the remote other particle. The effect does not involve any information exchange between the particles, which can be excluded by the experimental conditions. In the quantum field theoretical interpretation the non-local extended two-particle field obeys symmetry conditions which are imposed by the creation process of the two particles in the source: Decay of a zero-spin source particle requires opposite spin orientations and momenta of the two resulting particles (field quanta). This symmetry property is ascribed to the quantum field as a whole and it is maintained also for the two particles regardless their mutual distance. The field states are much more restricted in their symmetry as if the particles were not correlated. There is no transfer of energy or information between the particles; only symmetry properties of the two-particle quantum field are responsible for the counterintuitive outcome of the EPR experiments.

It is obvious that quantum field theory yields a more intuitive approach to the interpretation of experiments being crucial for quantum physical understanding than single particle quantum mechanics. The essential issues are the non-locality of quantum fields and the lack of identity of a particle. In this sense the quantum field appears as the most fundamental object of reality in our understanding of the world. Quantum field theory and its formalism are obviously closer to reality than any other physical theory invented so far.

Nevertheless, it is necessary to emphasize, that a quantum field is not an entity or object in the classical philosophical sense [16]. As always in quantum physics, measurements on a quantum field yield random results with certain probabilities at a certain time and at a particular position in space. Consequently, unlike in a classical field, there are no fixed physical values which are attributed to a position in space but rather field operators  $\hat{\Psi}(\mathbf{r})$  and  $\hat{\Psi}^+(\mathbf{r}')$  describing in an abstract formalism a type of measurement. In order to get measurable quantities including probabilities for the occurrence of a certain measurement result at the position  $\mathbf{r}$  or  $\mathbf{r}'$ , the field operators must operate on the quantum state of the field, an abstract Hilbert vector (8.25), which is not defined at a particular position in space but rather attributed to the field as a whole. In this sense a quantum field may be better described in terms of a non-local, position dependent structure of relations. This view on the physical reality is sometimes called *structural realism* [16].

### 8.3.4 Electrons in Crystals: Back to the Single Particle Approximation

Condensed matter, in particular solids, are systems ideally suited for a theoretical description in terms of fermion many-body field theory. In a solid there are approximately  $10^{23}$  electrons per  $\text{cm}^3$  in chemical bonds or freely moving on a background of atomic nuclei with equal positive charge. Consequently the ideal solid is neutral as a whole. Mass is essentially incorporated in the tiny atomic nuclei, while the spatial extension of a solid is determined by the electrons. As electrons are much lighter, about 2000 times, than protons and neutrons in the nuclei, the theoretical treatment of the many-body system of a solid is performed on the basis of the so-called *Born–Oppenheimer approximation* (adiabatic approximation). Because of their comparatively small mass the electrons follow nearly momentarily (adiabatically) the movement of the heavy atomic nuclei. The dynamics of both systems, that of the electrons and that of the nuclei, are nearly decoupled. In a first step, we can theoretically treat both systems independently in good approximation and then introduce the interaction between electrons and nuclei as a perturbation in the next step (Sect. 8.4).

In this approximation, we assume the nuclei to be fixed in space. Their positive charges generate the binding potential for electrons due to attractive Coulomb forces. In the present context, we limit our consideration to crystals, an important class of materials in solid matter physics. In crystals the atom, respectively nuclei positions span a 3-dimensional (3D) periodic lattice. According to (6.225) the nuclei positions might be described by vectors  $\mathbf{r}_n$ . Hereby,  $n = (m, n, p)$  is a triple of numbers which count the number of basis vectors  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ , of the lattice to reach a particular atom(nucleus) starting from an atom which defines the zero position in the lattice. With this notation the periodic potential of the spatially fixed nuclei (rigid lattice) is represented as

$$V_G(\mathbf{r}) = V_G(\mathbf{r} + m\mathbf{a} + n\mathbf{b} + p\mathbf{c}) = V_G(\mathbf{r} + \mathbf{r}_n). \quad (8.164)$$

This lattice potential  $V_G$  acts on each electron in the crystal; therefore it must be inserted into the square bracket of the many-body Hamiltonian (8.115). Integration over the whole system volume in (8.115) guarantees that each electron is exposed to the potential. Apart from its interaction with the rigid lattice potential each electron is subject to Coulomb repulsion to neighbouring electrons. Two electrons at positions  $\mathbf{r}$  and  $\mathbf{r}'$  experience the Coulomb interaction  $e^2/4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|$ . Since the electron density in the single-particle Schrödinger theory is given by  $\rho(r) = e\psi^*\psi$ , the total interaction energy between the electrons (non-quantized) is represented as

$$E_{ww} = \frac{1}{2} \int d^3r d^3r' \psi^*(\mathbf{r}')\psi(\mathbf{r}) \frac{e^2}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|} \psi^*(\mathbf{r}')\psi(\mathbf{r}). \quad (8.165)$$

The integral includes the pair-wise interaction between two electrons at  $\mathbf{r}$  and  $\mathbf{r}'$ . Double counting of each electron in this expression is balanced by the prefactor 1/2. In this “classical” expression the wave functions commute with each other. Upon quantizing the many-body field the wave functions are replaced by non-commuting field operators. Hereby the question arises in what order these operators must be arranged. The following conditions must be fulfilled: The Coulomb interaction must vanish for a single particle field state, i.e. for a system with only one electron. Consequently the destruction operators  $\hat{\psi}(\mathbf{r})$  and  $\hat{\psi}(\mathbf{r}')$  must be arranged on the right side and the creation operators on the left side of the Coulomb energy term. Furthermore, the position coordinates  $\mathbf{r}$  and  $\mathbf{r}'$  must appear in opposite order in the creation and destruction operators. Otherwise we would not obtain the interaction energy (8.165) of the Schrödinger equation by calculating the expectation value of the interaction operator in the quantized field formalism.

With these ingredients the Hamilton operator of the many-electron (interacting) field in a crystal is obtained as

$$\begin{aligned} \hat{H} = & \int d^3r \hat{\psi}^+(\mathbf{r}) \left[ -\frac{\hbar^2}{2m}\Delta + V_G(\mathbf{r}) \right] \hat{\psi}(\mathbf{r}) \\ & + \int d^3r d^3r' \hat{\psi}^+(\mathbf{r})\hat{\psi}^+(\mathbf{r}') \frac{e^2}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|} \hat{\psi}(\mathbf{r}')\hat{\psi}(\mathbf{r}). \end{aligned} \quad (8.166)$$

The first integral describes, apart from the electron kinetic energy, the Coulomb attraction between each electron and the positive atomic nuclei expressed by the lattice potential  $V_G$ . The second integral adds up all energy contributions due to the Coulomb repulsion between the electrons. The many-particle operator  $\hat{H}$  (8.166) operates on electronic many-particle states of the type (8.137c). With  $\hat{a}_i$  being the adjoint operator (8.30) to  $\hat{a}_i^\dagger$ , the energy expectation value of the field is obtained as

$$\langle \phi | \hat{H} | \phi \rangle = \langle 0 | \hat{a}_N^\dagger, \dots, \hat{a}_2 \hat{a}_1 | \hat{H} | \hat{a}_1^\dagger \hat{a}_2^\dagger, \dots, \hat{a}_N^\dagger | 0 \rangle. \quad (8.167)$$

An optimum solution to the many-body problem is possible by applying the variational method (Sect. 6.2) as approximation and minimizing the expectation value (8.167):

$$\langle \phi | \hat{H} | \phi \rangle = \text{minimum}. \quad (8.168a)$$

The normalization of the field states must be respected as a constraint condition:

$$\langle \phi | \phi \rangle = 1. \quad (8.168b)$$

The calculation starts with the expansion of the field operators  $\hat{\psi}^+$  and  $\hat{\psi}$  in the Hamiltonian  $\hat{H}$  (8.166) in terms of orthogonal eigenfunctions  $\varphi_i(\mathbf{r})$  (8.128). Hereby the field operators are transformed into a sum of destruction ( $\hat{a}_i$ ) and creation ( $\hat{a}_i^+$ ) operators. The functions  $\varphi_i(\mathbf{r})$  and  $\varphi_i^*(\mathbf{r})$  are arbitrarily chosen, so far not in detail defined set of orthogonal functions. Their special form will be determined by the minimization procedure (8.168a), (8.168b). The outcome of this minimization calculation will be the optimum set of eigenfunctions (single-particle wave functions) for our many-particle problem. The expansion  $\hat{\psi}(\mathbf{r}) = \sum_i \hat{a}_i \varphi_i(\mathbf{r})$  inserted into the field Hamilton operator (8.166) yields:

$$\begin{aligned} \hat{H} = & \sum_{ij} \hat{a}_i^+ \hat{a}_j \int d^3r \varphi_i^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V_G(\mathbf{r}) \right] \varphi_j(\mathbf{r}) \\ & + \frac{1}{2} \sum_{ijkl} \hat{a}_i^+ \hat{a}_j \hat{a}_k^+ \hat{a}_l \int d^3r d^3r' \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}') \frac{e^2}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|} \varphi_k^*(\mathbf{r}') \varphi_l(\mathbf{r}). \end{aligned} \quad (8.169)$$

In the calculation of the expectation value  $\langle \phi | \hat{H} | \phi \rangle$ , which must be minimized, matrix elements of the following type are obtained:

$$\langle 0 | \hat{a}_N, \dots, \hat{a}_2 \hat{a}_1 | \hat{a}_i^+ \hat{a}_j | \hat{a}_1^+ \hat{a}_2^+, \dots, \hat{a}_N^+ | 0 \rangle, \quad (8.170a)$$

$$\langle 0 | \hat{a}_N, \dots, \hat{a}_2 \hat{a}_1 | \hat{a}_i^+ \hat{a}_j^+ \hat{a}_k \hat{a}_l | \hat{a}_1^+ \hat{a}_2^+, \dots, \hat{a}_N^+ | 0 \rangle. \quad (8.170b)$$

The evaluation by use of the anti-commutation relations (8.127a)–(8.127c) is left to the reader. Problem 8.6, where also some hints are given, is devoted to this evaluation. The following expression for the expectation value of the field energy results:

$$\begin{aligned} \langle \phi | \hat{H} | \phi \rangle = & \sum_i \int d^3r \varphi_i^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m} \Delta + V_G(\mathbf{r}) \right] \varphi_j(\mathbf{r}) \\ & + \frac{1}{2} \sum_{ij} \int d^3r d^3r' \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}) \frac{e^2}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|} \varphi_j^*(\mathbf{r}') \varphi_j(\mathbf{r}') \\ & - \frac{1}{2} \sum_{ij} \int d^3r d^3r' \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}') \frac{e^2}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|} \varphi_j^*(\mathbf{r}) \varphi_j(\mathbf{r}'). \end{aligned} \quad (8.171)$$

The first sum contains the expectation values of the Schrödinger single-particle energies of non-interacting electrons in the states  $|\varphi_i(\mathbf{r})\rangle$  (single-particle theory). Because of  $e\varphi_i^*\varphi_i$  being the electronic charge density the second sum adds up the Coulomb repulsion contributions between two electrons at the positions  $\mathbf{r}$  and  $\mathbf{r}'$ . This term is due to the classical Coulomb electron–electron interaction. In contrast, the third sum is of inherently quantum mechanical origin; it can not be understood on the basis of classical considerations. It contains state amplitudes (wave functions)  $\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r}')$  at different space positions. The term might be thought as originating from an interchange of the electrons at  $\mathbf{r}$  and  $\mathbf{r}'$  in the Coulomb interaction term (second sum). It is a direct consequence of the anti-symmetry of fermion many-particle states. Its contribution to the total field energy is called *Coulomb exchange interaction*.

To proceed with the minimization of (8.168a), (8.168b), the arbitrary eigenfunctions  $\varphi_i^*(\mathbf{r})$  are assumed to be variables in the variational calculus. As described in (8.106), the derivations with respect to these functions  $\partial/\partial\varphi_i^*(\mathbf{r})$  are calculated.

Furthermore, the constraint condition (8.168b) has to be taken into account. As is common in the variational calculus the constraint relation (8.168a) is multiplied by the so-called Lagrange multiplier, a free parameter, and added to the quantity (8.171) to be minimized. The sum of both is minimized. As Lagrange multiplier, we choose  $E$  (for energy, as will be seen to be adequate) and derivate with respect to  $\varphi_i^*(\mathbf{r})$ :

$$\frac{\partial}{\partial\varphi_i^*(\mathbf{r})}\left\{\langle\phi|\hat{H}|\phi\rangle+E\langle\phi|\phi\rangle\right\}=0. \quad (8.172a)$$

The constraint condition reads

$$\langle\phi|\phi\rangle=\int d^3r\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r})=1, \quad (8.172b)$$

and the minimization yields:

$$\begin{aligned} &\left[-\frac{\hbar^2}{2m}\Delta+V_G(\mathbf{r})\right]\varphi_i(\mathbf{r})+\varphi_i(\mathbf{r})\left[\sum_j\int d^3r'\frac{e^2}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{r}'|}\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r}')\right] \\ &-\sum_j\int d^3r'\varphi_i(\mathbf{r}')\frac{e^2}{4\pi\varepsilon_0|\mathbf{r}-\mathbf{r}'|}\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r})=E\varphi_i(\mathbf{r}). \end{aligned} \quad (8.173)$$

Neglecting the two sum terms on the left side equation (8.173) represents the familiar Schrödinger equation for the single-particle eigenfunctions  $\varphi_i(\mathbf{r})$  in the periodic potential  $V_G(\mathbf{r})$ . The Lagrange multiplier  $E$  is the energy eigenvalue in this formulation. The two sum terms on the left, however, transform the problem of the determination of the functions  $\varphi_i(\mathbf{r})$  into a complex integro-differential equation. This so-called *Hartree–Fock equation* can only be solved by using modern supercomputers. There are meanwhile sophisticated *self consistent field theoretical techniques* which can be treated on these computers. A breakthrough in this field was achieved by the *Charge Density Functional (CDF) theory* by which the electronic structure

of molecules, crystals, solid and liquid matter in general as well as of nanostructures can be calculated in good agreement with experimental data.

The two sum terms in (8.173) shall be considered a little bit more in detail. The first term is easily understood:  $e\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r}')$  represents the electronic charge density at the position  $\mathbf{r}'$ . Consequently, this term describes the Coulomb interaction between an electron in the state  $\varphi_i(\mathbf{r})$  and all other electrons in the states  $\varphi_j(\mathbf{r}')$  at positions  $\mathbf{r}'$ . The sum term in square brackets might be written as a potential contribution  $v(\mathbf{r})$  still containing the wave functions  $\varphi_j^*$  and  $\varphi_j$  which should be determined. With well guessed trial wave functions, one might approximately replace this term by a fixed position dependent potential, similar to  $V_G(\mathbf{r})$ . Apart from the second sum term we, then, regain the single-particle Schrödinger equation, but now with an effective potential  $V_{\text{eff}}(\mathbf{r}) = V_G(\mathbf{r}) + v(\mathbf{r})$  instead of the nuclear potential  $V_G(\mathbf{r})$ :

$$\left[ -\frac{\hbar^2}{2m}\Delta + V_{\text{eff}}(\mathbf{r}) \right] \varphi_i(\mathbf{r}) = E\varphi_i(\mathbf{r}). \quad (8.174)$$

In this approximation, the potential  $V_{\text{eff}}$  contains both the action of the positive atomic nuclei and that of all other electrons except the one for which the Schrödinger equation is solved. One can describe this effective potential as that of the atomic nuclei which are shielded by all other electrons except the one considered. Unlike the considered so-called *Coulomb term* the third term on the left side of (8.173) can not be reduced to an effective shielding potential, since the wave function at the position  $\mathbf{r}$  appears in the sum rather than in front of it. This inherently quantum mechanical term is called *exchange term*. It is often neglected in rough approximation or approximated using physically reasonable assumptions. These approximations also lead to a single-particle Schrödinger equation. This is a remarkable fact. The highly complex many-particle problem of solid matter has been reduced to a single-particle problem, of course, with strong simplifications for the Coulomb and the exchange terms in (8.173). It is obvious that many important properties of matter as for example, ferromagnetism or the energies of unoccupied electronic states can not be described quantitatively in this rough approximation.

### 8.3.5 The Band Model: Metals and Semiconductors

Nevertheless the single-particle Schrödinger equation (8.174) allows an understanding of essential properties of solids. The approach to understand a solid, in particular a crystal, on the basis of the Schrödinger equation (8.174) will be discussed more in detail in the following.

The starting point of our consideration is the effective potential  $V_{\text{eff}}(\mathbf{r})$  generated by the regular arrangement of the positive atomic nuclei and the electrons surrounding them. Now we consider one single electron which moves in this effective potential of electronically screened atomic nuclei. The corresponding single-electron states

are described by the wave functions  $\varphi_i(\mathbf{r})$ . Because of the Pauli exclusion principle, these states can be occupied by an electron only once (with spin degeneracy two-fold). In the sense of the present single-particle approximation, all electronic states [wave functions  $\varphi_i(\mathbf{r})$ ] and the corresponding energies  $E_i$  calculated by means of the Schrödinger equation (8.174) are, then, successively occupied by one electron up to a maximum energy, the Fermi energy  $E_F$  (Sect. 5.6.3).

In the next step, we calculate the electronic energy levels for an electron in the periodic potential of a crystal. It is emphasized that not only the potential of the positive nuclei  $V_G(\mathbf{r})$  but also the screened effective potential  $V_{\text{eff}}(\mathbf{r})$ , called *core* potential, has the translational symmetry of the periodic crystal lattice, that is:

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{eff}}(\mathbf{r} + \mathbf{r}_n). \quad (8.175)$$

The periodicity of the potential implies some interesting symmetry properties of the electronic states and the energy eigenvalues of the electrons. A periodic function can be expanded in a Fourier series:

$$V_{\text{eff}}(\mathbf{r}) = \sum_{\mathbf{G}} V_{\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}}. \quad (8.176)$$

To guarantee the translational invariance of  $V_{\text{eff}}(\mathbf{r})$  the vectors  $\mathbf{G}$  in the expansion must fulfill the following condition with respect to the lattice vectors  $\mathbf{r}_n = m\mathbf{a} + n\mathbf{b} + p\mathbf{c}$ :

$$\mathbf{G} \cdot \mathbf{r}_n = 2\pi\nu, \quad \nu \text{ integer}. \quad (8.177)$$

This is just the relation (6.236) which has been introduced for the definition of the reciprocal lattice in connection with the discussion of particle scattering on crystals (Sect. 6.6.4). The vectors  $\mathbf{G}$  in (8.176), thus, assume the manifold of reciprocal lattice vectors

$$\mathbf{G}_{hkl} = h\mathbf{g}_1 + k\mathbf{g}_2 + l\mathbf{g}_3. \quad (8.178)$$

The basis vectors  $\mathbf{g}_i$  of the reciprocal lattice are defined as in (6.237) in Sect. 6.6.4. For a 1D lattice with the interatomic distance  $a$  in real space, the reciprocal lattice vectors are, of course,  $G_h = h2\pi/a$ .

To solve the Schrödinger equation (8.174), we expand the wave function in (8.174) in terms of plane waves:

$$\varphi_i(r) = \sum_{\mathbf{k}} C_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (8.179)$$

For macroscopic crystal dimensions the wave vectors  $\mathbf{k}$  are distributed quasi-continuously (Sect. 3.6.1). Expansion (8.179) as well as the potential expansion (8.176) is inserted into the Schrödinger equation (8.174) and we obtain

$$\sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m} C_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} + \sum_{\mathbf{k}' \mathbf{G}} C_{\mathbf{k}'} V_{\mathbf{G}} e^{i(\mathbf{k}' + \mathbf{G}) \cdot \mathbf{r}} = E \sum_{\mathbf{k}} C_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (8.180)$$

After renaming the summation indices it follows:

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \left[ \left( \frac{\hbar^2 k^2}{2m} - E \right) C_{\mathbf{k}} + \sum_{\mathbf{G}} V_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} \right] = 0. \quad (8.181)$$

This relation must be fulfilled for any position  $\mathbf{r}$ . Consequently, the expression in square brackets must vanish, since it does not depend on  $\mathbf{r}$ , that is,

$$\left( \frac{\hbar^2 k^2}{2m} - E \right) C_{\mathbf{k}} + \sum_{\mathbf{G}} V_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} = 0. \quad (8.182)$$

This set of algebraic equations is an alternative representation of the Schrödinger equation (8.174) in wave vector space. The set of equations couples expansion coefficients  $C_{\mathbf{k}}$  of  $\varphi_i(\mathbf{r})$  with each other, the  $\mathbf{k}$  vectors of which differ by only reciprocal wave vectors  $\mathbf{G}$ , i.e.  $C_{\mathbf{k}}$  couples with  $C_{\mathbf{k}-\mathbf{G}}, C_{\mathbf{k}-\mathbf{G}'}, C_{\mathbf{k}-\mathbf{G}''}, \dots$

Due to the lattice periodicity the initially continuous problem (concerning  $\mathbf{k}$  vector distribution) simplifies to a problem with  $N$  discrete equations, where  $N$  is the finite number of elementary cells in real space. Each equation is attributed to a  $\mathbf{k}$  vector of a particular elementary cell of reciprocal space denoted by  $\mathbf{G}, \mathbf{G}', \mathbf{G}'', \dots$ . Each of the  $N$  equations has a solution which can be represented as a superposition of plane waves with wave vectors  $\mathbf{k}$  differing only by reciprocal lattice vectors  $\mathbf{G}$ . Thus, the eigenvalues of the Schrödinger equation (8.174) can be indexed by wave vector indices  $\mathbf{k}$ :  $E_{\mathbf{k}} = E(\mathbf{k})$ . The wave function (8.179) belonging to a particular energy eigenvalue  $E_{\mathbf{k}}$  can, therefore, be represented ( $i \rightarrow \mathbf{k}$ ) as

$$\varphi_{\mathbf{k}} = \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G}) \cdot \mathbf{r}} = \left( \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G} \cdot \mathbf{r}} \right) e^{-i\mathbf{k} \cdot \mathbf{r}} = u_{\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}}. \quad (8.183)$$

The function  $u_{\mathbf{k}}(\mathbf{r})$  is a Fourier series with discrete summation indices  $\mathbf{G}$ , the reciprocal lattice vectors: consequently this function has the translational periodicity of the crystal lattice. As in Sect. 3.6.1 for a macroscopic crystal, we can require periodic boundary conditions with  $k_i = 0, \pm 2\pi/L, \dots, \pm n_i 2\pi/L$  ( $L$  is macroscopic side length of crystal cube). Consequently, the Schrödinger equation (8.174) for an electron in the periodic crystal lattice potential is solved by modulated plane waves:

$$\varphi_{\mathbf{k}}(r) = u_{\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (8.184a)$$

The modulation amplitude has lattice periodicity:

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{r}_{\mathbf{n}}). \quad (8.184b)$$

After its discoverer, the wave function (8.184a), (8.184b) is called *Bloch wave* and the particular property of that function *Bloch theorem*.

From the properties of Bloch waves, we can derive further symmetry relations, as for example:

$$\begin{aligned}\varphi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) &= \sum_{\mathbf{G}'} C_{\mathbf{k}+\mathbf{G}-\mathbf{G}'} e^{-i\mathbf{G}' \cdot \mathbf{r}} e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{r}} \\ &= \left( \sum_{\mathbf{G}''} C_{\mathbf{k}-\mathbf{G}''} e^{-i\mathbf{G}'' \cdot \mathbf{r}} \right) e^{i\mathbf{k} \cdot \mathbf{r}} = \varphi_{\mathbf{k}}(\mathbf{r}),\end{aligned}\quad (8.185a)$$

that is,

$$\varphi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \varphi_{\mathbf{k}}(\mathbf{r}). \quad (8.185b)$$

Bloch waves with wave vectors differing by a reciprocal lattice vector are equal.

Be  $\hat{H}$  the single-particle Hamilton operator in (8.174) (not the many body field operator usually used in this chapter), then (8.174) might be written as

$$\hat{H}\varphi_{\mathbf{k}} = E(\mathbf{k})\varphi_{\mathbf{k}}. \quad (8.186a)$$

For the same problem, but shifted by a reciprocal lattice vector  $\mathbf{G}$  the Schrödinger equation is

$$\hat{H}\varphi_{\mathbf{k}+\mathbf{G}} = E(\mathbf{k} + \mathbf{G})\varphi_{\mathbf{k}+\mathbf{G}}, \quad (8.186b)$$

and by using (8.185b):

$$\hat{H}\varphi_{\mathbf{k}} = E(\mathbf{k} + \mathbf{G})\varphi_{\mathbf{k}}. \quad (8.186c)$$

Combined with (8.186a), we conclude:

$$E(\mathbf{k}) = E(\mathbf{k} + \mathbf{G}). \quad (8.187)$$

In the crystal potential with lattice translational periodicity, the single electron energy eigenvalues  $E(\mathbf{k})$  are periodic in the reciprocal space which is spanned by the reciprocal lattice vectors  $\mathbf{G}_{hkl}$ .

The periodicity interval, the elementary cell of the reciprocal  $\mathbf{k}$  space is called *Brillouin zone*. Its zero point at the center of the cell, also the zero point of reciprocal space, might be any reciprocal lattice point for an infinitely extended lattice. This central zero point of the Brillouin zone is usually called  $\Gamma$  point. For more details, we refer to textbooks of solid state physics [17]. For simplicity reasons, we limit the discussion to a monatomic 1D crystal with atomic distance  $a$  (atomic chain). Then, points of the 1D reciprocal lattice have a distance from each other of  $G = 2\pi/a$  and the 1D Brillouin zone centered around the zero point of the lattice has the two boundaries at  $\pm\pi/a$  (Fig. 8.7). At first, we assume the periodic potential  $V_{\text{eff}}(\mathbf{r})$  to be negligibly small, but the translational symmetry of the potential shall remain. In this case the solutions  $\varphi_k(x)$  of the Schrödinger equation (8.174) are plane waves  $\exp(i\mathbf{k}x)$ , the energies of which form a parabola  $E = \hbar^2 k^2 / 2m$ . According to (8.187), the lattice periodicity of the potential (in spite of the negligibly small potential)

requires a total energy dispersion which is periodic in  $k$  with a periodicity interval of the reciprocal lattice distance  $G$ . Therefore, the energy parabola must be repeated along the  $k$  axis, each time at a distance  $G$  (Fig. 8.7a). At the border lines of the Brillouin zones at  $\pm\pi/a, \pm3\pi/a, \pm5\pi/a$ , etc. parabolas from neighbouring Brillouin zones cut each other. The corresponding energy eigenvalues at these  $k$  values are degenerate. Consequently the two solutions of the Schrödinger equation belonging to the different parabolas are equally important. The most general solution, for example, at the point  $k = G/2$  (right border of the 1st Brillouin zone), is a superposition of the two waves with wave numbers  $k = G/2$  and  $k = (G/2) - G = -(G/2)$ . There are two possible superpositions with opposite sign between the partial waves:

$$\varphi_+ \propto (e^{iGx/2} + e^{-iGx/2}) \propto \cos\left(\pi \frac{x}{a}\right), \quad (8.188a)$$

$$\varphi_- \propto (e^{iGx/2} - e^{-iGx/2}) \propto \sin\left(\pi \frac{x}{a}\right). \quad (8.188b)$$

As is seen in Fig. 8.8a, the probability density  $\varphi_+^* \varphi_+$  accumulates electronic charge at the positions of the atomic nuclei while  $\varphi_-^* \varphi_-$  accumulates negative charge between the positive atomic cores. Both situations are different from that of a negligible effective potential where the electronic charge is homogeneously distributed along the atomic chain. In contrast, for an effective potential with finite height the two wave functions  $\varphi_+(x)$  and  $\varphi_-(x)$  being related to different charge distributions with respect to the atom positions correspond to different electron energies. The wave function  $\varphi_+(x)$  belongs to an energy somewhat below the free electron energy at negligible potential. For the wave function  $\varphi_-(x)$ , the electron energy is shifted upwards in comparison to that of a free electron. The two degenerate single electron energy levels at the Brillouin zone boundaries  $k = \pm G/2$  split and form a *forbidden band* on the energy scale (Fig. 8.8d). Within this forbidden band, no electron energy levels are found.

Therefore, the spectrum of single electron energies in a periodic potential consists of a sequence of alternating *allowed* and *forbidden energy bands* (Fig. 8.7b). Generally, the width of the allowed bands increases with growing energy (Fig. 8.7b). Within an allowed band, the energy dispersion  $E(k)$  exhibits oscillatory behavior with the periodicity of the reciprocal  $k$  space. In most cases it is, therefore, sufficient to limit the information about electronic energy bands to merely the 1st Brillouin zone centered around the  $\Gamma$  point. Both at the lower and at the upper edge of an allowed band the  $E(k)$  dependence is parabolic in good approximation, that is, the energy  $E$  is proportional to  $\pm k^2$ .

For free electrons, the energy dispersion  $E = \hbar^2 k^2 / 2m$  is parabolic with  $m$  as the free electron mass. Since  $E(k)$  of an electron in a crystal is also parabolic at the lower and upper edges of an electronic band, we can introduce a so-called *effective electron mass*  $m^*$  in those energy ranges of a band. Crystal electrons near the band edges, thus, behave under the action of a driving force (electric or magnetic field) as if they had a mass  $m^*$  which is determined by the curvature of the energy band

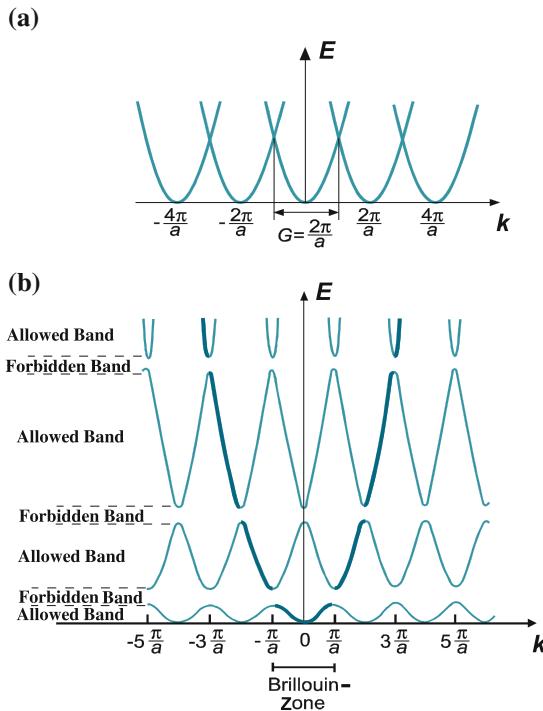
$E(k)$ . Corresponding to the curvatures of the energy parabolas at  $k = \pi/a$  or  $k = 0$  (upper and lower band edges) the effective electronic masses  $m^*$  are defined by

$$E(k) = \frac{\hbar^2 k^2}{2m^*} + E_0, \quad E_0 = E_{\max} \text{ or } E_{\min}, \quad (8.189)$$

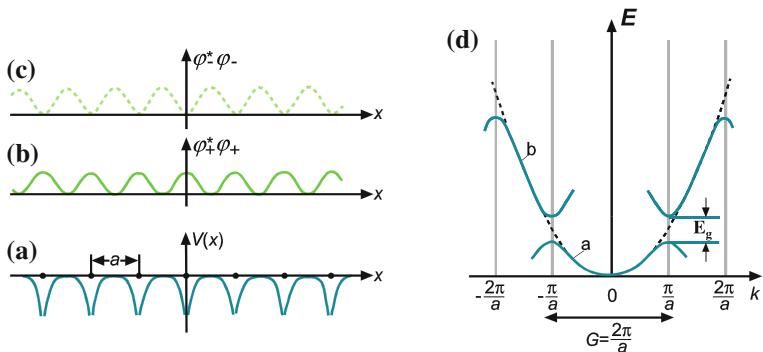
respectively,

$$m^* = \hbar^2 \left( \frac{d^2 E}{dk^2} \right)^{-1}. \quad (8.190a)$$

At the lower band edge the electronic effective mass  $m^*$  is positive as we are familiar with from a free electron in vacuum. At the upper edge, the band curvature is negative (Fig. 8.7b) and a negative effective electronic mass results. In this energy range, crystal electrons behave differently from our familiar picture (see below).



**Fig. 8.7** **a, b** Origin of electronic band structure of a periodic crystal lattice, one-dimensional with positive atomic cores at a linear distance  $a$  (linear atomic chain). **a** Free electron parabolas (single-electron approximation) periodically extended in the reciprocal space of wave numbers  $k$ . This  $E(k)$  dispersion is obtained in the limit of vanishing lattice potential (empty lattice). **b** Energy dispersion curves  $E(k)$  for an electron in a 1D lattice with atomic distance  $a$ , extended into reciprocal space beyond the 1st Brillouin zone. In the free electron approximation allowed and forbidden energy bands of electronic states are obtained. Parts of the free electron parabola are plotted in *thick solid line*

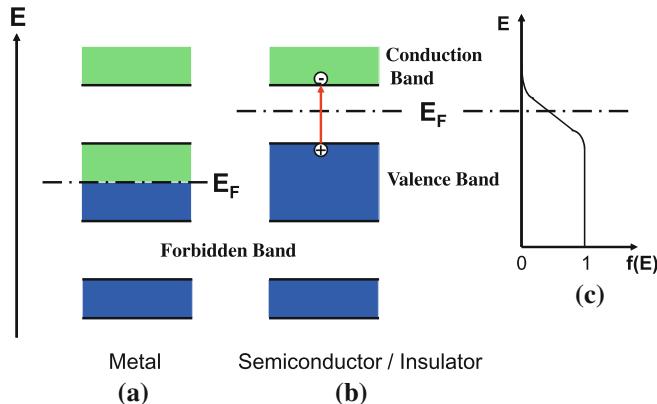


**Fig. 8.8** **a–d** Potential  $V(x)$  and electronic wave functions for an electron in the periodic 1D potential (atomic distance  $a$ ). **a** Qualitative plot of potential energy  $V(x)$ . Points indicate the positions of positive ion cores. **b** Probability density  $\rho_+ = \varphi_+^* \varphi_+$  of the standing electron wave, which results from Bragg reflection at  $k = \pm\pi/a$  at the upper edge of the allowed band (part **a**) in Fig. 8.8d (band maximum). **c** Probability density  $\rho_- = \varphi_-^* \varphi_-$  of the standing electron wave at the lower edge of the allowed band (part **b**) in Fig. 8.8d (band minimum). **d** Splitting of the free electron parabola (broken line) at the borders of the 1st Brillouin zone at  $k = \pm\pi/a$ ;  $E_g$  energetic width of the forbidden band

For a realistic 3D crystal with 3D reciprocal space (Sect. 6.6.4), the single-electron energy levels form periodic energy surfaces  $E(\mathbf{k})$  in this reciprocal  $\mathbf{k}$  space. Their maxima and minima are again described by paraboloids. In this case, the effective electron mass must be described most generally by a 3D tensor with different mass components  $m_{ij}$  for the different directions in  $\mathbf{k}$  space:

$$m_{ij} = \hbar^2 \left( \frac{\partial^2 E}{\partial k_i \partial k_j} \right)^{-1}. \quad (8.190b)$$

In the frame of the described single-electron approximation of solids all electronic states given by the band structure  $E(\mathbf{k})$  (Fig. 8.7b for one dimension) are now filled by one single electron up to a maximum energy, or by two electrons in case of spin degenerate bands. Depending on the available number of electrons per atom the energetically highest occupied energy level might be located within an allowed energy band (Fig. 8.9a) or it might be identical with the upper edge of an allowed band (Fig. 8.9b). In the band structure representation of Fig. 8.9 no notice has been taken of the periodic behavior of  $E(k)$  as a function of wave number  $k$ , only allowed and forbidden energy ranges are marked on the energy scale. This simplified electronic band structure is frequently used in semiconductor physics and electrical engineering. When an allowed band is partially filled as in Fig. 8.9a the highest filled energy level is identical with the Fermi energy  $E_F$ . Electrons in states at or somewhat below  $E_F$  can take up infinitesimal amounts of energy  $\delta E$  in an electric field; they can be accelerated by the field. More precisely, the electronic wave packets can move and can carry an electric current. Such materials are *metals*.



**Fig. 8.9** a–c Energetic location of the Fermi energy  $E_F$  in the band structure of a metal (a) and a semiconductor (b). a In a metal  $E_F$  lies within an allowed band and separates occupied (blue) from empty (green) electronic states. b In a semiconductor or an insulator  $E_F$  lies in the forbidden band between valence and conduction band. The excitation of an electron hole couple is indicated. c Fermi occupation distribution  $f(E)$  plotted in relation to semiconductor/insulator band scheme

If for a material the highest energy level of an allowed band is filled upon arranging the electrons on single-particle states, these electrons can not take up infinitesimal amounts  $\delta E$  of energy in an electric field. Below the highest occupied level all other states are occupied and Pauli principle forbids further occupation. Above the highest occupied level there are no electronic states to be occupied because of the forbidden band. At low temperatures, electrons can not be accelerated in an electric field; they can not carry an electric current. The material is an electric *insulator*.

For relatively narrow forbidden bands (0.1–3 eV typically), electrons can be excited thermally or optically by light irradiation from the initially fully occupied band into the empty band above. In the initially empty band, the few excited electrons find enough empty states in their direct energetic neighborhood; they can take up infinitesimal energy amounts in an electric field and can carry a current. These materials, in which quasi-free electrons can be generated by excitation, are *semiconductors*. The initially empty band is called *conduction band*, while the nearly occupied band (without excitation fully occupied) is called *valence band*. The most important semiconductors are Si, Ge, GaAs, InAs, GaN, ... .

Because of the excitation of carriers over the forbidden band being necessary for the conduction process semiconductors exhibit a concentration of free electrons and an electrical conductivity which increases exponentially with temperature [17]. Electrons which have been excited into the conduction band leave behind empty electronic states in the valence band. These empty states are called *defect electrons* or *holes*. In solid state physics, these holes at the upper edge of the valence band are shown to behave as positive charge carriers in an external electric field [17]. They have a positive effective mass, the amount of which is equal to that of the missing

valence electron in this state (8.190a), (8.190b). Remember that an electron at the upper valence band edge has a negative effective mass.

Upon such an excitation of electrons from the valence into the conduction band the number of electrons in the conduction band is equal to the number of holes in the valence band. The statistical distribution of electrons and holes, that is, of unoccupied electronic states on the energy scale, is governed by the Fermi statistics  $f(E)$  (Sect. 5.6.3). The probability for occupation and for non-occupation of an electronic state is symmetrical with respect to the Fermi energy  $E_F$  (Fig. 5.13). To guarantee this property of  $f(E)$  in presence of a forbidden band between conduction electrons and holes the Fermi energy (level),  $E_F$  must be located near midgap of the forbidden band (Fig. 8.9c). Slight deviations from midgap are due to different electronic state densities at the lower conduction band and upper valence band edge. The state densities usually differ because of different band curvatures (effective masses).

Because of the parabolic band dispersion at the band edges the densities of states near these edges are calculated as for the free electron gas in three dimensions (Sect. 3.6.1). In the corresponding square root terms,  $\sqrt{E}$  the free electron mass must only be replaced by the effective mass of a conduction electron or the effective mass of a hole in the valence band.

The effective mass (8.190a), (8.190b) rather than the free electron mass also determines the dynamics of a conduction electron or a hole in the valence band. This is evident from a consideration of the acceleration of an electron in the conduction band. The movement of the electron is described in terms of its wave packet (Sect. 3.2) which can be thought to be built up by Bloch waves (8.184a). The velocity of the wave packet is given by its group velocity

$$\mathbf{v} = \nabla_{\mathbf{k}}\omega(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}E(\mathbf{k}), \quad (8.191)$$

$E(\mathbf{k})$  is the energy dispersion surface of the band structure near the lower conduction band edge. An external electric field  $\mathcal{E}$  can move an electron only through empty states of the band structure, that is, along  $E(\mathbf{k})$ . Hereby, the electron energy changes by

$$\delta E = \nabla_{\mathbf{k}}E(\mathbf{k})\delta\mathbf{k}. \quad (8.192)$$

Using the correspondence principle (Sect. 3.4), we express this energy change by its classical analogue

$$\delta E = -e\mathcal{E} \cdot \mathbf{v}\delta t, \quad (8.193a)$$

respectively,

$$\dot{\mathbf{p}} = \hbar\dot{\mathbf{k}} = -e\mathcal{E}. \quad (8.193b)$$

From (8.191)–(8.193b), we derive the vector component  $v_i$  of the electron acceleration to be

$$\dot{v}_i = \frac{1}{\hbar} \frac{d}{dt} (\nabla_{\mathbf{k}} E) = \frac{1}{\hbar} \sum_j \frac{\partial^2 E}{\partial k_i \partial k_j} \dot{k}_j = \frac{1}{\hbar^2} \sum_j \frac{\partial^2 E}{\partial k_i \partial k_j} (-e\mathcal{E}_j). \quad (8.194)$$

This is a semi-classical dynamic equation for a crystal electron; it is fully analogous to the classical equation  $m^* \dot{v} = -e\mathcal{E}$ . Only the effective mass appears in its tensor representation (8.190b).

Finally, it must be mentioned that semiconductors can be doped. As an example, we consider arsenic (As) atoms which might substitute Si atoms in low concentration in a Si crystal. Technologically this substitution is achieved by ion implantation, diffusion or during the growth process itself. A pentavalent As atom incorporated in a Si crystal on the site of a missing tetravalent Si atom has one additional valence electron which does not participate in the four covalent chemical bonds to neighbouring Si atoms. This additional electron is weakly bonded to the As core (typical binding energies in the 30 meV range); it can easily be excited, already at room temperature, into the conduction band of the Si crystal, where it participates in the electrical conductance. Since As easily donates one outer valence electron into the conduction band, this type of pentavalent defect atom in Si is called a *donor*. For most semiconductors, the forbidden band is too broad to enable enough “intrinsic” conduction by thermal excitation. For most applications in electronics, doping is required. For enhanced hole conduction, so-called *acceptor* dopants are incorporated into Si and other semiconductors. These are trivalent atoms as for example, boron, where one outer valence electron is missing in comparison with the four electrons needed for complete bonding in the lattice. These acceptor atoms, therefore, easily accept one electron from the valence band of the semiconductor. The additional creation of a hole increases the hole conductance.

For a more thorough discussion of these topics, the reader should refer to textbooks of solid state physics [4, 17].

## 8.4 Quantized Lattice Waves: Phonons

In Sect. 8.3.4, we have seen, that in a crystal the dynamics of the heavy atomic cores is largely decoupled from that of the light electrons. Within the frame of the Born–Oppenheimer approximation, both systems are separately described and their interaction is treated as a perturbation. Essential properties of the many-electron system of the solid can, thus, be obtained by considering the electron dynamics within the rigid periodic potential of the atomic cores.

On the other hand, the much slower dynamics of the heavy atomic cores is described separately within the potential generated by the chemical bonds. The chemical bonds between the atoms originate from the outer valence electrons of the atoms, independently of the particular type of bond (covalent, ionic etc.). The potential for the atomic core dynamics is generated by part of the crystal electrons and it has the translational symmetry of the crystal. In a stable periodic crystal, the

atomic cores are fixed in bonding potentials. Therefore, we expect, in first approximation, the atomic cores to perform periodic harmonic oscillations in their binding potentials, as was described for a harmonic oscillator in general (Sect. 4.4). Since an oscillating atom affects its neighbouring atoms, the oscillations of different atoms are mutually coupled. Consequently, the dynamics of the atomic cores in a crystal lattice is expected to be described by collective excitations extended over large areas or the whole crystal. The adequate picture is that of waves in which the single atomic cores are the oscillating objects. This field picture is familiar to us from classical continuum theory and elastic waves such as sound waves in a solid. Since in quantum mechanics oscillatory motions must be quantized (Sect. 4.4), the oscillation field of the vibrating atomic cores, so-called lattice waves, must also be quantized.

To put these considerations into a mathematical formalism, we introduce a simple one-dimensional (1D) model of a crystal, the so-called *linear monatomic chain* (Fig. 8.10). Identical atoms are chemically bonded in a 1D arrangement with an interatomic distance  $a$ . The bonding of the atoms is described in first approximation by a parabolic potential, that is, the force on an atom  $v$  is proportional to its displacement  $u_v$  from the equilibrium position. The proportionality constant might be denoted by a spring force constant  $f$ . Taking into account the forces exerted by the two next neighbouring atoms from left and right on atom  $v$  with mass  $M$  the classical dynamic equation for this atom reads

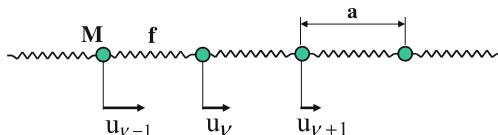
$$M\ddot{u}_v(t) = f[u_{v+1}(t) - u_v(t)] - f[u_v(t) - u_{v-1}(t)], \quad (8.195a)$$

that is,

$$M\ddot{u}_v(t) = f(u_{v+1} + u_{v-1} - 2u_v). \quad (8.195b)$$

The linear atomic chain is assumed to consist of  $N$  atoms, that is, it has a length  $L = Na$ . Because of the macroscopic dimension of  $L$  periodic boundary conditions (Sect. 3.6.1) can be used in the further treatment, that is, the atomic displacements  $u_v(t)$  obey the relation

$$u_v(t) = u_{v+N}(t) \quad \text{with } -\frac{N}{2} \leq v < \frac{N}{2}. \quad (8.195c)$$



**Fig. 8.10** Model of monatomic linear chain, consisting of atoms with mass  $M$  at the periodicity distance  $a$ .  $u_v$  are atomic displacements within an excited lattice wave. Chemical bonding forces are modeled by springs (force constant  $f$ )

The displacements  $u_v(t)$  are defined at fixed discrete lattice sites  $x_v$ , the equilibrium position of the atom  $v$ . Consequently,  $u_v(x_v, t)$  is a vibration field which has physically meaningful values only at discrete spatial coordinates  $x_v$ , in contrast to the electromagnetic or the Schrödinger particle field being defined continuously in space.

To solve the system of differential equations (8.195a)–(8.195c), we make an ansatz with plane waves which propagate along the linear chain:

$$u_v(t) = u_0(\kappa) e^{-i\omega(\kappa)t} e^{i\kappa v a} = \frac{1}{\sqrt{N}} A_\kappa(t) e^{i\kappa v a}. \quad (8.196a)$$

$x_v = va$  are the atomic positions and the factor  $1/\sqrt{N}$  in

$$u_0(\kappa) e^{-i\omega(\kappa)t} = \frac{1}{\sqrt{N}} A_\kappa(t) \quad (8.196b)$$

guarantees orthonormality of the ansatz (8.196a) in the sense of

$$\sum_v^N \left( \frac{1}{\sqrt{N}} e^{i\kappa v a} \right)^* \left( \frac{1}{\sqrt{N}} e^{i\kappa v a} \right) = 1, \quad (8.197a)$$

$$\sum_v^N \frac{1}{N} e^{i(\kappa - \kappa') v a} = 0 \quad \text{for } \kappa \neq \kappa', \quad (8.197b)$$

that is,

$$\sum_v^N \frac{1}{\sqrt{N}} e^{-i\kappa v a} \frac{1}{\sqrt{N}} e^{-i\kappa' v a} = \delta_{\kappa, \kappa'}. \quad (8.197c)$$

After inserting the displacement  $u_v(t)$  (8.196a), (8.196b) into the dynamic equation (8.195a)–(8.195c), we obtain:

$$M \ddot{A}_\kappa e^{i\kappa v a} = f(A_\kappa e^{i\kappa a} + A_\kappa e^{-i\kappa a} - 2A_\kappa), \quad (8.198a)$$

respectively,

$$\ddot{A}_\kappa = \frac{f}{M} 2(\cos \kappa a - 1) A_\kappa = -4 \frac{f}{M} \sin^2 \frac{\kappa a}{2}. \quad (8.198b)$$

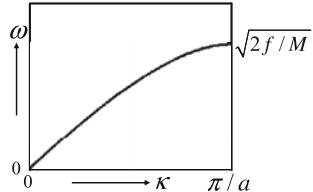
This is the differential equation of a harmonic oscillator (Sect. 4.4):

$$\ddot{A}_\kappa(t) + \omega_\kappa^2 A_\kappa(t) = 0 \quad (8.199a)$$

with

$$\omega_\kappa^2 = 4 \frac{f}{M} \sin^2 \frac{\kappa a}{2}. \quad (8.199b)$$

**Fig. 8.11** Dispersion curve  $\omega(\kappa)$  of lattice waves of a monatomic linear chain (Fig. 8.10)



The dynamic equations of the monatomic linear chain decouples into  $N$  different oscillator differential equations. These equations are numerated according to discrete wave numbers  $\kappa$ . For large  $N$  values, however,  $\kappa$  is distributed quasi-continuously. The frequencies of the oscillators as a function of the wave number  $\kappa$  obey the dispersion law (8.199b), which is plotted in Fig. 8.11. Because of the translational symmetry of the linear atomic chain the dispersion relation (8.199b) is periodic in the reciprocal space of wave numbers  $\kappa$ , with a periodicity interval  $2\pi/a$  (Brillouin zone). This is analogous to the behavior of electron waves in Sect. 8.3.4.

The field of lattice waves decouples into  $N$  harmonic oscillators. These oscillators, of course, obey the laws of quantum mechanics (Sect. 4.4) and we directly conclude that according to (4.122) the energy of an oscillator with quantum number  $\kappa$  is:

$$E_\kappa = \left( n_\kappa + \frac{1}{2} \right) \hbar \omega_\kappa \quad (8.200)$$

$n_\kappa$  indicates the excitation state of the oscillator  $\kappa$ . The total energy of the field is obtained by summing up the single oscillator contributions (8.200):

$$E = \sum_\kappa \left( n_\kappa + \frac{1}{2} \right) \hbar \omega_\kappa. \quad (8.201)$$

This field energy of lattice waves is totally analogous to the energy of the quantized light field (8.26b). Therefore, the excitation number  $n_\kappa$  of the oscillator  $\kappa$  can also be interpreted as the number of bosonic excitations of the field of lattice waves of a crystal. Correspondingly, we attribute to the field of lattice waves a type of excitations, that is, (quasi)particles which are called *phonons*. The name is derived from the Greek word  $\varphi\omega\nu\eta'$  (phone) meaning sound. The underlying reason for the name is the nature of lattice waves with frequencies  $\omega(\kappa)$  at small wave numbers  $\kappa < 1/a$  (large wavelengths  $\lambda > a$ ), where the sine dependence of the dispersion (8.199b) can be approximated by a linear function. In this wave number limit, the lattice waves are identical with the sound waves of classical continuum mechanics [17].

In order to deepen the concept of phonons as quasi-particles of the lattice wave field, we quantize the wave field according to the established rules. For this purpose, the dynamic equations (8.195a)–(8.195c) must be transferred into the Hamilton formalism. From (8.195a)–(8.195c), the total energy of the lattice wave field representing simultaneously the Hamilton function is obtained as

$$H = \sum_{\nu} \frac{p_{\nu}^2}{2M} + \frac{1}{2} f \sum_{\nu} (u_{\nu} - u_{\nu+1})^2, \quad (8.202)$$

$p_{\nu} = M\dot{u}_{\nu}$  is the momentum of the vibrating atom  $\nu$ . Using (8.196a), (8.196b), the most general expressions for the atomic displacements  $u_{\nu}$  and the momenta  $p_{\nu}$ , respectively, are expansions in Fourier series:

$$u_{\nu}(t) = \sum_{\kappa} \frac{1}{\sqrt{N}} [A_{\kappa}(t)e^{i\kappa\nu a} + A_{\kappa}^*(t)e^{-i\kappa\nu a}], \quad (8.203a)$$

$$p_{\nu}(t) = M\dot{u}_{\nu} = \sum_{\kappa} \frac{i\omega_{\kappa} M}{\sqrt{N}} [-A_{\kappa}(t)e^{i\kappa\nu a} + A_{\kappa}^*(t)e^{-i\kappa\nu a}]. \quad (8.203b)$$

The series terms with  $A_{\kappa}^*$  must be added to those with  $A_{\kappa}$  in order to guarantee real valued displacements  $u_{\nu}$  as is required for a realistic atomic oscillation. This is analogous to the electromagnetic field, where the vector potential  $\mathbf{A}$  is a real valued quantity.

As is easily seen, (8.202) yields the correct Hamilton equations by differentiation:

$$\dot{u}_{\nu} = \frac{\partial H}{\partial p_{\nu}} = \frac{1}{M} p_{\nu}, \quad (8.204a)$$

$$\dot{p}_{\nu} = -\frac{\partial H}{\partial u_{\nu}} = f(u_{\nu+1} + u_{\nu-1} - 2u_{\nu}). \quad (8.204b)$$

Because of  $\omega_{\kappa} = \omega_{-\kappa}$  and by use of (8.196a), (8.196b), we can express the displacements and the momenta also as

$$u_{\nu}(t) = \sum_{\kappa} \frac{1}{\sqrt{N}} (A_{\kappa} + A_{-\kappa}^*) e^{i\kappa\nu a}, \quad (8.205a)$$

$$p_{\nu}(t) = \sum_{\kappa} \frac{-i\omega_{\kappa} M}{\sqrt{N}} (A_{\kappa} - A_{-\kappa}^*) e^{i\kappa\nu a}, \quad (8.205b)$$

or after Fourier transformation:

$$A_{\kappa} + A_{-\kappa}^* = \frac{1}{\sqrt{N}} \sum_{\nu} e^{-i\kappa\nu a} u_{\nu}(t). \quad (8.205c)$$

Already from the Hamilton equations (8.204a), (8.204b), we could evaluate  $p_{\nu}$  and  $u_{\nu}$  as the canonical variables to be replaced by operators, which obey the bosonic commutation relations:

$$\hat{p}_{\nu} \hat{u}_{\mu} - \hat{u}_{\mu} \hat{p}_{\nu} = \frac{\hbar}{i} \delta_{\mu\nu}, \quad (8.206a)$$

$$\hat{u}_{\mu} \hat{u}_{\nu} - \hat{u}_{\nu} \hat{u}_{\mu} = 0, \quad (8.206b)$$

$$\hat{p}_\mu \hat{p}_\nu - \hat{p}_\nu \hat{p}_\mu = 0. \quad (8.206c)$$

According to (8.203a), (8.203b) and (8.205a)–(8.205c), the quantities  $A_\kappa(t)$  and  $A_\kappa^*(t)$  also become operators and their sequence order is important upon multiplication. Having this in mind, we insert (8.203a), (8.203b) and (8.205a)–(8.205c), respectively, into the Hamilton function (8.202). Using the dispersion relation (8.199b) and the normalization condition (8.197a), we obtain:

$$H = \sum_\kappa M\omega_\kappa^2 [A_\kappa^*(t)A_\kappa(t) + A_\kappa(t)A_\kappa^*(t)]. \quad (8.207)$$

The analogy to the electromagnetic field (8.14) is obvious. Similarly as in (8.206a)–(8.206c), we quantize the Hamiltonian (8.207) by replacing  $A_\kappa(t)$  and  $A_\kappa^*(t)$  by the corresponding operators  $\hat{A}_\kappa$  and  $\hat{A}_\kappa^+$ . This transforms the Hamilton function into a Hamilton field operator  $\hat{H}$ . Using the definition

$$\hat{A}_\kappa(t) = \sqrt{\frac{\hbar}{2M\omega_\kappa}} \hat{c}_\kappa(t) \quad (8.208)$$

we introduce dimensionless amplitudes, respectively field operators  $\hat{c}_\kappa$  and  $\hat{c}_\kappa^+$ , and the Hamilton operator (8.207) can be expressed as

$$\hat{H} = \sum_\kappa \hbar\omega_\kappa \frac{1}{2} [\hat{c}_\kappa^+ \hat{c}_\kappa + \hat{c}_\kappa \hat{c}_\kappa^+]. \quad (8.209)$$

Using (8.205a)–(8.205c) and (8.208), we derive relations between the field operators  $\hat{c}_\kappa$  and  $\hat{c}_\kappa^+$  and the canonical variables (operators)  $\hat{u}_\nu$  and  $\hat{p}_\nu$ :

$$\hat{u}_\nu(t) = \sum_\kappa \sqrt{\frac{\hbar}{2MN\omega_\kappa}} e^{i\kappa\nu a} (\hat{c}_\kappa + \hat{c}_{-\kappa}^+), \quad (8.210a)$$

$$\hat{p}_\nu(t) = \sum_\kappa (-i) \sqrt{\frac{\hbar M\omega_\kappa}{2N}} e^{i\kappa\nu a} (\hat{c}_\kappa - \hat{c}_{-\kappa}^+). \quad (8.210b)$$

By Fourier transformation and adding and subtracting the expressions, we obtain the inverse relations ( $\omega_\kappa = \omega_{-\kappa}$ ):

$$\hat{c}_\kappa = \sum_\nu \left( \sqrt{\frac{M\omega_\kappa}{2\hbar N}} \hat{u}_\nu + i \sqrt{\frac{1}{2\hbar MN\omega_\kappa}} \hat{p}_\nu \right) e^{-i\kappa\nu a}, \quad (8.211a)$$

$$\hat{c}_\kappa^+ = \sum_\nu \left( \sqrt{\frac{M\omega_\kappa}{2\hbar N}} \hat{u}_\nu - i \sqrt{\frac{1}{2\hbar MN\omega_\kappa}} \hat{p}_\nu \right) e^{i\kappa\nu a}. \quad (8.211b)$$

By means of the commutation relations (8.206a)–(8.206c), the following commutation rules for  $\hat{c}_\kappa$  and  $\hat{c}_\kappa^+$  are derived:

$$[\hat{c}_\kappa, \hat{c}_{\kappa'}^+] = \hat{c}_\kappa \hat{c}_{\kappa'}^+ - \hat{c}_{\kappa'}^+ \hat{c}_\kappa = \delta_{\kappa, \kappa'}, \quad (8.212a)$$

$$[\hat{c}_\kappa, \hat{c}_{\kappa'}] = [\hat{c}_\kappa^+, \hat{c}_{\kappa'}^+] = 0. \quad (8.212b)$$

These commutation rules are identical with those of the bosonic photon field (8.20a), (8.20b) in Sect. 8.2.

Therefore,  $\hat{c}_\kappa$  and  $\hat{c}_\kappa^+$  are destruction and creation operators, which destruct or create an excitation, that is, a quasi-particle or phonon with wave vector  $\kappa$  in the field of lattice waves. Analogously to the quantized electromagnetic field (Sect. 8.2) the Hamilton operator of the phonon field is obtained from (8.209) by using the commutators (8.212a), (8.212b):

$$\hat{H} = \sum_\kappa \hbar \omega_\kappa \hat{c}_\kappa^+ \hat{c}_\kappa + \frac{1}{2} \sum_\kappa \hbar \omega_\kappa. \quad (8.213)$$

As for the light field, the phonon field of lattice waves has a ground state (vacuum) energy

$$E_0 = \frac{1}{2} \sum_\kappa \hbar \omega_\kappa. \quad (8.214a)$$

This energy must be attributed to the crystal lattice even in case that no lattice waves are excited anymore in the ground state  $|0\rangle$ . In analogy to the light field [see (8.26b)],  $\hat{n}_\kappa = \hat{c}_\kappa^+ \hat{c}_\kappa$  represents the phonon number operator; it indicates the number of phonons with wave number  $\kappa$  being excited in the field state  $|\phi\rangle = |\dots, n_{\kappa-1}, n_\kappa, n_{\kappa+1}, \dots\rangle$ . For this multi-phonon state  $|\phi\rangle$ , the field Hamiltonian (8.213) has the eigenvalue (8.201), the total field energy.

As for the electromagnetic field, the most general bosonic multi-phonon state can be generated by multiple operation of phonon creation operators  $\hat{c}_\kappa^+$  on the ground state  $|0\rangle$ :

$$|\phi\rangle = |\dots, n_\kappa, \dots\rangle = \prod_\kappa \frac{1}{\sqrt{n_\kappa!}} (\hat{c}_\kappa^+)^{n_\kappa} |0\rangle. \quad (8.214b)$$

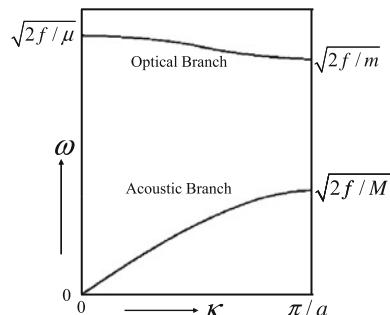
In general, all relations for boson creation and destruction operators, for example, (8.33a), (8.33b), are identical for photon and phonon fields.

In spite of the identical formal description of the electromagnetic field (Sect. 8.2) and the phonon field (because of their bosonic character), there is an essential difference concerning the physical nature of both fields. Photons as excitations of the electromagnetic field are real elementary particles (Sect. 5.6.4). They do exist in vacuum, separately from any medium carrying them. Their existence as  $\gamma$  quanta is documented in elementary particle physics by scattering experiments in high energy accelerators. On the other hand, phonons are collective excitations which are built

up from vibrations of crystal atoms. They do not exist in free space but are rather linked to the medium of a crystal. Phonons are therefore called quasi-particles. For both types of field excitations, however, the laws of quantum mechanics are valid. Therefore, both excitations obey the same type of a boson quantum field theory.

A further difference concerns the space symmetry of the two fields. The electromagnetic field in vacuum has infinitesimal translational symmetry and the space is homogeneous and isotropic. Correspondingly, the light dispersion relation  $\omega_L = cq$  is valid in reciprocal space without any further symmetry requirements with regard to special wave vector directions  $\mathbf{q}$ . Lattice waves and their quanta, the phonons, however, exist on a crystal lattice. This lattice has translational symmetry, but with discrete periodicity intervals. In real space, the lattice is described by the lattice vectors  $x_v = va$  in the 1D case and by  $\mathbf{r}_n = m\mathbf{a} + n\mathbf{b} + p\mathbf{c}$  in a 3D crystal. Correspondingly, the reciprocal space of wave vectors has also translational symmetry with discrete reciprocal wave vectors  $\mathbf{G}_{hkl}$  (8.178). The phonon dispersion relation  $\omega(\kappa)$  is, therefore, also periodic in the reciprocal space (Fig. 8.11), similarly as the electronic energy bands (8.187) in Sect. 8.3.4.

We want to extend our discussion of lattice waves and phonons from the simple case of a monatomic linear chain (Figs. 8.10 and 8.11) to a diatomic chain. Even though this model has little in common with a real solid it is frequently used for the description of a crystal with two different atoms per unit cell. We consider again a linear chain in which all nearest neighbors are connected by identical springs with force constant  $f$ . The unit cell contains two different atoms with masses  $M$  (heavy) and  $m$  (light). In this case, two vibration equations of the type (8.195a)–(8.195c) exist, one for each mass. Since mass  $M$  is coupled to its neighbouring mass  $m$  and vice versa, the two differential equations are coupled via the corresponding atomic displacements and the solution yields two different dispersion branches of possible lattice waves (Fig. 8.12) [17]. Phonons of the energetically lower branch



**Fig. 8.12** Dispersion curves  $\omega(\kappa)$  of lattice waves of a diatomic linear chain with alternately arranged atoms of mass  $M$  (heavy) and mass  $m$  (light); atomic distance  $a/2$ , periodicity interval  $a$ .  $\mu$  is the reduced mass defined by  $\mu^{-1} = M^{-1} + m^{-1}$ . In the optical branch the two atoms with masses  $M$  and  $m$  vibrate against each other for small wave numbers  $\kappa$ , while in the acoustic branch the atoms move in the same direction. At the Brillouin zone boundary  $\pi/a$  only the light atom ( $m$ ) vibrates in the optical branch and only the heavy one ( $M$ ) in the acoustic branch. In the reciprocal space  $\kappa$  the curves  $\omega(\kappa)$  are extended periodically

behave similarly as those of the monatomic chain (Fig. 8.11). They have vanishing frequencies for the wave number limit  $\kappa \rightarrow 0$ . In this low wave number range, where  $\omega$  is proportional to  $\kappa$  (large wavelength  $\lambda \gg a$ ), neighbouring atoms move with similar displacements. In this regime, a description in terms of a continuum theory is adequate, in which moving volume elements  $\Delta V$  with many atoms are considered [17]. The corresponding lattice waves are common sound waves. Consequently, this lower phonon branch in Fig. 8.12 is called *acoustic branch*.

In contrast, neighbouring atoms move in opposite direction for lattice waves of the energetically higher dispersion branch in Fig. 8.12. Correspondingly, the phonon frequency at  $\kappa = 0$  is obtained as  $\sqrt{2f/\mu}$  with  $\mu$  as the reduced mass of  $M$  and  $m$ . Interestingly enough, for both phonon dispersion branches only one type of atoms is in motion for waves at the Brillouin zone boundary at  $\kappa = \pi$  (Fig. 8.12). In the acoustic branch only the heavy atoms oscillate, that is, only the heavy atom mass  $M$  determines the phonon frequency  $\sqrt{2f/M}$ . In the upper dispersion branch in Fig. 8.12, only the light atoms ( $m$ ) oscillate for  $\kappa = \pi$ .

In an ionic crystal such as NaCl or KJ, the two types of atoms are positively and negatively charged ions. In that case, the lattice waves of the upper dispersion branch are related to oscillating electric dipole moments, which couple to electromagnetic waves of the infrared spectral range (for  $\kappa \approx 0$ ). In a generalized way, the upper phonon dispersion branches are called *optical branches* also for larger wave numbers and for crystals which are not infrared active.

Real crystals are three-dimensional and periodic with lattice vectors  $\mathbf{r}_n$ . Correspondingly, the reciprocal space is three-dimensional with wave vectors  $\kappa$  and 3D Brillouin zones at distances  $\mathbf{G}_{hkl}$  (Sect. 8.2.2). Unlike curves as in Fig. 8.12, the phonon dispersion branches are then acoustic and optical dispersion surfaces  $\omega(\kappa)$  in reciprocal  $\kappa$  space. In 3D real space, the crystal atoms can oscillate parallel or normal to the propagation direction (wave vector) of the lattice wave. Correspondingly, there are so-called longitudinal ( $\mathbf{u}_\kappa \parallel \kappa$ ) and transverse ( $\mathbf{u}_\kappa \perp \kappa$ ) oscillation modes, respectively dispersion branches [17]. In general wave propagation directions longitudinal and transverse lattice waves might be mixed. In any case, phonon modes in a 3D crystal must be characterized by their frequency  $\omega_{\kappa,\lambda}$  with  $\kappa$  the 3D wave vector of the lattice wave and  $\lambda$  as a symbol for the polarization (normal or parallel to  $\kappa$ ). As an example,  $\omega_{\kappa,TO}$  denotes a phonon with wave vector  $\kappa$  of the transverse optical (TO) branch.

Phonons are excitations of the lattice wave field. Waves with sharp, tightly distributed wave vectors are delocalized over the whole crystal. Similarly, as for electron waves (Sect. 3.2), there are also locally restricted oscillatory excitations of the lattice. They must be described in terms of wave packets with a limited spectrum of phonon wave vectors  $\kappa$ . In this description, the particle character of phonons is obvious. We need this particle picture for the description of thermal quantities as for example, the specific heat or the thermal conductivity. The adequate picture, there, is based on a gas of phonons, that is, quasi-particles.

### 8.4.1 Phonon–Phonon Interaction

The assumption of a parabolic potential for the atom dynamics in a crystal is only a first approximation for the chemical bonding forces. In better approximation, we must allow for exponents higher than two in the atomic displacements  $u_v$  from the equilibrium position. The next higher approximation beyond the parabolic one of Sect. 8.4 takes into account anharmonic potential terms up to cubic order in the atomic displacements. The harmonic Hamilton function (8.202) is, then, replaced by

$$H = \sum_v \frac{p_v^2}{2M} + \frac{1}{2} f \sum_{v=0}^N (u_v - u_{v+1})^2 + \frac{1}{3!} g \sum_{v=0}^N (u_v - u_{v+1})^3. \quad (8.215)$$

The constant  $g$  describes the deviation from the harmonic potential. This anharmonic  $g$  term (in total denoted as  $h$  term) is a small perturbation of the harmonic Hamiltonian (8.202), which can be treated within the frame of perturbation theory (Sects. 6.3 and 6.4). At first, we must transform the classical variables into operators of the quantum field formalism [18]. In particular, the variables  $p_\pi$ ,  $u_v$ ,  $H$  and  $h$  are replaced by the operators  $\hat{p}_v$ ,  $\hat{u}_v$ ,  $\hat{H}$  and  $\hat{h}$ . According to perturbation theory matrix elements of the type  $\langle \phi' | \hat{h} | \phi \rangle$  with  $\hat{h}$  as the anharmonic perturbation operator and  $|\phi\rangle$  as a quantum field state (8.214a), (8.214b) must be calculated.

According to (8.215), the perturbation operator is

$$\hat{h} = \frac{1}{3!} g \sum_{v=0}^N (\hat{u}_v - \hat{u}_{v+1})^3. \quad (8.216)$$

The displacement operators  $\hat{u}_v$  are defined by (8.210a), that is,

$$\hat{u}_{v+1} = \sum_\kappa \sqrt{\frac{\hbar}{2MN\omega_\kappa}} e^{i\kappa va} (\hat{c}_\kappa + \hat{c}_{-\kappa}^+) e^{i\kappa a}. \quad (8.217)$$

This representation yields:

$$\begin{aligned} \hat{u}_v - \hat{u}_{v+1} &= \sum_\kappa \sqrt{\frac{\hbar}{2MN\omega_\kappa}} e^{i\kappa va} (1 - e^{i\kappa a}) (\hat{c}_\kappa + \hat{c}_{-\kappa}^+) \\ &= \sum_\kappa K_\kappa e^{i\kappa va} (\hat{c}_\kappa + \hat{c}_{-\kappa}^+), \end{aligned} \quad (8.218)$$

$K_\kappa$  is an abbreviation for the  $v$ -independent prefactors. From (8.216) and (8.218), the perturbation operator is obtained as

$$\begin{aligned}\hat{h} &= \frac{1}{3!} g \sum_{\nu=0}^N (\hat{u}_\nu - \hat{u}_{\nu+1})^3 \\ &= \frac{1}{3!} g \sum_{\nu=0}^N \sum_{\kappa\kappa'\kappa''} K_\kappa K_{\kappa'} K_{\kappa''} e^{i\nu a(\kappa+\kappa'+\kappa'')} \\ &\quad \cdot (\hat{c}_\kappa + \hat{c}_{-\kappa}^+) (\hat{c}_{\kappa'} + \hat{c}_{-\kappa'}^+) (\hat{c}_{\kappa''} + \hat{c}_{-\kappa''}^+),\end{aligned}\quad (8.219a)$$

and after multiplication of the bracket terms (operator order must be respected):

$$\begin{aligned}\hat{h} &= \frac{1}{3!} g \sum_{\nu=0}^N \sum_{\kappa\kappa'\kappa''} K_\kappa K_{\kappa'} K_{\kappa''} e^{i\nu a(\kappa+\kappa'+\kappa'')} \\ &\quad \cdot (\hat{c}_\kappa \hat{c}_{\kappa'} \hat{c}_{\kappa''} + \hat{c}_{-\kappa}^+ \hat{c}_{-\kappa'}^+ \hat{c}_{-\kappa''}^+ + \hat{c}_\kappa \hat{c}_{-\kappa'}^+ \hat{c}_{\kappa''} + \hat{c}_{-\kappa}^+ \hat{c}_{\kappa'} \hat{c}_{\kappa''} \\ &\quad + \hat{c}_\kappa \hat{c}_{\kappa'} \hat{c}_{-\kappa''}^+ + \hat{c}_{-\kappa}^+ \hat{c}_{-\kappa'}^+ \hat{c}_{\kappa''} + \hat{c}_\kappa \hat{c}_{-\kappa'}^+ \hat{c}_{-\kappa''}^+ + \hat{c}_{-\kappa}^+ \hat{c}_{\kappa'} \hat{c}_{-\kappa''}^+).\end{aligned}\quad (8.219b)$$

The sum of triple products of creation and destruction operators in (8.219b) creates and destructs single phonons in a phonon field state (8.214b) such that the phonon occupation numbers  $n_\kappa$  are changed. For the calculation of the perturbation matrix elements  $\langle \phi' | \hat{h} | \phi \rangle$ , we must, therefore, use the most general field state  $|\phi\rangle$  which is a superposition of many-particle states of the type (8.214b):

$$|\phi\rangle = \sum_{n_1, \dots, n_\kappa, \dots, n_\infty} \alpha(n_1, \dots, n_\kappa, \dots) | \dots, n_1, \dots, n_\kappa, \dots \rangle. \quad (8.220)$$

$\alpha(n_1, n_2, \dots, n_\kappa, \dots)$  are phase factors or amplitudes which indicate the contribution of a particular many-particle state  $| \dots, n_1, \dots, n_\kappa, \dots \rangle$  to the most general field state. Thus, the single many-particle states  $| \dots, n_1, \dots, n_\kappa, \dots \rangle$  span an infinitely dimensional Hilbert space which is called *Fock space* [see Sect. 8.3, (8.124)]. Field states of the type (8.220) can be represented within this space.

For the evaluation of the perturbation matrix elements we consider, as an example, the action of the operator sequence  $\hat{c}_\kappa \hat{c}_{-\kappa'}^+ \hat{c}_{\kappa''}$  in (8.219b) on a many-particle state  $| \dots, n_\kappa, \dots \rangle$ . Hereby, we take into account the relation  $n_\kappa = n_{-\kappa}$  which is easily proven (Problem 8.6). Using (8.33a), (8.33b), we obtain:

$$\begin{aligned}&\hat{c}_\kappa \hat{c}_{-\kappa'}^+ \hat{c}_{\kappa''} | \dots, n_\kappa, n_{\kappa'}, n_{\kappa''}, \dots \rangle \\ &= \hat{c}_\kappa \hat{c}_{-\kappa'}^+ \sqrt{n_{\kappa''}} | \dots, n_\kappa, n_{\kappa'}, (n_{\kappa''} - 1), \dots \rangle \\ &= \sqrt{n_\kappa (n_{\kappa'} + 1) n_{\kappa''}} | \dots, (n_\kappa - 1), (n_{\kappa'} + 1), (n_{\kappa''} - 1), \dots \rangle.\end{aligned}\quad (8.221)$$

Because of the orthogonality of the many-particle states the matrix element  $\langle \phi | \hat{c}_\kappa \hat{c}_{-\kappa'}^+ \hat{c}_{\kappa''} | \dots, n_\kappa, n_{\kappa'}, n_{\kappa''}, \dots \rangle$  vanishes unless the  $\langle \phi |$  bra has the following phonon occupations:

$$\langle \phi | = \langle \dots, (n_{\kappa''} - 1), (n_{\kappa'} + 1), (n_\kappa - 1), \dots |. \quad (8.222)$$

The term (8.221) of the perturbation operator  $\hat{h}$  (8.219b) indeed destroys two phonons with wave vectors (numbers)  $\kappa$  and  $\kappa''$  and creates a phonon with  $\kappa'$ . Within the frame of the time-dependent perturbation method (Sect. 6.4) or in scattering theory within the frame of the Born approximation (Sect. 6.6.2), namely, the perturbation operator (8.219b), that is, also the operator sequence  $\hat{c}_\kappa \hat{c}_{-\kappa'}^\dagger \hat{c}_{\kappa''}$ , mediate between an initial ket  $|i\rangle$  and a final bra  $\langle f|$ . In the present case,  $|i\rangle$  is represented by the general many-particle state (8.220) and  $\langle f|$  must be equal to (8.222) for non-vanishing transition amplitudes.

The physical interpretation of the described formalism is straight-forward: A scattering process with participation of three phonons is described. Two phonons  $(\kappa, \kappa'')$ , maybe assumed as wave packets centered around  $\kappa$  and  $\kappa''$ , hit each other and are scattered into a new phonon with wave vector  $\kappa'$ . Without any further comment, we have transferred the 1D wave numbers  $\kappa$  into 3D wave vectors  $\kappa$ , because the formalism described so far can be rewritten without any essential changes into the three dimensions of a real crystal.

Scattering processes between real particles, classical ones and quantum mechanical elementary particles, generally obey strict conservation rules. Is this also true for quasi-particles as phonons? The simple answer is given by the exponential term of the perturbation operator (8.219a), (8.219b). In analogy to (6.230)–(6.233), this term can be written:

$$\sum_{v=0}^N e^{iv a(\kappa + \kappa' + \kappa'')} = \sum_{v=0}^N [e^{ia(\kappa + \kappa' + \kappa'')}]^v. \quad (8.223)$$

The further evaluation is performed by means of the sum rule for the geometrical series as in Sect. 6.6.4. In analogy to (6.232) and Fig. 6.21 the term (8.223) yields sharp high intensity bands on a negligible background only for

$$(\kappa + \kappa' + \kappa'')a = k2\pi, \quad \text{for integer } k. \quad (8.224a)$$

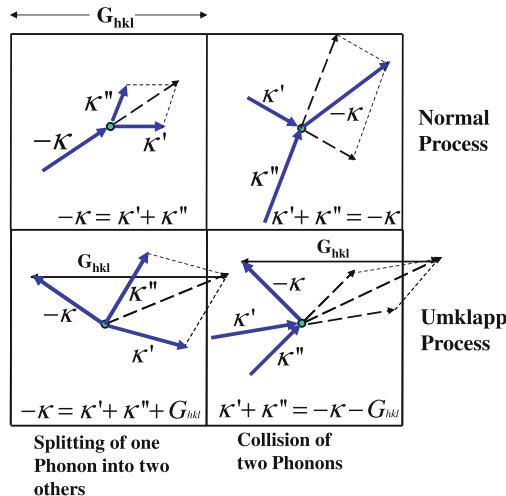
Note that  $2\pi/a$  is the 1D reciprocal lattice vector.

The generalization of (8.224a) to a 3D crystal with 3D reciprocal lattice vectors  $\mathbf{G}_{hkl}$  (8.178) yields the condition

$$\kappa + \kappa' + \kappa'' = \mathbf{G}_{hkl}. \quad (8.224b)$$

For the special case  $\mathbf{G}_{hkl} = \mathbf{0}$ , the familiar momentum ( $\hbar\mathbf{k}$ ) conservation for scattering of real elementary particles is obtained after multiplication of (8.224b) with  $\hbar$ . In this picture, two particles merge in a scattering process to form a third new particle. Then the momenta of the three particles must add up to zero. The new particle takes over the momenta of the two initial particles.

In contrast to this classical momentum conservation (8.224b) describes a more general conservation rule. All (quasi-)momenta  $\hbar\kappa$  must sum up to a reciprocal lattice vector  $\mathbf{G}_{hkl}$  rather than to zero. This is a consequence of the translational symmetry of



**Fig. 8.13** Graph representation of important phonon scattering processes by means of phonon wave vectors (blue arrows) in the 1st Brillouin zone;  $\mathbf{G}_{hkl}$  reciprocal lattice vector. In the decay process (left) the phonon  $\kappa$  is split into two new phonons  $\kappa'$  and  $\kappa''$ . In the phonon collision process (right) two colliding phonons create one new phonon. In normal processes (above) all phonon wave vectors fit in the 1st Brillouin zone and the direction of phonon related transport (energy, momentum, etc.) is not changed by the scattering process. In so-called umklapp processes (below) the wave vector resulting from the scattering process reaches beyond the 1st Brillouin zone. Addition of a reciprocal lattice vector  $\mathbf{G}_{hkl}$  inverts the initial phonon transport direction

the crystal lattice to which the quasi-particles phonons are linked. Therefore, (8.224b) is called wave vector conservation or multiplied by  $\hbar$  *quasi-momentum conservation*.

Analogous discussions can be made for the other triple products of creation and destruction operators in the perturbation operator  $\hat{h}$  (8.219b). The evaluation of the perturbation matrix elements yields a variety of different phonon scattering processes, e.g. the destruction of three phonons with simultaneous creation of three new phonons, collision of two phonons with creation of a new phonon (discussed above) or the splitting of one phonon into two new phonons (Fig. 8.13).

For all these processes, the  $\kappa$  conservation law (8.224b) must hold. Phonon scattering processes split into two categories. For phonons with relatively small wave vectors  $|\kappa|$ , which fit into the Brillouin zone, the wave vector conservation rule (8.224b) can be fulfilled with  $\mathbf{G}_{hkl} = \mathbf{0}$  (Fig. 8.13, upper part). A phonon, which splits into two new ones, transfers its forward quasi-momentum (wave vector) to the two new excitations or two colliding phonons transfer their quasi-momenta to the new created phonon. Since phonons carry a quasi-particle energy  $\hbar\omega_\kappa$  (8.213), the total energy is transported in the same direction as initially before the scattering process. This type of process is called a *normal process*.

A different type of scattering occurs for short wavelength phonons with large wave vectors which reach at or extend over the Brillouin zone boundary (Fig. 8.13, lower part). For fulfilling the wave vector conservation rule (8.224b),

a reciprocal wave vector  $\mathbf{G}_{hkl}$  might be involved. In this case, the wave vectors of the new phonons resulting from the scattering process are directed in opposite direction of the initial ones. The direction of the energy transport is reversed by this type of scattering (Fig. 8.13, lower part). These processes are, therefore called *umklapp processes* (umklapp means reversal in German).

Some physical implications of the phonon–phonon interaction might be considered a little bit more in detail. In a real crystal, the interatomic bonding potentials are harmonic only in first approximation. The anharmonic potential contributions always give rise to phonon scattering, i.e. phonons never exist for an arbitrary time period. Phonons usually assumed as wave packets sharply centered around a mean wave vector  $\kappa$  can travel only along a limited mean free path  $l_\kappa = v_\kappa \tau$  between two scattering processes. They have a finite mean life time  $\tau$  during which they travel with their group velocity  $\mathbf{v}_\kappa = \nabla_\kappa \omega(\kappa)$ . Similarly as for a classical Boltzmann gas or a gas of photons in the electromagnetic field (Sects. 5.6.3 and 8.2.1), we use the particle picture to describe macroscopic thermodynamical properties of matter based on phonons. Phonons carry a particle energy  $\hbar\omega_\kappa$  and, thus, contribute for example, to the specific heat, the heat conductivity or the thermoelectric power of matter.

Of particular interest is the heat conductivity due to phonons, since umklapp processes play a major role. Heat conduction occurs between two spatially separated regions being on different temperature. Because of the finite mean free path, we can define a local phonon temperature for spatial ranges, the volume of which  $\Delta V$  is large as compared with the mean free path ( $\Delta V \gg l_\kappa^3$ ). Because of the phonon collision processes local thermal equilibrium can be established within these regions. Within such a volume  $\Delta V$ , the phonon states are occupied according to the (equilibrium) Bose statistics (Sect. 5.6.3) which defines the local temperature  $T$  within this volume. Consequently the phonon states in neighbouring volumes with differing temperatures  $T_1 \neq T_2$  are occupied differently. Higher temperature means occupation of energetically higher phonon states. This occupation difference causes a phonon current from regions of higher temperature into those of lower temperature. Since phonons carry a particle energy, the phonon current implicates a heat current, too. The heat current, of course, undergoes scattering processes due to phonon–phonon interaction. It is immediately evident that normal phonon scattering processes of phonons with small wave vectors and  $\mathbf{G}_{hkl} = \mathbf{0}$  do not change the direction of heat transport (Fig. 8.13). There is no loss of thermal energy during phonon propagation from the warmer to the cooler region. Normal phonon scattering processes can not explain the finite heat resistance of matter. Only umklapp processes cause a reversal of the direction of phonon propagation and, thus, a finite heat resistance.

As an interesting consequence heat resistance of phonons occurs only at higher temperatures. Only at higher temperatures enough phonon states with sufficiently high energy and sufficiently large wave vectors  $\kappa$  are occupied to induce enough umklapp processes for a finite heat resistance. Only in this case the heat resistance of a solid is preferentially determined by phonon–phonon scattering. At low temperatures, typically below 10 K, only normal phonon–phonon scattering occurs. This would give rise to an infinitely high heat conductivity. At temperatures below about 10 K other scattering processes limit the phonon heat conductivity of a solid. Phonons as

wave-like excitations of the solid are scattered, similarly as light and electron waves, on defects, interfaces and inner surfaces. These scattering processes are responsible for the finite phonon heat resistance at low temperatures.

In very pure and perfect crystals, a somewhat strange phenomenon is observed: At low temperatures heat conductivity depends on size, dimensions and the quality of the surface. In the described case, phonon–phonon and phonon scattering on bulk defects do not anymore limit a heat current. Scattering on inner surfaces becomes the determining factor. This phenomenon is of particular interest in connection with nanoscaled crystalline materials. Systematic studies on such materials are performed with respect to thermoelectric applications. For a deeper treatment of heat conductivity in solids, the reader should refer to text books on solid state physics [17].

### 8.4.2 Electron–Phonon Interaction

Within the frame of the Born Oppenheimer approximation the extremely different masses of electrons and atomic nuclei (cores) allows a separate theoretical treatment of the dynamics of both systems: Electrons move in the rigid periodic lattice of the atomic cores (Sect. 8.3.4) while the atomic cores vibrate in the approximately harmonic potential of electron induced chemical bonds (Sect. 8.4). In a better description, beyond Born Oppenheimer approximation, however, lattice vibrations, that is, phonons disturb the ideal lattice periodicity of a crystal. The electrons do not move anymore in an ideal lattice with translational symmetry. The first approximation for an adequate description is in terms of a small perturbation of the ideal rigid atomic lattice due to lattice waves. Due to the atomic displacements  $\mathbf{u}_v(t)$ , there is a time dependent perturbation  $\delta V(\mathbf{r}, t)$  of the periodic lattice potential  $V_{\text{eff}}(\mathbf{r})$  (8.175), (8.176), which must be inserted into the field Hamilton operator for electrons (8.115). Within the frame of time-dependent scattering theory (Sect. 6.4.1), a field perturbation operator must be derived, which acts on field states of the total field consisting of electronic and phonon states:

$$|\phi\rangle = |\dots, n_j, \dots, n_\kappa, \dots\rangle = |\dots, n_j, \dots\rangle |\dots, n_\kappa, \dots\rangle. \quad (8.225)$$

In this combined field state  $|\phi\rangle$ , the electronic (fermionic) occupation numbers are denoted by  $n_j = 0, 1$  with  $j$  containing the wave vector  $k$  and the spin quantum number  $s$ . The phonon part of the field is characterized by the phonon occupation numbers  $n_\kappa$  with  $\kappa$  containing the 3D phonon wave vector and the polarization (atomic displacement direction relative to wave vector, transverse or longitudinal).

According to Sect. 8.3.2, respectively the relations (8.142a), (8.142b), (8.143a), (8.143b), the field operator  $\hat{H}_{ep}$  describing the perturbation of electrons by phonons is derived from the Schrödinger single-particle energy expectation value due to the potential perturbation  $\delta V(\mathbf{r}, t)$  as

$$\hat{H}_{ep} = \int d^3r \hat{\psi}^+(\mathbf{r}) \delta \hat{V}(\mathbf{r}, t) \hat{\psi}(\mathbf{r}), \quad (8.226)$$

$\hat{\psi}^+$  and  $\hat{\psi}$  are the field operators of the fermionic electron field part (8.128). They are derived from the corresponding electronic wave functions  $\psi^*$ ,  $\psi$ .  $\delta V(\mathbf{r}, t)$  is the perturbation of the rigid periodic lattice potential  $V_{\text{eff}}(\mathbf{r})$  (8.175) due to phonons. Phonons induce wave-like displacements  $\mathbf{u}_v(\mathbf{r}_v, t)$  of the atoms at the lattice sites  $\mathbf{r}_v$ . Corresponding field operators are derived from the phonon creation and destruction operators (8.210a). While the displacements  $\mathbf{u}_v(\mathbf{r}_v, t)$  are originally defined only at lattice sites  $\mathbf{r}_v$ , that is, the equilibrium positions of atoms, the action of this local atom displacement on the electrons extends over the whole space continuously. In the present context, the perturbation potential  $\delta V(\mathbf{r})$  in (8.226) is, therefore, more adequately derived from a continuous displacement field  $\mathbf{u}(\mathbf{r})$  rather than from the discontinuous field  $\mathbf{u}_v(\mathbf{r}_v, t)$ . For small displacements  $\mathbf{u}(\mathbf{r})$ , the perturbation of the periodic electronic potential  $V_{\text{eff}}(\mathbf{r})$  might be approximated by

$$\delta V = V_{\text{eff}}[\mathbf{r} + \mathbf{u}(\mathbf{r})] - V_{\text{eff}} = (\nabla V_{\text{eff}}) \cdot \mathbf{u}(\mathbf{r}). \quad (8.227)$$

The vector

$$\nabla V_{\text{eff}}(\mathbf{r}) = \mathbf{D}, \quad (8.228)$$

describes deformations of the periodic lattice potential due to atomic displacements. It is called *deformation potential*. This quantity can be determined experimentally from changes of the electronic band structure (Sect. 8.3.4) due to mechanical crystal deformations. Theoretical calculations of (8.228) are based on methods of many-body solid state physics [4].

In analogy to (8.210a), the displacement field  $\mathbf{u}(\mathbf{r})$  [ $\hat{=}\mathbf{u}_v(\mathbf{r}_v, t)$ ] is represented in terms of phonon creation and destruction operators as

$$\hat{\mathbf{u}}(\mathbf{r}) = \sum_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \mathbf{n}_{\kappa} e^{i\kappa \cdot \mathbf{r}} (\hat{c}_{\kappa} + \hat{c}_{-\kappa}^+), \quad (8.229)$$

$\mathbf{n}_{\kappa}$  is the unity vector of the atomic vibration direction in case of a phonon excitation ( $\kappa, \omega_{\kappa}$ ). Therefrom the perturbation operator (8.227) is derived as

$$\delta \hat{V} = \sum_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} (\mathbf{D} \cdot \mathbf{n}_{\kappa}) e^{i\kappa \cdot \mathbf{r}} (\hat{c}_{\kappa} + \hat{c}_{-\kappa}^+). \quad (8.230)$$

Like in (8.128) the field operators  $\hat{\psi}^+$  and  $\hat{\psi}$  for electrons are expanded in terms of single-particle wave functions  $\varphi_i(\mathbf{r})$ . A reasonable choice for metal electrons or quasi-free electrons in a semiconductor conduction band with an effective mass  $m^*$  is electronic plane waves with a wave vector  $\mathbf{k}$ . The spin quantum number is neglected

because of spin degeneracy or for simplicity reasons. This yields the following representation of the multi-electron field operator:

$$\hat{\psi}(\mathbf{r}) = \sum_{\mathbf{k}} \hat{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (8.231)$$

We introduce the following abbreviation for the deformation potential

$$\mathbf{D} \cdot \mathbf{n}_{\kappa} = D_{\kappa} \quad (8.232)$$

and obtain by use of (8.230) and (8.231) the following expression for the electron–phonon interaction (perturbation) operator:

$$\begin{aligned} \hat{H}_{ep} &= \int d^3r \sum_{\mathbf{k}\mathbf{k}'} \sum_{\kappa} D_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \hat{a}_{\mathbf{k}'}^+ e^{-i\mathbf{k}'\cdot\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}} (\hat{c}_{\kappa} + \hat{c}_{-\kappa}^+) \hat{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \sum_{\mathbf{k}\mathbf{k}'\kappa} D_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \hat{a}_{\mathbf{k}'}^+ \hat{a}_{\kappa} (\hat{c}_{\kappa} + \hat{c}_{-\kappa}^+) \int d^3r e^{i(\mathbf{k}-\mathbf{k}'+\kappa)\cdot\mathbf{r}}. \end{aligned} \quad (8.233)$$

For a large macroscopic volume, the integral extends into infinity and yields a  $\delta$ -function. Consequently, a finite interaction is only obtained under the condition:

$$\mathbf{k} - \mathbf{k}' + \kappa = 0, \quad \text{that is} \quad (8.234a)$$

$$\mathbf{k}' = \mathbf{k} + \kappa. \quad (8.234b)$$

This simplifies the expression for the interaction operator as follows:

$$\hat{H}_{ep} = \sum_{\mathbf{k}\kappa} D_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \hat{a}_{\mathbf{k}+\kappa}^+ \hat{a}_{\mathbf{k}} (\hat{c}_{\kappa} + \hat{c}_{-\kappa}^+). \quad (8.235)$$

In the sum of the phonon creation operators  $\hat{c}_{-\kappa}^+$ , we reverse the sign from  $\kappa$  to  $-\kappa$  and obtain

$$\hat{H}_{ep} = \sum_{\mathbf{k}\kappa} D_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \hat{a}_{\mathbf{k}+\kappa}^+ \hat{a}_{\mathbf{k}} \hat{c}_{\kappa} + \sum_{\mathbf{k}\kappa} D_{\kappa} \sqrt{\frac{\hbar}{2MN\omega_{\kappa}}} \hat{a}_{\mathbf{k}-\kappa}^+ \hat{a}_{\mathbf{k}} \hat{c}_{\kappa}^+. \quad (8.236)$$

The phonon creation and destruction operators  $\hat{c}_{\kappa}^+$  and  $\hat{c}_{\kappa}$  originate from the vibrational amplitudes  $A_{\kappa}(t)$  and  $A_{\kappa}^*(t)$  (8.196a), (8.196b), respectively, the corresponding operators (8.208). Therefore, they are time-dependent (Heisenberg picture):

$$\hat{c}_\kappa(t) = \hat{c}_\kappa(0)e^{-i\omega_\kappa t}, \quad (8.237a)$$

$$\hat{c}_\kappa^+(t) = \hat{c}_\kappa^+(0)e^{i\omega_\kappa t}. \quad (8.237b)$$

Analogous time-dependencies are also given for photon creation and destruction operators  $\hat{b}_{\mathbf{q}\lambda}^+$ ,  $\hat{b}_{\mathbf{q}\lambda}$  deduced from the light field amplitudes (8.10), (8.16), (8.19) as well as for electron destruction and creation operators:

$$\hat{a}_j(t) = \hat{a}_j(0)e^{-i\varepsilon_j t/\hbar}, \quad (8.238a)$$

$$\hat{a}_j^+(t) = \hat{a}_j^+(0)e^{i\varepsilon_j t/\hbar}. \quad (8.238b)$$

Consequently, the electron–phonon interaction operator (8.236) is also time-dependent:

$$\begin{aligned} \hat{H}_{ep} = \sum_{\mathbf{k}\kappa} D_\kappa \sqrt{\frac{\hbar}{2MN\omega_\kappa}} & [\hat{a}_{\mathbf{k}+\kappa}^+(0)\hat{a}_\mathbf{k}(0)\hat{c}_\kappa(0)e^{i\frac{1}{\hbar}(\varepsilon_{\mathbf{k}+\kappa}-\varepsilon_\mathbf{k}-\hbar\omega_\kappa)t} \\ & + \hat{a}_{\mathbf{k}-\kappa}^+(0)\hat{a}_\mathbf{k}(0)\hat{c}_\kappa^+(0)e^{i\frac{1}{\hbar}(\varepsilon_{\mathbf{k}-\kappa}-\varepsilon_\mathbf{k}+\hbar\omega_\kappa)t}]. \end{aligned} \quad (8.239)$$

The perturbation of the many-body electron–phonon state (8.225) by means of the interaction potential  $\hat{H}_{ep}$  (8.239) is periodic in time.

For the calculation of the transition rates (probabilities) between different electron and phonon states, Fermi’s golden rule (6.111a), (6.111b) can be used (Sect. 6.4.1). Hereby, the transition rate  $R_{fi}$  from an initial state  $|i\rangle$  into the final state  $|f\rangle$  is given by

$$R_{fi} = \frac{2\pi}{\hbar} |\langle f | \hat{H}_{ep} | i \rangle|^2 \delta(E_f^{\text{tot}} - E_i^{\text{tot}}), \quad (8.240)$$

$|i\rangle$  and  $|f\rangle$  are initial and final states of the combined system electrons plus phonons (8.225). The  $\delta$ -function with a finite value (integral) only for  $E_f^{\text{tot}} = E_i^{\text{tot}}$  is obtained for long enough perturbation times [long in comparison to the reciprocal frequency in (8.239)]. In this case, the  $\delta$ -function expresses energy conservation during state transitions within the electron–phonon system. The  $\delta$ -function, of course, yields a finite value in (8.240), (8.239) after integration only in case that the exponential factors in front of the time coordinate  $t$  (8.239) vanish. Therefore, apart from  $\mathbf{k}$ ,  $\kappa$  conservation (8.234a), (8.234b) also the following energy conservation rules are valid for the electron–phonon interaction:

$$\varepsilon_{\mathbf{k}+\kappa} = \varepsilon_\mathbf{k} + \hbar\omega_\kappa, \quad (8.241a)$$

$$\varepsilon_{\mathbf{k}-\kappa} = \varepsilon_\mathbf{k} - \hbar\omega_\kappa. \quad (8.241b)$$

### 8.4.3 Absorption and Emission of Phonons

After having verified  $\mathbf{k}$  vector and energy conservation for electron–phonon interaction, we must analyse the interaction matrix elements  $\langle f | \hat{H}_{ep} | i \rangle$  more in detail to better understand the results of Fermi’s golden rule (8.240) applied to electron–phonon scattering processes. Because of (8.236) and (8.239), there are two different operator combinations which operate on the possible initial electron–phonon states  $|i\rangle = |\dots, n_{\mathbf{k}}, \dots; \dots, n_{\boldsymbol{\kappa}}, \dots\rangle$ :

$$\begin{aligned} & \hat{a}_{\mathbf{k}+\boldsymbol{\kappa}}^+ \hat{a}_{\mathbf{k}} \hat{c}_{\boldsymbol{\kappa}} |\dots, n_{\mathbf{k}}, \dots, n_{\boldsymbol{\kappa}}, \dots\rangle \\ & \propto |\dots, (n_{\mathbf{k}} - 1), \dots, (n_{\mathbf{k}+\boldsymbol{\kappa}} + 1), \dots, (n_{\boldsymbol{\kappa}} - 1), \dots\rangle, \end{aligned} \quad (8.242a)$$

and

$$\begin{aligned} & \hat{a}_{\mathbf{k}-\boldsymbol{\kappa}}^+ \hat{a}_{\mathbf{k}} \hat{c}_{\boldsymbol{\kappa}}^+ |\dots, n_{\mathbf{k}}, \dots, n_{\boldsymbol{\kappa}}, \dots\rangle \\ & \propto |\dots, (n_{\mathbf{k}-\boldsymbol{\kappa}} + 1), \dots, (n_{\mathbf{k}} - 1), \dots, (n_{\boldsymbol{\kappa}} + 1), \dots\rangle. \end{aligned} \quad (8.242b)$$

Since electrons are fermions ( $n_{\mathbf{k}} = 0, 1$ ), non-negligible values are obtained in (8.242a) only for  $n_{\mathbf{k}} = 1$  (occupied state) and  $n_{\mathbf{k}+\boldsymbol{\kappa}}$  in the initial state. Note that the interaction matrix elements in (8.240) are different from zero only for final states being identical with the states resulting from the operations (8.242a), (8.242b).

Therefore, the operator action in (8.242a) is interpreted as the destruction of an electron with wave vector  $\mathbf{k}$  and simultaneous creation of an electron in the state  $(\mathbf{k} + \boldsymbol{\kappa})$ . Simultaneously, a phonon with wave vector  $\boldsymbol{\kappa}$  is destroyed in the phonon field. This process can also be described as scattering of an electron from  $\mathbf{k}$  to  $(\mathbf{k} + \boldsymbol{\kappa})$  on a phonon which is destroyed by transferring its wave vector  $\boldsymbol{\kappa}$  to the electron. This process can only occur, if the single electron state  $\mathbf{k}$  was initially occupied and the resulting final state  $(\mathbf{k} + \boldsymbol{\kappa})$  was empty.

An analogous discussion of (8.242b) yields the following interpretation: An electron in the single electron state  $\mathbf{k}$  (initially occupied) is scattered into the state  $(\mathbf{k} - \boldsymbol{\kappa})$  (initially empty) with simultaneous creation of a phonon with wave vector  $\boldsymbol{\kappa}$ . The electron has a lower energy ( $= \hbar^2 k^2 / 2m$ ) after the scattering process than before. This energy is not lost; it is transferred to the phonon created in the scattering process.

To summarize the conservation rules: In the scattering process described by (8.234a), (8.234b) and (8.241a) an electron changes its wave vector from  $\mathbf{k}$  into  $(\mathbf{k} + \boldsymbol{\kappa})$  by scattering on a phonon with wave vector  $\boldsymbol{\kappa}$ . Hereby, the phonon is destroyed and transfers its wave vector to the electron. Simultaneously the energy  $\hbar\omega_{\boldsymbol{\kappa}}$  of the destroyed phonon is transferred to the electron (8.241a). This process is described as *phonon absorption* by the electron system.

Scattering processes of the type (8.241b), respectively (8.242b) involve a wave vector change of the electron from  $\mathbf{k}$  to  $(\mathbf{k} - \boldsymbol{\kappa})$ . Simultaneously the electron energy is decreased by the energy  $\hbar\omega_{\boldsymbol{\kappa}}$  of a phonon with wave vector  $\boldsymbol{\kappa}$  which is created

in the process. This process is called *phonon emission*. A phonon is emitted by the many-electron system.

We can express the phenomena in somewhat different words: Because of the time-dependent perturbation  $\hat{H}_{ep}$  (8.239), electron–phonon many-particle states (8.225) are no longer eigenstates of the total Hamilton operator of the fields. Transitions between the electron and the phonon system occur. Electrons are scattered by phonons or vice versa phonons by electrons. Hereby, phonons are created (emitted) or destroyed (absorbed). The processes are governed by the following conservation laws for energy and wave vector:

Phonon absorption:

$$\varepsilon_{\mathbf{k}'} = \varepsilon_{\mathbf{k}+\kappa} = \varepsilon_{\mathbf{k}} + \hbar\omega_{\kappa}, \quad (8.243a)$$

$$\mathbf{k}' = \mathbf{k} + \kappa. \quad (8.243b)$$

Phonon emission:

$$\varepsilon_{\mathbf{k}'} = \varepsilon_{\mathbf{k}-\kappa} = \varepsilon_{\mathbf{k}} - \hbar\omega_{\kappa}, \quad (8.244a)$$

$$\mathbf{k}' = \mathbf{k} - \kappa. \quad (8.244b)$$

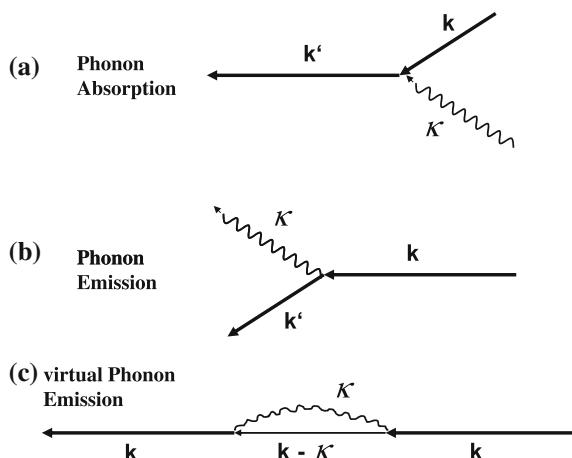
Note that after multiplication with  $\hbar$  (8.243b), (8.244b) might also be interpreted as quasi-momentum conservation rules. Hereby, the electron has a real mechanical momentum  $\hbar\mathbf{k}$ ,  $\hbar\mathbf{k}'$ , while the phonon has a quasi-momentum  $\hbar\kappa$ , which is determined apart from a reciprocal lattice vector  $\mathbf{G}_{hkl}$  (Sect. 8.4.1). As is shown in Problem 8.5, a phonon does not have a real mechanical momentum. All real momenta of the vibrating atoms add up to zero.

The described phonon–electron scattering processes are instructively depicted by means of simple graphs (Fig. 8.14). Arrows indicating direction and total amount by length represent the wave vectors of the involved particles. Electrons are represented in solid line while phonons are plotted as wavy arrows. The time sequence is expressed by the direction of the arrows, commonly propagating from right to left.

Electron–phonon scattering processes are of predominant importance for the understanding of the *electrical resistance* of solids. In an ideally periodic crystal lattice, electronic Bloch waves (8.184a), (8.184b) are stationary solutions to the Schrödinger equation (Sect. 8.3.5). These waves continue to exist and can, therefore, not explain the decay of an electric current, that is, a finite electrical resistance. Electrical resistance must be caused by deviations from the ideal periodicity, be it by local lattice defects or by vibrations of the atoms around their equilibrium positions. The latter deviations from periodicity are connected with phonon excitations. Thus, phonons are perturbations which induce transitions between electronic Bloch states; they are scattering centers in the sense of electron–phonon scattering discussed above. Electrons accelerated by an external electric field carry an electric current (Sect. 8.3.5). Phonon absorption and phonon emission change the direction and the amount of the electron momenta in field direction. This appears as a macroscopic electrical resistance.

The number of excited phonons in a crystal increases with increasing temperature (Bose statistics, Sect. 5.6.3). Therefore, phonon scattering is the major reason for the electrical resistance at temperatures around room temperature and above. At lower temperatures, at around 100 K and below, there is not sufficient phonon excitation and electron scattering on defects prevails. Defect scattering becomes the determining factor for the limitation of electrical current transport. For further details, the reader should refer to textbooks of solid state physics [4, 17].

There is another interesting effect connected with phonon emission and absorption: Electrons propagating through the crystal lattice might emit a phonon and might reabsorb it after a certain time  $\tau$  (Fig. 8.14c). The length of the time interval  $\tau$  is limited by the energy uncertainty  $\varepsilon_k \tau \cong \hbar$ , where  $\varepsilon_k$  is the phonon energy. After this so-called *virtual phonon emission* process the electron continues to propagate with its initial wave vector. A detailed quantum field theoretical treatment of this phenomenon [4] shows that electrons can slightly decrease their energy in the crystal lattice by virtual phonon emission processes. Their effective mass (Sect. 8.3), on the other hand, is somewhat increased. The effect can easily be understood in a qualitative way: Permanent emission and absorption of phonons means shift of the positive atomic cores from their equilibrium positions. Due to the Coulomb interaction with the positive cores, the electron generates a somewhat lower potential during its propagation through the lattice. Its energy is lowered by a tiny amount. But because of the permanent phonon emission and absorption (scattering) processes a tiny resistance against propagation is built up. This increases the effective mass of the electron.



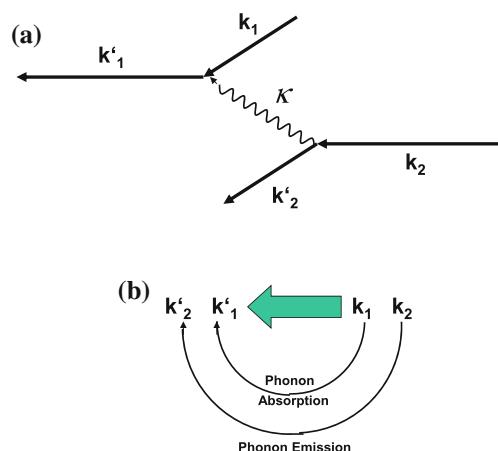
**Fig. 8.14** Graph representation of different electron phonon scattering processes;  $\mathbf{k}$  and  $\mathbf{k}'$  are electron wave vectors;  $\kappa$  is a phonon wave vector (*wiggly line*). **a** In the phonon absorption process an electron is scattered from  $\mathbf{k}$  into  $\mathbf{k}'$ . **b** In the phonon emission process an electron is scattered from  $\mathbf{k}$  into  $\mathbf{k}'$  with simultaneous emission of a phonon  $\kappa$ . **c** In the virtual phonon emission process an electron emits a phonon  $\kappa$  during its propagation through the crystal and reabsorbs it after a time  $\tau$ , which obeys energy uncertainty  $\varepsilon_k \tau \approx \hbar$ . Then the electron continues to propagate with its initial wave vector  $\mathbf{k}$

An electron propagating through the crystal lattice carries with it a cloud of phonons. The electron is dressed by phonons, similarly as an electron in the light field might be dressed by photons (Sect. 8.2.3). The compound particle, electron plus phonon cloud, is called *polaron*. The polaron mass is dependent on the electron velocity, since the heavy atoms can not follow the fast electron movement at higher velocities. Phonon emission and absorption is suppressed and the electron propagates with its bare effective mass  $m^*$ .

#### 8.4.4 Field Quanta Mediate Forces Between Particles

From the discussion of phonon emission and absorption by electrons, we are led to a fundamental phenomenon of quantum field theory, which explains such different phenomena as superconductivity in solids and particle interaction and transformation in elementary particle physics (Sect. 5.6.4): Interactions between particles (field excitations) are mediated by the exchange of other field quanta. Looking at the graphs of phonon absorption and emission (Fig. 8.14) we immediately realize that the phonon (wavy arrow) emitted in the emission process in Fig. 8.14b might shortly afterwards be absorbed by another electron in the process of Fig. 8.14a. Both electrons, the phonon emitting and the phonon absorbing one, change their momentum ( $\mathbf{k}$  vector) and their energy. This combination of phonon emission and absorption by two different electrons is shown in terms of a graph in Fig. 8.15a. A phonon with wave vector  $\kappa$  is emitted by an electron (2) which is scattered from  $\mathbf{k}_2$  to  $\mathbf{k}'_2$ . The same phonon is absorbed by electron (1) which is simultaneously scattered from  $\mathbf{k}_1$  to  $\mathbf{k}'_1$ . One might interpret the total process as scattering of two electrons on each other mediated by exchange of a phonon (Fig. 8.15b).

**Fig. 8.15** **a**, **b** Scattering of two electrons  $\mathbf{k}_1$  and  $\mathbf{k}_2$  on each other. The scattering process is mediated by phonon (wave vector  $\kappa$ ) exchange. **a** graph representation by means of wave vectors of the participating particles. **b** Scheme of the electronic transition from  $\mathbf{k}_1, \mathbf{k}_2$  to  $\mathbf{k}'_1, \mathbf{k}'_2$  by phonon absorption and emission



To substantiate the picture of phonon mediated electron–electron scattering, we consider the energy balance for the two electrons in their initial and final states. From (8.243a), (8.243b), (8.244a) and (8.244b), one obtains:

$$\varepsilon_{\mathbf{k}'_1} + \varepsilon_{\mathbf{k}'_2} = \varepsilon_{\mathbf{k}_1 - \boldsymbol{\kappa}} + \varepsilon_{\mathbf{k}_2 + \boldsymbol{\kappa}} = \varepsilon_{\mathbf{k}_1} - \hbar\omega_{\boldsymbol{\kappa}} + \varepsilon_{\mathbf{k}_2} + \hbar\omega_{\boldsymbol{\kappa}} = \varepsilon_{\mathbf{k}_1} + \varepsilon_{\mathbf{k}_2}. \quad (8.245)$$

In addition to this energy conservation, we derive the following relations for the wave vectors from the graph diagram in Fig. 8.15a:

$$\mathbf{k}_1 = \mathbf{k}'_1 + \boldsymbol{\kappa}, \quad (8.246a)$$

$$\mathbf{k}_2 + \boldsymbol{\kappa} = \mathbf{k}'_2, \quad \text{that is} \quad (8.246b)$$

$$\mathbf{k}'_1 + \mathbf{k}'_2 = \mathbf{k}_1 + \mathbf{k}_2. \quad (8.246c)$$

On the basis of (8.245) and (8.246c) and “hiding” the participation of an exchanged phonon the process plotted in Fig. 8.15 is interpreted as an electron–electron scattering (interaction) process  $\mathbf{k}_1, \mathbf{k}_2 \rightarrow \mathbf{k}'_1, \mathbf{k}'_2$ . As expected for scattering of two particles on each other, energy and momentum are conserved.

The electron–phonon interaction operator (8.239), which contains both phonon emission and absorption, can indeed be transformed into an operator which describes only electron–electron interaction. Phonon specific variables are, then, integrated into an interaction amplitude  $V_{\mathbf{k}', \mathbf{k}, \boldsymbol{\kappa}}$  [4].

We rewrite the wave vectors of the electrons in Fig. 8.15a:  $\mathbf{k}_1, \mathbf{k}_2$  are denoted by  $\mathbf{k}, \mathbf{k}'$ .

Then, the final states of the electrons after the scattering processes are  $\mathbf{k} + \boldsymbol{\kappa}$ ,  $\mathbf{k}' - \boldsymbol{\kappa}$  and because of (8.246a)–(8.246c) the wave vector change due to the interaction between the electrons  $\mathbf{k}$  and  $\mathbf{k}'$  can be expressed as:

$$\mathbf{k} \rightarrow \mathbf{k} + \boldsymbol{\kappa}, \quad \mathbf{k}' \rightarrow \mathbf{k}' - \boldsymbol{\kappa}. \quad (8.247)$$

Electrons with  $\mathbf{k}$  and  $\mathbf{k}'$  are destroyed and electrons with  $\mathbf{k} + \boldsymbol{\kappa}$  and with  $\mathbf{k}' - \boldsymbol{\kappa}$  are created. These processes occur for all possible electron and phonon wave vector combinations. Without any detailed calculation [4] we can, thus, guess the mathematical expression of the interaction operator for electron–electron scattering mediated by phonons:

$$\hat{H}_{ee} = - \sum_{\mathbf{k}\mathbf{k}'\boldsymbol{\kappa}} V_{\mathbf{k}, \mathbf{k}', \boldsymbol{\kappa}} \hat{a}_{\mathbf{k}+\boldsymbol{\kappa}}^+ \hat{a}_{\mathbf{k}} \hat{a}_{\mathbf{k}'-\boldsymbol{\kappa}}^+ \hat{a}_{\mathbf{k}'}. \quad (8.248)$$

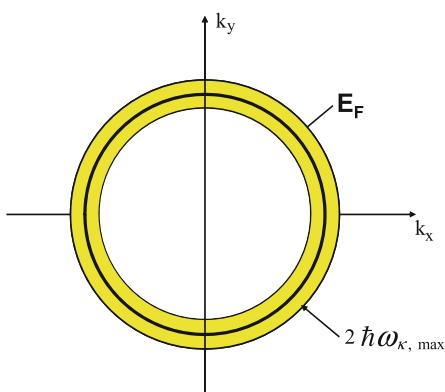
The action of this operator on multi-particle states (8.137a) or on states of the combined electron–phonon system of a crystal (8.225) exactly reflects the process of phonon mediated electron–electron scattering as depicted in Fig. 8.15. Hereby, only electron states are subject to the operator action; the phonon part of the process is globally hidden in the interaction amplitudes (coupling matrix elements)  $V_{\mathbf{k}, \mathbf{k}', \boldsymbol{\kappa}}$ .

A negative sign is attributed to the coupling matrix elements  $V_{\mathbf{k}, \mathbf{k}', \kappa}$  to make the attractive character of the electron–electron interaction obvious. Already in connection with the polaron (Sect. 8.4.3) an electron appeared to be accompanied by a phonon cloud which generates a deeper potential in the neighborhood. In the present case, this potential modification near an electron produces an attractive force on a second electron. In other words, during its propagation an electron slightly deforms the crystal lattice nearby by attracting the neighbouring positive atomic cores (described in terms of a phonon cloud). The resulting enhanced positive charge near the electron attracts a second electron. To illustrate this phenomenon, one frequently uses the ball/mattress paradigm. Two balls rolling on a mattress lower their potential energy by forming a common deepening in the mattress. The balls are attracted into the same hole. The formation of two separate holes would cost more energy.

There are some further interesting conclusions from (8.248). Electron collisions can transfer electrons only from occupied states to empty states. In metals, the Fermi level  $E_F$  (spherical surface for free electrons) separates occupied from empty states (Sect. 8.3.5, Fig. 8.9). While typical Fermi energies  $E_F$  are in the order of 5 eV [17], maximum phonon energies reach values of about 50–100 meV. Therefore, the phonon mediated electron–electron interaction is restricted to a thin spherical shell with a thickness  $2\hbar\omega_{\kappa, \text{max}}$  of about 100 meV around the Fermi sphere in reciprocal space (Fig. 8.16).

At first glance one might wonder, why the weak, phonon mediated electron–electron attraction has some physical effect in comparison to the extremely strong Coulomb repulsion between electrons. The answer becomes clear by considering the different reach of influence of both interactions. Only electrons in states near the Fermi surface  $E_F(\mathbf{k})$  can carry an electric current, since they can be accelerated into empty states. In metals, these electrons have a typical Fermi velocity  $v_F$  of about  $10^8$  cm/s [17]. In contrast, lattice dynamics determined by the heavy atomic cores is much slower. The maximum lattice deformation due to phonon emission and absorption occurs at a distance of typically  $2\pi v_F/\omega_{\kappa, \text{max}}$  (phonon vibration period  $2\pi/\omega_{\kappa, \text{max}} \approx 10^{-13}$  s) behind a propagating electron. At this distance from

**Fig. 8.16** Schematic plot (2D cut in  $k_x$ ,  $k_y$  plane) of the region in reciprocal space (yellow ring), in which electron phonon interaction is possible. Only within a spherical energy shell of thickness  $2\hbar\omega_{\kappa, \text{max}}$  ( $\omega_{\kappa, \text{max}}$  maximum phonon frequency) around the Fermi sphere  $E_F$  electrons can transfer energy to phonons or vice versa



the first electron, there is the phonon induced potential minimum for the second attracted electron. This distance between the two correlated (interacting) electrons is estimated to about 100 nm. This is the typical interaction range of the phonon mediated electron–electron interaction. Because of the high free electron concentration in metals of about  $10^{22} \text{ cm}^{-3}$  the Coulomb repulsion between electrons is screened out over such large distances and is insignificant for electrons being apart by around 100 nm [17].

The phonon mediated electron attraction is indeed observed beside the Coulomb repulsion. It is the underlying physical reason for the interesting phenomenon of *superconductivity*. In many materials, the interaction Hamiltonian (8.248) couples two electrons with opposite spins with each other. These so-called *Cooper pairs* of highly correlated electrons behave as bosons because of the opposite electronic spins. They can propagate through the solid without scattering on defects or phonons, that is, without any electrical resistance [4, 17].

The phonon mediated electron–electron interaction described by (8.248) can be assumed as a paradigm for all kinds of particle interaction, even up to the relativistic range of elementary particle interaction (Sect. 5.6.4). In the relativistic quantum field theories of elementary particles, the interaction between two particles (Fig. 5.17), for example, between the nucleons proton ( $p$ ) and neutron ( $n$ ), is explained by the exchange of bosonic field quanta. Within the framework of quantum field theories, an action at distance between particles, as in the classical gravitation theory of Newton or in classical Coulomb interaction between two charged particles, does not exist. In the latter case, the Coulomb interaction between two electrons is attributed to an exchange of photons in quantum electrodynamics (Sect. 8.2).

Electrons in vacuum behave similarly as electrons in the phonon field of a solid. Similarly as polarons and their phonon cloud (Sect. 8.4.3) electrons in vacuum interact with the quantized electromagnetic field (Sect. 8.2). Even in vacuum the electromagnetic field is present in its vacuum ground state. Similarly, as we have seen already in Sect. 8.2 for atoms in a light field electrons are accompanied by a photon cloud. According to quantum electrodynamics they are dressed by photons. Electrons in vacuum can emit and absorb photons by interaction with the ground state of the electromagnetic field.

Coulomb repulsion between two negatively charged electrons is attributed in this picture to the emission of a photon by one electron and its absorption by the second electron. The force between the two electrons is ascribed to the recoil momentum of the photon  $p = \Delta E/c$ , where the energy  $\Delta E$  can be estimated by the potential  $V(r)$  of the two electronic charges at a distance  $r$  [ $V(r) \approx \Delta E$ ]. According to the time-energy uncertainty (3.22) the photon propagation time between the two electrons is estimated by

$$\Delta t \approx \hbar/\Delta E . \quad (8.249)$$

The photon, thus, travels along the distance

$$r \approx c\Delta t = \hbar c/\Delta E \quad (8.250)$$

between the two electrons. From (8.249) and (8.250) the potential  $V(r)$  between the two electrons is obtained as

$$V(r) \propto \Delta E = \hbar c / r. \quad (8.251)$$

This simplified discussion of photon exchange between two electrons yields the well known  $1/r$  dependence of the electrostatic Coulomb potential between two charges.

Feynman [19] has illustrated particle interaction by exchange of field quanta in a simple way in terms of the chemical bonding of the  $\text{H}_2^+$  molecule (Sect. 6.2.3). In the model description of this simple molecule, the chemical bond between the two protons (H nuclei) is due to the delocalized electron between the nuclei. As is described in Sect. 6.2.3, the delocalized electron might be imagined as tunneling back and forth between the two protons. Hereby, the double tunneling amplitude  $2|t_{LR}|$  (6.58c) represents the energetic difference between the bonding and the antibonding state (energy level) of the  $\text{H}_2^+$  molecule. Thus, the electron exchange between the two protons causes a proton–proton interaction with a total energy lowering. The energy lowering represents the binding energy  $B$  of the molecule; it amounts approximately to  $|t_{LR}|$  (6.58a)–(6.58c). Following now Feynman’s arguments, we assume the back and forth tunneling electron to move with the amplitude of a free particle. In this approximation, its propagation is described by a free spherical wave  $r^{-1}[\exp(ipr/\hbar)]$ . According to Sect. 6.2.3,  $t_{LR}$  is not only a measure for the binding energy  $B$  but it is simultaneously the tunneling probability amplitude of the electron. Therefore,  $|t_{LR}|$  might be approximated by the amplitude of the spherical wave and the binding energy, that is, the attractive potential, of the two protons at a distance  $R$  from each other can approximately be written as

$$B \propto \frac{1}{R} e^{ipR/\hbar}. \quad (8.252)$$

Elementary particles move with relativistic velocities. The momentum  $p$  of the particle mediating the interaction in (8.252) must, therefore, obey the relativistic momentum energy relation (2.2)

$$E^2 = p^2 c^2 + \mu^2 c^4, \quad (8.253)$$

$\mu$  is the mass of the particle mediating the interaction (here the electron being exchanged between the protons). The energy  $E$  is negligibly small in comparison with the rest (mass) energy  $\mu c^2$  of the tunneling particle. By neglecting this energy ( $E \cong 0$ ), the particle momentum  $p$  becomes an imaginary quantity:

$$p \simeq i\mu c. \quad (8.254)$$

By inserting this imaginary momentum into (8.252), we obtain the following expression for the binding energy which arises from the exchange of the  $\mu$  particle:

$$B \propto \frac{1}{R} e^{-(\mu c/\hbar)R}. \quad (8.255)$$

We have started with the electron mediated attraction of the two protons in the  $\text{H}_2^+$  molecule, but have afterwards tacitly generalized the calculation to the general case of two particles which interact with each other by exchange of a third particle with mass  $\mu$ . In this general sense, the interaction potential (8.255) (binding energy) is understood.

The quantity  $\mu c/\hbar$  in the exponent of (8.255) is a decay factor (inverse decay length) which determines the range of influence of the interaction. For  $R \gg (\mu c/\hbar)^{-1}$  the two particles are not subject to an interaction mediated by exchange of the  $\mu$  particle. The bigger the mass  $\mu$  of the mediating particle is, the shorter is the range of influence of the interaction.

The potential  $B(R)$  is called *Yukawa potential* [20] named after its inventor Yukawa, who used (8.255) for the description of the nucleon interaction (proton, neutron) in an atomic nucleus by exchange of  $\pi$  mesons (Sect. 5.6.4). One can easily follow Yukawa's arguments by using Table 5.2 to estimate the mass  $(\mu_\pi c^2)$  of the  $\pi$  meson to be about 0.14 GeV. This mass yields a decay factor  $\mu_\pi c/\hbar$  in (8.255) of about  $10^{13} \text{ cm}^{-1}$ . The inverse value nicely matches the typical diameter of an atomic nucleus. The short range spatial interaction range of the strong forces between nucleons stabilizing the atomic nucleus is indeed found (Sect. 5.6.4).

The theory of quantized fields has created a fundamentally new understanding of the interaction between particles. While in classical physics gravitational forces and Coulomb interaction between charged particles is described by laws being based on action at distance, in quantum field theory particle-particle interaction is ascribed to an exchange of field quanta, a third type of particles. All particles are excitations of field states. Because of the exchange of particles any particle-particle interaction in nature can not propagate faster than light. There is no instantaneous interaction as in classical physics.

Furthermore, one might wonder that phenomena being related to electrons in solids obey the same laws of quantum field theory which underly the behavior of elementary particles up to energies and dimensions of black holes and the big bang. The rules and the underlying logic of quantized fields are obviously very general, at least the formalism which has been invented for their mathematical description (see also Sect. 8.3.3).

## References

1. M. Peskin, D. Schröder, *Introduction to Quantum Field Theory* (Addison Wesley, Reading, 1997)
2. C. Ytzykson, J.B. Zuber, *Quantum Field Theory* (Dover, New York, 2006)
3. N.N. Bogoliubov, D.V. Shirkov, *Introduction to Theory of Quantized Fields* (Interscience, New York, 1959)
4. H. Haken, *Quantenfeldtheorie des Festkörpers* (B.G. Teubner, Stuttgart, 1993)
5. Ch. Kittel, *Quantum Theory of Solids*, 4th edn. (Wiley, New York, 1967)
6. M. Born, W. Heisenberg, P. Jordan, Z. Phys. **35**, 557 (1926)
7. P. Jordan, Z. Phys. **45**, 765 (1927)

8. S. Dürr, T. Nonn, G. Rempe, *Nature* **395**, 33 (1998)
9. C.S. Adams, M. Sigel, J. Mlynek, *Phys. Rep.* **240**, 143 (1994)
10. S. Dürr, G. Rempe, *Adv. At. Mol. Opt. Phys.* **42**, 29 (2000)
11. H.B.G. Casimir, D. Polder, *Phys. Rev.* **73**, 360 (1948)
12. S.K. Lamoreaux, *Phys. Rev. Lett.* **78**, 5 (1997)
13. G. Bressi, G. Carugno, R. Onofrio, G. Ruoso, *Phys. Rev. Lett.* **88**, 041804 (2002)
14. F. Schrödinger, Aspekte zu den spukhaften Quanten-Korrelationen, in *Beobachter und Lebenswelt*, ed. by H. Reinalter (Druck- und Verlagshaus Thaur GmbH, Wien, 1996) p. 69
15. A. Enders, G. Nimtz, *Phys. Rev. E* **48**, 632 (1993) and *Phys. Rev. B* **47**, 9605 (1993)
16. St. French, J. Ladyman, *Remodelling Structural Realism. Quantum Physics and the Metaphysics of Structure*. *Synthese* **136**, 31 (2003)
17. H. Ibach, H. Lüth, *Solid State Physics*, 4th edn. (Springer, Berlin, 2009)
18. Th.A. Bak (ed.), *Aarhus Summer School Lectures 1963, Phonons and Phonon Interactions* (W.A. Benjamin, New York, 1964)
19. R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. III (Addison-Wesley, Reading, 1966)
20. H. Yukawa, *Proc. Phys. Maths. Soc. Jpn.* **17**, 230 (1935)
21. E.T. Jaynes, F.W. Cummings, *Proc. IEEE* **51**, 89 (1963)

# Chapter 9

## Synopsis

Quantum theory, as it appears at the end of this book, is a consistent and elegant theoretical system or concept, which underlies all science branches, from elemental particle physics, condensed matter physics, chemistry, biology, nano-engineering up to astrophysics and cosmology. In its fundamental aspects, the particle-wave duality and the non-deterministic behavior on the atomic and sub-atomic level, quantum theory describes reality in such a satisfactory way, that no experimental results or observations are known so far which are in contradiction to the theory. In this sense, I call it a hyper-theory.

Quantum theory, on the other hand, has many counter-intuitive aspects, which makes it difficult for the beginner to get familiar with. If the theory appears strange, it is mainly because we try to describe it with words of our everyday life. Reality outside of our human brain and our senses is obviously different from what we are able to imagine and recognize with senses and brain developed (within certain limits) during biological evolution.

At this point, we are able to analyze the counter-intuitive aspects of quantum physics somewhat more from a higher level, in a kind of synopsis.

There is the **collapse of the wave function**  $\psi$  which describes the instantaneous change from  $\psi$  into an eigenstate of the measured observable (operator) during the measurement process. Problems in understanding this collapse arise from a misunderstanding of the nature of the wave function  $\psi$ . It is certainly uncomfortable to accept that a measurement changes the real state of a system to such a dramatic extent that all possible measurement results, before and after that measurement would be completely different. But the wave function is not the physical system itself, it is a mathematical representation of knowledge about the system. This knowledge allows probabilistic predictions about the values of observables, which have no reality until they have been determined in the measurement. The measurement takes the information out of the system and brings a particular value of the observable into reality. The collapse of the wave function is not a real physical process, it merely describes how a possible measurement outcome described in probabilistic terms becomes a real

measured value (the same for the quantum state). The adequate formal description of this process is in terms of entanglement between sample and probe Hilbert spaces.

Classical physics describes our world on a macroscopic level without these weird quantum mechanical phenomena as particle-wave duality, superposition, entanglement etc. A large macroscopic object is “here” or “there”. An atom can be in a superposition state, “suspended” between different positions. It can pass a double slit in a superposition state, moving through both openings at the same time. The result is quantum interference which is never observed in the macroscopic world.

Where is the border line between the most fundamental description of nature in terms of quantum physics and the adequate approximation by classical mechanics for macroscopic systems? The exploration of the connection between classical and quantum physics is a major field of present experimental and theoretical research. The phenomenon which tends to blur quantum effects in the macroscopic world, to destroy quantum interference and coherence is **decoherence**. Decoherence causes the difference between atoms and (Schrödinger’s) “cats”. The decoherence process depends on the number and variety of interactions of a quantum system with its environment. The bigger the object the more it is coupled to other degrees of freedom including its environment, the faster it loses quantum coherence by entanglement with internal and environmental degrees of freedom. It, thus, seems that the border line between classical and quantum mechanical behavior is just a matter of technology. Quantum physical properties can be preserved by protecting the system against interactions with the environment. This means for a mesoscopic or macroscopic system, that billions of atoms had to be protected against interactions with all kinds of particles, photons, electrons, other atoms etc. Extreme shielding tools in ultrahigh vacuum would be required, against all kind of radiation (visible, X-rays, high energy particles, ...) and thermal excitation (low temperature). But after all this protection there is one interaction which can scarcely be switched off, the coupling to gravity (gravitational waves). The coupling of mesoscopic masses to gravitational fluctuations probably constitutes the ultimate cause of decoherence. Again, the paramount importance of a quantum-gravitation theory, non-existing so far, is apparent for our understanding of nature.

Another, maybe the most counter-intuitive aspect of quantum physics is **non-locality**. Two microscopic quantum systems interacting with each other generally end up in a non-separable state. The properties of each system can not be described independently of each other due to entanglement. This holds whatever the distance between the components of the entangled state is. Entanglement, thus, naturally leads to non-locality. In the quantum regime, physics at one place cannot be described independently of what happens in another disconnected part of the universe. Meanwhile, this non-locality aspect of quantum physics has been proven experimentally by two-particle coincidence measurements (Bell’s inequality relations).

While in the earlier single and two particle quantum mechanics non-locality is hard to accept by our classical mind, the subsequently developed **quantum field theory** is inherently non-local. A many-particle field is extended over the whole space.

The structure and dynamics of the field is determined by boundary conditions. The quantum field around a double slit is different from that of a single slit. Excitations of the field, electrons as particles, follow the shape of the field and form interference patterns regardless whether they arrive at the detector screen in high density or one after each other in extremely low density without having the chance of mutual interaction. Within the quantum field theoretical view particle wave duality is a consequence of the particle nature as an excitation of the quantum field.

Also the outcome of coincidence measurements regarding the Einstein, Podolsky, Rosen (EPR) paradox and Bell's inequality relation, respectively, can be ascribed to the non-local character of the two-particle field which imposes strong symmetry restrictions on spin and momentum of the two particles being emitted from the common source. No energy or information transfer between the particles is necessary to explain the strong correlations which are observed.

The quantum field picture is obviously closer to reality than single particle quantum mechanics. Therefore, for a more intuitive understanding of non-locality it would be advantageous to start the introduction into quantum physics from a field theoretical view point rather than from one-particle Schrödinger mechanics. But this approach seems to me pedagogically quite difficult. In this book, I chose the standard way from one-particle dynamics to quantum fields.

One last remark concerning present research: Modern experimental techniques, in particular in the fields of quantum optics and nano-physics, on the basis of fabrication procedures developed for integrated semiconductor circuits, enable the realization of fundamental experiments which played an essential role as **thought experiments** in the historical development of quantum theory. Some instructive examples are given:

- (i) Extremely sensitive semiconductor pixel detectors allow the detection of single electrons and photons. Thus, the probability interpretation of the wave function can be demonstrated experimentally by counting electrons on a screen after passage of a double slit.
- (ii) The disappearance of quantum interference in a double slit experiment due to read-out of which-way information is demonstrated experimentally in various ways, both in quantum optics (Sect. 8.2.4) and by nanoelectronic circuits (Sect. 7.2.3).
- (iii) Scanning electron tunneling microscopy (STM), for the first time, shows real pictures of atomic orbitals which were so far only existent as mathematical entities (Sect. 6.2.3).
- (iv) By means of fast and sensitive particle and photon detectors the Einstein–Podolsky–Rosen (EPR) paradoxon could be approached experimentally. The non-locality of quantum physics could be demonstrated along the arguments of Bell's inequality relation.
- (v) The fabrication of semiconductor nanostructures and various experimental measurement techniques (transport, magnetotransport, photon interaction etc.) allowed the preparation of artificial atoms and molecules as well as quantum bits. The quantum mechanical laws governing these structures are proven experimentally.

The recently opened field of quantum information gives great promise for further fundamental research on quantum phenomena and the quantum-classical frontier in physics, which has attracted so much interest since the birth of quantum mechanics. This is far beyond the practical advantages of building a quantum computer, if this will ever be possible.

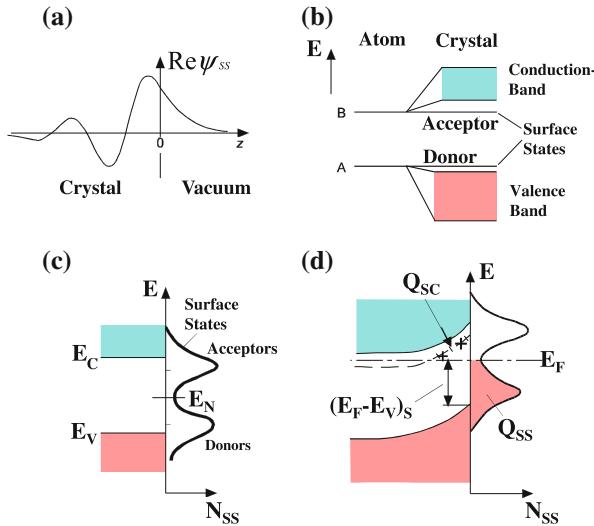
# Appendix A

## Interfaces and Heterostructures

In classical solid state physics, generally condensed matter volumes with more than  $10^{22}$  atoms per  $\text{cm}^3$  are considered. Then, surface effects being related to about  $10^{15}$  atoms per  $\text{cm}^2$  can be neglected. Surface effects must, however, certainly be taken notice of for nanostructures, where on a dimension scale of several 10 nm essential properties, in particular the electronic structure, are determined by surface atoms. Note that a GaAs nano-column with dimensions  $50 \text{ nm} \times 50 \text{ nm} \times 90 \text{ nm}$  has about  $0.3 \times 10^6$  surface atoms in comparison to  $11 \times 10^6$  bulk atoms. A similar situation is given for semiconductor heterostructures, where essential electronic properties are determined by the interfaces between different semiconductor layers with thicknesses in the nanometer range.

We start our discussion with the simplest possible interface, the 2D surface of an extended solid against vacuum. At such an interface, chemical bonds are broken; a surface atom has less bonding partners than a bulk atom. Consequently, the electronic structure near the surface is modified as compared with a region deep in the bulk of the solid. A similar situation is given at the transition between two semiconductors in a heterostructure (Sect. 8.3.4). Surface and interface atoms usually assume equilibrium positions different from those in the bulk. Even the 2D surface periodicity of a crystal surface might differ from that in the bulk. So-called reconstructions with 2D superstructures are observed on surfaces. Such reconstructions are generally described by notations as  $(2 \times 1)$  or  $(7 \times 7)$ . The numbers in brackets indicate the ratios between the lengths of the primitive translational vectors of the 2D superstructure and those of the corresponding bulk elementary cell.

The modified atomic surface structure with different atomic positions, of course, results in new electronic states in comparison with the bulk electronic band structure (Sect. 8.3.4), so-called surface or interface states. Their wave functions are localized near the surface within spatial ranges of nanometers. The wave amplitude of these interface states decays exponentially into the vacuum and/or the bulk of the solid over a length in the nanometer range (Fig. A.1a). The spectral distribution of these single-electron surface states on the energy scale differs from the bulk distribution of states described by the band structure  $E(\mathbf{k})$  (Sect. 8.3.4). This is easily seen for a free surface. Because of less neighbors, the wave functions of surface atoms



**Fig. A.1 a–d** Electronic surface states on a semiconductor surface. **a** Wave function  $\text{Re } \psi_{\text{ss}}$  of a surface state localized at the surface at  $z = 0$  (qualitatively). **b** Splitting of two atomic levels  $A$  and  $B$  into bulk valence and conduction band due to covalent interaction of atomic valence electrons in the crystal. The acceptor- and donor-like surface state levels split off less from the bulk bands because of less interaction with fewer neighbouring atoms than in the bulk. **c** Because of the 2D translational symmetry of the surface electronic surface state levels exhibit dispersion along  $\mathbf{k}_{\parallel}$  and form broader bands in the surface state density  $N_{\text{ss}}$ ;  $E_N$  neutrality level of the surface states. **d** Because of charge neutrality near the surface the Fermi level  $E_F$  cuts the surface state density  $N_{\text{ss}}$  near  $E_N$ . This causes an upwards band bending in the  $n$ -doped semiconductor, i.e. a depletion space charge layer with positively charged donor ion cores

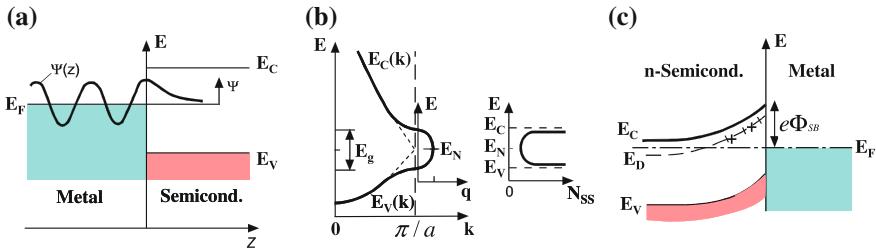
have less overlap to neighbouring wave functions. Therefore, splittings and shifts of atomic levels (in comparison to the free atom) of surface atoms due to interaction with neighbors is less pronounced than for bulk atoms (Sect. 6.2.3). For a surface atom important energy levels as those, which are responsible for the formation of the conduction and the valence band of a semiconductor, are closer in energy to those of the free atom than to the states of the bulk bands (Fig. A.1b).

Each atomic orbital, which is involved in chemical bonding and, thus, gives rise to one of the electronic bands of the band structure, is also the origin of a surface state level  $E_{\text{ss}}$ . Surface states, which are derived from the bulk band structure in this sense, so-called *intrinsic surface states* (in contrast to *extrinsic surface states* being derived from surface defects or adsorbate atoms), have 2D translational symmetry along the surface of an ideal crystal. Therefore, their wave functions  $\psi_{\text{ss}}$  have Bloch wave character with regard to position coordinates  $\mathbf{r}_{\parallel}$  and wave vectors  $\mathbf{k}_{\parallel}$  parallel to the surface (Sect. 8.3.4):

$$\psi_{\text{ss}}(\mathbf{r}_{\parallel}, z) = u_{\mathbf{k}_{\parallel}}(\mathbf{r}_{\parallel}, z) \exp(i\mathbf{k}_{\parallel} \cdot \mathbf{r}_{\parallel}), \quad (\text{A.1})$$

But along a coordinate normal to the surface the amplitude of the surface state wave function (A.1) decays exponentially into the vacuum and into the crystal bulk. Because of the Bloch character parallel to the surface the energy eigenvalues of intrinsic surface states are periodic on the 2D translational surface lattice and the corresponding 2D reciprocal lattice ( $\mathbf{k}_{\parallel}$ ), respectively. In analogy to the 1D and 3D case (Sect. 8.3.4), a 2D electronic band structure  $E(\mathbf{k}_{\parallel})$  with dispersion along the  $\mathbf{k}_{\parallel}$  directions of the 2D surface reciprocal lattice (Fig. 8.7) is generated. As a result, the surface state density  $N_{\text{SS}}$  plotted versus energy  $E$  generally forms broad surface state bands (Fig. A.1c). Surface states are derived from bulk electronic states, in the present case from conduction and valence band states. Consequently, they have different charging character depending on their origin. Surface states derived from the conduction band are usually located close to the lower conduction band edge (Fig. A.1b, c) and, as conduction band states, they are negatively charged if occupied by an electron. Unoccupied they are neutral. In analogy to the nomenclature for bulk doping atoms, these conduction band derived surface states are called *surface acceptor states*. Note, however, that bulk acceptor states energetically lie slightly above the upper valence band edge (in contrast to surface acceptors lying close to conduction band). Surface states being derived from the valence band are called *surface donor states*; their charging character is that of bulk donors. They are neutral if occupied by an electron and positively charged as empty states. Therefore, somewhere midst in the surface state distribution there exists a so-called neutrality level  $E_N$  (Fig. A.1c) which separates donor-like from acceptor-like surface states. If the Fermi level  $E_F$  crosses the neutrality level  $E_N$ , the surfaces as a whole are neutral; they do not carry any charge. If  $E_F$  is located slightly above  $E_N$ , the surface states carry a negative surface charge. For  $E_F < E_N$  the surface states are positively charged.

In case of charged surface states ( $E_F \neq E_N$ ), there must be an opposite electrical charge in the spatial region just below the solid surface in order to guarantee overall neutrality of the surface region. The energetic position of  $E_F$  in the bulk of the semiconductor is determined by the bulk doping. In an  $n$ -doped semiconductor,  $E_F$  is located slightly below the lower conduction band edge (Fig. A.1d). On such an  $n$ -doped semiconductor, therefore, the bulk electronic bands bend upwards to enable an as small as possible charging of the surface states ( $E_F$  as close as possible to  $E_N$ ). A so-called *electron depletion zone* is built up below the surface, in which the free conduction band electrons are depleted and the spatially fixed bulk donor atoms are ionized (positively charged). The amount of band bending is given by the condition, that the negative charge  $Q_{\text{SS}}$  in the surface states is compensated by the positive space charge  $Q_{\text{SC}}$  ( $Q_{\text{SS}} = Q_{\text{SC}}$ ). Consequently, the Fermi level  $E_F$  must be located slightly above the neutrality level  $E_N$ . For common surface state densities  $N_{\text{SS}}$  in the order of  $10^{15} \text{ cm}^{-2}(\text{eV})^{-1}$  the deviation ( $E_N - E_F$ ) is tiny, that is, in the  $10^{-2}$  to  $10^{-3}$  eV range. The Fermi level is essentially fixed (*pinned*) in energy near the neutrality level  $E_N$ . Upon a change of temperature or bulk doping the Fermi level remains essentially fixed on the energy scale with respect to the bulk band edges. The pinning position ( $E_N - E_F$ ) at the surface is characteristic for a particular surface state distribution and, thus, for a particular surface; it is called *surface potential*.



**Fig. A.2** a–c Metal induced interface states (MIGS = Metal Induced Gap States) at the interface of a metal-semiconductor junction. **a** Origin of MIGS: Free electron Bloch states of the metal leak into the semiconductor and in the energy range of the forbidden band they decay exponentially into the semiconductor. The corresponding “Bloch tails” must be represented by a superposition of valence and conduction band states of the semiconductor bulk;  $E_V$  and  $E_C$  valence and conduction band edge, respectively. **b** In the simple model of a monatomic linear (1D) chain described by the real valued band structure  $E_V(k)$  and  $E_C(k)$  of valence and conduction band ( $E_g$  forbidden band) the exponentially decaying interface states (Fig. A.1a) with imaginary wave vector  $\kappa = iq$  fill the energy range of the forbidden band;  $E_N$  neutrality level,  $N_{ss}$  interface state density of MIGS. **c** In a metal-semiconductor junction (Schottky contact) the Fermi level  $E_F$  must cross the density of MIGS near  $E_N$ . This causes the formation of a depletion space charge layer with Schottky barrier  $e\Phi_{SB}$ , in analogy to Fig. A.1d

These concepts developed for the vacuum/solid interface (surface) can easily be transferred to the ideal metal/semiconductor interface (Schottky barrier, Sect. 6.1.1) or hetero-interfaces in semiconductor heterostructures, in which two or several different semiconductor layers as AlAs/GaAs or InGaAs/InP are epitaxially deposited on each other.

In an ideal metal/semiconductor junction, the delocalized Bloch waves of the metal free electron states (potential well with standing wave states occupied up to  $E_F$ ) tunnel into the semiconductor in the energy range of the forbidden band (Sect. 8.3.4), where no semiconductor states exist (Fig. A.2a). The tunneling Bloch waves, therefore decay exponentially into the semiconductor. The corresponding “Bloch tails” must be represented mathematically in terms of Fourier series of valence and conduction band wave functions, since no other semiconductor eigenstates exist in the energetic neighborhood of the forbidden band. Similarly as in the case of surface states, these *metal-induced interface states* (MIGS = metal induced gap states) have partially donor or acceptor charging character depending on the participation of more valence or more conduction band states in their Fourier representation. Like surface states the MIGS are spatially localized directly near the interface. Because of the varying charging character (donor or acceptor like) along the energy scale there exists a charge neutrality level  $E_N$ , which separates more donor like (lower half of forbidden band) from more acceptor like interface (MIGS) states (Fig. A.2b). Because of similar participation of valence and conduction band states in the formation of MIGS the charge neutrality level  $E_N$  is usually located near mid-gap of the semiconductor. For sufficiently high density of MIGS (usually given) the Fermi level  $E_F$  is pinned near  $E_N$ . Both for an *n*-doped and a *p*-doped semiconductor band

bending results. For an  $n$ -doped semiconductor the resulting depletion space charge layer with the Schottky barrier  $e\Phi_{SB}$  is plotted in Fig. A.2c (see also Fig. 6.1). In the ideal case, the height of the Schottky barrier is again characteristic for a particular metal/semiconductor junction.

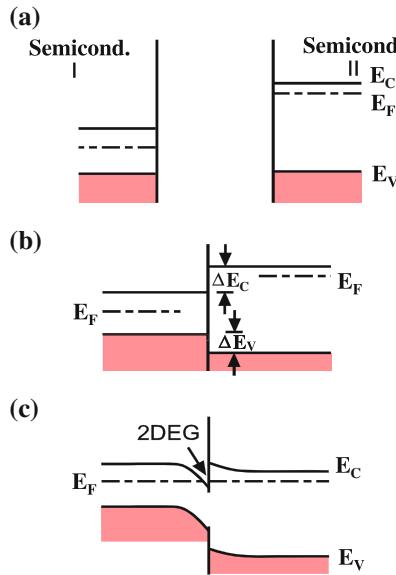
Similar rules as for a vacuum/solid interface (surface) and for a Schottky barrier also govern the relative adjustment of the electronic band schemes of two semiconductors in a heterostructure. If two semiconductors with different band gap have been grown epitaxially on each other, there are two energy ranges  $\Delta E_L$  and  $\Delta E_V$ , called *conduction band* and *valence band offset* (discontinuity), where the continuum of conduction band, respectively valence band states of one semiconductor (e.g., GaAs with smaller gap) touches the forbidden band of the other semiconductor with larger gap (e.g., AlAs). As in the case of the Schottky contact (Fig. A.2a), the Bloch states of the semiconductor with small gap tunnel into the forbidden band of the semiconductor with large gap. This is true for the empty conduction band states as well as for the occupied valence band states. Similar arguments as for the Schottky contact require, that the neutrality levels  $E_N$  of both semiconductors (those of the interface states) must match for an electrically neutral interface region. Slight deviations from the exact  $E_N$  matching result from charge transfers between interface atoms on both sides. The matching of the neutrality levels of the interface states of both semiconductors determines the band offsets  $\Delta E_L$  and  $\Delta E_V$  as characteristic quantities of a particular semiconductor hetero-junction.

In both semiconductors I and II, in Fig. A.3, the energetic position of the Fermi level  $E_F$  within the band gap is determined by the bulk doping. Furthermore, in thermal equilibrium  $E_F$  must have the same value on both sides of the interface. In order to fulfill both conditions, specifically fixed band offsets  $\Delta E_L$  and  $\Delta E_V$  with simultaneously given  $E_F$  positions by doping, band bending in both semiconductors with accompanying space charge zones must occur.

An example, which is particularly important in quantum electronics, is shown in Fig. A.3c: An  $n$ -doped semiconductor with large band gap has been grown on a nearly intrinsic or unintentionally doped semiconductor with smaller gap. Because of the upwards bent conduction band in the semiconductor with large band gap (e.g., AlAs) bulk donors near the junction are ionized and deliver their valence electrons into the triangular potential well on the other side of the junction in the semiconductor with small gap (e.g., GaAs). The triangular potential well results from the downwards band bending in that semiconductor. The conduction electrons are confined in this narrow well with a thickness between 1 and 2 nm; they form a 2D electron gas (2DEG) with free electron movement parallel to the interface. The electronic energy eigenvalues are, therefore, represented as

$$E_j = \varepsilon_j + \frac{\hbar^2 k_{\parallel}^2}{2m_{\parallel}^*}, \quad (\text{A.2})$$

$\varepsilon_j$  are the eigenvalues resulting from the quantization normal to the interface within the triangular quantum well.  $k_{\parallel}$  is a wave vector parallel to the interface (free move-



**Fig. A.3 a–c** Formation of a semiconductor heterojunction with a 2-dimensional electron gas (2DEG) at its interface. **a** Band structures of the two spatially separated semiconductors I (intrinsic or weakly  $n$ -doped) and II (highly  $n$ -doped). **b** The two semiconductors in spatial contact, but not in thermal equilibrium: the two Fermi levels  $E_F$  are different and given by the corresponding bulk doping. This situation is only possible in a gedanken experiment. The band discontinuities (offsets)  $\Delta E_V$  and  $\Delta E_C$  are determined in good approximation by matching the neutrality levels of the interface states in material I and II. **c** Both semiconductors in contact and thermal equilibrium (realistic situation); ideal interface free of interface states in the energy range of the common forbidden band. The band bending in both semiconductors is determined by the bulk doping on both sides and the interface specific band off-sets  $\Delta E_V$  and  $\Delta E_C$ . In the case shown here a 2DEG is formed at the interface in the semiconductor with smaller gap

ment of electrons within 2DEG) and  $m_{||}^*$  the corresponding electronic effective mass. The semiconductor with small band gap, in which the free electrons of the 2DEG move, is not intentionally doped. The dopant defect atoms, which deliver the free electrons in the 2DEG, are in the neighbouring semiconductor spatially separated from the 2DEG conduction electrons. Therefore, defect scattering as an important source of electrical resistance (Sect. 6.6.3) is strongly suppressed, even at high doping concentrations. At low temperatures, the quasi-free electrons of this interface 2DEG reach mobilities  $\mu$ , which exceed that of commonly bulk doped semiconductors by orders of magnitude. In AlGaAs/GaAs heterostructures, for example,  $\mu$  values up to  $10^7 \text{ cm}^2/\text{Vs}$  are reached.

Semiconductor heterostructures with high mobility 2DEGs described here are the most important building blocks for nanostructures and devices to study quantum effects. Furthermore, these heterostructures, in particular in the material systems AlGaAs/GaAs, InGaAs/InP, AlGaN/GaN, are the basis for the fabrication of

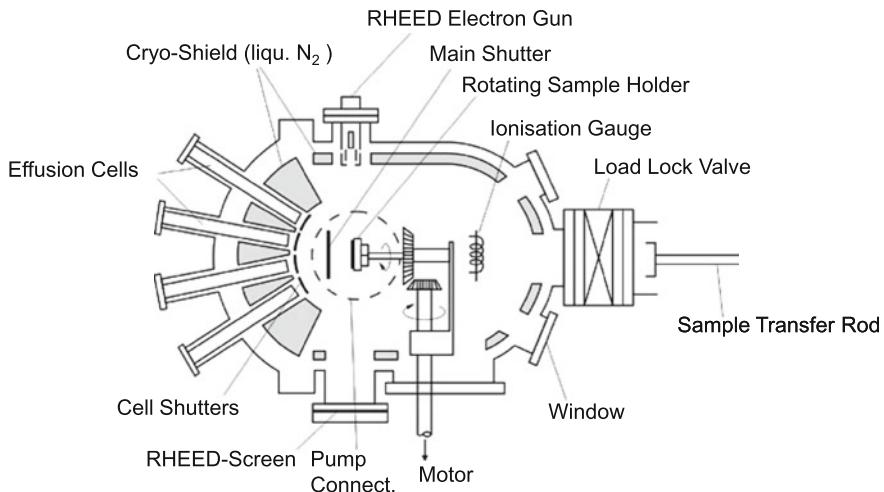
the fastest, low noise field effect transistors (FETs). These heterostructure FETs are sometimes called HEMT (high electron mobility transistor); they have revolutionized the field of modern high-frequency electronics. Therefore, heterostructures with a 2DEG at their interface as in Fig. A.3c are frequently called HEMT structures.

## Appendix B

# Preparation of Semiconductor Nanostructures

The possibility to fabricate nanostructures with typical dimensions between 10 nm and several 100 nm has opened new ways to many experiments regarding the fundamentals of quantum physics and to quantum electronic devices. For the preparation of nanostructures, there are two important basic experimental techniques: Epitaxy allows the crystalline growth (sometimes also deposition of non-crystalline materials) of different materials in layer stacks (heterostructures) with layer thicknesses down to a few nanometers. Lithographic techniques, preferentially with light and/or electron beams allow lateral structuring of epitaxially grown layer stacks down to the nanometer scale. In the present context, we restrict our discussion to semiconductor material systems, because this material class is most important for quantum electronics.

The most important epitaxy methods for semiconductor heterostructures are Molecular Beam Epitaxy (MBE) and Metal Organic Vapor Phase Epitaxy (MOVPE). In MBE source materials, for example, As, Al and Ga for GaAs are evaporated from heated crucibles (effusion cells) and deposited on a substrate such as GaAs or Si. This process runs far from thermal equilibrium in an ultrahigh vacuum (UHV) vessel at background pressures of about  $10^{-8}$  Pa (Fig. B.1). UHV conditions are necessary for obtaining the required high purity standard of semiconducting materials. For maintaining the low pressure UHV conditions during the epitaxy process, the interior wall of the vessel is covered with a cryoshield cooled by liquid nitrogen ( $N_2$ ). The shield has openings for the electrically heated evaporation crucibles and analytical tools for the control of the process. Mechanically driven, computer controlled shutters close and open the effusion cells (crucibles) and switch the molecular beams of source materials on and off according to sophisticated computer programs. The source materials are deposited on a heated substrate, for example, a GaAs wafer, which rotates during the epitaxy for the purpose of homogeneous deposition. For the deposition of III–V, semiconductors typical growth temperatures are between 500 and 700 °C. Elevated substrate temperatures during growth are necessary for a sufficiently high atomic surface mobility which enables the built-in of atoms in the growing surface at the “correct” site. For growth rates of about 1  $\mu\text{m}/\text{h}$ , that is, 0.3 nm/s and shutter switching times below 1 s, atomically sharp transitions between



**Fig. B.1** Scheme of a molecular beam epitaxy (MBE) chamber for the fabrication of III–V semiconductor heterostructures: UHV vessel with cryo-shield and integrated deposition and characterization tools

two different materials deposited on each other can be achieved (Fig. 3.15). According to Fig. B.1, an MBE chamber is commonly equipped with an ionization gauge and an electron gun in combination with an opposite fluorescent screen. Hereby a high energy (1000–2000 eV) electron beam can be reflected under grazing incidence on the growing semiconductor surface and an electron diffraction pattern is observed on the fluorescent screen. This technique, called RHEED (Reflection High Energy Electron Diffraction) is frequently used to control the epitaxial growth process in-situ.

While the MBE method is important preferentially in fundamental research, in industrial applications the MOVPE technique is favored, since it does not require UHV conditions. In MOVPE, explained here for the example of GaAs growth, solid GaAs layers are deposited from gaseous source materials (called precursors) containing Ga and As. As precursors mostly AsH<sub>3</sub> (arsine) and the metal-organic gas trimethyl-gallium [TMG = Ga(CH<sub>3</sub>)<sub>3</sub>] are used. The total reaction, which runs along complex intermediate steps, can be expressed as  $[Ga(CH_3)_3]_{gas} + [AsH_3]_{gas} \rightarrow [GaAs]_{solid} + [3CH_4]_{gas}$ . AsH<sub>3</sub> is directly supplied into the quartz reactor from a gas bottle through a gas flow valve. The metal-organic component TMG is contained in a bottle, in which a certain TMG vapor pressure is established by a surrounding heat bath. Hydrogen as a carrier gas flows through the bottle and transports the TMG to the growth reactor. In the quartz reactor the substrate, a GaAs wafer rotates on a heated susceptor. In contrast to MBE layer growth in MOVPE is a complex interplay of complex processes and chemical reactions: Diffusive transport of the precursors to the heated substrate, decomposition of the source materials in the gas phase and on the surface of the growing layer, incorporation of the Ga and As atoms into the layer

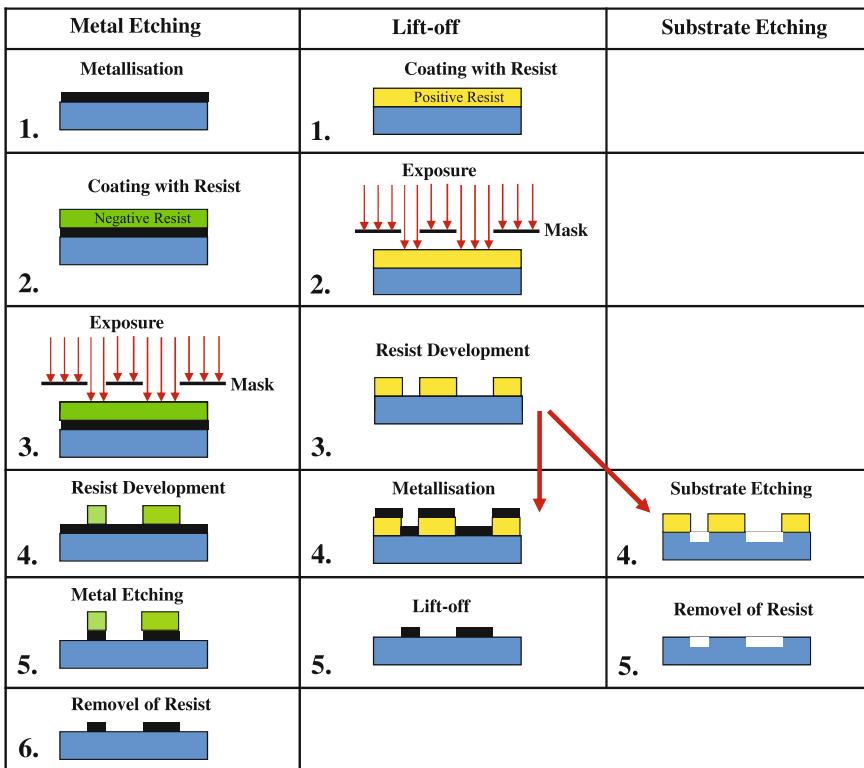
and finally off-diffusion of the gaseous reaction products from the growing layer into the exhaust gas flow. Beside the essential chemical gas phase and surface reactions also flow dynamical issues play an equally important role for the achievement of optimum growth conditions in MOVPE.

In contrast to MBE, the MOVPE technique offers an important advantage for the growth of nanostructures: non-directional growth. There is no preferential deposition direction as in MBE the molecular beam direction. Furthermore, the specific type of the involved chemical surface reactions allows *selective growth* in MOVPE. On a substrate wafer covered with a laterally structured mask, for example, a SiO<sub>2</sub> film with well defined holes, the MOVPE growth parameters (temperature, III–V source ratio, gas flow etc.) can be adjusted such that film growth occurs only in the mask openings on the free wafer surface, while on the less reactive mask surface crystal growth is suppressed. This technique is frequently used for the preparation of nanostructures already during the growth process.

Analogous growth properties like those of GaAs are also given for other III–V compounds as for example, AlGaAs, InP or InGaAs. As precursor materials, PH<sub>3</sub> (phosphine), TMIn, TMAI and similar metal-organic gases are used. The growth parameters are, of course, somewhat different from those of GaAs.

By means of MBE and MOVPE heterostructure systems with layer thicknesses down to below 1 nm and atomically sharp transitions can easily be grown. In order to “machine” 3D nanostructures an additional lateral structuring of the layer systems is necessary. This is achieved by lithography (Greek:  $\lambda\iota\vartheta\sigma\varsigma$  = stone,  $\gamma\rho\alpha\varphi\epsilon\iota\upsilon$  = write). In addition to the nanostructuring in  $z$  direction by epitaxy structuring in  $x$  and  $y$  direction parallel to the layer sequence is done lithographically. Lateral structuring down into the 100 nm range is performed by *optical lithography*, that is, by use of illumination techniques with visible and UV light. Nanostructuring down to the 5–10 nm scale presently requires *electron beam (e-beam) lithography* in electron optical columns. Both lithographic techniques are similar insofar as a photoresist is irradiated either by light or by an electron beam. The irradiation changes the chemical and structural properties of the resist such that its solubility in an organic solvent is modified. In optical lithography the irradiation of the resist is performed through a patterned mask, either by contact illumination or by optical projection. The desired pattern is thus transferred as a whole to the resist in parallel, that is, in one single illumination step. Electron beam lithography, on the other hand, is a serial illumination technique, where a focused electron beam is scanned computer controlled over the resist film. In this way, the beam writes the pattern step by step into the resist. The whole process runs in a high precision electron microscope column, similarly as in a scanning electron microscope. In comparison with optical lithography, electron beam lithography is much more costly and time-consuming. For industrial applications present research aims therefore at replacing electron beam lithography in nanostructure science by parallel illumination techniques with light of extremely short wavelength (EUV = extreme UV).

In order to transfer lateral structures of the irradiated photoresist into semiconductor layer systems or onto a wafer, several etching processes are applied. The most important transfer methods are compiled in Fig. B.2. Two important processes



**Fig. B.2** Schematic representation of important lithographical structuring methods. The illumination steps (exposure) for metal etching (3) and for the lift-off process (2) are performed in parallel in optical lithography by exposure through masks (as shown) or by projection, and alternatively in series in scanning electron beam lithography. A negative photoresist becomes insoluble in illuminated areas, while a positive resist is soluble there in the subsequent development step

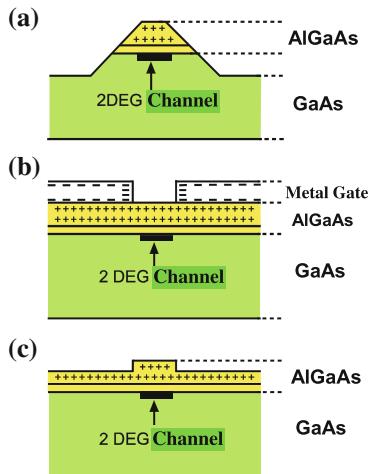
are described, the preparation of lateral metal structures on a semiconductor wafer and direct etching of structures into a wafer. In case of metal structuring on a wafer two techniques are commonly used, *direct metal etching* and the so-called *lift-off process*. In metal etching (left column in Fig. B.2) patterns are etched into the metal film which has been deposited in advance by evaporation or sputtering (step 1). The metal film with a typical thickness of 100 nm is then spin-coated in a centrifuge with an organic photoresist (frequently PMMA = poly-methyl metacrylate) (step 2). In the described process, a so-called negative resist is used, in which the locally illuminated areas—be it by light through a mask or by an electron beam (step 3)—exhibit a decreased solubility in the subsequent development process (step 4). The decreased solubility of the resist in the illuminated areas arises from irradiation induced polymer formation. In the development process (step 4) an organic solvent removes the non-illuminated areas of the resist layer. The remaining resist structures then protect the underlying metal film against metal etching (step 5). Depending on the require-

ments concerning edge sharpness the etching itself is performed by wet chemistry in solution (HCl, H<sub>2</sub>O<sub>2</sub>, etc.) or by ion bombardment in a plasma discharge (O<sub>2</sub>), sometimes in combination with more or less chemically reactive species (HF, HBr). The latter technique called **Reactive Ion Etching (RIE)** is performed in a vacuum chamber (RIE chamber). After the metal etching nanostructures consisting of metal and the resist on top remain on the substrate. In the last final step 6, the resist caps are removed by heating in an oven and the desired metal nanostructures are found on the wafer.

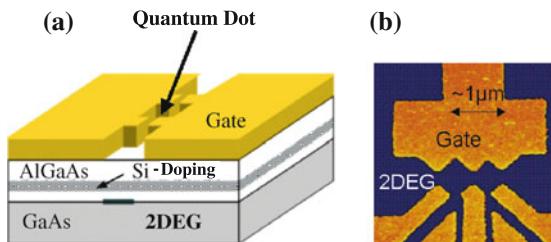
Since III–V semiconductors and metals exhibit similar etching properties, the hitherto described method is not well suited for the generation of metallic nanostructures on III–V substrates. For this purpose, the so-called *lift-off process* is preferred. The III–V wafer surface is covered with a positive photo-resist (step 1). This type of resist is chemically modified by illumination, be it by light or by an electron beam, such that the illuminated areas become more soluble due to breaking of polymer bonds. Subsequently the positive resist is illuminated by light or by an electron beam (step 2). After development, that is, the removal of the resist in the illuminated areas (step 3) the wafer is covered by a structured (with pattern) resist film. The subsequent deposition of metal covers the resist and the open wafer areas equally (step 4). In the lift-off process, the metal covered resist structures are now removed in a solvent (step 5) and only the metal structures remain on the wafer, similarly as in the direct metal etching process.

Positive resist can also be used to transfer patterns directly into the semiconductor wafer. After illumination of the resist (step 2) and removal of the illuminated areas (step 3) only one etching process (step 4), mostly by RIE, is used to generate the required pattern of holes or nanostructures extending from the wafer surface like a mesa (Spanish: table, named after the mesa table mountains in the American desert). In the last 5th step, the resist having remained on the top of the mesas is removed. This structuring method is extremely useful for the preparation of 3-dimensional quantum boxes (dots) based on semiconductor multiple heterostructures. An epitaxially grown layer stack with two energetic barriers for electrons, for example, AlAs embedded in GaAs (Fig. 3.15), creates a 2-dimensional quantum confinement by means of the two energetic barriers (Sect. 3.6.5). Confinement in the third direction, that is, formation of the quantum dot, is achieved by lithographic structuring of nano-scale mesa-like columns (Fig. 5.18a).

The described structuring methods are particularly important with regard to HEMT structures (Appendix A) with an embedded 2DEG at the heterointerface, as for example, between GaAs and AlGaAs. Examples for the preparation of 0D quantum dots or 1D conducting channels in a 2DEG are represented in Fig. B.3. The described substrate etching (Fig. B.2) is used in Fig. B.3a to shape a mesa from an AlGaAs/GaAs HEMT structure, which confines the 2DEG laterally. Within the mesa, the 2DEG does not extend to the very border, because there is a depletion space charge layer (Appendix A) just below the mesa surface. A similar local confinement of the 2DEG can be achieved by a laterally structured metal gate contact deposited on top of the semiconductor heterostructure (Fig. B.3b). Below the metal gate a Schottky contact with depletion zone is formed, which depletes the 2DEG



**Fig. B.3** a–c Different methods to determine tight, laterally restricted areas in a 2-dimensional electron gas (2DEG), in which 2D conductivity is retained: definition of a 0D quantum dot or a 1D conduction channel (wire). **a** Lithographically prepared mesa on an AlGaAs/GaAs heterostructure. **b** Split-gate metal contacts with electron depletion zone of the Schottky contact below. **c** Local thinning of the top AlGaAs layer decreases the donor density and shifts the depletion zone locally into the 2DEG, i.e., forms a local insulating barrier



**Fig. B.4** a, b Quantum dots in a 2DEG of an AlGaAs/GaAs heterostructure realized by split-gate technology [1]. **a** Schematic plot of the preparation of a single quantum dot. The Si dopant of the AlGaAs layer is plotted as a layer doping. **b** Scanning electron tunneling (SET) micrograph of a double quantum dot structure on an AlGaAs/GaAs heterostructure realized by split-gate technology. The two quantum dots are formed in the 2DEG by the action of the metallic saw tooth gate contact and the opposite gate finger contacts on top of the heterostructure

locally from electrons. This structure frequently called split gate arrangement allows the variation of the extension of the conducting 2DEG channel (or quantum dot) by changing the applied gate voltage, that is, the extension of the depletion zone. In Fig. B.4, a quantum dot (schematically) and a double quantum dot (scanning electron micrograph) fabricated by the described technique are represented.

A 2DEG in an AlGaAs/GaAs heterostructure can exist under a surface only at a distance from the surface which extends deeper than the thickness of the depletion

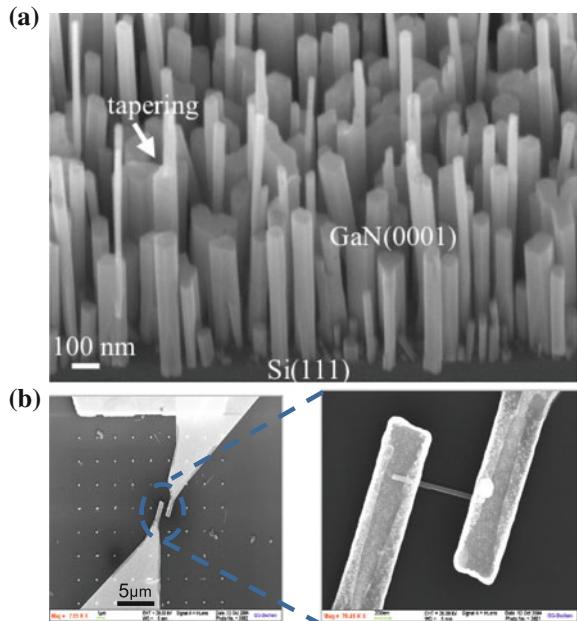
space charge layer (Fig. A.1d). A 2DEG located within reach of the surface depletion zone would be emptied from electrons and the corresponding spatial range would be insulating. This effect is also used to laterally structure a 2DEG. By means of lithographically structured masks and subsequent etching steps, the uppermost AlGaAs layer of the HEMT structure is thinned in those areas, where insulating barriers should confine the conducting 2DEG (Fig. B.3c). Only in sufficiently thick areas of the HEMT structure the 2DEG continues to exist.

Thinning of the uppermost AlGaAs layer of a HEMT structure for generating local insulating barriers can be replaced by other methods which destroy the semiconductor in those areas. One possibility is the implantation of ions in high dose, another one is the oxidation of the semiconductor into some depth to destroy the underlying 2DEG. To perform these techniques on the nanometer scale scanning tunneling microscopes (Fig. 3.10) or scanning force microscopes have been applied recently (Fig. 5.21). By means of a scanning force microscope with metallic tip a water film covering the surface can be anodically oxidized locally near the tip. For this purpose, the tip must be biased as anode versus ground.

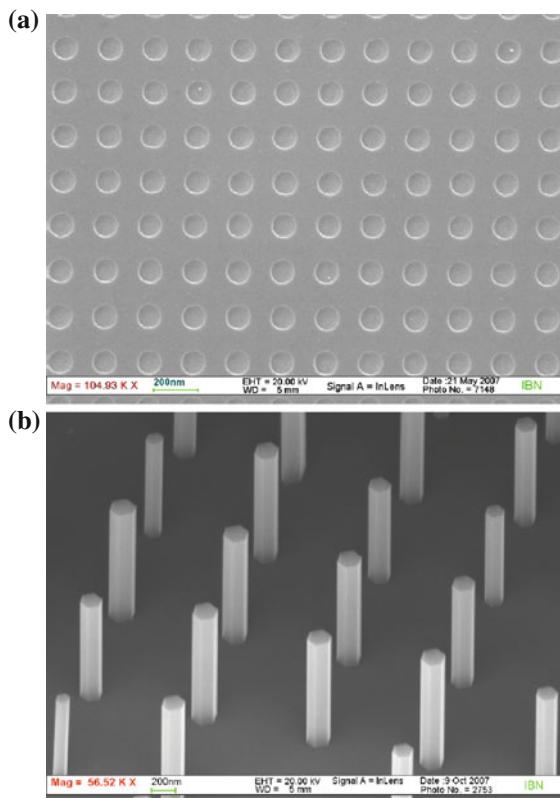
All methods for the preparation of nanostructures discussed so far start with macroscopic and/or solid films and use lithographic techniques to machine the nanoscale structures into the solid (top-down approach). In recent years, so-called bottom-up techniques have gained much interest, in which the self-organization principle of nature, as we know it from crystal growth, is successfully applied. Under adequate physical and chemical conditions, nanostructures grow by self-organization of organic nanoparticles and fiber structures. Particularly in semiconductor physics the self-organized epitaxial growth of semiconductor whiskers (wires, columns) has gained importance for the study of quantum transport and for future device applications. After deposition of metallic nanoparticles with dimensions of 10–100 nm on a semiconductor wafer, these particles act as crystallization nuclei and local catalysts in the subsequent epitaxial growth of whiskers both in MBE and MOVPE. In the MOVPE process, the metallic nuclei locally enhance the catalytic decomposition of the gaseous precursors and induce epitaxial growth under special growth conditions only on the nucleus. This leads to whisker growth; 10–100 nm thick nanocolumns grow with a length up to micrometers. The nanocolumns might be harvested from the growth substrate in an ultrasonic bath and spread on a high resistive substrate, commonly a Si/SiO<sub>2</sub> substrate (for use as a gate). After lithographical preparation of metallic contacts (Fig. B.5b) these samples can be used for the study of quantum transport.

In the case of group III-nitrides (GaN, InN etc.) whisker growth occurs in MBE under nitrogen rich conditions even without the use of metallic nuclei (Fig. B.5) [2]. It is of particular interest that during whisker growth in MBE or MOVPE the source materials can be exchanged to generate vertical heterostructures in the nanocolumns. Also axial overgrowth of columns having been grown in a first step is possible. In this way core/shell nanowires are produced. The whole heterostructure technology of layer growth can be transferred to the growth of semiconductor nanocolumns. This allows the simple fabrication of nano-devices such as resonance tunneling diodes (two barriers within a nanocolumn), nano-HEMTS or nano-Aharonov–Bohm rings.

**Fig. B.5 a, b** Scanning electron micrograph of an arrangement of GaN whiskers grown by molecular beam epitaxy (MBE) under nitrogen rich conditions on a Si (111) substrate (a) and of one single whisker which has been removed from the substrate and is placed on a new Si/SiO<sub>2</sub> host substrate (b). Electron beam lithographically structured markers (*bright spots*) were deposited in advance on the host substrate. The whisker is electrically contacted by metal contacts (*large bright areas*) which are also prepared by electron beam lithography [2]



For eventual applications in quantum electronics, an ordered growth of semiconductor nanocolumns with respect to thickness, length and local arrangement is desirable. This can be achieved by depositing metal dots of homogeneous thickness on well defined sites on the wafer. Well ordered arrangement of nanocolumns can also be achieved by selective growth of columns through mask holes, which have been lithographically prepared on the wafer in advance. As an example, Fig. B.6 shows ordered selective growth of GaAs nanocolumns with a diameter of about 150 nm on GaAs (111) [B.3]. The mask consisting of an inorganic negative resist (HSQ = Hydrogen Silesquioxane) was fabricated by electron beam lithography. The GaAs columns were grown in MOVPE at a growth temperature of 750 °C using AsH<sub>3</sub> and TMG (Trimethyl Gallium) as precursors. Once more it should be emphasized that selective growth is essential as it is possible preferentially in MOVPE, where the non-reactive mask surface does not catalytically dissociate the precursors and, thus, GaAs growth is suppressed. Growth of semiconductor nanocolumns has meanwhile been demonstrated in MBE, MOVPE and other vapor phase deposition techniques for the important semiconductors Si, SiGe, GaAs, InAs, GaN, InN, ZnO and others. Hence, interesting new developments are expected that may lead to a semiconductor-based quantum electronics.



**Fig. B.6** **a, b** Ordered selective growth of GaAs nanocolumns on a masked GaAs (111) substrate (scanning electron micrograph). **a** HSQ (hydrogen silesquioxane) mask prepared on a GaAs (111) wafer by electron beam lithography. **b** GaAs nanocolumns (diameter approx. 150 nm) grown by metal-organic vapor phase epitaxy (MOVPE) by use of an HSQ mask [3]

## References

1. L.P. Kouwenhoven, D.G. Austing, S. Tarucha, Rep. Prog. Phys. **64**, 701 (2001)
2. R. Calarco, M. Marso, R. Meijers, T. Richter, N. Aykanat, T. Thillosen, T. Stoica, H. Lüth, NanoLetters **5**, 981 (2005)
3. V. Klinger, J. Wensorra (Research Centre Jülich), private communication

# Appendix C

## The Reduced Density Matrix

A major motivation for defining a reduced density matrix is the description of a quantum system (1) which is embedded in another system (2), maybe the environment of system (1). This embedment means that the two systems are quantum correlated; they are entangled and can not, in principle, be separated from each other. The total combined system (1) + (2) might be in a pure quantum state with maximum possible knowledge for an outside observer. The standard terminology for this total system is that of a “closed system”.

We now assume, that the observer has only access to the first system (1), he can only perform measurements on system (1) but not on system (2). The question arises, what kind of information can he obtain by measuring an observable  $A$  only in the subsystem (1), while the environment, system (2) being entangled with (1), is either inaccessible, can not be completely measured or is simply of no interest. In this situation subsystem (1) is frequently called an open system.

For the formal description, we ascribe two orthonormal sets of eigenfunctions  $|n\rangle \dots |m\rangle \dots$  and  $|i\rangle \dots |j\rangle \dots$  to the subsystems (1) and (2), respectively. The most general state of the total (closed) system can, then, be written as

$$|\psi\rangle = \sum_{ni} c_{ni} |n\rangle |i\rangle \quad (\text{C.1})$$

Since this state vector can not be separated into a product containing only the set  $|n\rangle$  of the subsystem (1) and the set  $|i\rangle$  of the (environment) subsystem (2), the (open) system (1) is entangled with its environment (2).

According to Sect. 7.3.2, where the density matrix  $\hat{\rho}$  is defined, the expectation value of an observable  $A$  (operator  $\hat{A}$ ) can be expressed as

$$\langle A \rangle = \text{Tr}(\hat{\rho} \hat{A}) \quad (\text{C.2})$$

Now,  $A$  shall be an observable which is measured only in subsystem (1). The corresponding operator  $\hat{A}$  does not operate on states in the (environment) system (2); it does not change them. Correspondingly  $\hat{A}$  can be expressed as

$$\hat{A} = \hat{A}_1 \hat{I}_2 \quad (\text{C.3})$$

where  $\hat{I}_2$  is the unity operator for eigenstates  $|i\rangle, |j\rangle, \dots$  of the environment system (2). With the density matrix of the closed total system

$$\hat{\rho} = |\psi\rangle\langle\psi| = \sum_{ni,mj} c_{ni} c_{mj}^* |n\rangle|i\rangle\langle j|\langle m| \quad (\text{C.4})$$

the operator product in (C.2) follows as

$$\hat{\rho} \hat{A} = \sum_{ni,mj} c_{ni} c_{mj}^* |n\rangle|i\rangle\langle j|\langle m| \hat{A}_1 \hat{I}_2 \quad (\text{C.5})$$

To obtain the expectation value (C.2), we calculate the trace of (C.5) by multiplying from the left (bra) and from the right (ket) with the orthogonal set  $|i'\rangle|n'\rangle$  of the closed total system and by subsequently adding up the diagonal elements of the obtained matrix:

$$\text{Tr}(\hat{\rho} \hat{A}) = \sum_{ni,mj,n'i'} \langle n' | i' | c_{ni} c_{mj}^* | n \rangle | i \rangle \langle j | \langle m | \hat{A}_1 \hat{I}_2 | i' \rangle | n' \rangle. \quad (\text{C.6})$$

Since  $\hat{A}_1$  and  $\hat{I}_2$  operate separately in the two subspaces, we can express the matrix of  $\hat{A}_1 \hat{I}_2$  as  $\langle m | \hat{A}_1 | n' \rangle \langle j | \hat{I}_2 | i' \rangle$  and obtain

$$\begin{aligned} \text{Tr}(\hat{\rho} \hat{A}) &= \sum_{ni,mj,n'i'} \langle n' | i' | c_{ni} c_{mj}^* | n \rangle | i \rangle \langle j | \hat{I}_2 | i' \rangle | m | \hat{A}_1 | n' \rangle \\ &= \sum_{ni,mj,n'i'} \langle n' | c_{ni} c_{mj}^* | n \rangle \langle i' | i \rangle \delta_{ij} \langle m | \hat{A}_1 | n' \rangle \\ &= \sum_{n'} \langle n' | \left( \sum_{nim} c_{ni} c_{mi}^* | n \rangle \langle m | \right) \hat{A}_1 | n' \rangle. \end{aligned} \quad (\text{C.7})$$

We define the reduced density matrix (bracket expression) as

$$\hat{\rho}_{\text{red}} = \sum_{nmi} c_{ni} c_{mi}^* | n \rangle \langle m | \quad (\text{C.8})$$

and obtain the expectation value (C.2) as

$$\text{Tr}(\hat{\rho} \hat{A}) = \text{Tr}_1(\hat{\rho}_{\text{red}} \hat{A}_1). \quad (\text{C.9})$$

The calculation reduces to forming the trace only in the open subsystem (1), that is,  $\text{Tr}_1$ , and using only the part  $\hat{A}_1$  of the operator which operates only in this subsystem (1). Instead of the full density matrix  $\hat{\rho}$ , however, the reduced density matrix (C.8) enters the expectation value (C.9). In this reduced density matrix, one has traced

all degrees of freedom of the environment system (2) which is not affected by the operator  $\hat{A}$  (measurement of observable  $A$ ). This description of the reduced density matrix as  $\text{Tr}_2 \hat{\rho}$  is easily shown by using (C.4) and calculating

$$\begin{aligned}\text{Tr}_2 \hat{\rho} &= \sum_{ni, mj, i'} \langle i' | c_{ni} c_{mj}^* | n \rangle | i \rangle \langle j | \langle m | | i' \rangle = \sum_{ni, mj, i'} c_{ni} c_{mj}^* | n \rangle \langle i' | i \rangle \langle j | i' \rangle \langle m | \\ &= \sum_{nmi} c_{ni} c_{mi}^* | n \rangle \langle m | = \hat{\rho}_{\text{red}}.\end{aligned}\quad (\text{C.10})$$

This is the representation of the reduced density matrix as it is used in (7.66a)–(7.66c). As was shown in Sect. 7.4, the exclusion of the environment (subsystem 2) from the consideration, that is, measurement of an observable  $A_1$  only in subsystem (1), transfers the open system (1), which is entangled with its environment, into a mixed quantum state.

# Problems

## Problems to Chap. 3

- 3.1 A particle with mass  $m$  is confined to a potential well (box) with infinitely high potential walls. Calculate the uncertainty for the position and momentum coordinates and demonstrate the validity of Heisenberg's uncertainty principle.
- 3.2 The wave function  $\psi(x)$  carries the average momentum  $\langle p \rangle$ . Prove that the wave function  $\psi(x) \exp(ip_0x/\hbar)$  has the average momentum  $\langle p \rangle + p_0$ .
- 3.3 An electron is locked in a 1-dimensional quantum well of width  $L$  and with infinitely high potential walls.
- Calculate the wave functions  $\psi_1(x, t)$  and  $\psi_2(x, t)$  of the ground state and the first excited state, respectively, as well as the corresponding energies.
  - Calculate the probability to find an electron in the superposition state of  $\psi_1$  and  $\psi_2$  as well as the average position coordinate  $\langle x \rangle$  in this state
- 3.4 Prove that the wave function

$$\psi(x, t) = A e^{i(kx - \omega t)} + B e^{i(-kx - \omega t)}$$

describes a particle flow being proportional to  $(A^2 - B^2)$ .

- 3.5 A wave function  $\psi(x)$  is discontinuous at the position  $x_0$ , i.e., there is a discontinuity with the two functional values  $\psi_1(x_0 - \varepsilon)$  and  $\psi_2(x_0 + \varepsilon)$  for  $\varepsilon \rightarrow 0$ . Prove that this discontinuity causes an infinite contribution  $\Delta\langle T \rangle \rightarrow \infty$  to the expectation value of the kinetic energy  $\hat{T} = \hbar^2 \hat{p}^2 / 2m$ .
- 3.6 In the simplest atom, the hydrogen (H) atom, an electron is bound to the positive nucleus (proton) by the attractive Coulomb force  $K = -e^2 / 4\pi\varepsilon_0 r^2$ . Write down the Schrödinger equation for the electron in the H atom and prove that it is solved by the ground state wave function  $\psi(r) = A \exp(-r/a_0)$ . What is the binding energy of the electron in this lowest energy (ground) state?
- 3.7 A particle with mass  $m$  is confined in a cubic box of length  $L$  on a quantized state  $\psi_n(\mathbf{r})$ . Calculate the force  $F = -\partial E / \partial L$ , if the walls of the box are slowly

moved inwards. During this process the particle shall stay in the same quantum state.

Compare the quantum mechanical result with the classical case that the particle has the energy  $E_n$ . For this purpose, calculate the particle's average velocity and the frequency of particle collisions with the walls. Therefrom the momentum transfer per collision and the resulting force on the walls can be calculated.

- 3.8 In the free electron gas model (potential box) the Fermi energy  $E_F$  can be defined as the energy, up to which at  $T = 0$  the single electron states of the potential box are occupied starting with the lowest energy state. Because of the Pauli principle each single electron state can be occupied only once.

Calculate Fermi wave length  $\lambda_F$  of electrons having the energy of the Fermi level  $E_F$ . The Fermi wave length  $\lambda_F$  shall be expressed as a function of the electron concentration for a 3D, 2D and 1D electron gas.

- 3.9 In semiconductor device applications often position ( $x$ ) dependent electron energies  $E(x)$  are plotted. Generally these plots describe space charge induced curved conduction and/or valence band edges  $E_C$ , respectively  $E_V$  of a semiconductor or semiconductor multilayer structure (see e.g. Figs. A.1d, A.2c, A.3c in Appendix A). Why are these so-called macropotential plots only approximations in view of the uncertainty principle? With the assumption that the electron energy near the band edges  $E_C$ ,  $E_V$  can be approximated by the energy of a free electron  $E = \frac{\hbar^2 k^2}{2m}$  estimate the validity limits  $\Delta E$ ,  $\Delta x$  for these  $E(x)$  plots. Compare with the spatial extension of a typical space charge layer in a semiconductor.

## Problems to Chap. 4

- 4.1 Calculate the inverse matrix  $\underline{\underline{M}}^{-1}$  of the matrix

$$\underline{\underline{M}} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 3 & 2 & 1 \end{pmatrix},$$

and show the validity of  $\underline{\underline{M}}\underline{\underline{M}}^{-1} = \underline{\underline{1}}$ .

- 4.2 Calculate the eigenvalues and the normalized eigenvectors of the matrix

$$\underline{\underline{M}} = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 4 \end{pmatrix}.$$

Are the eigenvectors orthogonal?

4.3 Consider the following matrix:

$$\underline{\Omega} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}.$$

- (a) Prove that the matrix is unitary.
- (b) Show that the matrix has the eigenvalues  $\exp(i\varphi)$  and  $\exp(-i\varphi)$ .
- (c) Calculate the corresponding eigenvectors and prove their orthogonality.

4.4 Prove the following commutator relations:

$$\begin{aligned} [\hat{A}, \hat{B}] + [\hat{B}, \hat{A}] &= 0, \\ [\hat{A}, \hat{B} + \hat{C}] &= [\hat{A}, \hat{B}] + [\hat{A}, \hat{C}], \\ [\hat{A}, \hat{B}\hat{C}] &= [\hat{A}, \hat{B}]\hat{C} + \hat{B}[\hat{A}, \hat{C}], \\ [\hat{A}, [\hat{B}, \hat{C}]] + [\hat{C}, [\hat{A}, \hat{B}]] + [\hat{B}, [\hat{C}, \hat{A}]] &= 0. \end{aligned}$$

4.5 Prove the following relation for the  $\delta$  function:

$$\delta(cx) = \frac{1}{|c|} \delta(x), \quad \text{for } c \neq 0 \text{ and real valued.}$$

- 4.6 Prove that the application of the operator  $(\alpha + \hat{b}^+)$  to an eigenstate of the harmonic oscillator yields a superposition of two states.  $\alpha$  is a normal  $C$  number and  $\hat{b}^+$  a step operator of the harmonic oscillator.
- 4.7 Prove by means of partial integration that the momentum operator  $\hat{p}_x = -i\hbar\partial/\partial x$  is Hermitian, i.e., that

$$\int \psi^* \hat{p}_x \varphi \, dx = \left( \int \varphi^* \hat{p}_x \psi \, dx \right)^*.$$

- (a) What condition must the wave function fulfill in the limit  $x \rightarrow \pm\infty$ ?
- (b) Show that in spherical coordinates the operator  $-i\hbar\partial/\partial r$  is not Hermitian. Prove furthermore that in spherical coordinates the momentum operator must be written as  $\hat{p}_r = -i\hbar \frac{1}{r} \frac{\partial}{\partial r} r$  to be Hermitian.
- 4.8 Solve the Schrödinger equation for a negative bonding  $\delta$  potential  $V(x) = -a\delta(x)$ .  
*Hint:* Integrate the Schrödinger equation around the zero point and search for a bound state with  $\psi(\pm x) \rightarrow 0$  for  $x \rightarrow \pm\infty$ .
- 4.9 Prove the Ehrenfest theorem

$$\frac{d}{dt} \langle \Omega \rangle = -\frac{i}{\hbar} \langle [\hat{\Omega}, \hat{H}] \rangle.$$

This theorem describes the time change of the expectation value  $\langle \hat{\Omega} \rangle$  of an operator  $\hat{\Omega}$  which does not explicitly depend on time as does  $\hat{H}$ .

*Hint:* consider the time derivative

$$\frac{d}{dt}\langle \hat{\Omega} \rangle = \frac{d}{dt}\langle \psi | \hat{\Omega} | \psi \rangle$$

and express  $\dot{\langle \psi |}$  and  $|\dot{\psi} \rangle$  by means of the Schrödinger equation.

Show furthermore that the classical relation  $\dot{x} = p/m$  is valid in quantum mechanics for the expectation values  $\langle x \rangle$  of position and  $\langle p \rangle$  of momentum.

## Problems to Chap. 5

- 5.1 Prove the following interesting properties of the Pauli spin operators  $\hat{\sigma}_i$ :
- (a)  $[\hat{\sigma}_i, \hat{\sigma}_j]_+ = \hat{\sigma}_i \hat{\sigma}_j + \hat{\sigma}_j \hat{\sigma}_i = 0$  for  $i \neq j$ . This operator is called anti-commutator; it plays an important role for the quantization of fermion fields (Sect. 8.3.1).
  - (b)  $\hat{\sigma}_i^2 = \hat{1}$ .
  - (c)  $[\hat{\sigma}_i, \hat{\sigma}_j]_+ = 2\delta_{ij}\hat{1}$ .
  - (d)  $(\mathbf{a} \cdot \hat{\boldsymbol{\sigma}})(\mathbf{b} \cdot \hat{\boldsymbol{\sigma}}) = \mathbf{a} \cdot \mathbf{b}\hat{1} + i(\mathbf{a} \times \mathbf{b}) \cdot \hat{\boldsymbol{\sigma}}$  with  $\mathbf{a}$  and  $\mathbf{b}$  being arbitrary 3D vectors.
- 5.2 A beam of fermions (spin 1/2 particles) passes along the  $y$  axis of two collinear Stern–Gerlach (SG) set-ups, which split the beam into two secondary beams with opposite spin orientation. In both SG set-ups the lower resulting beam is blocked. In the first SG apparatus the  $\mathbf{B}$  field is oriented in  $z$  direction, while the  $\mathbf{B}$  field is oriented in  $x$  direction in the second SG apparatus.  
What ratio of particles, which have passed the first SG set-up, is detected behind the second SG apparatus?
- 5.3 Prove that the operator  $\delta \mathbf{s} \cdot \hat{\mathbf{p}}$  with  $\hat{\mathbf{p}}$  as momentum operator produces an infinitesimal displacement  $\delta \mathbf{s}$  of the wave function.

*Hint:* Consider the relation

$$\langle x, y | \hat{1} - \frac{i}{\hbar} \delta \mathbf{s} \cdot \hat{\mathbf{p}} | \psi \rangle = \psi(x - \delta s_x, y - \delta s_y).$$

- 5.4 At the time  $t = 0$  an electron is in the spin state  $s_z = \hbar/2$  in a static magnetic field  $\mathbf{B}$ . What is the duration of a spin flip process after switching on an electromagnetic RF field with the correct (according to static magnetic field  $\mathbf{B}$ ) frequency and a magnetic field amplitude of 1 mT.
- 5.5 A helium (He) atom is in the ionized state  $\text{He}^+$ , i.e., one electron is missing in the electronic shell. Estimate the energy difference between the ground state and the 1st excited state of the one remaining valence electron. In the hydrogen atom the ground state binding energy amounts to 13.6 eV.

- 5.6 Calculate the Fermi energy  $E_F$  for a 2D electron gas (2DEG) with a 2D electron density of  $n = 10^{12} \text{ cm}^{-2}$  (Sect. 3.6.1). How large is the de Broglie wavelength of an electron at the Fermi energy?
- 5.7 A neutron star is the end product (final state) of the gravitational collapse (supernova) of a heavy star. Such a neutron star is a giant nucleus composed of neutrons. Neutrons (approx. proton mass) are fermions and obey the Fermi principle, i.e., they have a Fermi distribution of energies. Consider a neutron star of mass  $5 \times 10^{30} \text{ kg}$  and radius 10 km.
- Calculate the mass density of the neutron star and compare with mass densities on earth.
  - Calculate the average energy of the neutrons in the star and derive a mean temperature of the star.
- 5.8 Consider a one-dimensional conducting ring with radius  $r_0$  (as in Sect. 5.7.3).
- Prove that the stationary eigenstates of an electron in this ring are described by the wave function  $\psi(\varphi) = (2\pi r_0)^{-1/2} \exp(im\varphi)$  with  $m$  as angular momentum directional quantum number.
  - Calculate the quantum mechanical current of this one electron in the ring at negligible temperature as a function of  $\varphi$  and compare with the classical expression for an electron orbiting in the ring.

## Problems to Chap. 6

- 6.1 Solve the Schrödinger equation for an electron which is confined in a 2D rectangular potential box with length  $L$ . Calculate the energy eigenvalues and the wave functions. Now a small perturbation potential  $v = ay$  ( $a$  constant) along the  $y$  direction is switched on. How do the energies of the ground state and the first excited states change?
- 6.2 An electron is confined in a 1D triangular potential  $V(x) = C|x|$ . Calculate approximately the electronic energies of the ground state and of the first excited state. Use the variational method (Sect. 6.2) with the assumption of reasonably well guessed trial wave functions having one fitting parameter.
- 6.3 Prove that the perturbation operator  $(e/m)\mathbf{A} \cdot \hat{\mathbf{p}}$ , which is used for the description of electron light interaction, is Hermitian.
- 6.4 Light with randomly varying polarization is irradiated on an atomic system and induces electronic transitions from the initial state  $|i\rangle$  to the final state  $|f\rangle$ . The random polarization is described by a polarization unity vector  $\mathbf{e}$  randomly rotating around the propagation wave vector  $\mathbf{q}$  of the light wave. Prove that in dipole approximation the transition dipole matrix element for light with random polarization (left expression) is related to that with a spatially fixed polarization (right expression) as

$$\overline{|\langle f | \mathbf{r} \cdot \mathbf{e} | i \rangle|^2} = \frac{1}{3} |\langle f | \mathbf{r} | i \rangle|^2,$$

$\mathbf{r}$  is the oscillation vector (direction) of the electronic transition dipole. Note that the vector  $\langle f | \mathbf{r} | i \rangle$  usually possesses complex components.

- 6.5 Using Born's approximation calculate the differential scattering cross section for scattering of an electron on the spherical charge distribution  $\rho(r)$  of an atom. The result should be expressed in terms of the (Rutherford) scattering cross section for scattering on a point charge and a form factor  $f$  as function of wave vector transfer  $K$ . Compare  $f(K)$  for the case of a homogeneous charge distribution  $\rho(r) = \text{const}(r \leq R)$  with that of a Gaussian charge distribution  $\rho(r)$  of the same half width  $R$ .
- 6.6 Prove that the scattering amplitude  $f(\vartheta, \varphi)$  (6.211) for particle scattering on a spherical potential  $V(r)$  can be represented at low particle energies ( $k \rightarrow 0$ ), i.e., for negligible wave vector transfer  $K = 2k \sin(\vartheta/2) \rightarrow 0$ , as

$$f(\vartheta) \cong -\frac{m V_0 r_0^3}{\hbar^2}.$$

$V_0$  is the effective height and  $r_0$  the effective extension of the scattering potential. The scattering is assumed as essentially isotropic, i.e., not being dependent on  $\vartheta$ .

- 6.7 Prove that for particle scattering on a spherical potential  $V(r)$  the scattering amplitude  $f(\vartheta, \varphi)$  (6.211) at high particle energies has non-negligible values only under the condition

$$K r' \cos \vartheta' = K r' \xi < \pi.$$

*Hint:* Consider the oscillating term  $\exp(iKr'\xi)$  in (6.211) and conclude that the scattering amplitude in forward direction yields significant contributions only within an angle range  $\vartheta' < 1/kr_0$  where  $r_0$  is the effective range of influence of the potential.

## Problems to Chap. 7

- 7.1 The Hermitian operator  $\hat{A}$  expressed by the corresponding matrix  $\underline{\underline{A}}$  has the eigenvalues  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ .
- (a) Prove the following relations for a 2-dimensional Hilbert space (spin):

$$\begin{aligned} \text{Tr } \underline{\underline{A}} &= \lambda_1 + \lambda_2, \\ \text{Det } \underline{\underline{A}} &= \lambda_1 \lambda_2. \end{aligned}$$

- (b) Prove the trace and determinant relations of (a) for an  $n$ -dimensional Hilbert space.

7.2 Prove that the density matrix of spins (with  $\hbar/2$ ) realized by an ensemble of fermions can be expressed as

$$\hat{\rho} = \frac{1}{2}(\hat{1} + \mathbf{a} \cdot \hat{\sigma}),$$

where  $\mathbf{a}$  is a normal Euclidian 3D vector.

Furthermore, prove that  $\mathbf{a}$  describes the average spin polarization  $\langle \sigma \rangle$ .

*Hint:* Use the fact that each  $2 \times 2$  matrix can be expressed as a linear combination of  $\hat{1}$  and the Pauli matrices.

7.3 Using problem 7.2 show that the operator combination  $\hat{O}_A(\alpha)\hat{O}_B(\beta)$  defined in Sect. 7.2.1 by (7.21), (7.22a), (7.22b) and used in (7.23) describes the correlation between the measurement results spin orientation  $\alpha$  at Alice's detector and spin orientation  $\beta$  at Bob's detector. Both measurements are performed on two spin anti-correlated particles which are emitted in opposite direction from a common source (Fig. 7.5).

7.4 Using the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = \hat{H} |\psi\rangle$$

show that the time evolution of the density matrix  $\hat{\rho}$  both for pure states (7.48) and for mixed states (7.45) is described by

$$\frac{\partial}{\partial t} \hat{\rho}(t) = -\frac{i}{\hbar} [\hat{H}, \hat{\rho}].$$

Note that the result, the so-called *von Neumann equation*, resembles Heisenberg's dynamical equation for operators (4.104) apart from the different sign.

7.5 By means of the unitary propagator (operator)  $\hat{U}$  the time evolution of a quantum state can be expressed as

$$|\psi(t)\rangle = \hat{U} |\psi(0)\rangle.$$

Prove that the time evolution of the density matrix is obtained as

$$\hat{\rho}(t) = \hat{U} \hat{\rho}(0) \hat{U}^+.$$

7.6 According to John von Neumann the entropy  $S$  of quantum mechanical system can be expressed in terms of its density matrix  $\hat{\rho}$  as

$$S = -k \text{Tr}(\hat{\rho} \ln \hat{\rho}).$$

- (a) Calculate the entropy of the quantum-bit state

$$|\psi\rangle = a|0\rangle + b|1\rangle.$$

The states  $|0\rangle$  and  $|1\rangle$  of the Q-bit might be realized by the spin states  $|\uparrow\rangle$ ,  $|\downarrow\rangle$  or by ground and excited state  $|g\rangle$ ,  $|e\rangle$  of a 2-level atom.

- (b) Prove that during the time evolution (expressed by the propagator  $\hat{U}$ ) of a pure state, e.g., of the Q-bit state  $|\psi\rangle = a|0\rangle + b|1\rangle$ , the entropy does not change:

$$\frac{\partial}{\partial t} S = 0.$$

Pure states evaluate reversibly in time, a condition for processing Q-bits in a quantum computer.

- (c) A measurement on the Q-bit state transforms this pure state into a mixed state with the density matrix  $\hat{\rho} = |a|^2|0\rangle\langle 0| + |b|^2|1\rangle\langle 1|$ . Prove this result! Show that due to the measurement the Q-bit entropy changes from zero to

$$S = -k(|a|^2 \ln |a|^2 + |b|^2 \ln |b|^2).$$

- 7.7 In quantum information the Hadamard transformation (operator  $\hat{G}_H$ ) realized by an appropriate quantum gate plays an important role. The operator  $\hat{G}_H$  is defined by its action on a 2-level system as follows:

$$\hat{G}_H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle); \quad \hat{G}_H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Consider the action of  $\hat{G}_H$  on the entangled 2-particle state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B).$$

This state is used for experimentally checking the Bell inequalities (Sect. 7.2.1). Prove that the resulting state is again entangled.

- 7.8 A and B are two coupled and entangled 2-level quantum systems (spins, atoms, etc.).

As separate systems they are described by the states  $|0\rangle_A$ ,  $|1\rangle_A$ ,  $|0\rangle_B$ ,  $|1\rangle_B$ .

- (a) Calculate the density matrix  $\hat{\rho}$  in bra-ket and in 2D matrix notation for the entangled state of the two systems:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B).$$

- (b) System B might be considered as the quantum environment of system A on which the interest is focused. Neglect system B when measuring a certain variable in system A. For this purpose calculate the reduced density

matrix  $\hat{\rho}_{\text{red}} = \text{Tr}_B \hat{\rho}$ , by which expectation values in the open system  $A$  are calculated. Does  $\hat{\rho}_{\text{red}}$  describe a pure or a mixed ensemble (proof!)?

- (c) Calculate the von Neumann entropy  $S = -k \text{Tr}(\hat{\rho}_{\text{red}} \ln \hat{\rho}_{\text{red}})$  for the reduced density matrix, i.e., the open system  $A$ .

## Problems to Chap. 8

8.1 A closed electrical oscillator circuit without resistance contains a capacitance  $C$  and an inductivity  $L$  in series.

- (a) Determine the oscillator equation for the charge  $Q$  and calculate the resonance frequency  $\omega = (LC)^{-1/2}$ .
- (b) Calculate the total energy stored in the oscillator, with  $\Phi_B = LI$  as the magnetic flux in the inductivity and  $Q = CU$  as the charge stored in the capacity. This total energy stored in the oscillator represents the Hamilton function of the system.
- (c) Quantize the electrical oscillator by finding its canonically conjugate variables in analogy to  $\hat{x}, \hat{p}$  for the harmonic mechanical oscillator. Then, introduce commutation relations analogously to  $[\hat{p}, \hat{x}] = -i\hbar$ .
- (d) What variables of the electrical oscillator are incommensurable, i.e., can not be measured simultaneously? What is the minimum energy stored in the oscillator?

8.2 The particle number operator (8.147) is represented as

$$\hat{N} = \int d^3r \hat{\psi}^+(\mathbf{r}) \hat{\psi}(\mathbf{r}).$$

Prove the following relation both for fermion and for boson fields:

$$\hat{\psi}(\mathbf{r}) \hat{N} = (\hat{N} + 1) \hat{\psi}(\mathbf{r}).$$

8.3 Jordan–Wigner matrices

- (a) Prove that the fermionic anti-commutation relations (8.127a)–(8.127c) for creation and destruction operators are fulfilled by the so-called Jordan–Wigner matrices (JW matrices):

$$\hat{a}^+ = \frac{1}{2}(\hat{\sigma}_x + i\hat{\sigma}_y) \quad \text{and} \quad \hat{a} = \frac{1}{2}(\hat{\sigma}_x - i\hat{\sigma}_y),$$

$\hat{\sigma}_x$  and  $\hat{\sigma}_y$  are the Pauli spin matrices (5.116). In this representation the occupation of a single electron state is given as

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

- (b) Many-particle states  $|\Phi\rangle$  require indices at the different JW matrices which denote wave vector and spin quantum number of each particle by  $i = (\mathbf{k}, s)$ :

$$|\Phi\rangle = \hat{a}_1^+, \hat{a}_2^+, \dots, \hat{a}_i^+, \dots, \hat{a}_N^+ |0\rangle.$$

Write down the eigenvalue equation for the particle number operator  $\hat{N}$  (8.138) and calculate the eigenvalue by using the JW matrix formalism.

- (c) Using the JW matrix formalism prove the anti-symmetry of a two-fermion state.  
 (d) Write down the many-particle wave function of the Fermi sea of free electrons, in which all single electron states below the Fermi wave vector  $k_F$  are occupied and all states above  $k_F$  are empty.

8.4 Prove that the annihilation field operator  $\hat{\Psi}(\mathbf{r})$  transforms the one-particle state  $\hat{\Psi}^+(\mathbf{r}) |0\rangle$  of a fermion at position  $\mathbf{r}$  into the ground state  $|0\rangle$ . Calculate for this purpose the particle density at an arbitrary position  $\mathbf{r}'$  in the field by means of the density operator (8.148).

- 8.5 (a) Prove that the particle number operator  $\hat{N} = \sum_j \hat{a}_j^+ \hat{a}_j$  (8.138) commutes with the Hamilton field operator  $\hat{H} = \sum_j \varepsilon_j \hat{a}_j^+ \hat{a}_j$  (8.132) of a free (without external interaction) electron (fermion) field. What is the consequence for the particle number taking into account Sect. 4.3.5?  
 (b) What is the situation, when the electron field interacts with a phonon field via an interaction operator (8.236) of the type

$$\hat{H}_{ep} = \sum_{kk'} A_\kappa \hat{a}_{k+\kappa}^+ \hat{a}_k \hat{c}_\kappa + \sum_{kk'} A_\kappa \hat{a}_{k-\kappa}^+ \hat{a}_k \hat{c}_\kappa^+ \quad ?$$

$A_\kappa$  are complex numbers and  $\hat{c}_\kappa, \hat{c}_\kappa^+$  phonon annihilation and creation operators, respectively. Annihilation and creation operators of different quantum fields commute. Is the particle number in the electron field conserved ? Describe the interaction process of  $\hat{H}_{ep}$ .

8.6 Many-particle matrix elements

- (a) Using the fermionic anti-commutation rules for electrons (8.127a)–(8.127c) prove the following relation for the many-particle matrix element (8.170a):

$$\langle 0 | \hat{a}_N, \dots, \hat{a}_2 \hat{a}_1 | \hat{a}_i^+ \hat{a}_j | \hat{a}_1^+ \hat{a}_2^+, \dots, \hat{a}_N^+ | 0 \rangle = \begin{cases} \delta_{ij}; & j = 1, 2, \dots, N \\ 0, & \text{otherwise.} \end{cases}$$

Note that  $\hat{a}_j$  does not operate on an empty state; it can be shifted through the operator product and yields zero by operating on  $|0\rangle$ . Furthermore states with differing electron occupation are orthogonal.

- (b) In analogy prove the following relation for the matrix element (8.170b):

$$\begin{aligned} & \langle 0 | \hat{a}_N, \dots, \hat{a}_2 \hat{a}_1 | \hat{a}_i^+ \hat{a}_k^+ \hat{a}_j \hat{a}_l | \hat{a}_1^+ \hat{a}_2^+, \dots, \hat{a}_N^+ | 0 \rangle \\ &= \begin{cases} \delta_{il} \delta_{kj}; & i \neq k, l \neq j; i = 1, 2, \dots, N, k = 1, 2, \dots, N \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

*Hint:* Clarify the calculation steps by firstly considering states with only 1, 2, 3 electrons, i.e., expressions of the type  $\hat{a}_1^+ \hat{a}_2^+ |0\rangle$ , respectively  $\hat{a}_1^+ \hat{a}_2^+ \hat{a}_3^+ |0\rangle$ .

- 8.7 Prove for the monatomic linear chain, that the total mechanical momentum of all vibrating atoms in an excited lattice wave (8.203a), (8.203b) vanishes. Discuss the result in relation to phonon quasi-momentum  $\hbar\kappa$  ( $\kappa$  wave number).
- 8.8 Using phonon creation and destruction operators prove the following relations for particle numbers and particle number operators:

$$\hat{n}_\kappa = \hat{n}_{-\kappa} \quad \text{and} \quad n_\kappa = n_{-\kappa}.$$

# Index

- A**
- Acceptor, 439
  - Acoustic branch, 447
  - Action, 42
  - Addition energy, 79
  - Adiabatic approximation, 426
  - Aharanov–Bohm-effect, 155
  - Allowed energy band, 434
  - Angstrom, 12
  - Angular momentum, 131, 133
  - Anharmonic oscillator, 126
  - Anharmonic term, 126
  - Annihilation operator, 380
  - Anti-bonding state, 237
  - Anti-commutation relation, 414
  - Antiparticle, 190
  - Antisymmetric, 177
  - Antisymmetry, 174
  - Atomic core, 208
  - Atomic scattering factor, 303
- B**
- Band structure, 435
  - Bare states, 397
  - Baryons, 193
  - Bell inequality, 327
  - Bit, 360
  - Black-body, 184
  - Bloch theorem, 432
  - Bloch wave, 432
  - Bohr magneton, 163
  - Bonding state, 237
  - Born approximation, 295
  - Born–Oppenheimer approximation, 426
- Bose-Einstein condensation, 187
  - Bosons, 178
  - Bra, 109
  - Bra vector, 107
  - Bragg (diffraction) peaks, 304
  - Bragg method, 307
  - Bras, 107
  - Brillouin zone, 433
  - Buckyball molecule, 19
- C**
- Canonical momentum, 148
  - Casimir force, 406
  - Characteristic equation, 103
  - Charge Density Functional (CDF) theory, 429
  - Chromodynamics, 197
  - Closed quantum system, 354
  - Collapse of the wave function, 95, 467
  - Color, 194
  - Commensurable, 38, 96
  - Commutator, 96
  - Complementary, 38
  - Complete, 94
  - Complete, orthonormal system, 94
  - Conductance quantum, 85
  - Conduction band, 437
  - Conduction band offset, 475
  - Constant of the motion, 117
  - Continuity relation, 56
  - Cooper pairs, 463
  - Core, 431
  - Correspondence principle, 39
  - Coulomb blockade, 77, 80
  - Coulomb blockade oscillations, 80

Coulomb blockade staircase, 81  
 Coulomb exchange interaction, 429  
 Coulomb scattering, 297  
 Coulomb term, 430  
 Crystal lattice, 302  
 Cyclotron frequency, 147

**D**

Decoherence, 354, 468  
 Defect electrons, 437  
 Deformation potential, 454  
 Degeneracy, 94  
 Degenerate, 94  
 Delta function, 112  
 Density matrix, 343, 347  
 Density of states, 54  
 Dephasing, 357  
 Dephasing time, 357  
 Depletion layer, 221  
 Destruction operator, 380  
 Diamagnetism, 152  
 Differential scattering cross section, 290  
 Dipole approximation, 261  
 Dirac equation, 407  
 Dirac notation, 106  
 Dirac sea, 189  
 Direct metal etching, 482  
 Dispersion, 36  
 Distributions, 112  
 Donor, 439  
 $d$ -orbitals, 145  
 Double slit interference, 30, 334  
 Drain, 75  
 Dressed states, 397  
 Dual vector space, 91  
 Dynamic scattering theory, 296

**E**

Effective electron mass, 434  
 Eigenvalue, 45, 92  
 Electrical resistance, 458  
 Electrochemical potential, 78  
 Electron beam (e-beam) lithography, 481  
 Electron depletion zone, 473  
 Electron spin resonance (ESR), 282  
 Electron tunneling, 61  
 Electron–phonon interaction, 453  
 Electroweak interaction, 195  
 Elementary particles, 188  
 Entanglement, 27, 324  
 EPR paradox, 324, 425, 469  
 Even parity, 265  
 Evolutionary epistemology, 7

Ewald construction, 306  
 Exchange term, 430  
 Expectation value, 45  
 Extrinsic surface states, 472

**F**

Fermi distribution, 183  
 Fermi energy, 182  
 Fermions, 177  
 Fermi's golden rule, 258  
 Field operator, 418  
 Fixed boundary conditions, 52  
 Flux quantum, 157  
 Fock space, 449  
 Forbidden band, 434  
 Forbidden energy band, 434  
 $f$ -orbitals, 145  
 Forward current, 221  
 Fullerene, 19, 20

**G**

Gate, 75  
 Gauge invariant, 154  
 Gauge transformation, 154  
 Generator of infinitesimal rotations, 141  
 Gluons, 197  
 Grating function, 304  
 Gravitation, 196  
 Gravitons, 197  
 Green function, 293  
 Group velocity, 35

**H**

Hadrons, 192  
 Hamilton equation, 40  
 Hamilton function, 40  
 Hamilton operator, 44  
 Hamiltonian, 40, 44  
 Harmonic oscillator, 118  
 Hartree–Fock equation, 429  
 Heavy bosons, 197  
 Heisenberg equation, 117  
 HEMT structure, 477  
 Hermitian operator, 93  
 Hidden local variables, 326  
 Hidden variables, 32  
 Higgs boson  
   field, 198  
   mechanism, 198  
 Hilbert space, 89, 91  
 Holes, 437

- Hund's rule, 205  
Hydrogen (H) atom, 206  
Hypothetical realism, 7
- I**  
Incommensurable, 38  
Insulator, 437  
Interference transistor, 159  
Intrinsic surface states, 472  
Inverse matrix, 101  
Isospin, 192
- K**  
Ket, 107–109  
Ket vector, 107  
Kinematic scattering theory, 296  
Kinetic momentum, 148  
Klein–Gordon equation, 407  
Kronecker symbol, 90
- L**  
Laue equations, 304  
Leptons, 191  
Lift-off process, 482, 483  
Linear monatomic chain, 440  
Linear Stark effect, 253  
Lithography, 481  
Lorentz force, 147  
Low Energy Electron Diffraction (LEED), 16
- M**  
Magnetic flux quantum, 157  
Magnetic quantum number, 138  
Magneton, 163  
Matrix, 98  
Matrix mechanics, 99  
Maxwell equation, 376  
Mesons, 193  
Metal Organic Vapor Phase Epitaxy (MOVPE), 479  
Metal-induced interface states, 474  
Metals, 436  
Miller indices, 306  
Mixed states, 346  
Molecular Beam Epitaxy (MBE), 479  
Moment of inertia, 133
- N**  
Negative differential resistance, 75  
Neutrino, 191
- Neutron, 190  
NMR signal, 284  
Non-locality, 468  
Normal process, 451  
Normalisation, 31  
Nuclear magnetic resonance (NMR), 282  
Nuclear magneton, 283
- O**  
Observable, 44  
Odd parity, 265  
Open quantum system, 354  
Operators, 44  
Optical absorption, 260  
Optical branches, 447  
Optical emission, 260  
Optical lithography, 481  
Orientation quantum number, 138  
Orthodox model, 78
- P**  
Paramagnetism, 152  
Parity selection rule, 265  
Particle resonance, 192  
Pauli equation, 168  
Pauli exclusion principle, 178  
Pauli matrices, 166  
Periodic boundary conditions, 52  
Periodic table of elements, 209  
Phase-based nanoelectronics, 159  
Phonon absorption, 457  
Phonon emission, 458  
Phonon–electron interaction, 452  
Phonon–phonon interaction, 448  
Phonons, 442  
Photons, 197  
 $\pi/2$  pulse, 277  
 $\pi$  pulse, 277  
Pinned (Fermi level), 473  
Planck constant, 2, 11  
Polaron, 460  
 $p$ -orbitals, 145  
Positron, 190  
Precession frequency, 171  
Principal quantum number, 144, 203  
Principle of complementarity, 38  
Probabilities, 60  
Projection operator, 108  
Propagator, 115  
Pure states, 346

**Q**

Quadratic Stark effect, 253  
 Quantum bit, 360  
 Quantum field theory, 468  
 Quantum gate, 360  
 Quantum number operator, 124  
 Quantum point contact, 82, 83, 87, 339  
 Quantum state, 89  
 Quarks, 191  
 Quasi-momentum conservation, 451

**R**

Rabi frequency, 276, 279  
 Reactive Ion Etching (RIE), 483  
 Reciprocal space, 304  
 Reduced density matrix, 352  
 Reduction, 95  
 Reflection amplitude, 59  
 Reflection coefficients, 60  
 Reflection High Energy Electron Diffraction, 480  
 Resonance tunneling diode, 75  
 Resonant tunneling, 71  
 Rest energy, 188  
 Reverse bias, 221  
 Reverse current, 221  
 Rotating crystal method, 306  
 Rotating wave approximation, 274  
 Rotation matrix, 103

**S**

Scanning electron tunneling microscope, 65  
 Schrödinger equation, 46  
 Schrödinger's Cat, 359  
 Second quantization, 407  
 Selection rules, 265  
 Selective growth, 481  
 Self-consistent field theoretical techniques, 429  
 Self-adjoint operator, 93  
 Semiconductors, 437  
 Single electron tunneling, 75, 77  
 Slater determinant, 179  
 Slit distance, 25  
 s-orbitals, 145  
 Source, 75  
 Spin, 162  
 Split gate contact, 85  
 Spontaneous emission, 261, 393  
 Standard model, 188  
 Statistical operator, 343, 347  
 Step operators, 122  
 Strong interaction, 195  
 Structural realism, 426

**Structure factor, 307**

Sub-bands, 56  
 Subsystem, 351  
 Superconductivity, 463  
 Superposition principle, 316  
 Superposition state, 277  
 Superstructure, 240, 471  
 Surface acceptor states, 473  
 Surface donor states, 473  
 Surface potential, 473  
 Symmetric, 177  
 Symmetry, 174

**T**

Thermal de Broglie wavelength, 186  
 Thought experiments, 469  
 Time-independent Schrödinger equation, 48  
 Torque, 131, 132  
 Trace, 347  
 Transfer-matrix, 63  
 Transformations, 106  
 Transmission, 60  
 Transmission amplitude, 59, 64  
 Transmission-matrix, 67  
 Transposed matrices, 99  
 Tunneling probability, 64

**U**

Umklapp processes, 452  
 Uncertainty principle, 38  
 Uncertainty relation, 38  
 Unitary matrices, 106

**V**

Vacuum state, 381  
 Valence band, 437  
 Valence band offset, 475  
 Virtual phonon emission, 459  
 Von Neumann equation, 499

**W**

Wave function, 29  
 Wave mechanics, 99  
 Wave packet, 33  
 Which-way information, 24  
 WKB approximation, 219

**Y**

Yukawa potential, 299, 465