SOLID-STATE SCIENCES

M. I. Monastyrsky (Ed.)

Topology in Condensed Matter

Springer Series in

SOLID-STATE SCIENCES 150

Springer Series in

SOLID-STATE SCIENCES

Series Editors:

M. Cardona P. Fulde K. von Klitzing R. Merlin H.-J. Queisser H. Störmer

The Springer Series in Solid-State Sciences consists of fundamental scientific books prepared by leading researchers in the field. They strive to communicate, in a systematic and comprehensive way, the basic principles as well as new developments in theoretical and experimental solid-state physics.

136 **Nanoscale Phase Separation and Colossal Magnetoresistance**
The Physics of Manganites and Related Compounds
By E. Dagotto

137 **Quantum Transport in Submicron Devices**
A Theoretical Introduction
By W. Magnus and W. Schoenmaker

138 **Phase Separation in Soft Matter Physics**
Micellar Solutions, Microemulsions, Critical Phenomena
By P.K. Khabibullaev and A.A. Saidov

139 **Optical Response of Nanostructures**
Microscopic Nonlocal Theory
By K. Cho

140 **Fractal Concepts in Condensed Matter Physics**
By T. Nakayama and K. Yakubo

141 **Excitons in Low-Dimensional Semiconductors**
Theory, Numerical Methods, Applications By S. Glutsch

142 **Two-Dimensional Coulomb Liquids and Solids**
By Y. Monarkha and K. Kono

143 **X-Ray Multiple-Wave Diffraction**
Theory and Application
By S.-L. Chang

144 **Physics of Transition Metal Oxides**
By S. Maekawa, T. Tohyama, S.E. Barnes, S. Ishihara, W. Koshibae, and G. Khaliullin

145 **Point-Contact Spectroscopy**
By Y.G. Naidyuk and I.K. Yanson

146 **Optics of Semiconductors and Their Nanostructures**
Editors: H. Kalt and M. Hetterich

147 **Electron Scattering in Solid Matter**
A Theoretical and Computational Treatise
By J. Zabloudil, R. Hammerling, L. Szunyogh, and P. Weinberger

148 **Physical Acoustics in the Solid State**
By B. Lüthi

149 **Solitary Waves in Complex Dispersive Media**
Theory · Simulation · Applications
By V.Y. Belashov and S.V. Vladimirov

150 **Topology in Condensed Matter**
Editor: M.I. Monastyrsky

Volumes 90–135 are listed at the end of the book.

M.I. Monastyrsky (Ed.)

Topology in Condensed Matter

With 108 Figures

Springer

Professor Dr. Michail Ilych Monastyrsky
Institute of Theoretical and Experimental Physics
B. Cheremushkinskaya 25, 117259 Moscow, Russia
E-mail: monastyr@itep.ru

ISSN 0171-1873

ISBN-10 3-540-23406-3 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-23406-7 Springer Berlin Heidelberg New York

Library of Congress Control Number: 2005933347
Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at http://dnb.ddb.de.

Typesetting: by the authors and SPI Publisher Services using a Springer LaTeX macro package
Cover concept: eStudio Calamar Steinen

Printed on acid-free paper SPIN: 10976142 57/3141/SPI Publisher Services - 5 4 3 2 1 0

Preface

This volume is based on the talks and lectures given by the participants of the 3-month seminar program "Topology in Condensed Matter," which was held in the MPIPKS Dresden, 8 May–31 July 2002 under the scientific direction of Professors M. Kleman, S. Novikov, and myself.

The aim of this program was to discuss recent applications of topology to several areas in condensed matter physics and related fields like biology. The last 30 years of the development of modern physics affirmed two important ideas. The first is the efficient applications of topology in physics. One should mention applications in condensed matter, such as classification of defects and textures in liquid crystals and superfluid liquids, the role of entangibility in polymer physics and DNA structures. The second tendency is also very prevalent. Some important discoveries in particle physics and condensed matter led to new and unpredictable questions in pure mathematics. We refer to the invention of monopoles and instantons in quantum field theory, quasicrystals fluid membranes of high genus, fullerenes (C_{60}, C_{90}, etc.), and so on in condensed matter. The number of such applications in the last years has increased substantially.

The papers presented in this volume and the next one "Topology in Biology" reflect the spectrum of topics discussed. Besides original papers, a mini-course in topology for physicists and biologists was organized. These lectures will be published in the second volume. By the common opinion of participants the seminar was very successful. The organizers and participants are grateful to the MPIPKS for the generous sponsorship of the seminar with such an unusual spectrum of interest. Special thanks go to the director of MPIPKS P. Fulde, the head of the visitors program S. Flach, and the secretaries K. Lantch, M. Lochar, and C. Poenish. We acknowledge our gratitude to the entire staff of the Institute for their help in organizing the seminar and for making sure it ran smoothly. We express our gratitude to Dr. Aschcroft, who suggested publishing these lectures in Springer Verlag and assisted in the preparation of these volumes. We hope that such programs that bring together mathematicians, physicists, and biologists will be continued.

Moscow and Dresden
October 2005 *Michael Monastyrsky*

Contents

List of Contributors

T. Jannsen
Institute for Theoretical Physics
University of Nijmegen Toernooiveld
6525 ED Nijmegen, The Netherlands
ted@sci.kun.nl

R. Kerner
Laboratoire de Physique Théorique
des Liquides (LPTL)
Université Pierre et Marie Curie
CNRS URMR 7600 Tour 22
4-ème étage, Boite 142 4 Place
Jussieu 75005 Paris, France
textttrk@ccr.jussieu.fr

A.M. Kosevich
B.I. Verkin Institute for Low
Temperature Physics and
Engineering of National Academy of
Sciences of Ukraine,Ukraine
kosevich@ilt.kharkov.ua

O.D. Lavrentovich
Chemical Physics Interdisciplinary
Program and Liquid Crystal
Institute
Kent State University
Kent, OH 44242, USA
old@lci.kent.edu

A.Ya. Maltsev
L.D. Landau Institute for
Theoretical Physics
ul. Kosygina 2 Moscow, Russian
Federation
maltsev@itp.ac.ru

R. Mosseri
Groupe de Physique des Solides
CNRS UMR 7588
Université Pierre et Marie Curie
Paris 6 et Denis Diderot Paris 7, 2
place Jussieu 75251 Paris cedex 05,
France
mosseri@ccr.jussieu.fr

S.P. Novikov
Institute for Physical Science and
Technology (IPST)
University of Maryland
College Park, MD 20742-2431, USA
novikov@ipst.umd.edu

V.A. Osipov
Bogoliubov Laboratory of
Theoretical Physics
Joint Institute for Nuclear Research
141980 Dubna, Moscow region,
Russian Federation
osipov@thsun1.jinr.ru

I.I. Smalyukh
Chemical Physics Interdisciplinary Program and Liquid Crystal Institute
Kent State University
Kent, OH 44242, USA
smalyukh@lci.kent.edu

J. Yi
Institute für Theoretische Physik
Universität Regensburg
93040 Regensburg, Germany
juyeon.yi@physik-regensburg.de

B. Zhilinskii
Université du Littoral
UMR du CNRS 8101
59140 Dunkerque, France
zhilin@univ-littoral.fr

Introduction

M. Monastyrsky

Articles presented in this volume can be divided into several parts.

The first part consists of papers by A.M. Kosevich "Topology in the Electron Theory of Metals" and A.Ya. Maltsev and S.P. Novikov "Topology, Quasiperiodic Functions, and the Transport Phenomena." The authors study problems in the electron theory of metals, including the topology of Fermi surfaces, magnetic oscillations, and magnetoresistivity. These problems, important from a physical point of view, lead to new and interesting fields of investigations in two-dimensional and three-dimensional topology.

The second part consists of three papers: R. Kerner's "The Role of Topology in the Growth and Agglomeration," V. Osipov's "Topological Defects in Carbon Nanocrystals," and J. Yi's "Physics from Topology and Structures." These authors consider from different views the role of topology in such "hot" topics as the growth of atomic structures (like fullerenes, glasses, and so on), defects in carbon nanotubes, and instability in one-dimensional conductors. Most of the famous discoveries of the last few decades like quantum Hall effect, Aharonov–Bohm's effect, and Josephson junctions have a common topological background.

In the third part we included two papers: T. Jannsen's "Phason Dynamics in aperiodic Crystals" and B. Zhilinskii's "Hamiltonian Monodromy as Lattice Defect." These authors studied the lattices of defects in different systems like quasicrystals and crystals. T. Jannsen has studied dynamical properties of quasicrystals. B. Zhilinskii has found interesting similarities between different problems like defects in periodic lattices and quantum spectra in quantum many-particle systems.

The paper by R. Mosseri "Two-Qubit and Three-Qubit Geometry and Hopf Fibrations" belongs to a new and fast-developing field of quantum information. It is amusing that such a keystone of topology as Hopf fibration appears in this fresh field of investigation.

The final part of the volume is the paper by I. Smalyukh and O.D. Lavrentovich "Defects and Three-Dimensional Director Fields in Cholesteric

Lamellae" considers defects in cholesteric liquid crystals. The authors give a detailed picture, including some experimental results for textures and defects in three-dimensional patterns. This field promises very interesting applications of three-dimensional geometry and topology.

1

Topology in the Electron Theory of Metals

A.M. Kosevich

Summary. Topological aspect of the dynamics of electrons in crystals (band electrons) is discussed. The main peculiarities of such electrons are connected with the form of their isoenergy surfaces, which is different from those of the free electrons. It is shown that the behavior of the band electrons in metals at low temperatures under the influence of external electric and magnetic fields depends strongly on the topology of the Fermi surfaces (the isoenergetic surfaces for the Fermi energy). Various examples of such a dependence are described.

1.1 Introduction

A quantitative description of the processes and physical phenomena under investigation is impossible without using various mathematical methods. To examine the physical processes closely, physicists apply more and more refined mathematical techniques. However there are branches of mathematics that help to understand not only the details of the physical phenomena but also some general regularity connecting physical behavior in a large number of different experiments. Topology is one such branch, which links physics as a rather old science and topology as a newer one. This is the subject of the present book.

Topological methods are especially useful when equations of physical fields have a complicated mathematical structure and do not allow to arrive at simple general solutions. A description of general topological conclusions in the theory of fields, phase transitions, and the superfluid phase of ^{3}He and nonlinear dynamics can be found in books [1, 2]. Concerning the physics of condensed matter, only two aspects are usually discussed: first, the topology of the order parameter in systems subjected to the phase transition, and second, the classification of possible forms of nonlinear objects like solitons, vortices, dislocations, and so on.

This book has the goal of attracting attention to the topology of characteristic lines and surfaces in the theory of crystals (metals first). Isoenergetic surfaces are well-known examples of such surfaces in the dynamics of electrons or other elementary excitations in the crystals. The Fermi surface is

often described as the isoenergetic surface in $\boldsymbol{k}$-space. The Fermi surfaces and the electron trajectories in the magnetic field are geometrical objects whose topological properties are discussed from various points of view. Many results discussed in this chapter were developed by Lifshitz and his scientific school, and this fact is reflected in the list of references.

1.2 Dynamics of Conductivity Electrons and the Fermi Surface

According to the Drude–Sommerfeld model [4] the charge carriers in metals represent the degenerated gas of free (noninteracting) electrons. The foundation of the understanding of electron properties of metals is the band theory and Fermi–Dirac statistics based on the assumption that the interaction among conductivity electrons is weak. A justification for such a model assumption is the fact that at low temperatures only a small portion of electrons with energies close to the Fermi energy takes part in the heat motion and kinetics.

Thus we agree to consider electrons with energies close to the Fermi energy as an ideal gas of the Fermi particles. The main dynamical variable of a free particle is its momentum $\boldsymbol{p}$. In the semiclassical approximation the spatial position of an electron is determined by its coordinate $\boldsymbol{x}$. The coordinate $\boldsymbol{x}$ and the momentum $\boldsymbol{p}$ are a pair of canonically conjugated variables giving an instantaneous state of a particle in the classical mechanics. In the absence of external fields, the electron energy ε depends only on $\boldsymbol{p}$ [$\varepsilon = \varepsilon(\boldsymbol{p})$], and this dependence is called the *dispersion relation* or dispersion law. For the free electron, $\varepsilon = p^2/(2m_0)$, where m_0 is the electron mass, and then we discuss the quadratic dispersion relation. The geometrical image of the dispersion relation is associated with a surface of equal energies (isoenergetic surface). The isoenergetic surface is a surface in the $\boldsymbol{p}$-space, defined with the condition:

$$\varepsilon(\boldsymbol{p}) = \varepsilon = \text{const}. \tag{1.1}$$

This is a sphere of the radius $p = \sqrt{2m_0}$ for the free electron. The Fermi surface is an isoenergetic surface for the Fermi energy ε_{F}. This is the sphere of the radius $p_{\mathrm{F}} = \sqrt{2m_0\varepsilon_{\mathrm{F}}}$ for the free electron gas. At $T = 0$ all the states inside the Fermi surface are occupied. Therefore the volume inside the Fermi surface and the Fermi energy are unambiguously connected with the fixed number of electrons:

$$N = 2\frac{V\Omega(\varepsilon_{\mathrm{F}})}{(2\pi\hbar)^3} = \frac{8\pi V p_{\mathrm{F}}^3}{3(2\pi\hbar)^3} = \frac{\pi(2m_0)^{3/2}}{3(2\pi\hbar)^3}\varepsilon_{\mathrm{F}}^{3/2}, \tag{1.2}$$

where $\Omega(\varepsilon_{\mathrm{F}})$ is the volume in the $\boldsymbol{p}$-space inside the Fermi sphere and the factor 2 is included to take into account two possible electron spin orientations. As a result

$$p_{\mathrm{F}}^3 = 3\pi^2\frac{N}{V}\hbar^3. \tag{1.3}$$

If one writes $N/V = 1/a^3$ using the average interelectron distance a, then (1.3) can be replaced by the following good approximation:

$$p_{\mathrm{F}} = \hbar(\pi/a). \tag{1.4}$$

Note that estimation (4) does not include any parameter from the dispersion relation. This is a very important observation for the question whether the calculation in (1.2), based on the concept of an ideal gas of the particles with the quadratic dispersion relation, can be used for the electrons in a crystal lattice. The momentum $\boldsymbol{p}$ is not a convenient dynamic variable in a periodic structure; instead we have the quasimomentum and the energy of an elementary excitation becomes a periodic function with the period of the reciprocal lattice. Therefore, the dispersion relation for the electron in a metal should be complicated and an anisotropic function of the quasimomentum, and the Fermi surface takes a shape different from the sphere. Nevertheless the characteristic radius of the Fermi surface (4) is estimated correctly. As a matter of fact, in calculating (1.2) we are interested only in the number of states occupied by the electrons and determined by the number of degrees of the freedom of all the electrons. When we distribute these states all over the phase space "cells," we fill in some phase space, the volume invariant with respect to the choice of the description of one-particle states.

We put aside the question of what determines the shape of the Fermi surface for the electrons in the crystal and instead discuss possible manifestations of the shape of isoenergetic surfaces in the electron dynamics. The electron motion in a magnetic field turns out to be most sensitive to the shape mentioned.

Consider the motion of an electron with the dispersion relation $\varepsilon = \varepsilon(\boldsymbol{p})$ in a homogeneous magnetic field in a homogeneous magnetic field $\boldsymbol{B}$. A pair of the electron dynamics has the form

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = \frac{e}{c}[\boldsymbol{v}\boldsymbol{B}], \qquad \boldsymbol{v} \equiv \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\partial\varepsilon}{\partial\boldsymbol{p}}. \tag{1.5}$$

If the magnetic field $\boldsymbol{B}$ is directed along the z-axis, the following set of equations of motion is arrived at

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = \frac{eB}{c}v_y, \qquad \frac{\mathrm{d}p_y}{\mathrm{d}t} = -\frac{eB}{c}v_x, \qquad \frac{\mathrm{d}p_z}{\mathrm{d}t} = 0. \tag{1.6}$$

Equation (1.6) possesses two integrals of motion,

$$\varepsilon(\boldsymbol{p}) = \text{const.} \quad p_z = \text{const.} \tag{1.7}$$

Equation (1.7) determines the electron trajectory in the $\boldsymbol{p}$-space. It is a curve of the cut of the surface $\varepsilon(\boldsymbol{p}) = \text{const.}$ by the plane $p_z = \text{const.}$ (Fig. 1.1). As follows from (1.6) the projection of the electron trajectory in the $\boldsymbol{x}$-space is

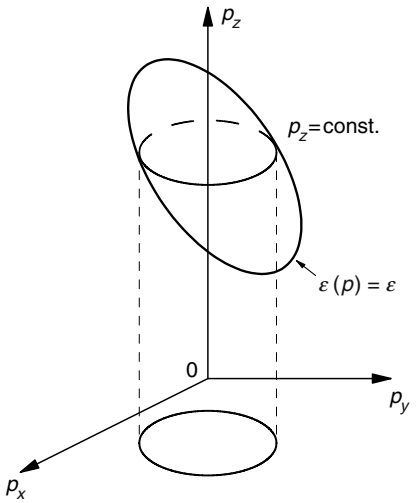

Fig. 1.1. The curve of the cut of the surface $\varepsilon(\boldsymbol{p}) = \text{constant}$ by the plane $p_z = \text{constant}$

similar to the trajectory in (1.7). To see the latter, we rewrite (1.7) as:

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = \frac{eB}{c}\frac{\mathrm{d}y}{\mathrm{d}t}, \qquad \frac{\mathrm{d}p_y}{\mathrm{d}t} = -\frac{eB}{c}\frac{\mathrm{d}x}{\mathrm{d}t}. \tag{1.8}$$

Actually the projection of the trajectory in the $\boldsymbol{x}$-space onto the plane xOy is similar to the trajectory in the $\boldsymbol{p}$-space, but turned 90° with respect to the coordinate axes Ox and Oy (Fig. 1.2). The coordinate directions of the velocity vector are shown in Fig. 1.2 as well.

Analysis of (1.6) and Fig. 1.2 leads us to the conclusion that the electron accomplishes a cyclic motion in the magnetic field rotating along a closed

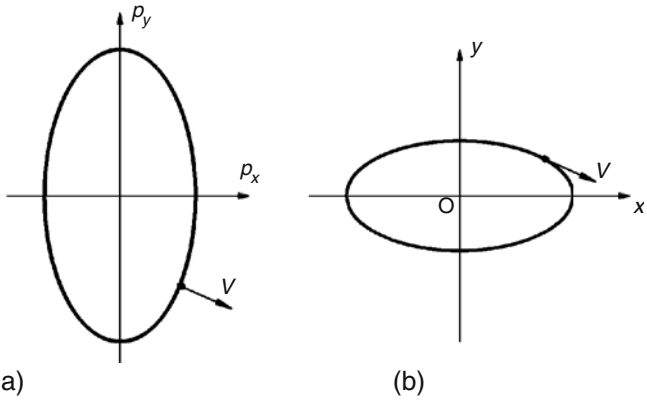

Fig. 1.2. The projection of the trajectory in the $\boldsymbol{p}$-space (**a**) and $\boldsymbol{x}$-space (**b**)

trajectory (it is the motion along the cycle for the free electron). It is easy to calculate the frequency of such a motion. Using (1.6) one can obtain the velocity of the electron along the trajectory in the $\boldsymbol{p}$-space

$$\frac{\mathrm{d}p_{\mathrm{l}}}{\mathrm{d}t} = \frac{eB}{c} v_{\perp}, \tag{1.9}$$

where $v_{\perp}$ is the projection of the electron velocity onto the plane perpendicular to the magnetic field $\boldsymbol{B}$. Equation (9) creates the following chain of equalities

$$\mathrm{d}t = \frac{c}{eB}\frac{\mathrm{d}p_{\mathrm{l}}}{v_{\perp}}, \qquad t = \frac{c}{eB}\int \frac{\mathrm{d}p_{\mathrm{l}}}{v_{\perp}}, \qquad \Delta t = \frac{c}{eB}\oint \frac{\mathrm{d}p_{\mathrm{l}}}{v_{\perp}}, \tag{1.10}$$

where the integrals are calculated along the electron trajectory, Δt is the period of cyclic electron motion, and the last integral is calculated along the total closed trajectory. To know Δt, one can determine the cyclic frequency of the electron as $\omega_{\mathrm{c}} = 2\pi/\Delta t$. This is the frequency of the cyclotron resonance in the magnetic field for the electron in a crystal. However such a frequency is usually presented in the form $\omega_{\mathrm{c}} = eB/mc$, where m is the electron mass. Therefore the electron cyclotron mass in the crystal depends on its trajectory and is a function of the electron state, but not a constant quantity. Let us find the geometric meaning of this electron characteristic. Let us calculate the change of the area on the plane $p_z = \text{const}$. surrounded with the electron trajectory when the electron energy increases by $\delta\varepsilon$:

$$\delta S = \oint \delta p_{\perp} \mathrm{d}p_{\mathrm{l}} = \oint \frac{\mathrm{d}p_{\mathrm{l}}}{v_{\perp}} \delta\varepsilon. \tag{1.11}$$

The integral is calculated along the closed trajectory. Comparing Δt in (1.10) with (1.11) one obtains

$$\Delta t = \frac{c}{eB}\frac{\partial S}{\partial \varepsilon}, \tag{1.12}$$

where the function $S = S(\varepsilon, p_z)$ is the above-mentioned area on the plane $p_z = \text{const}$. Thus the effective cyclotron mass of the electron is given by expression [3]

$$m_c = m_c(\varepsilon, p_z) = \frac{1}{2\pi}\frac{\partial S}{\partial \varepsilon}. \tag{1.13}$$

Equation (13) for the electrons in a metal changes to $\varepsilon = \varepsilon_{\mathrm{F}}$ on the Fermi surface. This means that the area of the cross-section of the Fermi surface with the plane $p_z = \text{const}$. has a direct physical sense and its derivative (13) is the quantity measured experimentally.

1.3 Geometry of the Fermi Surface in Crystal

The fact that the shape of an isoenergetic surface for the electron in metals differs from the spherical shape was understood simultaneous with the formulation of the principal statements of the band theory. In the monograph by Sommerfeld and Bethe [4] published in 1933, it has already been shown that the Fermi surface can pass through the entire reciprocal lattice intersecting the boundaries of the first Brillouin zone. Though there exists a great number of shapes of the Fermi surface (some of them are described in book [3]), one could not predict a shape of the Fermi surface in the metal considered. Harrison [5] was the first to propose a simple explanation of two aspects: (1) the appearance of the expected shape of the Fermi surface, and (2) the possibility of the coexistence of several Fermi surfaces in one metal (to be precise, several cavities or sheets of one not singly connected Fermi surface).

To illustrate Harrison's method, we consider a simple cubic lattice with the lattice constant a and use the cross-section of the reciprocal lattice by the plane $p_z = 0$. The center of the elementary cell coincides usually with the point $\boldsymbol{p} = 0$ (the main square in Fig. 1.3a and the square in Fig. 1.3b). However it can be located in another point (the dashed square in Fig. 1.3c). Start from the free-electron model when the Fermi surface is a sphere of the radius (4). The Fermi momentum has a magnitude of the order of the size of the Brillouin zone if the number of the free electrons equals the number of atoms. Draw a circle of such a radius around the point $\boldsymbol{p} = 0$. This is the cross-section of the Fermi surface. Taking into account that the electron energy is a periodic function in the $\boldsymbol{p}$-space, one ought to draw such circles around each of translation vectors of the reciprocal lattice. The Fermi surfaces intersect the boundaries of the Brillouin zones and each other. Every point of the intersection of the circles in Fig. 1.3a is a point of a degeneration. It is understood that the degeneration points appeared by the force of the primitive model used for consideration, and the circle form of the isoenergetic surfaces should be considered as a result of the zero approximation. Taking into account any physical circumstance in the next approximation, one removes the degeneration, and the intercrosses of the Fermi surfaces disappear. Instead several closed lines on the plane of the reciprocal lattice appear. It is convenient to situate a part of them in the elementary cells either of the first or the second type. Usually one refers them to as different (the first, second, third, and so on) Brillouin zones and considers them as various cavities or sheets of one Fermi surface.

A set of the Fermi surfaces in a strongly anisotropic crystal can differ from that drawn in Fig. 1.3a. In such a case new types of lines can appear; these are the so-called *open Fermi surfaces* (Fig. 1.4). Thus there are two topologically different types of the Fermi surfaces, namely, closed and open surfaces.

If the plane in Fig. 1.4 is perpendicular to the external magnetic field, the two types of sections of the Fermi surfaces are associated with the two types of electron trajectories in the magnetic field: closed (typical for the free

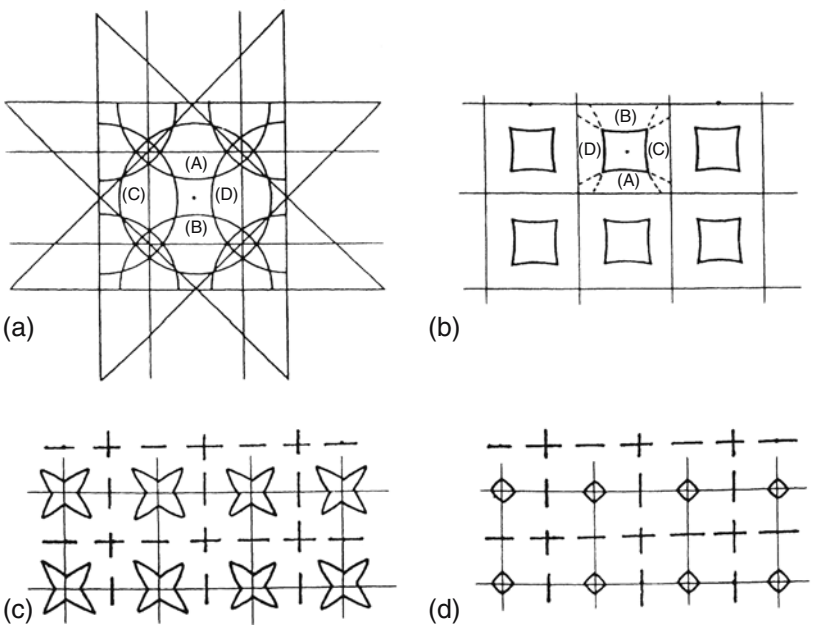

Fig. 1.3. The cross-sections of the Fermi surfaces

electron) and open (passing through the whole $\boldsymbol{p}$-space). The latter possibility means that the electron makes an infinite motion in the usual $\boldsymbol{x}$-space and can go on to infinity. It is clear that such a situation is possible only for the band electron in a crystal and manifests peculiarities of the electron dynamics caused by the topology of the Fermi surface.

Naturally only closed cross-sections can appear on the closed Fermi surface and therefore the electrons always move along cyclic trajectories in the

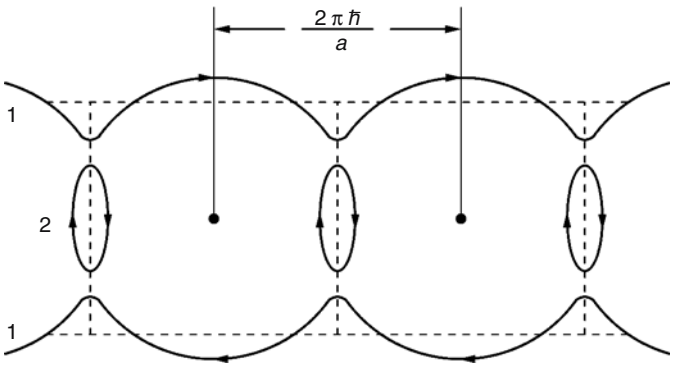

Fig. 1.4. Open Fermi surface

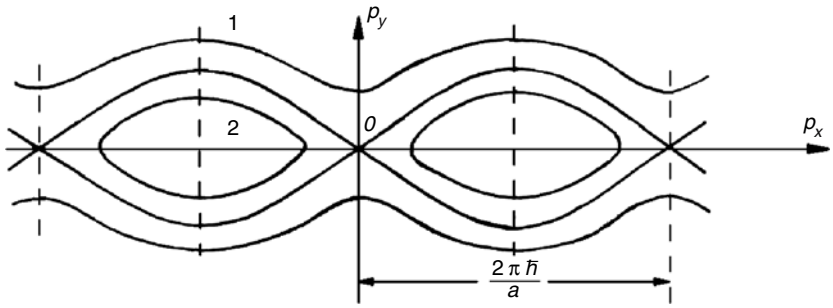

Fig. 1.5. Different cross-sections for the Fermi surface in the form of a corrugated cylinder

magnetic field. The open Fermi surface can create both closed and open sections. Different cross-sections are shown in Fig. 1.5 for the Fermi surface in the form of a corrugated cylinder. The direction of the "openness" of the cylinder is chosen along the axis p_x and the magnetic field is directed along the axis z. Lines of the two types determine the trajectories of the electron with different magnitudes of p_z. The trajectory lines of the first type are the open trajectories corresponding to the infinite motion of the electron. The lines of type 2 are the closed trajectories. Two types of the trajectories are separated with the open trajectory (*separatrice*) passing through a saddle point on the Fermi surface. The separatrice is characterized by the self-intersection points on the boundaries of the Brillouin zones. These points are singular because the group velocity in them changes into 0. The electron ought to stop at such a point, but it approaches the saddle point only asymptotically (at $t \to \infty$). The stop at the saddle point means that the electron mass turns into infinity. Thus the trajectory passing through the saddle point possesses a singular mass.

It has been shown earlier that the effective electron mass in the magnetic field is determined by (1.13) where S is the area of the cross-section of an isoenergetic surface considered as a function of ε. If one shifts the Fermi energy ε_{F} to the energy $\varepsilon = \varepsilon_{\mathrm{F}} + \Delta\varepsilon$ exceeding the first on the magnitude $\Delta\varepsilon$, the singular trajectory is replaced by a closed trajectory near it. The trajectories of different types near the saddle point are shown in Fig. 1.6a. The peculiarity of the trajectory of the second type is entirely connected with the existence of the saddle point near it, and therefore such a peculiarity is determined by the energy dependence of the part of the area ΔS lying in the vicinity of the saddle point. In the main approximation, the part of the trajectory under discussion can be considered as an hyperbola:

$$\frac{p_x^2}{2m_1} - \frac{p_y^2}{2m_2} = \Delta\varepsilon, \tag{1.14}$$

where the values $m_1, m_2 > 0$ characterize the curvature of the Fermi surface.

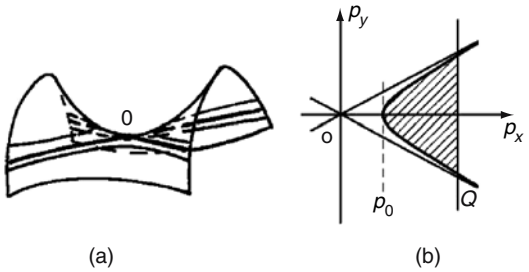

Fig. 1.6. The different trajectories near the saddle point (**a**); the vicinity with the hyperbola asymptotes (**b**)

It is clear that the trajectory passing through the saddle point can be replaced in this vicinity with the hyperbola asymptotes. The magnitude ΔS is calculated as the area bounded with the hyperbola and the straight line $p_z = Q = \text{const.}$, where Q is chosen on the small (but finite) distance from the saddle point (see Fig. 1.6b). Then

$$\Delta S = 2m_2 \int_{p_0}^{Q} \left(\frac{p_x^2}{2m_1} - \Delta\varepsilon \right)^2 \mathrm{d}p_x = 4\sqrt{m_1 m_2}\Delta\varepsilon \int_1^{x(\Delta\varepsilon)} \sqrt{x^2 - 1}\, \mathrm{d}x, \tag{1.15}$$

where $p_0 = \sqrt{2m_1\Delta\varepsilon}$ and $x(\Delta\varepsilon) = Q/\sqrt{2m_1\Delta\varepsilon}$. In the main approximation, the cyclotron mass of the electron on the trajectory close to the singular one is equal to

$$m = \frac{(m_1 m_2)^{1/2}}{\pi} \ln \frac{\varepsilon(Q)}{\Delta\varepsilon}, \tag{1.16}$$

where $\varepsilon(Q) = Q^2/(2m_1)$. Thus when the electron approaches the saddle point its effective mass grows logarithmically. Hence the electron cyclotron period grows to infinity, and the electron motion along the singular trajectory becomes similar to the motion along an open trajectory in the sense that the electron spends infinite time passing the trajectory.

1.4 Quantum Magnetic Oscillations and the Shape of the Fermi Surface

The problem of the shape of the Fermi surface concerns first of all with the de Haas–van Alphen effect. The de Haas–van Alphen effect is one of the most interesting macroscopic quantum phenomena. Its nature consists in an oscillatory dependence of the magnetization of a metal in the magnetic field. The effect was discovered in 1930 by Leiden physicists after whom it has been called so. Detailed description of the history and investigations of the effect over several decades, and the estimation of the role of the effect in the experimental study of the electron spectrum of metals can be found in Shoenberg's monograph [6]. The most important point for us is the fact that this is a purely

quantum phenomenon caused by the quantization of the electron motion in the magnetic field.

Let us perform a semiclassical quantization of the electron motion along the closed trajectory. It is known that the magnetic flux Φ through the area of the closed electron orbit is quantized as:

$$\Phi = \phi_0 n; \qquad n = 0, 1, 2, ..., \qquad \phi_0 = \frac{2\pi c\hbar}{e}, \tag{1.17}$$

where ϕ_0 is the magnetic flux quant.

If A is the area of the projection of the closed electron orbit on the plane xOy perpendicular to the magnetic field, one can write $\Phi = BA$. According to (1.8) the area A is proportional to the area of the cross-section of the isoenergetic surface $\varepsilon(\boldsymbol{p}) = \varepsilon = \text{const.}$ by the plane $p_z = \text{const.}$:

$$A = \left(\frac{c}{eB}\right)^2 S(\varepsilon, p_z).$$

Then the following rule of the quantization appears

$$S(\varepsilon, p_z) = \frac{2\pi\hbar eH}{c}(n + \gamma), \qquad n = 0, 1, 2, ..., \tag{1.18}$$

where γ is a parameter of the order of magnitude of the unit. Onsager [7] was the first to obtain the quantization rule in such a form. Formula (18) gives a dependence of the electron energy on the quantum quantities n and p_z.

Special consideration is needed in the case of quantization of the electron motion along the trajectories close to the separatrice in Fig. 1.4, which is actually close to the trajectory with a cross-section. If the magnetic field is slightly inclined to the z-axis in Fig. 1.4, the separatrice can transform into a closed line separating small closed trajectories of different sizes. Similar cross-sections of the Fermi surface are shown as diagonal lines in Fig. 1.7. The semiclassical quantization in such a case was studied by Azbel [8]. If trajectories 1 and 2 in Fig. 1.7 are close to each other, then the effective value γ depends strongly on the number n, and near the points of self-intersection the distance between the energy levels oscillates in the magnetic field.

Using (1.18), Lifshitz and Kosevich [9] calculated magnetization at low temperatures of the electron gas with arbitrary dispersion relation $\varepsilon(\boldsymbol{p})$. They obtained the oscillatory part of the magnetization of the metal and their result can be presented schematically in the form:

$$\Delta M_{\text{osc}} = M_0(B, T) \cos\left(\frac{c\hbar S_m(\varepsilon_{\text{F}})}{eB} - \gamma\right) \cos\left(\frac{\pi g}{2}\frac{m_{\text{c}}}{m_0}\right), \tag{1.19}$$

where M_0 is the oscillation amplitude depending smoothly on B, $S_m(\varepsilon_{\text{F}})$ is the area of the extremal cross-section of the Fermi surface by the plane $p_z = \text{const.}$, m_{c} is the effective electron mass in the metal, m_0 is the mass of the free electron, and g is the gyromagnetic relation determining the electron spin magneton ($g = 2$ for free electrons).

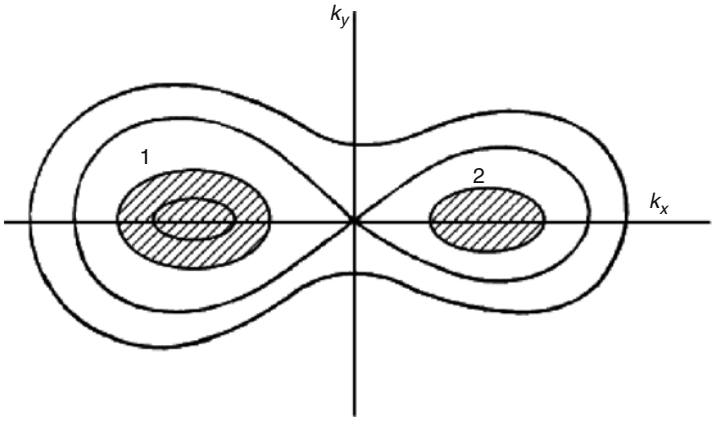

Fig. 1.7. Cross-section of the Fermi surface

The main characteristic of the de Haas–van Alphen effect is the period of the oscillations. The period in the reciprocal magnetic field is equal to

$$\Delta\left(\frac{1}{B}\right) = \frac{2\pi \mid e \mid \hbar}{cS_m}. \tag{1.20}$$

This period is independent of the magnetic field and temperature, and in the direct magnetic field it equals

$$\Delta(B) = \frac{2\pi e\hbar}{cS_m} B^2 \tag{1.21}$$

and is proportional to the squared magnetic field (Fig. 1.8).

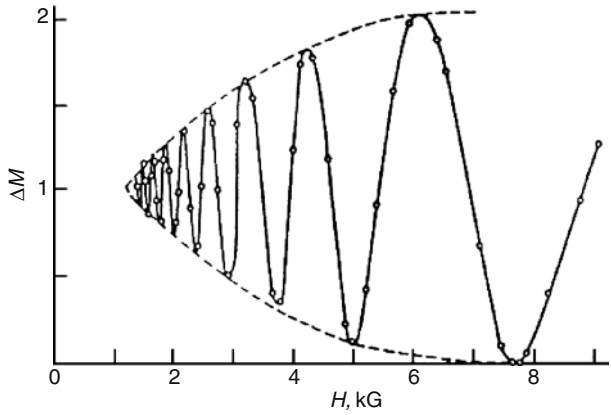

Fig. 1.8. The main characteristics of the de Haas–van Alphen effect, see (1.21)

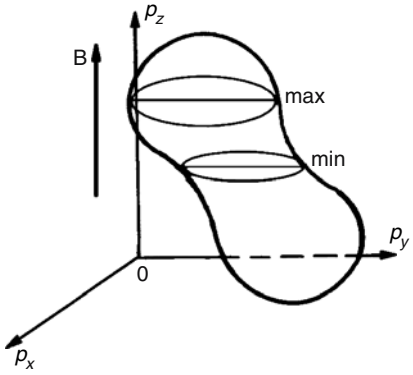

Fig. 1.9. The convex and nonconvex cross-sections of the Fermi surface

If the Fermi surface is convex there is only one extremal (maximal) cross-section, while there are several extremal cross-sections if it is nonconvex (Fig. 1.9). In the latter case the diagram of the dependence of $M_{\rm osc}$ on the magnetic field has the form of a sum of several oscillatory curves.

A number of cross-sections appear when the Fermi surface has the form of a corrugated cylinder. If the magnetic field $\boldsymbol{B}$ is directed along the cylinder axis, there are two types of extremal cross-sections (maximal and minimal). If the vector $\boldsymbol{B}$ forms an angle θ with respect to the direction of the cylinder axis (Fig. 1.10), there is a set of cross-sections depending continuously on the angle θ. It is easy to understand that the areas of the extremal cross-sections increase proportional to $\tan\theta$ $(0 < \theta < \pi/2)$, and this leads to the inversion dependence of the oscillation periods (20) on $\tan\theta$. At $\theta = \pi/2$ the cross-section passing through the entire $\boldsymbol{p}$-space appears and its contour is an open trajectory of type 1 in Fig. 1.5. Then the corresponding quantum oscillations disappear.

The temperature dependence of the oscillation amplitude has a simple form at not very low temperatures $T > \hbar\omega_{\rm c}$ when

$$M_0 \sim \exp\left(-\frac{2\pi^2 T}{\hbar\omega_{\rm c}}\right), \tag{1.22}$$

where $\omega_{\rm c} = eB/(m_{\rm c}c)$ is the cyclotron frequency. As a result, the three factors in (1.19) give a possibility of measuring experimentally the three most important characteristics of the electron system: (a) the oscillation periods allow us to obtain the extremal cross-sections of the Fermi surface at different directions of the magnetic field, (b) the temperature dependence (22) gives a possibility to determine the effective electron masses on the extremal orbits, and last (c) the third factor in (1.19) allows us to estimate the

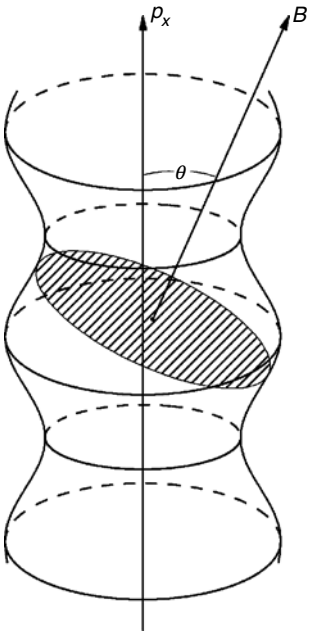

Fig. 1.10. The vector $\boldsymbol{B}$ forms an angle θ with respect to the direction of the cylinder axis

gyromagnetic relation for the electron in a metal. All the mentioned possibilities have been used with great success in many experiments, and formula (19) has been highly estimated by investigators of the electron properties of metals. Shoenberg called (1.19) the Lifshitz–Kosevich formula (LK formula), and this can be found in many papers and reviews, including the book [10].

From the geometrical point of view, point (a) is most interesting since it allows us to formulate a rigorous problem of reconstruction of the Fermi surface from experimental data for the oscillation periods measured for various directions of the magnetic field. Such a question was formulated as a mathematical inversion problem and was solved by Lifshitz and Pogorelov [11]. It was proved that knowledge of all central cross-sections of the convex Fermi surface defines the shape of the Fermi surface uniquely.

Besides the de Haas–van Alphen effect, there are other physical phenomena for the quantum magnetic oscillations and the Shubnikov–de Haas effect stands first among them. The Shubnikov–de Haas effect is the oscillatory dependence of the magnetoresistance of a metal in the magnetic field. The physical nature of the Shubnikov–de Haas oscillations is the same as in the case of de Haas–van Alphen effect, and corresponding measurements give the same periods of the oscillations [12]. It turns out that the magnetoresistance oscillations can be observed in organic conductors as well.

1.5 Magnetic Breakdown

When discussing the periods of the quantum magnetic oscillations in the case of the complicated Fermi surface, we assumed the electron motion along any trajectory independent of the fact that the other trajectories could lie close to the one under consideration. However, if the Fermi surface is nonconvex, there is always a whole domain of magnetic field orientations for which the semiclassical trajectories turn out to be close to each other. Then the quantum tunneling between the close trajectories becomes possible. The phenomenon caused by such a tunneling in the strong magnetic field is called *magnetic breakdown* [13].

The semiclassical nature of the magnetic breakdown can be easily explained using Figs. 1.4 and 1.11. Let the electron move along the small closed orbit of type 2 in Fig. 1.4 near the open orbit of type 1 at times. If the magnetic field exceeds some limiting magnitude B_0, the electron first "jumps" onto trajectory 1 and then on to the second orbit, and as a result moves along the closed trajectory of a large radius (both the small and the large closed trajectories are shown in Fig. 1.11).

It is possible to give another interpretation of this phenomenon. The electron that moves initially along the open trajectory of type 1 in Fig. 1.4 begins to move along the closed trajectory under the influence of a strong magnetic field. Therefore, the strong magnetic field can change both the sizes and the topology of the electron trajectories.

It is clear that the magnetic breakdown is a physical phenomenon deviating from the framework of the classical dynamics of the particle with a definite trajectory. Transitions among the various trajectories are characterized by some quantum probability. And the magnetic breakdown is a rather complicated quantum process, which can make a reconstruction of the electron spectrum of a metal. Detailed description of the magnetic breakdown can be found in book [3] and review [14].

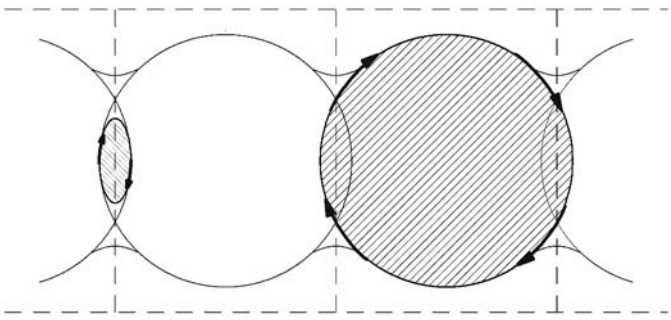

Fig. 1.11. The magnetic breakdown of the Fermi surface

Concerning oscillatory effects, the magnetic breakdown manifests itself either in jumps of the periods of oscillations, when the magnetic field grows up to the limiting breakdown magnitude and the electrons change the close orbits jumping from one to another, or in a disappearance of the oscillations, when the latter orbit has such a great area that the periods and amplitudes become too small for the experimental observation.

1.6 Band Electrons in the Electric Field and Bloch Oscillations

A number of shapes of the Fermi surfaces are caused by the periodic dependence of the electron energy on the momentum. Such a dependence leads to some peculiarities of the electron dynamics in $\boldsymbol{E}$.

In semiclassical approximation, the electron dynamics is governed by the equation:

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = e\boldsymbol{E} \tag{1.23}$$

with the usual definition of the velocity

$$\boldsymbol{v} = \frac{\partial \varepsilon}{\partial \boldsymbol{p}}, \tag{1.24}$$

where $\varepsilon = \varepsilon(\boldsymbol{p}) = \varepsilon(\boldsymbol{p} + \boldsymbol{G})$, and $\mathbf{G}$ is the vector of the reciprocal lattice in the $\boldsymbol{p}$-space.

Assume that the vector $\boldsymbol{E}$ is directed along the z-axis and restrict ourselves to the simplest relation $\varepsilon = \varepsilon(p_x)$:

$$\varepsilon = \varepsilon_0 \sin\frac{ap_x}{\hbar}, \qquad v = \frac{a\varepsilon_0}{\hbar}\cos\frac{ap_x}{\hbar}. \tag{1.25}$$

Then (1.23) gives $p_x = eEt$ and determines the periodic dependence of the electron velocity on time [15]:

$$v_x = v_0(t) \equiv \frac{a\varepsilon_0}{\hbar}\cos(\omega_{\mathrm{B}} t). \tag{1.26}$$

The frequency $\omega_{\mathrm{B}} = eEa/\hbar$ is called *Bloch frequency.*

At reasonable values of the electric field, the frequency of the Bloch oscillations of the electron in a metal is many orders of magnitude lower than the collision frequency of the electron even in extremely pure metals (in other words, the oscillation period is much greater than the relaxation time τ in the metal, and the amplitude of the Bloch oscillations is much greater than the electron mean free path). Therefore, in calculating the resistivity of conductors and in other similar cases, the periodic character of the electron motion need not be taken into account, and the electron motion over short path lengths can be assumed translational. It was assumed for a long time that the Bloch

oscillations are an extremely curious physical phenomenon but are of interest only from a theoretical standpoint.

The situation was fundamentally altered by the development of technology for fabricating extremely perfect semiconductor superlattices with structural periods much greater than the lattice constant. In such structures, the period of the reciprocal lattice is greatly diminished, and the electron energy spectrum separates into narrow sub-bands, with the result that the Bloch oscillations corresponding to them have rather high frequencies (the obvious condition $\omega\tau \gg 1$ becomes attainable); thus it becomes realistic to produce such oscillations and observe them experimentally.

Esaki and Tsu [16] were the first who attracted the attention of scientists to the fact that the periodic dependence of the electron energy on the quasimomentum could be observed in the semiconductor superlattice. Their idea can be easily formulated in the τ-approximation. If the relaxation is not taken into consideration, the dependence of the electron velocity on time is determined with (1.26) where a is the period of the superlattice now. If the relaxation is taken into account, the real change of the electron velocity can be written as

$$\mathrm{d}v_x = \exp(-t/\tau)\mathrm{d}v_0. \tag{1.27}$$

It follows from (1.27) that

$$v_x(t) = \int \exp(-t/\tau)\mathrm{d}v_0(t) = \int_0^t \frac{\mathrm{d}v_0}{\mathrm{d}t}\exp(-t'/\tau)\mathrm{d}t'. \tag{1.28}$$

Thus

$$v_x(t) = eE\int_0^t \frac{\partial^2\varepsilon}{\partial p_x^2}\exp(-t'/\tau)\mathrm{d}t' = \frac{eE}{m(0)}\int_0^t \cos(\omega_\mathrm{B} z)\exp(-z'/\tau)\mathrm{d}z', \tag{1.29}$$

where $m(0)^{-1} = \partial^2\varepsilon/\partial p_x^2$ is the reciprocal effective mass of the electron at $p_x = 0$.

For a large time ($t \gg \tau$), one can put $t = \infty$ at the upper limit of the integral in (1.29). Then the well-known Esaki and Tsu formula appears for the steady average velocity of the electron in the semiconductor superlattice as:

$$\langle v\rangle = v_x(\infty) = \frac{\omega_\mathrm{B}\tau}{1+(\omega_\mathrm{B}\tau)^2}\frac{a\varepsilon_0}{\hbar}. \tag{1.30}$$

In the weak fields (under the condition $\omega_\mathrm{B}\tau \ll 1$), the standard linear dependence of the average electron velocity on the electric field appears as

$$\langle v\rangle = \frac{e\tau}{m(0)}E. \tag{1.31}$$

Equation (1.31) explains the conductivity in the weak electric fields. In the strong electric fields ($\omega_\mathrm{B}\tau \gg 1$), which are possible in semiconductor materials, (1.30) gives the dependence of the average velocity on E, which is not

possible in metals: the average electron velocity decreases when the electric field increases:

$$\langle v \rangle = \frac{\varepsilon_0}{e\tau} \frac{1}{E}. \tag{1.32}$$

Equation (32) leads to a negative differential electroconductivity of the semiconductor superlattice.

The electrodynamics of semiconductor superlattices turns out now into an independent part of physics [17] and its general contents concern the topic of our discussion only slightly.

1.7 Topology of the Fermi Surfaces and Low-Temperature Magnetoresistivity of Metals

The quantum oscillations in metals are undoubtedly a very good demonstration of the quantum nature of magnetism, and their various manifestations confirm the complicated forms of the Fermi surfaces. However, only the closed electron trajectories contribute to the measured quantum oscillations. Dynamics of electrons moving along the open trajectories are studied by investigation of other phenomena. Among them, low-temperature galvano-magnetic phenomena occupy a notable place. The galvano-magnetic phenomena are very sensitive to the form of the electron energy spectrum and have been used as a simple and reliable spectroscopic method for the reconstruction of the topology of the Fermi surfaces. In recent years, the topological aspect of the galvano-magnetic phenomena has been subjected to thorough mathematical examination [18].

However, we are interested not in the details of the mathematical analysis but in the physical results connecting experimental observations with the geometry of the Fermi surfaces. The excellent review [19] exhausts the topic of the present section, and hence only general qualitative remarks are made here. It is only worthwhile to note that both the review [19] and the book [3] give a total account of the theory of the low-temperature galvano-magnetic properties of metals in strong magnetic fields when the topological peculiarities of the Fermi surfaces are displayed fully. This theory has been called *LAK theory* after the first letters of the names of its creators, Lifshitz, Azbel, and Kaganov.

If the magnetic field is strong enough, the cyclotron frequency is so large that at low temperatures (when the mean free path is rather long) the electron passes many times around any closed trajectory inside the Brillouin zone and, consequently, can intersect many elementary cells in the $\boldsymbol{p}$-space moving along the open trajectory. The latter explains the possibility of the manifestation of open trajectories in macroscopic properties of metals.

The electric field in a metal is very weak, and the dynamics of electrons is determined in the main approximation by the shape of its trajectory in the magnetic field. The electric field can be taken into account as a small perturbation. A detailed description of different situations and various possibilities

associated with the electron dynamics on the open Fermi surfaces is given in paper [20] and analyzed later in Abrikosov's monograph [22].

Let us examine the galvano-magnetic phenomena in a metal with the Fermi surface in the form of a corrugated cylinder with the direction of "openness" along the axis p_x in the magnetic field $\boldsymbol{B}=\boldsymbol{B}(\cos\theta, 0, \sin\theta)$ (see Fig. 1.10 or Fig. 1.12). A model dispersion relation corresponding to such a Fermi surface has the form:

$$\varepsilon(\boldsymbol{p}) = \Delta \sin^2 \frac{p_x b}{2} + \frac{p_y^2 + p_z^2}{2m} \tag{1.33}$$

supposing $\varepsilon_{\mathrm{F}} > \Delta$. The direction of the magnetic field is close to the z-axis.

The cross-sections of a weakly corrugated cylinder by the planes $p_{\mathrm{B}} = \boldsymbol{pB}/B = \text{const.}$ are shown in Fig. 1.12. When θ is other than $\pi/2$ all the cross-sections are closed. As θ approaches $\pi/2$ the closed orbits become strongly elongated orbits and their lengths can exceed the period of the reciprocal lattice $2\pi/b$. The period of motion along such an orbit increases infinitely

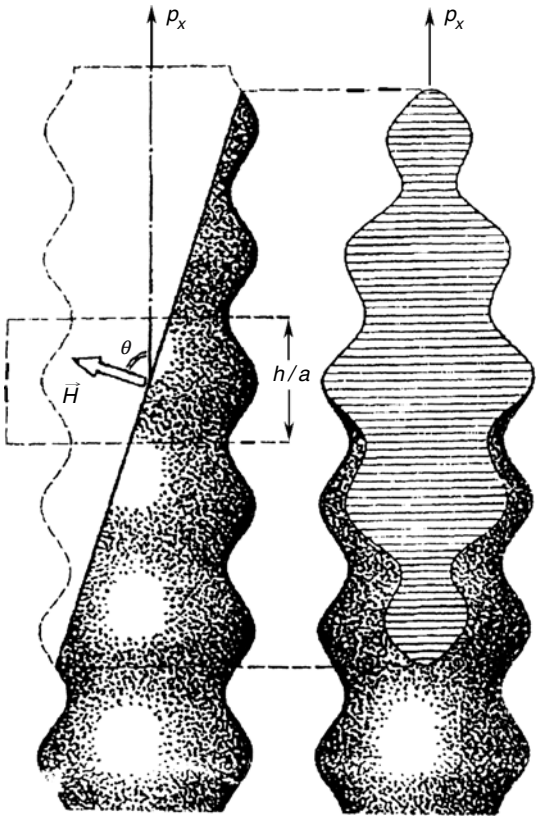

Fig. 1.12. Fermi surface in the form of a corrugated cylinder

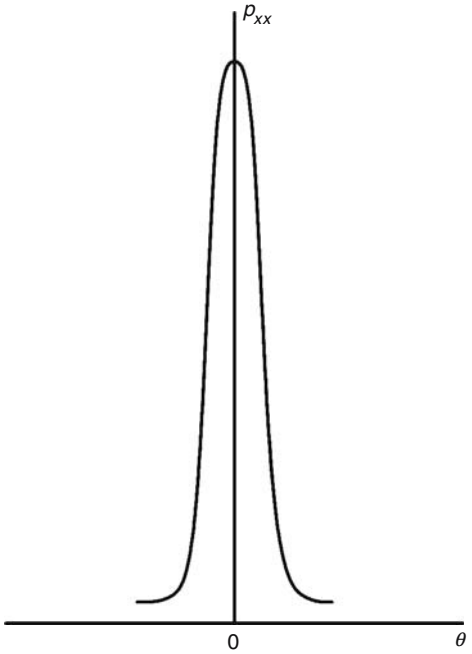

Fig. 1.13. Dependence of the magnetoresistivity on the orientation of magnetic field

when $\theta \to \pi/2$, and if $\theta = \pi/2$ the strongly elongated orbit "breaks" into two open orbits.

If the length of a long closed trajectory in the $\boldsymbol{x}$-space exceeds the mean free pass of the electron, such a trajectory contributes to kinetics as an open one. It is easy to estimate a limiting angle θ when the value $2\pi\hbar c/aeB\theta$ becomes equal to the mean free pass. Inside the interval of angles estimated by such a way, the elongated orbits belong to electrons that cannot realize their cyclotron motion. These electrons have $\langle v_x \rangle = 0$ and $\langle v_y \rangle \neq 0$. Thus a narrow maximum has to be expected for the dependence of the magnetoresistivity on the value and the orientation of the strong magnetic field, when $\boldsymbol{B}$ is perpendicular to the axis of the cylinder (Fig. 1.13). If the Fermi surface is more complicated than the corrugated cylinder (for example, similar to those shown in Fig. 1.14), then orientations of topological peculiarities can be presented using the stereographic projection of the magnetic field directions for which open trajectories are possible (Fig. 1.15). The center of the circle corresponds to the normal in Fig. 1.14, while the boundary of the circle to $\theta = \pi/2$. The hatched regions show where open trajectories can exist. The straight lines from the center to the boundary and the circle $\theta = \pi/2$ correspond to the open trajectories. The dashed lines close to the boundary circle separate regions where closed orbits are strongly elongated ones.

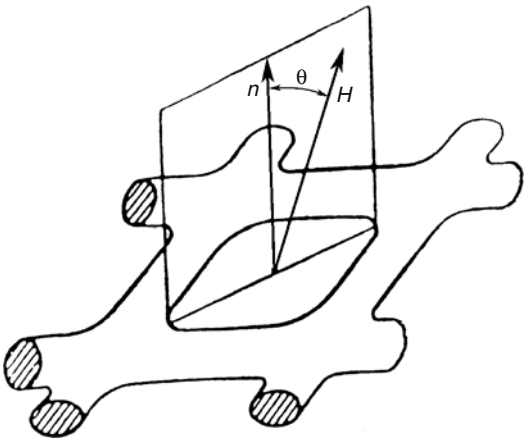

Fig. 1.14. Fermi surface

Additional information on details of the Fermi surface and effective masses of the electrons can be collected by analyzing the Azbel–Kaner cyclotron resonance [23, 24]. However, this effect is connected with high-frequency properties of metals and can be described in a separate publication. Despite understanding the great importance of the Azbel–Kaner effect in the electron theory of metals, only a mention of the theory is made here.

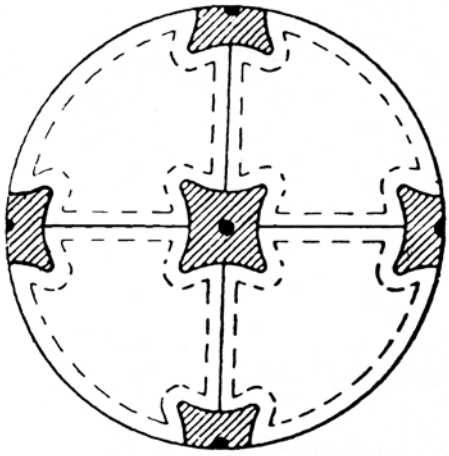

Fig. 1.15. The stereographic projection of the magnetic field directions

1.8 Berry's Phase and the Topology of the Electron Trajectories in the Magnetic Field

In the final part of the discussion of problems connected with the topology of electron orbits in the magnetic field, let us discuss the pure quantum effect of a geometrical Berry's phase [25] the question concerns the parameter γ in (1.18) influencing the phase of the oscillatory part of the magnetization (see (1.19)). A calculation of the parameter γ by the WKB method was made in papers [26] for the first time. Later it was performed strictly in [27] for the one-band electron spectrum. Paper [27] confirmed a specific contribution into γ of the electron trajectories with a self-intersection (like the separatrice in Fig. 1.5) described earlier by Azbel [8].

We describe in a few lines the physical circumstances that need to be taken into account when discussing Berry's phase. It is known that while a local value of a phase of the quantum wave function is not measured, only its gradients and a time derivative can be measured. Write the phase of the wave function in the form

$$\psi(\boldsymbol{x}, t) = |\psi| \exp(\mathrm{i}\varphi(\boldsymbol{x}, t)).$$

Suppose that $\psi(\boldsymbol{x}, t)$ is a single-valued function of coordinates and time. Though the phase $\varphi(\boldsymbol{x}, t)$ is not to be single-valued, its change along any closed circuit must be multiples of 2π. In the case of the electron motion in the magnetic field directed along the z-axis, the dependence of the phase on xy is important:

$$\varphi = \frac{p_z - \varepsilon t}{\hbar} + \varphi_0(x, y),$$

where $p_z = \text{const.}$, $\varepsilon = \text{const}$. The condition mentioned here has the form:

$$\oint \frac{\partial \varphi_0}{\partial x_\alpha} \mathrm{d}x_\alpha = 2\pi n, \qquad n = 0, 1, 2, ..., \quad (\alpha = 1, 2), \tag{1.34}$$

where the integration is performed along any closed circuit on the plane (x, y).

The phase of the semiclassical wave function is determined with a classical action S for the system under consideration $\varphi = (1/\hbar)S$. The action in the magnetic field acquires the addition

$$\delta S_0 = \frac{e}{c} \int_0^{\boldsymbol{x}} \boldsymbol{A} \, \mathrm{d}\mathbf{x}, \tag{1.35}$$

where $\boldsymbol{A}$ is the vector-potential. Consequently a closed electron trajectory Γ has to satisfy condition (34):

$$\frac{e}{\hbar c} \oint_\Gamma \boldsymbol{A} \, \mathrm{d}\boldsymbol{l} = 2\pi n, \qquad n = 0, 1, 2, ... \tag{1.36}$$

Equation (36) is equivalent to the semiclassical quantization of the magnetic flux (17). This relation is the physical basis of the Aharonov–Bohm effect [17] and explains the quantization of electron orbits in a singular field of the vector-potential [29]. The latter phenomenon has the following realization. An infinitely thin straight-line solenoid creates the magnetic flux localized along the z-axis and the vector-potential $\boldsymbol{A}$ in the surroundings. The vector-potential $\boldsymbol{A}$ has only one angle component $A_\theta = \Phi_0/(2\pi r)$, where $r^2 = x^2+y^2$. Such a vector-potential does not create any magnetic field in the surroundings ($\mathrm{rot}\,\boldsymbol{A} = 0, r \neq 0$). However the electron forced to move along the closed trajectory "feels" the existence of the magnetic flux if the trajectory encloses the z-axis.

This is a nonlocal topological effect. Any physical field of the magnetic origin does not exist at all points $r \neq 0$ and does not act on the electron. Only entire closed trajectory realizes the effect. A singularity of the trajectory Γ consists in the fact that the forced motion of the electron occurs on the two connected planes (x, y); its single connection is broken by the existence of the localized magnetic flux "perforating" the plane in the point $x = y = 0$. The Aharonov–Bohm effect was the first physical realization of a phenomenon associated with manifestation of Berry's phase.

Berry [25] made a very important observation that has led to the discovery with great consequences. Let the Hamiltonian and the action of the system depend on some set of continuous parameters, which can be called the *space of parameters* called the *space of parameters* $\boldsymbol{R}$. Then the wave function of parameters as well. If one speaks of the electron in a crystal, the Brillouin zone in the $\boldsymbol{k}$-space can serve as the space of parameters [30]. When the parameters $\boldsymbol{R}$ are slowly altered with increasing time and are returned to their original values at the moment $t = T$, the excursion of the system between times $t = 0$ and $t = T$ can be pictured as transport around a closed path in parameter space, with Hamiltonian $\mathcal{H}(\boldsymbol{R}(t))$ and such that $\boldsymbol{R}(T) = \boldsymbol{R}(0)$. Then the Hamiltonian is returned to its initial form and the system is returned to its initial state, apart from the phase of the wave function. The phase acquires the addition $\delta\beta$ not being connected with the classical action. The addition $\delta\beta$ was called *geometrical phase*; that is just Berry's phase.

The total phase change around any closed circuit in the space of parameters obeys a condition of the type of (1.34)

$$\delta\varphi \equiv \frac{1}{\hbar}\delta S_0 + \delta\beta = 2\pi n, \qquad n = 0, 1, 2, ... \tag{1.37}$$

It is not difficult to obtain a formal expression for Berry's phase. Let $\mathcal{H}(\boldsymbol{R}(t))$ be Hamiltonian and $|\psi(t)\rangle$ be the wave function or the vector of state of the system (we follow the original paper [25] and use the notations of Dirac). The evolution of the function $|\psi(t)\rangle$ is described by the Schroedinger equation:

$$\mathrm{i}\hbar \left| \frac{\partial\psi(t)}{\partial t} \right\rangle = \mathcal{H}(\boldsymbol{p}(t))\,|\psi(t)\rangle\,. \tag{1.38}$$

At any moment, the natural basis consists of the eigenstates $|n(\boldsymbol{R})\rangle$ (assumed discrete and not degenerated) of $\mathcal{H}$ for $\boldsymbol{R}$, which satisfy

$$E_n(\boldsymbol{R})|n(\boldsymbol{R})\rangle = \mathcal{H}(\boldsymbol{R})|n(\boldsymbol{R})\rangle \tag{1.39}$$

with energies $E_n(\mathbf{R})$.

Adiabatically, a system prepared in one such state $|n(\boldsymbol{R}(0))\rangle$ will evolve with $\mathcal{H}$ and so will be in the state $|n(\boldsymbol{R}(t))\rangle$ at t. Thus $|\psi(t)\rangle$ can be written as [31]:

$$|\psi(t)\rangle = \exp\left(\frac{-\mathrm{i}}{\hbar}\int_0^t E_n(\boldsymbol{R}(t))\mathrm{d}t'\right)\exp(\mathrm{i}\beta_n(t))|n(\mathbf{R}(t))\rangle, \tag{1.40}$$

where the first exponential is the familiar dynamical phase factor. The last exponential is single-valued in a parameter domain that includes the circuit in the parameter space. The object of attention should be the second exponential. The crucial point will be that its phase $\beta_n(t)$ is *nonintegrable*, cannot be written as a function of $\mathbf{R}$, and in particular is not single-valued under continuation around a circuit, i.e., $\beta_n(t) \neq \beta_n(0)$.

Direct substitution of (1.40) into (1.38) leads to the equation for $\beta_n(t)$:

$$\frac{\mathrm{d}\beta_n(t)}{\mathrm{d}t} = \mathrm{i}\left\langle n(R(t))\left|\frac{\partial n(\boldsymbol{R}(t))}{\partial t}\right.\right\rangle. \tag{1.41}$$

Calculate the total addition β_n around a closed circuit C:

$$\beta_n(T) = \mathrm{i}\oint_C \left\langle n(\boldsymbol{R})\left|\frac{\partial n(\boldsymbol{R})}{\partial \boldsymbol{R}}\right.\right\rangle \mathrm{d}\boldsymbol{R}. \tag{1.42}$$

Since the functions $|n\rangle$ are normalized ($\langle n|n\rangle = 1$), the value $\left\langle n\left|\frac{\partial n}{\partial t}\right.\right\rangle$ is imaginary, which guarantees that $\beta_n(t)$ is real. If $\beta_n(0) = 0$ then (1.42) defines the geometrical phase, i.e., Berry's phase.

Consider the geometrical phase of the band electron moving along some closed circuit Γ in the $\boldsymbol{p}$-space. The Brillouin zone in this case plays the role of the space of parameters and $\boldsymbol{R}$=$\boldsymbol{k}$. Let the electron have at the moment t the Bloch wave function in band s (we return to usual notations of the wave functions):

$$|s\rangle = u_{s\boldsymbol{k}}(\boldsymbol{x})\exp(\mathrm{i}\boldsymbol{k}\boldsymbol{x}), \tag{1.43}$$

where $u_{s\boldsymbol{k}}(\boldsymbol{x})$ is the periodic in the $\boldsymbol{x}$-space function and $\boldsymbol{k}$ is the wave vector. Bloch wave (43) and the energy $\varepsilon_s(\mathbf{k})$ are the eigenfunction and the eigenvalue of Eq. (1.39) at the moment t. The quasiwave vector $\mathbf{k}$ is included in this equation as a parameter. For example, in the magnetic field $\boldsymbol{k}(t) = \boldsymbol{k} - (e/\hbar c)\boldsymbol{A}(t)$. Therefore, when the physical conditions change adiabatically, one can write

$$|s,t\rangle = u_{s\boldsymbol{k}}(t)\exp(\mathrm{i}\boldsymbol{k}\boldsymbol{x}) \tag{1.44}$$

supposing that the quasiwave vector in the exponent $\exp(\mathrm{i}\boldsymbol{k}\boldsymbol{x})$ is independent of time. A slow dependence on time is contained in the Bloch amplitude and

energy $\varepsilon_s(\boldsymbol{k}(t))$. Taking into account the last fact, substitute (1.44) in (1.42):

$$\beta_n(T) = \mathrm{i} \oint_\Gamma \mathrm{d}\boldsymbol{k} \int u^*_{n\boldsymbol{k}}(x) \frac{\partial u_{n\boldsymbol{k}}(\boldsymbol{x})}{\partial \boldsymbol{x}} \mathrm{d}^3x. \tag{1.45}$$

This is just the geometric phase of the electron doing a cyclic motion in the magnetic field. It is interesting to note [27] that the integral

$$\Omega(\boldsymbol{k}) = \mathrm{i} \int u^*_{s\boldsymbol{k}}(\boldsymbol{x}) \frac{\partial u_{s\boldsymbol{k}}(\boldsymbol{x})}{\partial \boldsymbol{k}} \mathrm{d}^3x \tag{1.46}$$

coincides with the diagonal element of the following matrix

$$\Omega_{ss'}(\boldsymbol{k}) = \mathrm{i} \int u^*_{s\boldsymbol{k}}(\boldsymbol{x}) \nabla_{\boldsymbol{k}} u_{s\boldsymbol{k}}(\boldsymbol{x}) \mathrm{d}^3x, \tag{1.47}$$

which determines a so-called periodic part of the coordinate operator (in the $\boldsymbol{k}$ representation) responsible for interband transitions (see [32]):

$$\hat{\boldsymbol{r}} = \mathrm{i} \frac{\partial}{\partial \boldsymbol{k}} + \widehat{\Omega}(\boldsymbol{k}). \tag{1.48}$$

Thus it turns out that Berry's phase has attracted attention to the second term in the coordinate operator (48) in a space with periodic medium.

Now let us come back to the quantization conditions (37) and the discussion of the parameter γ in (1.18). If the orbit under study does not come close to any other trajectory with the same k_z, and its shape differs noticeably from an intersecting one, according to Zilberman [26] the parameter γ always has the value

$$\gamma = \frac{1}{2}. \tag{1.49}$$

It is this value that is commonly used in describing the oscillation phenomena in metals [6]. Therefore the quantization of the electron motion under conditions mentioned earlier should use the following value for δS_0 in (1.37)

$$\delta S_0 = 2\pi \left(n + \frac{1}{2} \right) \hbar. \tag{1.50}$$

Thus the parameter γ in (1.18) equals

$$\gamma = \frac{1}{2} - \frac{1}{2\pi} \Delta\beta, \tag{1.51}$$

where $\Delta\beta = \beta(T) - \beta(0)$ is the total addition to the geometric phase during the cyclic electron motion along a closed trajectory in the magnetic field.

Applications of (1.51) and conclusions connected with it are discussed in paper [34]. The difference of Berry's phase from 0 is usually due to degeneracy of electron states. There are two types of the degeneration connected with band contact and intersections of the isoenergetic surfaces. It is common knowledge that the contact of the bands in a metal can occur at symmetry

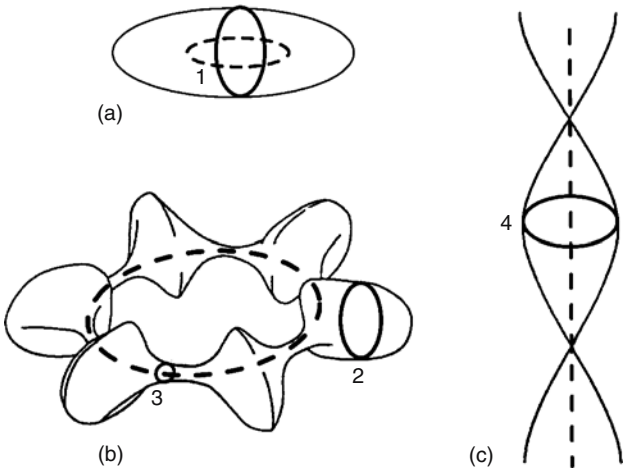

Fig. 1.16. Some examples of the Fermi surfaces in various metals

points and along the symmetry axis of the Brillouin zones. Besides, as shown by Herring [35], there are *lines of accidental contact* between the the bands in crystals. Analysis of the geometry of the Fermi surfaces in metals shows that lines of symmetry degeneration and accidental contacts would exist in many metals. In Fig. 1.16 examples of the Fermi surfaces in various metals are shown, which possess peculiarities discussed earlier.

To calculate the value γ it is convenient to use the result obtained as early as 1962 [33]. It has been found if the contour Γ encloses a line of the contact between the bands, and the energies of the bands separate linearly in $\boldsymbol{k}$ in the vicinity of the line, then

$$\Delta\beta = \pm\pi, \tag{1.52}$$

where the sign on the right-hand side is determined by the direction of the integration. The conditions demanded are satisfied if the line of the degeneration coincides with that of the accidental contact or with the threefold symmetry axis (lines 3 and 4 in Fig. 1.16). For the trajectories mentioned

$$\gamma = 0 \tag{1.53}$$

($\gamma = 1$ and $\gamma = 0$ are equivalent). If in the vicinity of the band-contact line the energy splitting of electron states is quadratic in the distance of the point $\boldsymbol{k}$ from the line, then

$$\Delta\beta = 0. \tag{1.54}$$

Therefore in such a case and for the trajectories not enclosing lines of degeneration

$$\gamma = \frac{1}{2}. \tag{1.55}$$

The results described depend neither on the form of $\varepsilon(\boldsymbol{k})$ nor on the shape and size of the electron trajectory and are topological in nature. It is due to the fact that the electron orbit links to the band-contact line, which is the line of singularities of the Bloch wave functions. If the linking is absent, the convenient result, $\gamma = 1/2$, holds. Measurements of the phase shift of the quantum oscillations in metals can provide a possibility of detecting band-contact lines.

Acknowledgments. The author is grateful to Oksana Charkina and Alexandr Kotlyar for their assistance in preparing the manuscript and indebted to Valentin Peschansky for useful discussions.

References

1. Nakahara, M.: Geometry, Topology and Physics. MIT Press, Cambridge, Massachusetts (1987)
2. Monastyrsky, M.: Topology of Gauge Fields and Condensed Matter. Hardcover, New York (March 1993)
3. Lifshitz, I.M., Azbel', M.Ya., Kaganov, M.I.: Electron Theory of Metals. Consultants Bureau, New York (1973)
4. Sommerfeld, A., Bethe, H.: Electrontheorie der Metalle. Springer, Berlin (1933)
5. Harrison, W.A.: Pseudopotentials in the Theory of Metals. Benjamin, Inc., New York (1966)
6. Shoenberg, D.: Magnetic Oscillations in Metals. Cambridge University Press, Cambridge, Massachusetts (1984)
7. Onsager, L.: Phil. Mag. **43**, 1006 (1952)
8. Azbel, M.Ya.: Soviet Phys. JETP **12**, 891 (1961)
9. Lifshitz, I.M., Kosevich, A.M.: Soviet Phys. JETP **2**, 636 (1956)
10. Wosnitza, J.: Fermi Surfaces of Low-Dimensional Organic Metals. Springer, Berlin (1996)
11. Lifshitz, I.M., Pogorelov, A.V.: Doklady Academii Nauk USSR **96**, 1143 (1954) (in Russian)
12. Kosevich, A.M., Andreev, V.V.: Soviet Phys. JETP **11**, 637 (1960)
13. Coken, M., Falikov, L.: Phys. Rev. Lett. **5**, 544 (1960)
14. Stark, R.W., Falikov, L.M.: Progress in Low Temperature Physics. Vol. 5, p. 235; North-Holland, Amsterdam (1967); Kaganov, M.I., Slutskin, A.A.: Phys. Rep. **98**, 189 (1983); Proshin, Ya.N., Useinov, N.Kh.: Uspekhi Fiz. Nauk **165**, 41 (1995) (in Russian)
15. Bloch, F.: Z. Phys. **52**, 555 (1928); Zener, C.: Proc. Roy. Soc., London, **A145**, 523 (1934)
16. Esaki, L., Tsu, R.: IBM J. Res. Dev. **14**, 61 (1970)
17. Bass, F.G., Bulgakov, A.A.: Kinetic and Electrodynamic Phenomena in Classical and Quantum Semiconductor Superlattices. Nova Science Publishers, Inc., New York (1998)
18. Novikov, S.P., Mal'tsev, A.Ya.: Physics-Uspekhi **41**, 231 (1998)
19. Kaganov, M.I., Peschansky, V.G.: Phys. Rep. **372**, 445 (2002)
20. Lifshitz, I.M., Peschansky, V.G.: JETP **35**, 1251 (1958) (in Russian)

21. Lifshitz, I.M., Peschansky, V.G.: JETP **38**, 188 (1960) (in Russian)
22. Abrikosov, A.A.: Fundamentals of the Theory of Metals. "Nauka", Moscow (1987) (Translated: North-Holland, Amsterdam) (1998)
23. Azbel, M.Ya., Kaner, E.A.: JETP **32**, 896 (1957) (in Russian) Phys. Chem. Solids **6**, 113 (1958)
24. Kaner, E.A.: Selected Papers. Kiev, Naukova Dumka (1989) (in Russian)
25. Berry, M.V.: Proc. R. Soc., London A **392**, 45 (1984)
26. Zilberman, G.E.: JETP **32**, 296 (1957); JETP **33**, 387 (1957) (in Russian)
27. Laura, M.Roth: Phys. Rev. **145**, 434 (1966)
28. Aharonov, Y., Bohm, D.: Phys. Rev. **115**, 485 (1959)
29. Kulik, I.O.: JETP Lett. **11**, 407 (1970) (in Russian)
30. Zak, J.: Phys. Rev. Lett. **62**, 2747 (1989)
31. Houston, W.V.: Phys. Rev. **57**, 184 (1940)
32. Lifshitz, E.M., Pitaevskii, L.P.: Statistical Physics, Pt. 2. Pergamon, Oxford (1986)
33. Blount, E.I.: Solid State Physics. Vol. 13, pp. 305–373, Academic Press, New York (1962)
34. Mikitik, G.P., Shalai, V.: Phys. Rev. Lett. **82**, 2147 (1999)
35. Herring, C.: Phys. Rev. **52**, 365 (1937)

2

Topology, Quasiperiodic Functions, and the Transport Phenomena

A.Ya. Maltsev and S.P. Novikov

Summary. In this chapter we give the basic concept of the "topological numbers" in the theory of quasiperiodic functions. Attention is especially paid to appearance of such quantities in transport phenomena, including galvanomagnetic phenomena in normal metals (Sect. 2.1) and the modulations of 2D electron gas (Sect. 2.3). We give a detailed introduction to both these areas and explain in a simple way the appearance of the "integral characteristics" in both these problems. Though this chapter cannot be considered a detailed survey in the area, it explains the main basic features of the corresponding phenomena.

2.1 Introduction

2.1.1 Galvanomagnetic Phenomena in Normal Metals: Classical Results, GSMF Limit

We first consider the transport phenomena connected with the geometry of quasiclassical electron trajectories in the magnetic field $\boldsymbol{B}$.

Let us start with the most fundamental case where this kind of phenomena appears in the conductivity of normal metals having complicated Fermi surfaces in the presence of a rather strong magnetic field. This classical part of the solid state physics was started by the Kharkov school of I.M. Lifshitz (Lifshitz, Azbel, Kaganov, Peschansky) in the 1950s and has become an essential part of conductivity theory in normal metals. In particular, they introduced the idea of the geometric strong magnetic field (GSMF) limit. Let us give here some small insight into this area. We start with the classical work of I.M. Lifshitz, M.Ya. Azbel and M.I. Kaganov [1], where the importance of topology of the Fermi surface for the conductivity was established. Namely, the difference between the "simple" Fermi surface (topological "sphere") (Fig. 2.1a) and more complicated surfaces where the nonclosed quasiclassical electron trajectories can arise was shown. In particular, detailed consideration of the "simple" Fermi surface and surfaces like "warped cylinder" (Fig. 2.1b) for the different directions of $\boldsymbol{B}$ was made.

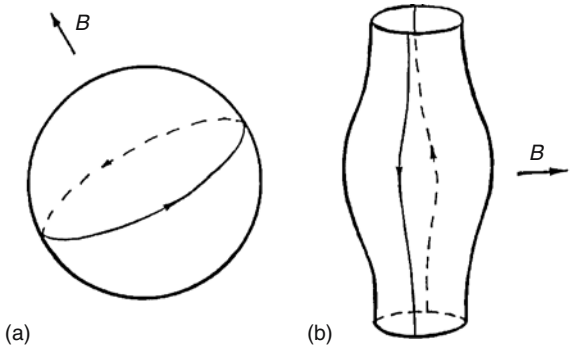

Fig. 2.1. The "simple" Fermi surface having the form of a sphere in the Brillouin zone and the periodic "warped cylinder" extending through an infinite number of Brillouin zones. The quasiclassical electron orbits in $\boldsymbol{p}$-space are also shown for a given direction of $\boldsymbol{B}$

Figure 2.1 represents the forms of the Fermi surfaces in $\boldsymbol{p}$-space and it should be remembered that only one Brillouin zone should be taken into account to get the right phase space volume for the electron states. The values of $\boldsymbol{p}$ which are different from any reciprocal lattice vector $n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ (where n_i are integers), are physically equivalent to each other and represent the same electron state. The Brillouin zone can then be considered as the parallelogram in the $\boldsymbol{p}$-space with the identified opposite sides on the boundary (Fig. 2.2).

Also the Fermi surfaces S_{F} will then be periodic in $\boldsymbol{p}$-space with periods $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$.

Remark. From a topological point of view, we consider the Brillouin zone as the compact three-dimensional torus $\mathbb{T}^3$. The corresponding Fermi surfaces will then also be compact surfaces of finite size embedded in $\mathbb{T}^3$.

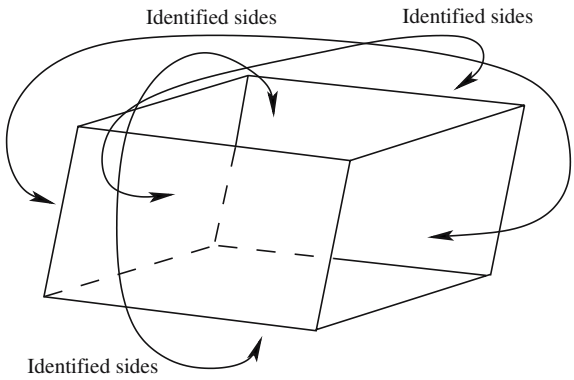

Fig. 2.2. The Brillouin zone in the quasimomentum ($\boldsymbol{p}$) space with the identified sides on the boundary

The presence of the homogeneous magnetic field $\boldsymbol{B}$ generates the evolution of electron states in the $\boldsymbol{p}$-space, which can be described by the dynamical system

$$\dot{\boldsymbol{p}} = \frac{e}{c}\left[\boldsymbol{v}_{\mathrm{gr}}(\boldsymbol{p}) \times \boldsymbol{B}\right] = \frac{e}{c}\left[\nabla\varepsilon(\boldsymbol{p}) \times \boldsymbol{B}\right], \tag{2.1}$$

where $\varepsilon(\boldsymbol{p})$ is the dependence of energy on the quasimomentum (dispersion relation) and $\boldsymbol{v}_{\mathrm{gr}}(\boldsymbol{p}) = \nabla\varepsilon(\boldsymbol{p})$ is the group velocity at the state $\boldsymbol{p}$. Both functions $\varepsilon(\boldsymbol{p})$ and $\boldsymbol{v}_{\mathrm{gr}}(\boldsymbol{p})$ are also periodic functions in $\boldsymbol{p}$-space and can be considered as one-valued functions in $\mathbb{T}^3$.

System (2.1) has two conservative integrals that are the electron energy and the component of $\boldsymbol{p}$ along the magnetic field. The electron trajectories can then be represented as the intersections of the constant energy surfaces $\varepsilon(\boldsymbol{p}) = \text{const.}$ with the planes orthogonal to $\boldsymbol{B}$ and only the Fermi level $\varepsilon(\boldsymbol{p}) = \varepsilon_{\mathrm{F}}$ is actually important for the conductivity. It easy to see then that global geometry of the "essential" electron trajectories will depend strongly on the form of Fermi surface in $\boldsymbol{p}$-space.

Coming back to the Fig. 2.1 we can see that the form of electron trajectories can be quite different for the Fermi surfaces similar to the Fermi surface shown in Fig. 2.1b, we can have periodic nonclosed electron trajectories (if $\boldsymbol{B}$ is orthogonal to vertical axis), while for the surface on Fig. 2.1a all the trajectories are just closed curves lying in one Brillouin zone for all directions of $\boldsymbol{B}$.

We now share that this global geometry plays the main role in the electron motion in the coordinate space also (despite the factorization in $\boldsymbol{p}$-space). Thus the electron wave-packet motion in $\boldsymbol{x}$-space ($\boldsymbol{x} = (x, y, z)$) can be found from the additional system

$$\dot{\boldsymbol{x}} = \boldsymbol{v}_{\mathrm{gr}}(\boldsymbol{p}(t)) = \nabla\varepsilon(\boldsymbol{p}(t))$$

for any trajectory in $\boldsymbol{p}$-space after the integration of system (2.1). The structure of system (2.1) permits to claim for example that the xy-projection of "electron motion" in $\boldsymbol{x}$-space has the same form as the trajectory in $\boldsymbol{p}$-space rotated by $\pi/2$. We can see then that the electron drift in $\boldsymbol{x}$-space in magnetic field is also very different for the trajectories shown in Fig. 2.3a, b due to the action of the crystal lattice.

The effect of this "geometrical drift" can be measured experimentally in the rather pure metallic monocrystals if the mean free electron motion time is big enough (such that the electron packet "feels" the geometric features of trajectory between the two scattering acts). The geometric picture requires then that the time between the two scatterings is much longer than the "passing time" through one Brillouin zone for the periodic trajectory and much longer than the "inverse cyclotron frequency" for closed trajectories.[1] For the

[1] This criterion can be actually more complicated for trajectories of more complicated form.

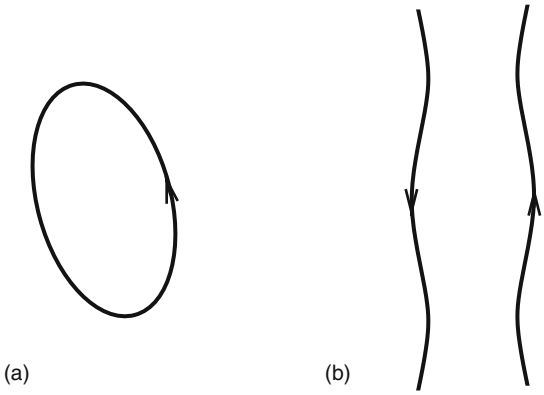

Fig. 2.3. Electron trajectories in $\boldsymbol{p}$-space given by the intersections of planes orthogonal to $\boldsymbol{B}$ for the Fermi surfaces shown in Fig. 2.1a, b for $\boldsymbol{B}$ orthogonal to vertical axis

approximation of effective mass m^* in crystal this condition can be roughly expressed as $\omega_B \tau \gg 1$, where $\omega_B = eB/m^*c$ is the formal cyclotron frequency and τ is the mean free electron motion time. Let us note that this requirement is satisfied better for big values of B and we consider the formal limit $B \to \infty$ in this chapter. We call this situation GSMF limit and consider the asymptotic of conductivity tensor for this case.[2]

We give here the asymptotic form of conductivity tensor obtained in [1] for the case of trajectories shown in Fig. 2.3a, b. Let us take the z-axis in the $\boldsymbol{x}$-space along the magnetic field $\boldsymbol{B}$. The axes x and y can be chosen arbitrarily for the case of Fig. 2.3a and we take the y-axis along the mean electron drift direction in $\boldsymbol{x}$-space for the case of Fig. 2.3b. (It is obvious that the x-axis will then be directed along the mean electron drift in $\boldsymbol{p}$-space in this situation). The asymptotic forms of the conductivity tensor can then be written as:

Case 1 (closed trajectories, Fig. 2.3a):

$$\sigma^{ik} \simeq \frac{ne^2\tau}{m^*} \begin{pmatrix} (\omega_B\tau)^{-2} & (\omega_B\tau)^{-1} & (\omega_B\tau)^{-1} \\ (\omega_B\tau)^{-1} & (\omega_B\tau)^{-2} & (\omega_B\tau)^{-1} \\ (\omega_B\tau)^{-1} & (\omega_B\tau)^{-1} & * \end{pmatrix}, \quad \omega_B\tau \gg 1 \tag{2.2}$$

[2] Formally another condition $\hbar\omega_B \ll \varepsilon_F$ should also be imposed on the magnetic field B. However, this condition is always satisfied for the real metals and all experimentally available magnetic fields (the upper limit is $B \sim 10^3$–10^4 T). So we do not pay special attention to this second restriction and assume that the limit $B \to \infty$ is considered in the "experimental sense," where the second condition is satisfied.

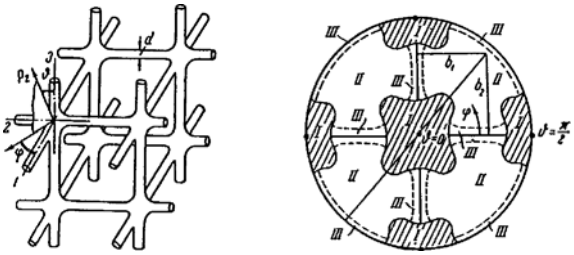

Fig. 2.4. The picture from [2] representing the "thin spatial net" and the corresponding directions of $\boldsymbol{B}$ on the unit sphere where the nonclosed electron trajectories exist

Case 2 (open periodic trajectories, Fig. 2.3b):

$$\sigma^{ik} \simeq \frac{ne^2\tau}{m^*} \begin{pmatrix} (\omega_B\tau)^{-2} & (\omega_B\tau)^{-1} & (\omega_B\tau)^{-1} \\ (\omega_B\tau)^{-1} & * & * \\ (\omega_B\tau)^{-1} & * & * \end{pmatrix}, \quad \omega_B\tau \gg 1, \tag{2.3}$$

where $*$ indicates some dimensionless constants of the order of 1.

We can see that conductivity reveals the strong anisotropy in the plane orthogonal to $\boldsymbol{B}$ in the second case, and the mean direction of the electron trajectory in $\boldsymbol{p}$-space (not in $\boldsymbol{x}$) can be measured experimentally as the zero eigen-direction of σ^{ik} for $B \to \infty$.

More general types of open electron trajectories are considered in [2, 3]. For example, the open trajectories that are not periodic are found in [2] for the "thin spatial net" (Fig. 2.4a). The open trajectories exist here only for the directions of $\boldsymbol{B}$ close to main crystallographic axes $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ (Fig. 2.4b). It was shown in [2] that the open trajectories lie in this case in the straight strips of finite width in the plane orthogonal to $\boldsymbol{B}$ and pass through them. The mean direction of open trajectories is given here by the intersections of plane orthogonal to $\boldsymbol{B}$ with the main crystallographic planes (xy), (yz), and (xz).

The form of conductivity tensor for this kind of trajectories used in [2] coincides with (2.3).

Some analytical dispersion relations are also considered in [3].[3] Let us mention here also the works [4–11] where different experimental (and theoretical) investigations for some real metals were made. Detailed consideration of these results can also be found in the survey articles [12, 13] and the book [14] (see also [15]).

[3] Actually this work contains some conceptual mistakes but it also gives some correct features concerning the existence of some open trajectories for these dispersion relations.

2.1.2 Modern Ideas: The GSMF Limit, Topology, and Dynamical Systems

From the physical point of view the problem arising here can be divided into two parts:

(1) The investigation of topology and dynamics of electron trajectories on the Fermi surface. Based on works [16–20] we call this Novikov problem.

(2) The problem of the relation of this dynamics with the physical properties of electric conductivity in the strong magnetic fields (GSMF limit) formulated essentially in the works of Lifshitz group [1–3, 12–14]).

The result of the Lifshitz group is based on the investigation of kinetic equation for the corresponding quasiparticles given in work [1] for the concrete examples. We had to generalize these results, which led us to the formulation of the GSMF limit in the following form: all essential properties of electrical conductivity (under certain restrictions) are determined by the geometry of the dynamical system on the Fermi surface for the limit of large values of $\boldsymbol{B}$.

It is worth noting that this part of investigation, including the GSMF-limit principle, was never mathematically rigorously investigated unlike the first part (the Novikov problem) where the investigation was made by the rigorous methods of differential topology. It appeared then that in the case of general position the electron trajectories have the integer topological invariants stable with respect to the small rotations of the magnetic field. These "topological quantum numbers" coincide for different trajectories (i.e., possess the "topological resonance" property). Due to this fact the "Topological quantum numbers" become macroscopic observable quantities. We state that there also exist very interesting cases of the so-called "chaotic trajectories". This type of trajectories is not yet completely investigated and it seems that new physical phenomena arise there.

Let us now describe in more detail the topological approach to the problem of general classification of all possible electron trajectories regardless the concrete features of the dispersion relation $\varepsilon(\boldsymbol{p})$ given by Novikov [16] (see also [21–23]). We formulate the Novikov problem here.

Novikov Problem

Let any smooth 3-periodic function $\varepsilon(\boldsymbol{p})$ be given in the three-dimensional space $\mathbb{R}^3$ (with arbitrary lattice of periods). Fix any nondegenerate energy level $\varepsilon(\boldsymbol{p}) = \text{const}$ (i.e., $\nabla\varepsilon(\boldsymbol{p}) \neq 0$ on this level) and consider the intersections of the corresponding smooth 3-periodic surface by any set of parallel planes in $\mathbb{R}^3$. Describe the global geometry of all possible nonsingular (open) trajectories that can arise in the intersections.

The term "the global geometry" means here first the asymptotic behavior of the trajectory when $t \to \pm\infty$ in the sense of dynamical systems. Let us also formulate here the Novikov conjecture about the generic nonsingular trajectories, which was proved later by his pupils.

Novikov Conjecture

The generic nonsingular open trajectories lie in the straight strips of finite width (in the plane orthogonal to $\boldsymbol{B}$*) and pass through them.*

In the process of proving of Novikov conjecture, the deeper properties of the generic open trajectories were actually revealed. They appeared to be stable with respect to the (small) rotations of the direction of $\boldsymbol{B}$. Moreover, it appeared that all the generic open orbits lie on some "warped planes" the quasimomenta space. All these "warped planes" have the integral mean direction (i.e., generated by two reciprocal lattice vectors) and are parallel on average to each other for a given direction of $\boldsymbol{B}$. These integral mean directions of "warped planes" appear to be rigid for small rotations of the direction of $\boldsymbol{B}$ and represent the "Topological quantum numbers" mentioned earlier.

Let us also emphasize that Novikov conjecture is connected with the generic open trajectories and cannot be valid in the special degenerate cases (Tsarev, Dynnikov) as we will see later.

There is also the natural question of what the generic case means in this situation. According to the Novikov conjecture the Hausdorff dimension of the set of directions of $\boldsymbol{B}$ on the unit sphere where the "nongeneric" open trajectories arise is strictly less than 1 for the generic Fermi surfaces (for some nongeneric Fermi surfaces this dimension can be greater than 1 as for example in the case of the surface $\cos x + \cos y + \cos z = 0$ (see [24, 25])).

Let us now give some historical review on the consideration of the Novikov problem in the topological school (Zorich, Dynnikov, Tsarev), where the basic theorems about the nonclosed trajectories were obtained. We provide here the main breakthroughs in this problem made in [17] (A.V. Zorich) and [20] (Dynnikov).

We first note that even for the rather complicated periodic Fermi surface, the electron trajectories will be quite simple if the direction of $\boldsymbol{B}$ is purely rational (with respect to reciprocal lattice), i.e., if the plane $\Pi(\boldsymbol{B})$ orthogonal to $\boldsymbol{B}$ contains two linearly independent reciprocal lattice vectors. This property can also be formulated in the form where the magnetic fluxes through the faces of elementary cell in the $\boldsymbol{x}$-space are proportional to each other with rational coefficients. In this situation the picture arising in $\Pi(\boldsymbol{B})$ is purely periodic and all open electron trajectories can also be just the periodic curves corresponding precisely to the case (2.3). However, the condition of rationality is completely unstable with respect to any small rotations of $\boldsymbol{B}$ such that the rational directions give just a set of measure 0 among all the directions of $\boldsymbol{B}$.

The remarkable fact proved by Zorich is that the open trajectories reveal the "topologically regular" properties even after the small rotations of the initial purely rational direction. That is, they lie in straight strips of finite width in accordance with the Novikov conjecture (but are not periodic anymore). Let us formulate this in a more precise form.

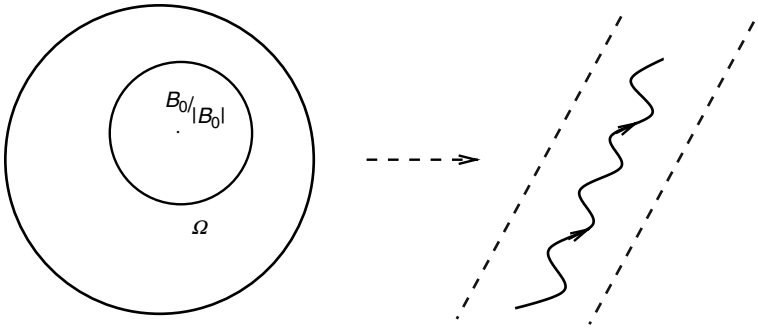

Fig. 2.5. The open region Ω around the purely rational direction $\boldsymbol{B}_0/B_0$ on the unit sphere and the general open trajectory lying in the straight strip of finite width in the plane orthogonal to $\boldsymbol{B}$ for $\boldsymbol{B}/B \in \Omega$

Theorem 1 (Zorich [17]). *Consider an arbitrary smooth Fermi surface and the rational direction of magnetic field $\boldsymbol{B}_0$ such that no singular trajectory connects two different (not equivalent modulo the reciprocal lattice) singular (stagnation) points of the system (2.1). Then there exists a small open region Ω on the unit sphere around direction $\boldsymbol{B}_0$ such that all open trajectories (if they exist) lie in straight strips of finite width in the plane orthogonal to $\boldsymbol{B}$ if $\boldsymbol{B}/B \in \Omega$* (Fig. 2.5).

It was also proved by Dynnikov that any trajectory of this kind passes through the corresponding strip and does not come back ([18, 19]).

Let us also mention that the additional topological condition in Theorem 1 has a generic form and generically does not impose anything on the direction $\boldsymbol{B}_0$.

In his theorem, of Zorich actually claims that all the rational directions of $\boldsymbol{B}$ can be extended to some "small open spots" on the unit sphere (parameterizing directions of $\boldsymbol{B}$) where we cannot have a situation more complicated than that represented in Fig. 2.5. This set already has the finite measure on the unit sphere and moreover we can conclude that any stable open trajectory can have only the form shown in Fig. 2.5 since the rational directions are dense everywhere on the unit sphere. The Zorich theorem, however, does not permit to state that this situation is the only possible one since the sizes of the "spots" become smaller and smaller for big rational numbers and we cannot claim that they cover all the unit spheres in a general situation.

The next important result was obtained by Dynnikov [20] who proved that the trajectories shown in Fig. 2.5 can be the only stable ones with respect to the small variation of the Fermi energy ε_{F} for a given dispersion relation $\varepsilon(\boldsymbol{p})$. We provide the exact form of the Dynnikov theorem in Sect. 2.2 where we will consider this aspect in more detail. We just state here that the methods developed in [20] permitted to prove later that all the cases of open trajectories different from those shown in Fig. 2.5 can appear only "with probability zero"

(i.e., for the directions of $\boldsymbol{B}$ from the set of measure zero on the unit sphere) for generic Fermi surfaces $S_{\mathrm{F}} : \varepsilon(\boldsymbol{p}) = \varepsilon_{\mathrm{F}}$ ([24,26]), which gave the final proof of Novikov conjecture for generic open trajectories.

The methods of proving Zorich and Dynnikov theorems gave the basis for the invention of the "topological quantum numbers" introduced in [28] by the present authors (see also the survey articles [29–31]) for conductivity in normal metals. Let us also state that another important property, called the "Topological Resonance," played a crucial role for physical phenomena in [28]. The main point of this property can be formulated as follows: all the trajectories having the form shown in Fig. 2.5 have the same mean direction in all the planes orthogonal to $\boldsymbol{B}$ for the generic directions of $\boldsymbol{B}$ (actually for any not purely rational direction of $\boldsymbol{B}$) and give the same form (2.3) of contribution to conductivity tensor in the same coordinate system. This important fact makes experimentally observable the integer-valued topological characteristics of the Fermi surface having the form of the integral planes of reciprocal lattice and corresponding "stability zones" on the unit sphere. We describe in detail these quantities in Sect. 2.2 of our paper. Our goal here is to give the main features of the corresponding picture, so we do not give all the details of the classification of all open trajectories for general Fermi surfaces. However, the picture we will describe serves as the "basic description" of conductivity phenomena and all the other possibilities can be considered as special additional features for the nongeneric directions of $\boldsymbol{B}$. Let us also state here that the final classification of open trajectories for generic Fermi surfaces was completed in general by Dynnikov in [27], which solves primarily the Novikov problem. The physical phenomena connected with different types of open trajectories can be found in detail in the survey articles [30, 31].

2.1.3 Transport in 2D Electron Gas and Topology of Quasiperiodic Functions

Let us now mention a few words about the so-called generalized Novikov problem in connection with the quasiperiodic functions on the plane with N quasiperiods. According to the standard definition the quasiperiodic function in $\mathbb{R}^m$ with N quasiperiods ($N \geq m$) is a restriction of a periodic function in $\mathbb{R}^N$ (with N periods) to any plane $\mathbb{R}^m \subset \mathbb{R}^N$ of dimension m linearly embedded in $\mathbb{R}^N$. In our situation we will always have $m = 2$ and the quasiperiodic functions on the plane will be the restrictions of the periodic functions in $\mathbb{R}^N$ to some 2D plane.

Generalized Novikov Problem

Describe the global geometry of open level curves of quasiperiodic function $f(\boldsymbol{r})$ on the plane with N quasiperiods.

It is easy to see that the generalized Novikov problem gives the Novikov problem for the electron trajectories if we put $N = 3$. Indeed, all the

trajectories in the planes orthogonal to $\boldsymbol{B}$ can be considered as the level curves of quasiperiodic functions $\varepsilon(\boldsymbol{p})|_{\Pi(\boldsymbol{B})}$ with three quasiperiods. As mentioned earlier, the general Novikov problem is solved primarily for $N = 3$. However, the case $N > 3$ becomes very complicated from the topological point of view and no general classification in this case exists at the moment. The only topological result existing now for the general Novikov problem is the analog of Zorich theorem (Theorem 1) for the case $N = 4$ [32] and the general situation is still under investigation.

In Section 2.3 we consider the applications of generalized Novikov problem connected with the "superlattice potentials" for the two-dimensional electron gas in the presence of orthogonal magnetic field. This kind of potentials is connected with modern techniques of "handmade" modulations of 2D electron gas such as the holographic illumination, "gate modulation", piezoelectric effect, etc. All such modulations are usually periodic in the plane and in many situations the level curves play an important role for the transport phenomena in such systems. The most important thing for us will be the conductivity phenomena in these 2D structures in the presence of orthogonal magnetic field $\boldsymbol{B}$. According to the quasiclassical approach the cyclotron electron orbits drift along the level curves of modulation potential in the magnetic field, which gives the "drift contribution" to conductivity in the plane. Among the works devoted to this approach we would like to mention here the article [33], where this method was introduced for the explanation of "commensurability oscillations" of conductivity in potential modulated just in one direction, and [34] where the same approach was used for the explanation of suppression of these oscillations by the second orthogonal modulation in the periodic case. Let us add that all these phenomena correspond to the long free electron motion time, which will now play the role of the "geometric limit" (not $B \to \infty$) in the second situation.

We will show that the generalized Novikov problem can also arise naturally in these structures if we consider the independent superposition of different periodic modulations. It can be proved that in this case we always obtain the quasiperiodic functions where the number of quasiperiods depends on the complexity of total modulation. The results in Novikov problem can then help to predict the form of the "drift conductivity" in the limit of long free electron motion time. In Sect. 2.3 we give the main features of the situation of superposition of several "1D modulations" where the potentials with a small number of quasiperiods can arise. The detailed consideration of this situation can be found in [35]. However, the Novikov problem also arises in a much more general case of arbitrary superpositions of more complicated (but periodic) structures.

Finally, we would like to mention that the quasiperiodic functions with a large number of quasiperiods can be a model for the random potentials on the plane. The corresponding Novikov problem arises in the percolation theory for such potentials. We will also discuss this situation at the end of Sect. 2.3.

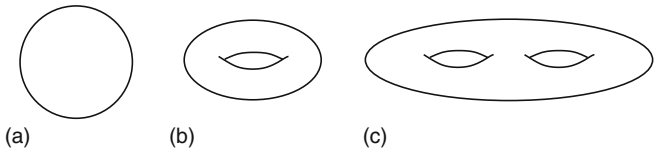

Fig. 2.6. The abstract surfaces (**a**), (**b**), and (**c**) with genuses 0, 1, and 2, respectively

2.2 The Classification of Fermi Surfaces and the "Topological Quantum Numbers"

Let us start with the definitions of genus and topological rank of the Fermi surface.

Definition 1. *Let us consider the phase space* $\mathbb{T}^3 = \mathbb{R}^3/L$ *introduced earlier. After the identification, every component of the Fermi surface becomes the smooth orientable two-dimensional surface embedded in* $\mathbb{T}^3$*. We can then introduce the standard genus of every component of the Fermi surface* $g = 0, 1, 2, ...$ *according to standard topological classification depending on whether this component is a topological sphere, torus, sphere with two holes, etc. (Fig. 2.6).*

Definition 2. *Let us introduce the topological rank* r *as the characteristic of the embedding of the Fermi surface in* $\mathbb{T}^3$*. It is much more convenient in this case to come back to the total* $\boldsymbol{p}$*-space and consider the connected components of the three-periodic surface in* $\mathbb{R}^3$*.*

(1) The Fermi surface has Rank 0 *if each of its connected component can be bounded by a sphere of finite radius.*

(2) The Fermi surface has Rank 1 *if each of its connected component can be bounded by the periodic cylinder of finite radius and there are components that cannot be bounded by the sphere.*

(3) The Fermi surface has Rank 2 *if each of its connected component that can be bounded by two parallel (integral) planes in* $\mathbb{R}^3$ *and there are components that cannot be bounded by a cylinder.*

(4) The Fermi surface has Rank 3 *if it contains components that cannot be bounded by two parallel planes in* $\mathbb{R}^3$*.*

Figure 2.7a, b, c, d represents the pieces of the Fermi surfaces in $\mathbb{R}^3$ with the topological ranks 0, 1, 2, and 3, respectively. As can be seen the genuses of the surfaces represented in Fig. 2.7a, b, c, d are also equal to 0, 1, 2, and 3, respectively. However, the genus and the Topological Rank are not necessary equal to each other in the general situation.

Let us discuss briefly the connection between the genus and the topological rank since this will play a crucial role in further consideration.

It is easy to see that the topological rank of the sphere can be only 0 and the Fermi surface consists in this case of the infinite set of the periodically repeated spheres $\mathbb{S}^2$ in $\mathbb{R}^3$.

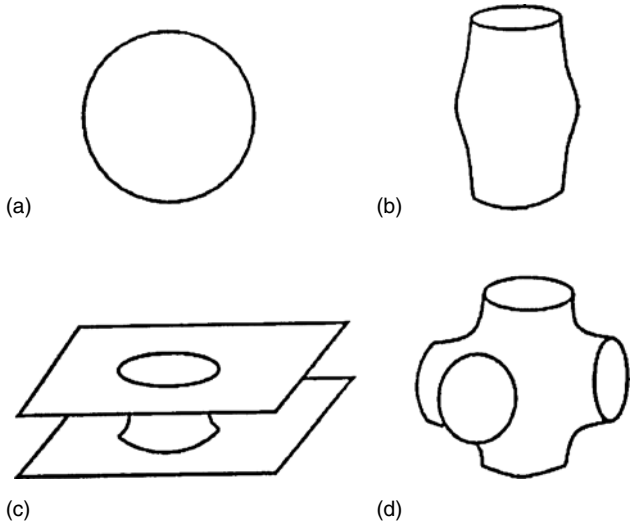

Fig. 2.7. The Fermi surfaces (**a**), (**b**), (**c**), (**d**) with topological ranks 0, 1, 2, and 3, respectively

The topological rank of the torus $\mathbb{T}^2$ can take three values $r = 0, 1, 2$. Indeed, it is easy to see that all three cases of periodically repeated tori $\mathbb{T}^2$ in $\mathbb{R}^3$ (Rank 0), periodically repeated "warped" integral cylinders (Rank 1), and the periodically repeated "warped" integral planes (Rank 2) give the topological two-dimensional tori $\mathbb{T}^2$ in $\mathbb{T}^3$ after the factorization (see Fig. 2.8).

It is not difficult to prove that these are the only possibilities that we can have for embedding of the two-dimensional torus $\mathbb{T}^2$ in $\mathbb{T}^3$. We note here that the mean direction of the "warped periodic cylinder" (embedding of Rank 1) can coincide with any reciprocal lattice vector $n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ in $\mathbb{R}^3$. Also the "directions" of the corresponding "warped planes" (embedding of Rank 2)

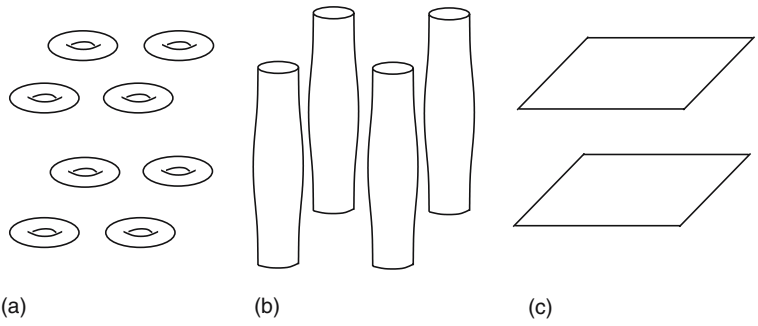

Fig. 2.8. The periodically repeated tori $\mathbb{T}^2$, periodically repeated "warped" integral cylinders, and the periodically repeated "warped" integral planes in $\mathbb{R}^3$

are always generated by two (linearly independent) reciprocal lattice vectors $m_1^{(1)}\boldsymbol{a}_1+m_2^{(1)}\boldsymbol{a}_2+m_3^{(1)}\boldsymbol{a}_3$ and $m_1^{(2)}\boldsymbol{a}_1+m_2^{(2)}\boldsymbol{a}_2+m_3^{(2)}\boldsymbol{a}_3$. We can thus see that both the embeddings of Rank 1 and Rank 2 of $\mathbb{T}^2$ in $\mathbb{T}^3$ are characterized by some integer numbers connected with the reciprocal lattice. Let us also make one more remark about the surfaces of Ranks 0, 1, and 2 in this case. Namely, the case $r = 2$ actually shows one difference from the cases $r = 0$ and 1, which is that the plane in $\mathbb{R}^3$ is not homologous to 0 in $\mathbb{T}^3$ (i.e., it does not restrict any domain of "lower energies") after the factorization. We can conclude that if these planes appear as the connected components of the physical Fermi surface (which is always homologous to 0), they should always come in pairs, Π_+ and Π_-, which are parallel to each other in $\mathbb{R}^3$. The factorization of Π_+ and Π_- gives then the two tori $\mathbb{T}^2_+$, $\mathbb{T}^2_-$ with the opposite homologous classes in $\mathbb{T}^3$.

It can be shown that the topological rank of any Fermi surface of genus 2 cannot exceed 2 also. The example of the corresponding embedding of such a component with maximal rank is shown in Fig. 2.7c and represents the two parallel planes connected by cylinders. We will not give the proof of this theorem here but just mention that this fact plays an important role in the classification of nonclosed electron trajectories on the Fermi surface of genus 2. Namely, it can be proved that the open trajectories on the Fermi surface of genus 2 cannot be actually more complicated than the trajectories on the surface of genus 1. In particular they always have the "topologically regular form" in the same way as on the Fermi surface of genus 1 (see Sect. 2.2). Also the same integral characteristics in the cases when this surface has Rank 1 or 2 as in the case of genus 1 can be introduced for genus 2 (actually for any genus if rank is equal to 1 or 2).

Finally we would like to mention that the topological rank of the components with genus $g \geq 3$ can take any value $r = 0, 1, 2, 3$.

Definition 3. *We call the open trajectory topologically regular (corresponding to "topologically integrable" case) if it lies within the straight line of finite width in $\Pi(\boldsymbol{B})$ and passes through it from $-\infty$ to ∞. We call all other open trajectories chaotic.*

Let us now discuss the connection between the geometry of the nonsingular electron orbits and the topological properties of the Fermi surface. We briefly consider here the simple cases of Fermi surfaces of Rank 0, 1, and 2 and then come to our basic case of general Fermi surfaces having the maximal rank $r = 3$. We then have the following situations:

(1) The Fermi surface has topological rank 0.

It is easy to note that in this simplest case all the components of the Fermi surface are compact (Figs. 2.7a, 2.8a) in $\mathbb{R}^3$ and there are no open trajectories at all.

(2) The Fermi surface has topological rank 1.

In this case we can have both open and compact electron trajectories. However the open trajectories (if they exist) should be quite simple in this case.

They can arise only if the magnetic field is orthogonal to the mean direction of one of the components of Rank 1 (periodic cylinder) and are periodic with the same integer mean direction (Figs. 2.7b, 2.8b). The corresponding sets of the directions $\boldsymbol{B}/B$ are just the one-dimensional curves and there cannot be open regions on the unit sphere for which we can find the open trajectories on the Fermi surface.

(3) The Fermi surface has topological rank 2.

It can be easily seen that this case gives much more possibilities for the existence of open orbits for different directions of the magnetic field. In particular, this is the first case where the open orbits can exist for the generic directions of $\boldsymbol{B}$. So, in this case we can have the whole regions on the unit sphere such that the open orbits present for any direction of $\boldsymbol{B}$ belong to the corresponding region. It is easy to see, however, that the open orbits also have quite a simple description in this case. Namely, any open orbit (if it exists) lies in the straight strip of the finite width for any direction of $\boldsymbol{B}$ not orthogonal to the integral planes given by the components of Rank 2. The boundaries of the corresponding strips in the planes $\Pi(\boldsymbol{B})$ (orthogonal to $\boldsymbol{B}$) will be given by the intersection of $\Pi(\boldsymbol{B})$ with the pairs of integral planes bounding the corresponding components of Rank 2. It can also be shown [18,19] that every open orbit passes through the strip from $-\infty$ to $+\infty$ and cannot turn back. We can then see that all the trajectories are "topologically regular" in this case also.

Based on the remarks given earlier, the contribution to the conductivity given by every family of orbits with the same mean direction reveals the strong anisotropy when $\omega_{\mathrm{B}}\tau \to \infty$ and coincides with the main order with formula (2.3) for the open periodic trajectories.

Trajectories of this type already have all the features of the general topologically integrable situation.

We start now with the most general and complicated case of arbitrary Fermi surface of topological rank 3.

We first describe a convenient procedure [26, 27] of reconstruction of the constant energy surface when the direction of $\boldsymbol{B}$ is fixed.

We will assume that the system (2.1) has generically only the nondegenerate singularities having the form of the nondegenerate poles or nondegenerate saddle points. The singular trajectories passing through the critical points (and the critical points themselves) divide the set of trajectories into different parts corresponding to different types of trajectories on the Fermi surface. We are not interested here in the geometry of compact electron trajectories in the "geometric limit" $\omega_{\mathrm{B}}\tau \to \infty$. It is not difficult to show that the pieces of the Fermi surface carrying the compact orbits can be either infinite or finite cylinders in $\mathbb{R}^3$ bounded by the singular trajectories (some of them may be just points of minimum or maximum) at the bottom and at the top (see Fig. 2.9).

Let us now remove all the parts containing the nonsingular compact trajectories from the Fermi surface. The remaining part,

$$S_{\mathrm{F}}/(\text{compact nonsingular trajectories}) = \cup_j S_j,$$

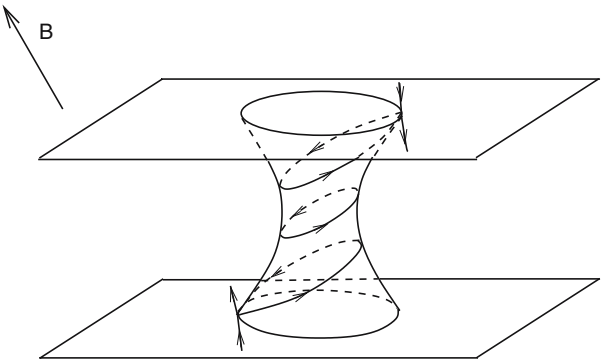

Fig. 2.9. The cylinder of compact trajectories bounded by the singular orbits (the simplest case of just one critical point on the singular trajectory)

is a union of the two manifolds S_j with boundaries ∂S_j, which are the compact singular trajectories. The generic type in this case is a separatrix orbit with just one critical point like in Fig. 2.9.

It is obvious that the open orbit will not be affected at all by the construction described here and the rest of the Fermi surface gives the same open orbits as all possible intersections with different planes orthogonal to $\boldsymbol{B}$.

Definition 4. *We call every piece S_j the* **"Carrier of open trajectories."**

Let us fill in the holes by topological 2D discs lying in the planes orthogonal to $\boldsymbol{B}$ and get the closed surfaces (see Fig. 2.10)

$$\bar{S}_j \;=\; S_j \cup (\text{2D discs}).$$

This procedure again gives the periodic surface $\overline{S}_\varepsilon$ after the reconstruction and we can define the "compactified carriers of open trajectories" both in $\mathbb{R}^3$ and $\mathbb{T}^3$.

It is obvious that the reconstructed surface can be used instead of the original Fermi surface for the determination of open trajectories. Let us ask a question: can the reconstructed surface be simpler than the original one?

The answer is positive and moreover it can be proved that "generically" the reconstructed surface consists of components of genus 1 only. This remarkable fact gives the very powerful instrument for the consideration of open trajectories on the arbitrary Fermi surface.

In fact, the proof of Theorem 1 was based on the statement that the genus of every compactified carrier of open orbits $\overline{S}_j$ is equal to 1 in this case.

Let us now formulate the theorem of Dynnikov [20], which made the second main breakthrough in the Novikov problem.

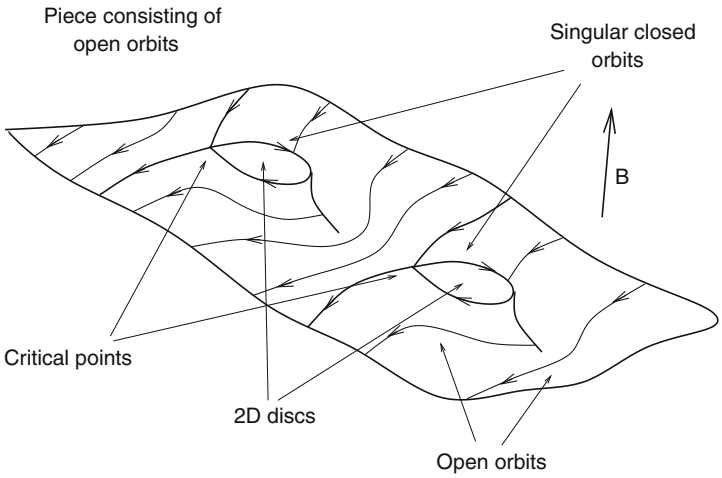

Fig. 2.10. The reconstructed constant energy surface with removed compact trajectories and the two-dimensional discs attached to the singular trajectories in the generic case of just one critical point on every singular trajectory

Theorem 2 (Dynnikov [20]). *Let a generic dispersion relation*

$$\varepsilon(\boldsymbol{p}) : \ \mathbb{T}^3 \to \mathbb{R}$$

be given such that for level $\varepsilon(\boldsymbol{p}) = \varepsilon_0$ *the genus* g *of some carrier of open trajectories* $\overline{S}_i$ *is greater than* 1. *Then there exists an open interval* $(\varepsilon_1, \varepsilon_2)$ *containing* ε_0 *such that for all* $\varepsilon \neq \varepsilon_0$ *in this interval the genus of the carrier of open trajectories is less than* g.

Theorem 2 claims that only the "topologically integrable case" can be stable with respect to the small variations of energy level also.

The formulated theorems permit us to reduce the consideration of open orbits in any stable situation to the case of the surfaces of genus 1 where the Fermi surface can have topological rank 0, 1, or 2 only. It is easy to see that the Rank 0 cannot appear just by definition of the reconstructed surface $\bar{S}_\varepsilon$ since it can contain only the compact trajectories. Rank 1 is possible in $\bar{S}_\varepsilon$ only for special directions of $\boldsymbol{B}$. Indeed, the component of Rank 1 has the mean integral direction in $\mathbb{R}^3$ and can contain the open (periodic) trajectories only if $\boldsymbol{B}$ is orthogonal to this integral vector in $\boldsymbol{p}$-space. The corresponding open trajectories are thus not absolutely stable with respect to the small rotations of $\boldsymbol{B}$ and cannot exist for the open region on the unit sphere.

We can then claim that the only generic situation for $\overline{S}_\varepsilon$ is a set of components of Rank 2, which are the periodic warped planes in this case. The corresponding electron trajectories can then belong just to "Topologically integrable" case being the intersections of planes orthogonal to $\boldsymbol{B}$ with the periodically deformed planes in the $\boldsymbol{p}$-space.

An important property of the compactified components of genus 1 arising for the generic directions of $\boldsymbol{B}$ is the following: they are all parallel on average to $\mathbb{R}^3$ and do not intersect each other. This property mentioned in [28] and called later the "topological resonance" plays an important role in the physical phenomena connected with geometry of open trajectories. In particular, all the stable topologically regular open trajectories in all planes orthogonal to $\boldsymbol{B}$ have the same mean direction and give the same form (2.3) of contribution to conductivity in the appropriate coordinate system common for all of them. This fact gives the experimental possibility of measuring the mean direction of noncompact topologically regular orbits both in $\boldsymbol{x}$ and $\boldsymbol{p}$ spaces from the anisotropy of conductivity tensor σ^{ik}.

We reiterate that the surface $\overline{S}_\varepsilon$ is the abstract construction depending on the direction of $\boldsymbol{B}$ and does not exist a priori in the Fermi surface $S_{\varepsilon_{\mathrm{F}}}$. The important fact, however, is the stability of the surface $\overline{S}_\varepsilon$ with respect to the small rotations of $\boldsymbol{B}$. This means in particular that the common direction of the components of Rank 2 is locally stable with respect to the small rotations of $\boldsymbol{B}$, which can then be found from the conductivity experiments. From the physical point of view, all the regions on the unit sphere where the stable open orbits exist can be represented as the "stability zones" Ω_α such that each zone corresponds to some integral plane Γ_α common to all the points of stability zone Ω_α. The plane Γ_α is then the integral plane in reciprocal lattice, which defines the mean directions of open orbits in $\boldsymbol{p}$-space for any direction of $\boldsymbol{B}$ belonging to Ω_α just as the intersection with the plane orthogonal to $\boldsymbol{B}$. As can be easily seen from the form of (2.3), this direction always coincides with the unique direction in $\mathbb{R}^3$ corresponding to the decrease of conductivity as $\omega_{\mathrm{B}}\tau \to \infty$.

The corresponding integral planes Γ_α can then be given by three integer numbers $(n_\alpha^1, n_\alpha^2, n_\alpha^3)$ (up to the common multiplier) from the equation

$$n_\alpha^1[\boldsymbol{x}]_1 + n_\alpha^2[\boldsymbol{x}]_2 + n_\alpha^3[\boldsymbol{x}]_3 = 0,$$

where $[\boldsymbol{x}]_i$ are the coordinates on the basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ of the reciprocal lattice, or equivalently

$$n_\alpha^1(\boldsymbol{x}, \boldsymbol{l}_1) + n_\alpha^2(\boldsymbol{x}, \boldsymbol{l}_2) + n_\alpha^3(\boldsymbol{x}, \boldsymbol{l}_3) = 0,$$

where $\{\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3\}$ is the basis of the initial lattice in the coordinate space.

We see then that the direction of conductivity decreasing $\hat{\eta} = (\eta_1, \eta_2, \eta_3)$ satisfies the relation

$$n_\alpha^1(\hat{\eta}, \boldsymbol{l}_1) + n_\alpha^2(\hat{\eta}, \boldsymbol{l}_2) + n_\alpha^3(\hat{\eta}, \boldsymbol{l}_3) = 0$$

for all the points of stability zone Ω_α, which makes possible the experimental observation of numbers $(n_\alpha^1, n_\alpha^2, n_\alpha^3)$.

The numbers $(n_\alpha^1, n_\alpha^2, n_\alpha^3)$ are called in [28] the "topological quantum numbers" of a dispersion relation in metal.

We can now consider the result of [2] about the "thin spatial net" as a particular case of this general theorem where the integer planes take the

simplest possibility of being the main planes xy, yz, xz. If we now introduce the "topological quantum numbers" for this situation, we will have only the triples $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, and $(0, 0, \pm 1)$ for this Fermi surface.

In general, we can state that the unit sphere should be divided into (open) parts where the open orbits are absent on the Fermi level for given directions of $\boldsymbol{B}$ and "stability zones" Ω_α where the open orbits exist on the Fermi level and have "topologically regular" form. Each stability zone corresponds to the triple of "topological quantum numbers" giving the integral direction of periodically deformed two-dimensional planes in $\bar{S}_{\varepsilon_{\mathrm{F}}}(\boldsymbol{B})$, which are swept by the zero eigen-vector of σ^{ik} for $\boldsymbol{B} \in \Omega_\alpha$.

We now state that the "topologically regular" trajectories are generic open trajectories, nonetheless they are not ideal for rather complicated Fermi surfaces. Namely, for rather complicated Fermi surfaces and the special directions of $\boldsymbol{B}$, the chaotic cases can also arise (Tsarev, Dynnikov).

It was first shown by Tsarev [36] that the more complicated chaotic open orbits can still exist on rather complicated Fermi surfaces S_{F}. An example of an open trajectory that does not lie in any finite strip of finite width was constructed. However, the trajectory had in this case the asymptotic direction of not even being restricted by any straight strip of finite width in the plane orthogonal to $\boldsymbol{B}$. The corresponding asymptotic behavior of conductivity should also reveal the strong anisotropy properties in the plane orthogonal to $\boldsymbol{B}$ although the exact form of σ^{ik} will be slightly different from (2.3) for this type of trajectories. For the same reason, the asymptotic direction of orbit can be measured experimentally in this case.

The more complicated examples of chaotic open orbits were constructed in [26] for the Fermi surface having genus 3. These types of open orbits do not have any asymptotic direction in the planes orthogonal to $\boldsymbol{B}$ and have a rather complicated form of "walking everywhere" in these planes.

The corresponding contribution to σ^{ik} is also very different for this kind of trajectories [37]. In particular, it appears that this contribution becomes 0 in all the directions including the direction of $\boldsymbol{B}$ for $B \to \infty$. The total conductivity tensor σ^{ik} has then only the contribution of compact electron trajectories in the conductivity along $\boldsymbol{B}$, which does not disappear when $B \to \infty$. The corresponding effect can be observed experimentally as the local minima of the longitudinal (i.e., parallel to $\boldsymbol{B}$) conductivity for the points of the unit sphere where this kind of trajectories can appear. A more detailed description of σ^{ik} in this case can be found in [37].

Let us add that Dynnikov proved recently that the measure of chaotic cases on the unit sphere is 0 for generic Fermi surfaces [26, 27]. The systematic investigation of the open orbits was completed in general after the works [17, 20, 26, 28] in [27]. In particular the total picture of different types of the open orbits for generic dispersion relations was presented. Let us formulate here the main results of [27] in the form of a Theorem.

Theorem 3 (Dynnikov [27]). *Let us fix the dispersion relation* $\varepsilon = \varepsilon(\boldsymbol{p})$ *and the direction of* $\boldsymbol{B}$ *of irrationality* 3 *and consider all the energy levels for* $\varepsilon_{\min} \le \varepsilon \le \varepsilon_{\max}$. *Then:*

(1) The open electron trajectories exist for all the energy values ε *belonging to the closed connected energy interval* $\varepsilon_1(\boldsymbol{B}) \le \varepsilon \le \varepsilon_2(\boldsymbol{B})$, *which can degenerate to just one energy level* $\varepsilon_1(\boldsymbol{B}) = \varepsilon_2(\boldsymbol{B}) = \varepsilon_0(\boldsymbol{B})$.

(2) For the case of the nontrivial energy interval the set of compactified carriers of open trajectories $\overline{S}_\varepsilon$ *is always a disjoint union of two-dimensional tori* $\mathbb{T}^2$ *in* $\mathbb{T}^3$ *for all* $\varepsilon_1(\boldsymbol{B}) \le \varepsilon \le \varepsilon_2(\boldsymbol{B})$. *All the tori* $\mathbb{T}^2$ *for all the energy levels do not intersect each other and have the same (up to the sign) indivisible homology class* $c \in H_2(\mathbb{T}^3, \mathbb{Z})$, $c \ne 0$. *The number of tori* $\mathbb{T}^2$ *is even for every fixed energy level and the corresponding covering* $\overline{S}_\varepsilon$ *in* $\mathbb{R}^3$ *is a locally stable family of parallel ("warped") integral planes* $\Pi_i^2 \subset \mathbb{R}^3$ *with common direction given by c. The form of* $\overline{S}_\varepsilon$ *described here is locally stable with the same homology class* $c \in H_2(\mathbb{T}^3)$ *under small rotations of* $\boldsymbol{B}$. *All the open electron trajectories at all the energy levels lie in the strips of finite width with the same direction and pass through them. The mean direction of the trajectories is given by the intersections of planes* $\Pi(\boldsymbol{B})$ *with the integral family* Π_i^2 *for the corresponding "stability zone" on the unit sphere.*

(3) The functions $\varepsilon_1(\boldsymbol{B})$, $\varepsilon_2(\boldsymbol{B})$ *defined for the directions of* $\boldsymbol{B}$ *of irrationality* 3 *can be continued on the unit sphere* S^2 *as the piecewise smooth functions such that* $\varepsilon_1(\boldsymbol{B}) \ge \varepsilon_2(\boldsymbol{B})$ *everywhere on the unit sphere.*

(4) For the case of trivial energy interval $\varepsilon_1 = \varepsilon_2 = \varepsilon_0$ *the corresponding open trajectories may be chaotic. The carrier of the chaotic open trajectory is homologous to 0 in* $H_2(\mathbb{T}^3, \mathbb{Z})$ *and has genus* ≥ 3. *For the generic energy level* $\varepsilon = \varepsilon_0$ *the corresponding directions of magnetic fields belong to the countable union of the codimension* 1 *subsets. Therefore a measure of this set is equal to 0 on* S^2.

We give here the results connected with generic directions of $\boldsymbol{B}$ and do not consider the special cases when $\boldsymbol{B}$ is purely or "partly" rational. The corresponding effects are actually simpler than formulated earlier and can be easily added to this general picture. Survey articles [27, 29–31] provide all the details (both from mathematical and physical point of view).

2.3 Quasiperiodic Modulations of 2D Electron Gas and the Generalized Novikov Problem

In this section we provide a general description about the quasiperiodic modulations of 2D electron gas and the main topological aspects for the special class of such structures. Let us first discuss about different modern modulation techniques and the quasiclassical electron behavior in such systems.

We first point here the holographic illumination of high-mobility 2D electron structures (AlGaA–GaAs heterojunctions) at temperatures $T \leq 4.2\,\mathrm{K}$ (see, for example, [38]). In these experiments the expanded laser beam was split into two parts, which gave an interference picture with the period a on the 2D sample. The illumination caused the additional ionization of atoms near the 2D junction, which remained for a rather long period of time after the illumination. During this relaxation time, the additional periodic potential $V(\boldsymbol{r}) = V(x)$, $V(x) = V(x + a)$ arose in the plane and the electron behavior was determined by the orthogonal magnetic field $\boldsymbol{B}$ and the potential $V(x)$.

The quasiclassical consideration for the case $|V(x)| \ll \varepsilon_{\mathrm{F}}$ was first considered by Beenakker [33] for the explanation of "commensurability oscillations" in such structures found in [38]. According to this approach the quasiclassical electrons near the Fermi level move around the cyclotron orbits in the magnetic field and drift due to potential $V(x)$ in the plane. Since only the electrons near Fermi level ε_{F} play the main role in conductivity, we can introduce the characteristic cyclotron radius $r_{\mathrm{B}} = m^* v_{\mathrm{F}}/eB$ for the Fermi velocity v_{F}. The corresponding drift of the electron orbits near the Fermi level will then be determined by the averaged effective potential $V_{\mathrm{B}}^{\mathrm{eff}}(x)$ given by the averaging of $V(\boldsymbol{r}) = V(x)$ over the cyclotron orbit with radius r_{B} centered at the point $\boldsymbol{r}$ (Fig. 2.11).

The potential $V_{\mathrm{B}}^{\mathrm{eff}}(x)$ is different from $V(x)$ but has the same symmetry and also depends only on x. The drift of the cyclotron orbits is along the level curves of $V_{\mathrm{B}}^{\mathrm{eff}}(x)$, which are very simple in this case (just the straight lines along the y-axis) and the corresponding velocity v_{drift} is proportional to the absolute value of gradient $|V_{\mathrm{B}}^{\mathrm{eff}}(x)|$ at each level curve. The analytic dependence of $|V_{\mathrm{B}}^{\mathrm{eff}}(x)|$ on the value of B (based on the commensurability of $2r_{\mathrm{B}}$ with the (integer number) $\times\, a$) was used in [33] for the explanation of the oscillations of conductivity along the fringes with the value of B.

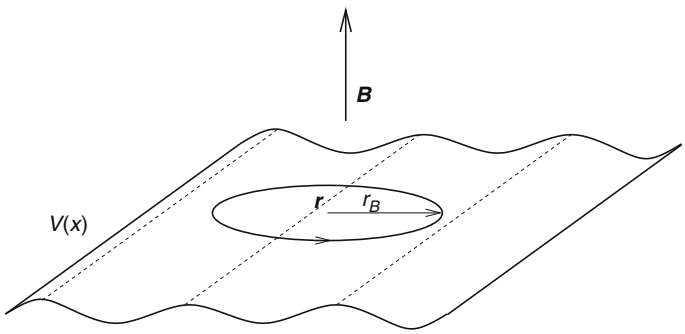

Fig. 2.11. The averaging of the the potential $V(x)$ over the cyclotron orbit with radius r_{B} centered at the point $\boldsymbol{r}$

In article [34] the situation with the double-modulated potentials made by the superposition of two interference pictures was also considered. The corresponding potential $V(\boldsymbol{r})$ is double periodic in $\mathbb{R}^2$ in this case and the same is true for potentials $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$. The consideration used the same quasiclassical approach for the potential $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$ based on the analysis of its level curves. It was then shown in [34] that the second modulation should suppress the commensurability oscillations in this case, which disappear completely for the equal intensities of two (orthogonal) interference pictures.

It is also obvious that all the open drift trajectories can be only periodic in the case of periodic $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$.

It seems that the situation with the quasiperiodic modulations of 2D electron gas did not appear in experiments. However, we think that this situation is also very natural for the technique described earlier and can be considered from the point of view of the generalized Novikov problem. The corresponding approach was developed in [35] for the special cases of superpositions of several (three and four) interference pictures on the plane. Nonetheless, as we already mentioned, the Novikov problem also arises actually for any picture given by superposition of several periodic pictures in the plane. The corresponding potentials can have many quasiperiods in this case and the Novikov problem can then reveal much more complicated (chaotic) properties than described in [35].

We next describe here just the main points of "topologically regular" behavior in the case of the superpositions of three and four interference pictures, which give the quasiperiodic potentials $V(\boldsymbol{r})$ and $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$ with three and four quasiperiods on the plane. Unlike the previous works we do not pay much attention to the analytic dependence on B and investigate mainly the geometric properties of conductivity in this situation.

Before we start the geometric consideration, we wish to also state that the holographic illumination is not a unique way of producing the superlattice potentials for the two-dimensional electron gas. Let us mention here the works [39–49] where the different techniques using the biasing of the specially made metallic gates and the piezoelectric effect were considered. Both 1D and 2D modulated potentials as well as more general periodic potentials with square and hexagonal geometry appeared in this situation. Actually these techniques give much more possibilities to produce the potentials of different types with the quasiperiodic properties.

Let us now have three independent interference pictures on the plane with three different generic directions of fringes η_1, η_2, η_3 and periods a_1, a_2, a_3 (see Fig. 2.12).

The total intensity $I(\boldsymbol{r})$ will be the sum of intensities

$$I(\boldsymbol{r}) = I_1(\boldsymbol{r}) + I_2(\boldsymbol{r}) + I_3(\boldsymbol{r})$$

of the independent interference pictures.

We assume that there are at least two noncoinciding directions (say η_1, η_2) among the set (η_1, η_2, η_3).

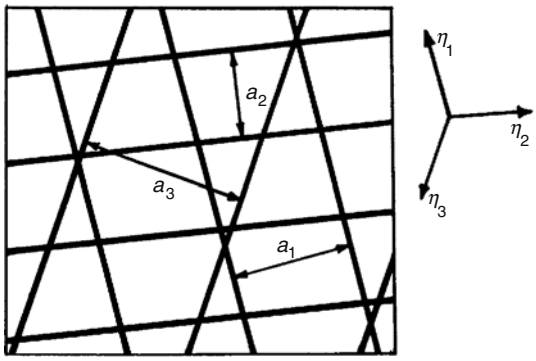

Fig. 2.12. The schematic sketch of the three independent interference pictures on the plane with different periods and intensities

It can be shown that the potentials $V(\boldsymbol{r})$ and $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$ can be represented in this situation as the quasiperiodic functions with three quasiperiods in the plane.

Let us now introduce the important definition of the "quasiperiodic group" acting on the potentials described earlier.

Definition 5. *Let us fix the directions η_1, η_2, η_3 and periods a_1, a_2, a_3 of the interference fringes in Fig. 2.12 and consider all independent parallel shifts of positions of different interference pictures in $\mathbb{R}^2$. All the potentials $V'(\boldsymbol{r})$ (and the corresponding $V_{\mathrm{B}}^{\mathrm{eff}\,\prime}(\boldsymbol{r})$) made in this way are related by the transformations of a quasiperiodic group.*

According to the definition the quasiperiodic group is a three-parametric Abelian group isomorphic to the three-dimensional torus $\mathbb{T}^3$ due to the periodicity of every interference picture.[4]

We state that potential $V(\boldsymbol{r})$ is generic if it has no periods in $\mathbb{R}^2$, is periodic if it has two linearly independent periods in $\mathbb{R}^2$, and is "partly periodic" if it has just one (up to the integer multiplier) period in $\mathbb{R}^2$.

It can also be shown that the quasiperiodic group does not change the "periodicity" of potentials $V(\boldsymbol{r})$, $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$.

The results for the Novikov problem can also be applied in this situation. We formulate here the main results for the generic potentials $V(\boldsymbol{r})$ (the special additional features can be found in [35]). Let us formulate here the theorem from [35] about the drift trajectories for the generic potentials of this kind based on the topological theorems for Novikov problem in 3-dimensional case (formulated earlier).

[4] It is obvious that the quasiperiodic group contains the ordinary translations as the algebraic subgroup.

Theorem 4 [35]. *Let us fix the value of B and consider the generic quasiperiodic potential $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$ made by three interference pictures and taking the values in some interval $\varepsilon_{\min}(B) \le V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r}) \le \varepsilon_{\max}(B)$. Then:*

(1) Open quasiclassical trajectories $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r}) = c$ always exist either in the connected energy interval

$$\varepsilon_1(B) \le c \le \varepsilon_2(B)$$

($\varepsilon_{\min}(B) < \varepsilon_1(B) < \varepsilon_2(B) < \varepsilon_{\max}(B)$) or just at one energy value $c = \varepsilon_0(B)$.

(2) For the case of the finite interval ($\varepsilon_1(B) < \varepsilon_2(B)$) all the nonsingular open trajectories correspond to topologically regular case, i.e., lie in the straight strips of the finite width and pass through them. All the strips have the same mean directions for all the energy levels $c \in [\varepsilon_1(B), \varepsilon_2(B)]$ such that all the open trajectories are on average parallel to each other for all values of c.

(3) The values $\varepsilon_1(B)$, $\varepsilon_2(B)$, or $\varepsilon_0(B)$ are the same for all the generic potentials connected by the "quasiperiodic group."

(4) For the case of the finite energy interval ($\varepsilon_1(B) < \varepsilon_2(B)$) all the nonsingular open trajectories also have the same mean direction for all the generic potentials connected by the "quasiperiodic group" transformations.

We again see that the "topologically regular" open trajectories are also generic for this situation as seen earlier.

Let us now consider the asymptotic behavior of conductivity tensor when $\tau \to \infty$ (mean free electron motion time). We consider here only the "topologically regular" case. Let us point out that the full conductivity tensor can be represented as the sum of two terms

$$\sigma_0^{ik}(B) = \sigma_0^{ik}(B) + \Delta\sigma^{ik}(B).$$

In the approximation of the drifting cyclotron orbits, the parts $\sigma_0^{ik}(B)$ and $\Delta\sigma^{ik}(B)$ can be interpreted as caused by the (infinitesimally small) difference in the electron distribution function on the same cyclotron orbit (weak angular dependence) and the (infinitesimally small) difference in the occupation of different trajectories by the centers of cyclotron orbits at different points of $\mathbb{R}^2$ (on the same energy level) as the linear response to the (infinitesimally) small external field $\boldsymbol{E}$, respectively.

The first part $\sigma_0^{ik}(B)$ has the standard asymptotic form:

$$\sigma_0^{ik}(B) \sim \frac{ne^2\tau}{m^{\mathrm{eff}}} \begin{pmatrix} (\omega_{\mathrm{B}}\tau)^{-2} & (\omega_{\mathrm{B}}\tau)^{-1} \\ (\omega_{\mathrm{B}}\tau)^{-1} & (\omega_{\mathrm{B}}\tau)^{-2} \end{pmatrix}$$

for $\omega_{\mathrm{B}}\tau \gg 1$ due to the weak angular dependence ($\sim 1/\omega_{\mathrm{B}}\tau$) of the distribution function on the same cyclotron orbit. We then have that the corresponding longitudinal conductivity decreases for $\tau \to \infty$ in all the directions in $\mathbb{R}^2$ and the corresponding condition is just $\omega_{\mathrm{B}}\tau \gg 1$ in this case.

For the part $\Delta\sigma^{ik}(B)$ the limit $\tau \to \infty$ should, however, be considered as the condition that every trajectory is passed for a rather long time by the

drifting cyclotron orbits to reveal its global geometry. Thus another parameter τ/τ_0, where τ_0 is the characteristic time of completion of close trajectories, should be used in this case and we should put the condition $\tau/\tau_0 \gg 1$ to have the asymptotic regime for $\Delta\sigma^{ik}(B)$. In this situation the difference between the open and closed trajectories plays the main role, and the asymptotic behavior of conductivity can be calculated in the form analogous to that used in [1–3] for the case of normal metals. That is:

$$\Delta\sigma^{ik}(B) \sim \frac{ne^2\tau}{m^{\text{eff}}} \begin{pmatrix} (\tau_0/\tau)^2 & \tau_0/\tau \\ \tau_0/\tau & (\tau_0/\tau)^2 \end{pmatrix}$$

in the case of closed trajectories and

$$\Delta\sigma^{ik}(B) \sim \frac{ne^2\tau}{m^{\text{eff}}} \begin{pmatrix} * & \tau_0/\tau \\ \tau_0/\tau & (\tau_0/\tau)^2 \end{pmatrix}$$

($* \sim 1$) for the case of open topologically regular trajectories if the x-axis coincides with the mean direction of trajectories.

The condition $\tau/\tau_0 \gg 1$ is much stronger than $\omega_B\tau \gg 1$ in the situation described here according to the definition of the slow drift of the cyclotron orbits. We can keep then just this condition in our further considerations and assume that the main part of conductivity is given by $\Delta\sigma^{ik}(B)$ in this limit. It is also obvious that the magnetic field B should not be "very strong" in this case.

Based on these remarks, we can now write the main part of the conductivity tensor $\sigma^{ik}(B)$ in the limit $\tau \to \infty$ for the case of topologically regular open orbits. Let us take the x-axis along the mean direction of open orbits and the y-axis orthogonal to x. The asymptotic form of σ^{ik}, $i, k = 1, 2$ can then be written as:

$$\sigma^{ik} \sim \frac{ne^2\tau}{m^{\text{eff}}} \begin{pmatrix} * & \tau_0/\tau \\ \tau_0/\tau & (\tau_0/\tau)^2 \end{pmatrix}, \quad \tau_0/\tau \to 0, \tag{2.4}$$

where $*$ is some value of the order of 1 (constant as $\tau_0/\tau \to 0$).

The asymptotic form of σ^{ik} makes possible the experimental observation of the mean direction of topologically regular open trajectories if the value τ/τ_0 is rather big.

Let us now introduce the "topological numbers" characterizing the regular open trajectories analogous to those introduced in [28] for the case of normal metals. We will first give the topological definition of these numbers using the action of the "quasiperiodic group" on the quasiperiodic potentials [35].

We assume that we have the "topologically integrable" situation where the topologically regular open trajectories exist in some finite energy interval $\varepsilon_1(B) \le c \le \varepsilon_2(B)$. According to Theorem 4 the values $\varepsilon_1(B)$, $\varepsilon_2(B)$ and the mean directions of open trajectories are the same for all the potentials constructed from our potential with the aid of the "quasiperiodic group." It also follows from the topological picture that all the topologically regular

trajectories are absolutely stable under the action of the "quasiperiodic group" for the generic $V_{\mathrm{B}}^{\mathrm{eff}\,\prime}(\boldsymbol{r})$ and can just "crawl" in the plane for the continuous action of such transformations.

We take the first interference picture (η_1, a_1) and shift continuously the interference fringes in the direction orthogonal to η_1 to the distance a_1 keeping two other pictures unchanged. At the end we will have the same potentials $V(x, y)$ and $V_{\mathrm{B}}^{\mathrm{eff}}(x, y)$ due to the periodicity of the first interference picture with period a_1. Let us fix now some energy level $c \in (\varepsilon_1(B), \varepsilon_2(B))$ and look at the evolution of nonsingular open trajectories (for $V_{\mathrm{B}}^{\mathrm{eff}}(x, y)$) while making our transformation. We know that we should have the parallel open trajectories in the plane each time and the initial picture should coincide with the final according to the construction. The form of trajectories can change during the process but their mean direction will be the same according to Theorem 4 ("topological resonance").

We can then claim that every open trajectory will be "shifted" to another open trajectory of the same picture by our continuous transformation. It is not difficult to prove that all the trajectories will then be shifted by the same number of positions n_1 (positive or negative), which depends on the potential $V_{\mathrm{B}}^{\mathrm{eff}}(x, y)$ (Fig. 2.13).

The number n_1 is always even since all the trajectories appear by pairs with the opposite drift directions.

Let us now do the same with the second and the third sets of the interference fringes and get an integer triple (n_1, n_2, n_3), which is a topological characteristic of potential $V_{\mathrm{B}}^{\mathrm{eff}}(x, y)$ (the "positive" direction of the numeration of trajectories should be the same for all these transformations).

The triple (n_1, n_2, n_3) can be represented as:

$$(n_1, n_2, n_3) \;=\; M\,(m_1, m_2, m_3),$$

where $M \in \mathbb{Z}$ and (m_1, m_2, m_3) is the indivisible integer triple.

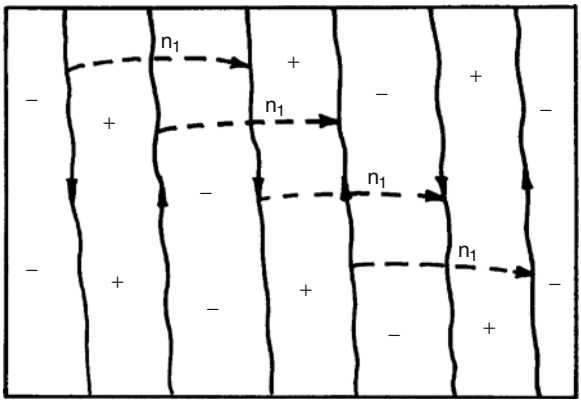

Fig. 2.13. The shift of "topologically regular" trajectories by a continuous transformation generated by the special path in the "quasiperiodic group"

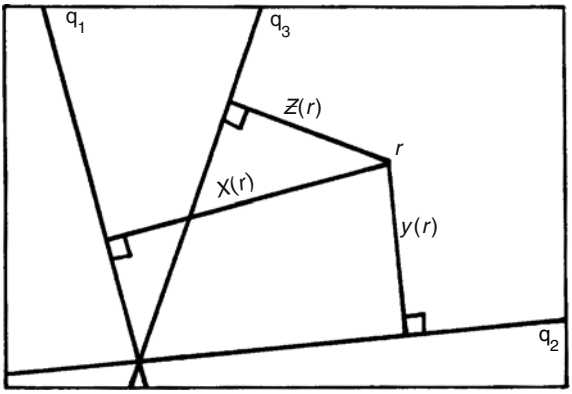

Fig. 2.14. The functions $X(\boldsymbol{r})$, $Y(\boldsymbol{r})$, and $Z(\boldsymbol{r})$ on the plane

The numbers (m_1, m_2, m_3) (defined up to the common sign) play now the role of "topological numbers" for this situation. For direct experimental observation of these numbers, the connection between these numbers and the mean direction of the "topologically regular" trajectories can play an important role. This connection is described as follows.

Let us draw three straight lines q_1, q_2, q_3 with the directions η_1, η_2, η_3 (Fig. 2.12) and choose the "positive" and "negative" half-planes for every line q_i on the plane. Let us now consider three linear functions $X(\boldsymbol{r})$, $Y(\boldsymbol{r})$, $Z(\boldsymbol{r})$ on the plane that are the distances from the point $\boldsymbol{r}$ to the lines q_1, q_2, q_3 with the signs "+" or "−" depending on the half-plane for the corresponding line q_i (Fig. 2.14). Let us choose here the signs "+" or "−" such that the gradients of $X(\boldsymbol{r})$, $Y(\boldsymbol{r})$, $Z(\boldsymbol{r})$ coincide with directions of shifts of the corresponding interference pictures in the definition of (m_1, m_2, m_3).

Theorem 5 [35]. *Consider the functions*

$$X'(\boldsymbol{r}) = X(\boldsymbol{r})/a_1, \quad Y'(\boldsymbol{r}) = Y(\boldsymbol{r})/a_2, \quad Z'(\boldsymbol{r}) = Z(\boldsymbol{r})/a_3$$

in $\mathbb{R}^2$. *The mean direction of the regular open trajectories is given by the linear equation:*

$$m_1 X'(x,y) + m_2 Y'(x,y) + m_3 Z'(x,y) = 0, \tag{2.5}$$

where (m_1, m_2, m_3) *is the indivisible integer triple introduced earlier.*

Let us now describe the situation with four independent sets of interference fringes in the plane (see also [35]). In general we get here the quasiperiodic potentials $V(\boldsymbol{r})$, $V_{\mathrm{B}}^{\mathrm{eff}}(\boldsymbol{r})$ with four quasiperiods. The situation in this case is more complicated than in the case $N = 3$ and no general classification of open trajectories exists at the time. At the moment only the theorem analogous to Zorich result can be formulated in this situation [32]. According to the Novikov theorem we can claim that the "small perturbations" of purely

periodic potentials having four quasiperiods have the "topologically regular" level curves like in the previous case.

The purely periodic potentials $V(\boldsymbol{r})$ give the same dense set in the space of parameters η_1, η_2, η_3, η_4, a_1, a_2, a_3, a_4 and can be found in any small open region of this space. The Novikov theorem claims then that every potential of this kind can be surrounded by the "small open ball" in the space of parameters η_1, η_2, η_3, η_4, a_1, a_2, a_3, a_4 where the open level curves will always demonstrate the "topologically regular" behavior. The set of potentials thus obtained has finite measure among all potentials and the "topologically regular" open trajectories can be found with finite probability also in this case. However, we do not claim here that the chaotic behavior has measure 0 for four quasiperiods and moreover we also expect the nonzero probability for the chaotic trajectories in this more complicated case.

The topologically regular cases demonstrate here the same "regularity properties" as in the previous case including the "Topological numbers." Thus, we can introduce in the same way the action of the quasiperiodic group on the space of potentials with four quasiperiods and define in the same way the four tuples (m_1, m_2, m_3, m_4) of integer numbers characterizing the topologically regular cases in this situation.

Also, the analogous theorem about mean directions of the regular trajectories can be formulated in this case. Namely, if we introduce the functions $X(\boldsymbol{r})$, $Y(\boldsymbol{r})$, $Z(\boldsymbol{r})$, $W(\boldsymbol{r})$ in the same way as for the case of three quasiperiods (above) and the corresponding functions

$$X'(\boldsymbol{r}) = X(\boldsymbol{r})/a_1, \ \ \ldots \ \ , W'(\boldsymbol{r}) = W(\boldsymbol{r})/a_4,$$

we can write the equation for the mean direction of open trajectories on the plane in the form:

$$m_1 X'(\boldsymbol{r}) + m_2 Y'(\boldsymbol{r}) + m_3 Z'(\boldsymbol{r}) + m_4 W'(\boldsymbol{r}) = 0.$$

The numbers (m_1, m_2, m_3, m_4) are stable with respect to the small variations of η_1, η_2, η_3, η_4, a_1, a_2, a_3, a_4 (and the intensities of the interference pictures I_1, I_2, I_3, I_4) and correspond again to some "stability zones" in this space of parameters.

A brief mention is now made about the limit of Novikov problem for large values of N. The following problem can be formulated as:

Give a description of global geometry of the open level curves of quasiperiodic function $V(\boldsymbol{r})$ in the limit of large numbers of quasiperiods.

We can claim that the open level curves should exist here also in the connected energy interval $[\varepsilon_1, \varepsilon_2]$ on the energy scale, which can degenerate just to one point ε_0.[5] We expect that the "topologically regular" open trajectories can also exist in this case. However the probability of "chaotic behavior" should increase for the cases of large N, which is closer now to random potential situation. The corresponding behavior can be considered then as the

[5] The proof given in [24] for the case of 3 quasiperiods works actually for any N.

"percolation problem" in special models of random potentials given by quasiperiodic approximations. Certainly, this model can be quite different from the others. Nevertheless, we expect a similar behavior of the chaotic trajectories for rather big N also in this rather special model. This area, however, is still under investigation.

References

1. Lifshitz, I.M., Azbel, M.Ya., Kaganov, M.I.: Sov. Phys. JETP **4**, 41 (1957)
2. Lifshitz, I.M., Peschansky, V.G.: Sov. Phys. JETP **8**, 875 (1959)
3. Lifshitz, I.M., Peschansky, V.G.: Sov. Phys. JETP **11**, 137 (1960)
4. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **8**, 383 (1959)
5. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **9**, 311(1959)
6. Alexeevsky, N.E., Gaidukov, Yu.P.: I.M.Lifshitz, V.G.Peschansky. Sov. Phys. JETP **12**:5, 837 (1960)
7. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **10**, 481 (1960)
8. Gaidukov, Yu.P.: Sov. Phys. JETP **10**, 913 (1960)
9. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **14**(2), 256 (1962)
10. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **15**(1), 49 (1962)
11. Alexeevsky, N.E., Gaidukov, Yu.P.: Sov. Phys. JETP **16**(6), 1481 (1963)
12. Lifshitz, I.M., Kaganov, M.I.: Sov. Phys. Usp. **2**, 831 (1960)
13. Lifshitz, I.M., Kaganov, M.I.: Sov. Phys. Usp. **5**, 411 (1962)
14. Lifshitz, I.M., Azbel, M.Ya., Kaganov, M.I.: Electron Theory of Metals. Moscow, Nauka (1971) (Translated: Consultants Bureau, New York, 1973)
15. Abrikosov, A.A.: Fundamentals of the Theory of Metals. "Nauka", Moscow (1987) (Translated: North-Holland, Amsterdam, 1998)
16. Novikov, S.P.: Russian Math. Surveys **37**, 1 (1982)
17. Zorich, A.V.: Russian Math. Surveys **39**, 287 (1984)
18. Dynnikov, I.A.: Russian Math. Surveys **57**, 172 (1992)
19. Dynnikov, I.A.: Russian Math. Surveys **58** (1993)
20. Dynnikov, I.A.: "A proof of Novikov's conjecture on semiclassical motion of electron." Math. Notes **53**:5, 495 (1993)
21. Novikov, S.P.: Proc. Steklov Inst. Math. 1 (1986)
22. Novikov, S.P.: "Quasiperiodic structures in topology". Proc. Conference "Topological Methods in Mathematics", dedicated to the 60th birthday of J.Milnor, June 15–22, S.U.N.Y. Stony Brook, 1991. Publish of Perish, Houston, TX, pp. 223–233 (1993)
23. Novikov, S.P.: Proc. Conf. of Geometry Tel Aviv University (1995) December 15–26, 1993
24. Dynnikov, I.A.: PhD Theses, Moscow State University, Dept. of Math. and Mech., Scientific Supervisor – S.P. Novikov, Moscow (1996)
25. Leo, R.D.: PhD Theses. University of Maryland. Department of Math., Scientific Supervisor – S.P. Novikov, College Park, MD 20742, USA
26. Dynnikov, I.A.: "Semiclassical motion of the electron. A proof of the Novikov conjecture in general position and counterexamples." Editors: V.M.Buchstaber, S.P.Novikov. Advances in the Mathematical Sciences. Solitons, Geometry, and Topology: On the Crossroad. American Mathematical Society Translations, Series 2, Vol. 179 (1997)

27. Dynnikov, I.A.: Russian Math. Surveys **54**, 21 (1999)
28. Novikov, S.P., Maltsev, A.Ya.: ZhETP Lett. **63**, 855 (1996)
29. Novikov, S.P., Maltsev, A.Ya.: Physics-Uspekhi **41**(3), 231 (1998)
30. Maltsev, A.Ya., Novikov, S.P.: ArXiv: math-ph/0301033, Bulletin of Braz. Math. Soc., New Series 34 (1), 171–210 (2003)
31. Maltsev, A.Ya., Novikov, S.P.: ArXiv: cond-mat/0304471
32. Novikov, S.P.: Russian Math. Surveys **54** (3), 1031 (1999)
33. Beenakker, C.W.J.: Phys. Rev. Lett. **62**, 2020 (1989)
34. Grant, D.E., Long, A.R., Davies, J.H.: Phys. Rev. B **61** (13), 127 (2000)
35. Maltsev, A.Ya.: ArXiv: cond-mat/0302014
36. Tsarev, S.P.: Private communication. (1992–93)
37. Maltsev, A.Ya.: ZhETP **85**, 934 (1997)
38. Weiss, D., Klitzing, K.V., Ploog, K., Weimann, G.: Europhys. Lett. **8** (2), 179 (1989)
39. Alves, E.S., Beton, P.H., Henini, M.: L. Eaves, P.C. Main, O.H. Hughes, G.A. Toombs, S.P. Beaumont, C.D.W. Wilkinson. J. Phys. Condens. Matter **1**, 8257 (1989)
40. Ismail, K., Antoniadis, D.A., Smith, H.I., Liu, C.T.: K. Nakamura, D.C. Tsui. J. Vac. Sci. Technol. B **7**, 2000 (1989)
41. Ismail, K., Smith III, T.P., Masselink, W.T., Smith, H.I.: Appl. Phys. Lett. **55**, 2766 (1989)
42. Fang, H., Stiles, P.J.: Phys. Rev. B **41** (10), 171 (1990)
43. Toriumi, A., Ismail, K., Burkhardt, M., Antoniadis, D.A., Smith, H.I.: Phys. Rev. B **41** (12), 346 (1990)
44. Puechner, R.A., Ma, J., Mezenner, R., Liu, W.-P., Kriman, A.M., Maracas, G.N., Bernstein, G., Ferry, D.K., Chu, P. Wieder, H.H., Newman, P.: Surf. Sci. **228**, 520 (1990)
45. Weiss, D., Klitzing, K.V., Ploog, K.: and G. Weimann, Surf. Sci. **229**, 88 (1990)
46. Gerhardts, R.R., Weiss, D., Wulf, U.: Phys. Rev. B **43**, 5192 (1991)
47. Davies, J.H., Larkin, I.A.: Phys. Rev. B **49**, 4800 (1994)
48. Larkin, I.A., Davies, J.H., Long, A.R., Cusco, R.: Phys. Rev. B **56** (15), 242 (1997)
49. Davies, J.H., Petticrew, D.E., Long, A.R.: Phys. Rev. B **58** (10), 789 (1998)
50. Dynnikov, I.A.: "Surfaces in 3-Torus: Geometry of plane sections." Proc. ECM2 (Budapest, July 1996). Progress in Mathematics, Vol. 168, pp. 162–177, Birkhauser, Basel, 1998
51. Dynnikov, I.A., Maltsev, A.Ya.: JETP **85**, 205 (1997)
52. Zorich, A.V.: Proc. "Geometric Study of Foliations"/ed. T. Mizutani et al. World Scientific, Singapore: 479–498 (1994) (Tokyo, November 1993)

3

The Role of Topology in Growth and Agglomeration

R. Kerner

Summary. We describe several models of growth of atomic structures, mostly by agglomeration from gas or from liquid. In many physical situations, like the formation of fullerenes, carbon nanotubes and onions, as well as in growth of quasicrystals or glass formation, a very important role is played by topology of local configurations. This chapter shows how some physical properties of certain forms of condensed matter can be derived from simple topological and geometrical considerations concerning the structure of local configuration space of atoms and molecules.

3.1 Introduction

The role of topology in physics of condensed matter has been steadily growing since the second half of the twentieth century. It is often essential in the analysis of geometrical structures, which enables us to derive subsequently many physical properties of various systems. Like in classical problems of mechanics, the analysis is divided into two fundamental parts: first *kinematics*, which defines the space of possible nonredundant motions of the system under investigation, second *dynamics*, which defines the relationships between the forces and the motions they provoke. In the case of crystalline lattices, the task is greatly simplified, because the kinematical part is largely constrained: once the lattice structure is settled, each atom has a well-defined phase space in which it can move, and its motions can be described by harmonic (or anharmonic) oscillator regime. Also the number of closest and next-to-the closest neighbors is strictly the same for the same atoms (it may vary from one kind of atom to another in many-component lattices, but the complication is not very great). This enables one to analyze mutual forces and interactions between the neighbors, and many excellent models, with the Ising model and its derivatives as best examples, have been constructed and have brought a very good understanding of various physical phenomena in solids.

The situation is radically different in amorphous solids and glasses. Here the number of neighbors is never well defined – only the averages can be taken as reliable parameters, whereas local situations of single atoms represent an

infinitely rich variety. Moreover, in order to foresee the local configuration that will represent (a local) free energy minimum, we must know an approximate expression for free energy there; but in order to define it, we should know with how many neighbors the given atom interacts, thus making the whole problem hopelessly intricate. The situation has some similarity to general relativity, in which the behavior of matter defines the geometry of the space time, but in order to describe the behavior of matter correctly (i.e., in order to solve the equations of motion), one has to know first what is the underlying geometry of space time. It is well known that except for very special cases, one can get only approximate solutions, starting from simplified situations in which one imposes the geometry first, then solves the (approximate) equations of motion, then modifies the geometry solving Einstein field equations with the matter source behavior obtained in previous stage, and so on.

As in any situation when a direct solution is not at hand, we must content ourselves with some sort of approximations. In the case of amorphous networks this means that we should abandon the exact description in terms of well-defined positions of individual atoms, introducing statistical description in terms of average values of the most important data characterizing typical local situations in which atoms can be found. The probabilities of finding particular short-range and medium-range configurations should depend on the energy stored in bonds and atoms involved, and also on the temperature in which the growth is occurs, through corresponding Boltzmann factors.

In this chapter we present several examples of this approach, leading to fairly good physical predictions concerning fullerenes and structural glasses. We start by recalling simple and powerful topological laws ruling two-dimensional networks. Euler's theorem which relates between them the numbers of summits, edges, and faces in convex polyhedra, provides a very strong constraint on any networks, in the plane, or on a sphere. The connection with local curvature gives a hint of the energy barriers resulting from purely topological properties of local configurations in networks. Applying these simple ideas, and generalizing them to three-dimensional cases, we are able to set forth several models of stochastic agglomeration and growth of random or self-organized networks leading to fairly good physical predictions concerning, among others, fullerenes and structural glasses.

3.2 Topology and Geometry of Polygon Tilings and Networks

Among all possible tilings of the Euclidean plane, we consider only the very simple case of *three-coordinate* networks with constant edge length. These tilings are formed exclusively of *equilateral* polygons; at each vertex, three polygons meet. It results from Euler's formula that the average number of sides of a polygon it this network must be equal to 6:

$$\frac{1}{\langle N_c \rangle} + \frac{1}{\langle N_f \rangle} = \frac{1}{2}. \tag{3.1}$$

This simple and important result is derived as a limiting case of Euler's formula relating the number of summits, edges, and faces of any convex polyhedron topologically equivalent to a sphere:

$$S - E + F = 2, \tag{3.2}$$

where S is the number of summits, E the number of edges, and F the number of faces. The right-hand side of this equation is called Euler's characteristic of a sphere. The same number for a torus is 0, and for a "pretzel" with two holes, it would be -2.

The proof of Euler's formula is well known and based on induction. More complicated polyhedra can always be produced from simpler ones by performing several elementary operations, like slicing one of the summits, thus creating a new face and new edges and summits, or adding a new summit or an edge by dividing one of the faces into two. In each of these cases the result of formula (3.2) remains exactly the same. For example, if we divide one of the faces into two, joining with a straight line (new edge) two of its summits, then the number of summits remains unchanged ($S \to S$, the number of edges E grows by 1, but so does the number of faces F, and $S - E + F$ remains constant). If we slice one of the summits at which M edges meet, we create a *new face* with M new summits (so the number of summits grows by $M - 1$, because the original one has disappeared), and the number of faces has grown by 1; finally, the number of edges has grown by M, all the M edges of the newly created face. Again, the expression $S - E + F$ remains equal to 2, as before.

In general, a two-dimensional sphere can be approached by a polyhedron inscribed in it, containing a certain number of k-sided faces F_k, with $k = 3, 4, ...N_f^{\max}$, ($N_f^{\max}$ denoting the faces (polygons) of maximal number of sides; in common cases $N_f^{\max}$ is rarely higher than 8); with S_m the number of m-coordinate summits (vertices), also with $m = 3, 4, ...N_c^{\max}$. The total number of faces F and the total number of summits S is given by

$$F = \sum_{k=3}^{N_f^{\max}} F_k\,, \qquad S = \sum_{m=3}^{N_c^{\max}} S_m. \tag{3.3}$$

Let us define the *average coordinate number* and the *average coordination number* as follows:

$$\langle N_f \rangle = \frac{1}{F} \sum_{k=3}^{N_f^{\max}} k\, F_k\,, \qquad \langle N_c \rangle = \frac{1}{S} \sum_{m=3}^{N_c^{\max}} m\, S_m. \tag{3.4}$$

The number of edges can be computed in two different ways: either we count m times all the m-coordinate summits (vertices), $\sum S_m$, and each edge will be counted twice, because it always belongs simultaneously to two summits; or we can count k times all the k-sided faces, and again, each edge will

be counted twice, because it is always shared by two adjacent faces. Therefore we get:

$$2E = \sum kF_k = \sum mS_m \tag{3.5}$$

and we can write

$$F\,\langle N_f\rangle = S\,\langle N_c\rangle = 2\,E. \tag{3.6}$$

This relation enables us to express Euler's formula in terms of only one variable, say F, substituting

$$S = \frac{\langle N_f\rangle}{\langle N_c\rangle}\,F\,, \qquad E = \frac{1}{2}\,\langle N_f\rangle\,F,$$

thus obtaining

$$S - E + F = \left[\,\frac{\langle N_f\rangle}{\langle N_c\rangle} - \frac{1}{2}\,\langle N_f\rangle + 1\right] F = X, \tag{3.7}$$

where X is the Euler–Poincaré characteristic determining the topology of the underlying surface, 2 for a sphere, 0 for a torus, -2 for a "pretzel" with two holes, etc.; in general, $X = 2 - h$, with h equal to the number of holes. A more symmetric form of formula (3.7) can be obtained dividing its both sides by $\langle N_f\rangle\,F$:

$$\frac{1}{\langle N_c\rangle} + \frac{1}{\langle N_f\rangle} = \frac{1}{2} + \frac{X}{\langle N_f\rangle\,F} \tag{3.8}$$

or, by virtue of the identity (3.6),

$$\frac{1}{\langle N_f\rangle} + \frac{1}{\langle N_c\rangle} = \frac{1}{2} + \frac{X}{\langle N_c\rangle\,S}. \tag{3.9}$$

We have changed the order of the right-hand side expression on purpose: now it is clear that both formulae prove the existence of *duality* between the faces (F) and summits (S). Any solution of (3.8) generates a solution of (3.9) by interchanging the numbers F and S, and correspondingly, $\langle N_f\rangle$ with $\langle N_c\rangle$. The two corresponding solutions are called *dual tilings of the sphere*. In the case of *homogeneous* tilings, with all summits, edges, and faces identical, one can replace the averages by exact numbers, which become all integers. The solutions have been known since antiquity and are called *platonic regular polyhedra* (Fig. 3.1). These are the following *integer* solutions of (3.8) (with $X = 2$):

$$\frac{1}{N_c} + \frac{1}{N_f} = \frac{1}{2} + \frac{2}{N_f\,F} \tag{3.10}$$

$$N_c = 3, \qquad N_f = 3, \qquad F = 4 \quad \text{the tetrahedron},$$

$$N_c = 3, \qquad N_f = 4, \qquad F = 6 \quad \text{the cube},$$

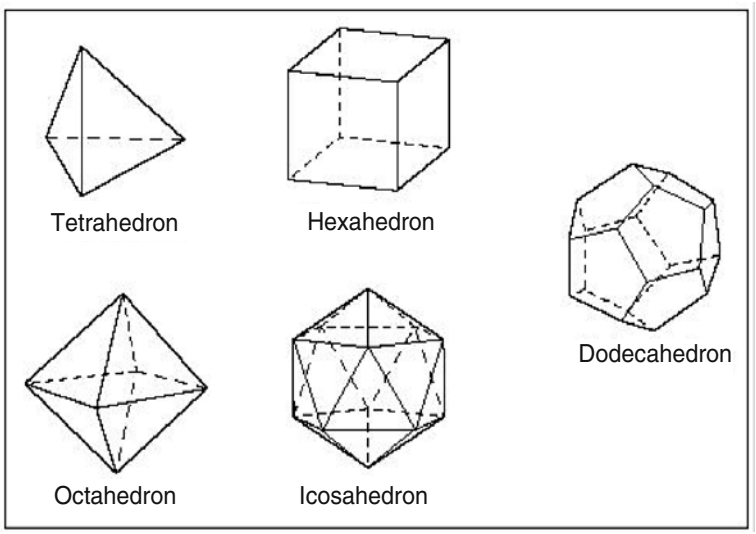

Fig. 3.1. The five Platonic polyhedra

$$N_c = 4, \quad N_f = 3, \quad F = 8 \quad \text{the octahedron,}$$
$$N_c = 3, \quad N_f = 5, \quad F = 12 \quad \text{the dodecahedron,}$$
$$N_c = 5, \quad N_f = 3, \quad F = 20 \quad \text{the icosahedron.}$$

According to this definition, the cube and the octahedron are dual to each other, as is the dodecahedron with the icosahedron; the tetrahedron is dual to itself. Of course, by admitting more than one kind of faces and more than one kind of summits (vortices), one can create infinitely many convex polyhedra, respecting Euler's formula [1].

A similar homogeneous tiling of a torus with regular identical polygons should verify relation (3.10) with $X = 0$. Here are the only three possible solutions:

$$N_c = 3, \quad N_f = 6, \quad \text{three-coordinate hexagonal lattice,}$$
$$N_c = 6, \quad N_f = 3, \quad \text{six-coordinate triangular lattice,}$$
$$N_c = 4, \quad N_f = 4, \quad \text{four-coordinate square lattice.}$$

It is worthwhile to note at this point that when the number of faces becomes very large, i.e., when $F \to \infty$, be it on a sphere or on another surface with more complicated topology, formula (3.10) will lead to the same solution as for the torus, because the last term on the right-hand side becomes negligible. Therefore the homogeneous regular tilings of the torus are the same as those of any manifold topologically equivalent with the Euclidean plane.

On the torus, the number of faces F does not appear in the formula, and is therefore arbitrary; on the Euclidean plane F is simply infinite.

This result is of a purely topological character; however, it also remains valid for metric (Euclidean) geometry. The important corollary to the last result can be drawn from the assumption that all polygons are equilateral and equiangular. Then we see that the only infinite homogeneous tilings of the plane must be composed of perfect squares, triangles, or hexagons, and all are three coordinate. By the same token we prove another important result, namely that the only angles that can be found in crystalline lattices are 60°, 90° , and 120°. What is amazing is that this result remains valid also in three dimensions, although there is a new free parameter, the number of faces meeting at one edge (which in two dimensions is fixed and always equal to 2) [2].

Combining topology with simple properties of Euclidean geometry enables one to find all possible *homogeneous* plane tilings made of perfect (i.e., equilateral and equiangular) polygons. A network is called *homogeneous* if all its vertices are identical; but now different equilateral polygons may meet at each of the vertices [3].

Let us start with three-coordinate lattices. At each vertex, three perfect polygons meet; if their number of sides is k_1, k_2 , and k_3, the sum of their respective angles must be equal to 2π. The angles being that of perfect polygons, this condition leads to the equation

$$\frac{(k_1-2)\,\pi}{k_1}+\frac{(k_2-2)\,\pi}{k_2}+\frac{(k_3-2)\,\pi}{k_3}=2\,\pi. \tag{3.11}$$

This equation has only four solutions in integers k_1, k_2, k_3:

$$(6,6,6),\quad (4,8,8),\quad (3,12,12),\quad (4,6,12).$$

Similarly, one can easily find that there exist only four homogeneous perfect polygon tilings with coordination number $N_c = 4$. They satisfy the condition

$$\frac{(k_1-2)\,\pi}{k_1}+\frac{(k_2-2)\,\pi}{k_2}+\frac{(k_3-2)\,\pi}{k_3}+\frac{(k_4-2)}{k_4}=2\,\pi. \tag{3.12}$$

The corresponding tilings are displayed in Fig. 3.2. It is remarkable that despite the variety of polygon types and their relative display, the overall symmetry of resulting lattices always falls into the well-known crystalline symmetries: cubic, hexagonal, or rhombohedric.

The validity of Euler's formula for the plane (3.1) can also be checked by another method—introducing a very useful concept of *relative frequencies* of polygons of a given number of sides. Consider an arbitrary planar network formed with polygons of various types, not necessarily equiangular (in fact, they may also be nonequilateral). Let $N_f^{\max}$ be the maximal number of faces in polygons encountered in the network, and let P_k denote the relative

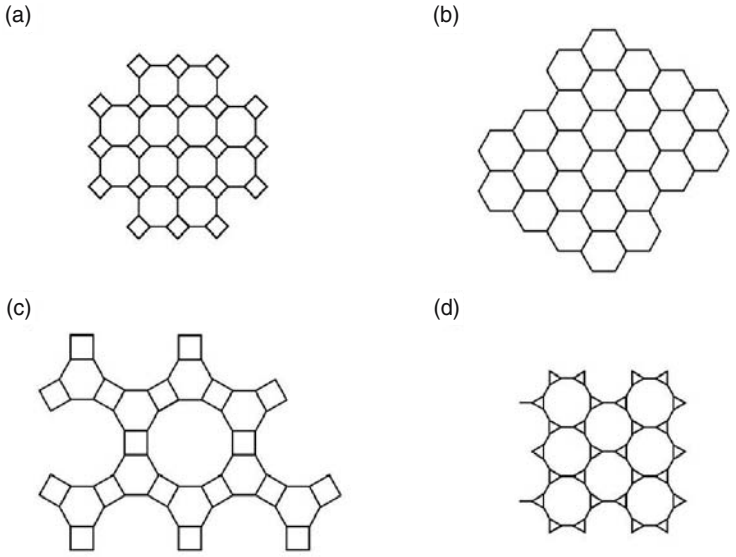

Fig. 3.2. Four three-coordinate homogeneous perfect planar tilings

frequency of occurence of k-sided polygons in the network. By definition,

$$\sum_{k=3}^{N_f^{\max}} P_k = 1. \tag{3.13}$$

With the variables P_k denoting the *probabilities* of finding a k-sided polygon among all the polygons in the network, the normalization condition (3.13) leaves only $N_f^{\max} - 1$ independent variables. In infinite planar networks, these are still subjected to another constraint resulting from Euler's formula:

$$\sum_{k=3}^{N_f^{\max}} k\, P_k = 1 = \langle N_f \rangle = \frac{2\,\langle N_c \rangle}{\langle N_c \rangle - 2}, \tag{3.14}$$

which is just another form of relationship (3.1). Looking at Figs. 3.3 and 3.4, we can easily compute the probabilities and check formula (3.14), in which $\langle N_c \rangle$ can be replaced by N_c because we consider homogeneous tilings. The case of pure hexagonal lattice is obvious, because $\langle N_f \rangle = N_f = 6$, $P_6 = 1$, $N_c = 3$, and the formula is obviously satisfied. For more complicated networks, we have for the three-coordinate polygon tiling $(4, 8, 8)$, $P_4 = \frac{1}{2}$ and $P_8 = \frac{1}{2}$; again, $\langle N_f \rangle = 4\,P_4 + 8\,P_8 = \frac{4}{2} + \frac{8}{2} = 6$. For the tiling with vertices $(4, 6, 12)$ we have $P_4 = \frac{1}{2}$, $P_6 = \frac{1}{3}$, and $P_{12} = \frac{1}{6}$; again, as it easy to check $\langle N_f \rangle = 6$.

If a three-coordinate lattice on a plane is formed exclusively with three types of polygons—pentagons, hexagons, and heptagons—then the statistics

of these polygons obeys one simple rule, which states that $P_5 = P_7$, whereas P_6 remains as the only free parameter. The proof is obvious: as we know, in an infinite planar lattice, when $N_c = 3$, we must have $\langle N_f \rangle = 6$. In the case considered here, this means that

$$5\,P_5 + 6\,P_6 + 7\,P_7 = 6. \tag{3.15}$$

But here $P_6 = 1 - P_5 - P_7$, which after substitution leads directly to $P_7 - P_5 = 0$.

One can arrive at the same conclusion in a more qualitative manner. We already know that a three-coordinate lattice formed exclusively by perfect hexagons is an admissible infinite tiling of Euclidean plane. A pentagon or a heptagon may be inserted in such a lattice, with their angles departing from the perfect ones (and with a slight deformation of angles of the adjacent hexagons). The sum of the angles in a pentagon is equal to $3\pi = 540^{\circ}$, and divided by 5 gives the average of $\frac{3\pi}{5} = 108^{\circ}$, the angle of perfect pentagon. This is 12° less than the 120° of a perfect hexagon. When found in a common vertex in which three polygons of the network meet, it creates an *angular deficit* of 12°, which can be geometrically interpreted as local positive curvature. Of course on a plane it is impossible, and the angles cannot keep their "perfect" values—the polygons have to adapt themselves in order to make the sum of three angles be equal to 2π. In a heptagon, the sum of the angles is equal to $5\pi = 900^{\circ}$, thus creating an *angular excess* at each of the seven angles equal to $\frac{7\pi}{5} - \frac{2\pi}{3}$, equivalent with creating negative local curvature in a vertex where a heptagon meets two hexagons. Each heptagon contributes to seven vertices, whereas each pentagon is shared by five vertices. It is easy to see that the sums of local angular excesses and angular deficits created by these two polygons mutually cancel themselves:

$$7 \times \left(\frac{5\pi}{7} - \frac{4\pi}{6} \right) + 5 \times \left(\frac{3\pi}{5} - \frac{4\pi}{6} \right) = 0. \tag{3.16}$$

It can be said that inserting an equal amount of pentagons and heptagons into a regular hexagonal lattice, which can also be viewed as creating local *defects* in a perfect tiling, does not alter globally the "flat" topology of the plane. This means that a plane can be covered not only by hexagons forming a well-known regular pattern, but also with pentagons and heptagons, provided that their numbers are strictly equal. It can be done in an ordered manner, thus creating one of the admitted crystalline symmetries, or in a chaotic way, leading to an amorphous (aperiodic) structure, like shown in Fig. 3.3.

In what follows, we use these topological and geometrical effects in a model describing actual growth and formation of real atomic structures.

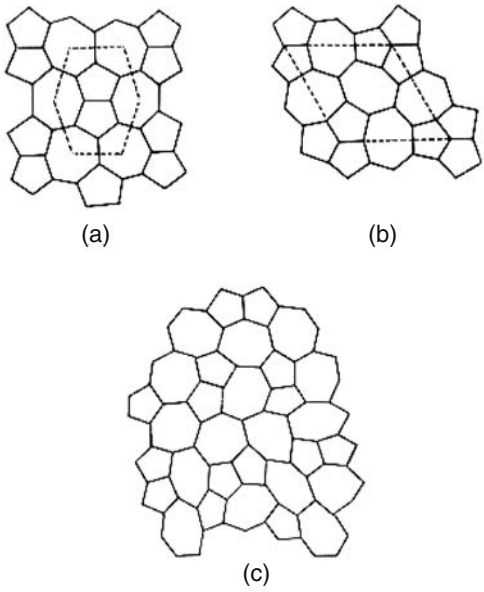

Fig. 3.3. Three possible planar lattices made of five-sided and seven-sided equilateral polygons: (a) and (b) display crystalline symmetry, (c) is a random lattice.

3.3 Dynamical Model of Polygon Agglomeration in Two Dimensions

An important category of crystalline or amorphous atomic structures have a pronounced one-dimensional or two-dimensional character. Physical properties of many polymers and liquid crystals (especially the so-called *nematics*) can be explained successfully with the well-known one-dimensional models, of which the Ising model is most celebrated. Two-dimensional structures are also common in many physical systems, in particular in high-temperature supraconductivity. Monoatomic and monomolecular layers play a growing role in modern technology. They are often obtained by special growth and agglomeration techniques, and the knowledge of the geometry of local configurations as well as the energy landscape resulting from the interactions between the closest and next-to-closest neighbors is essential for understanding the resulting structures [4].

The effects certain combinations of polygons provoke by creating local curvature, positive or negative, must have an important influence on the energy stored in the network. One may express the same idea by stating that creating local curvature must cost some extra effort as compared with totally flat lattice tiling a plane. The energy costs would be different, of course, if

the support on which the network is created were no more flat: then it might happen that a locally curved two-dimensional network is better adapted and costs less energy than a flat one.

Let us consider the simplest case in two dimensions, with a three-coordinate network consisting exclusively of five-sided, six-sided, and seven-sided equilateral polygons. As we know, on the Euclidean plane tri-coordinate and purely hexagonal lattice displays perfect translational symmetry and can be extended to infinity. In contrast, pentagons and heptagons found in such a lattice represent local *defects*; it is therefore logical to assume that their presence in the otherwise hexagonal lattice creates local stress equivalent to certain energy cost. Whatever its value, we know that in order to be able to produce an infinite tiling of a plane, the total number of pentagons must be equal to that of heptagons.

Now, if someone wishes to produce a tiling of the plane using completely at random these three types of equilateral polygons, the numbers of pentagons and heptagons would asymptotically tend to the same limit, with certain (arbitrary) amount of hexagons, the only relevant global characterization of this random network. However, if one wishes to analyze the network in more detail, one should look for correlations between the closest polygons. As can be easily seen in Fig. 3.3, with the same global probabilities $P_5 = P_7 = \frac{1}{2}$ one can produce quite different results. The difference can be immediately felt if one considers the statistics of *doublets*, i.e., pairs of edge-sharing polygons. One finds easily that in the lattice (a) one has $P_{55} = 8.33\%$, $P_{57} = 66\%, 67\%$, and $P_{77} = 25\%$, whereas in the lattice (b) one has $P_{55} = 25\%$, $P_{57} = 45\%$, and $P_{77} = 30\%$. Of course, in this particular example there are no doublets containing a hexagon, because here $P_6 = 0$. In a completely random lattice the six probabilities of doublets would behave as the corresponding binomial distribution:

$$(P_5 + P_6 + P_7)^2 = (P_5^2 + 2P_5P_6 + 2P_5P_7 + P_6^2 + 2P_6P_7 + P_7^2) = 1, \quad (3.17)$$

which can be interpreted as a doublet frequency distribution:

$$\begin{gathered} P_{55} = P_5^2, \quad P_{56} = 2\,P_5P_6, \quad P_{57} = 2\,P_5P_7, \\ P_{66} = P_6^2, \quad P_{67} = 2P_6P_7, \quad P_{77} = P_7^2. \end{gathered} \quad (3.18)$$

By construction, these probabilities are normalized to 1.

But if the real energy barriers exist for creation of such pairs, they will alter this ideally random distribution of doublets. Let us evaluate such a distribution, assuming that a real network is produced at finite temperature T. Following [5], let us consider the simplest linear approximation, in which the energy cost of creating a defect is the same for a pentagon as for a heptagon, denoted by ΔE in standard units kT, where k is Boltzmann's constant and T the absolute temperature in K (kelvin). We also assume that these energies add up following the curvature effect created by two neighboring defects. This

means that two edge-sharing pentagons must cost the energy $2\,\Delta E$, two edge-sharing heptagons also cost $2\,\Delta E$, but a pentagon sharing an edge with a heptagon cancels its curvatures, and the corresponding energy is roughly the same as for two hexagons, i.e., 0 as compared with the "ideal" hexagonal lattice, as for any two edge-sharing hexagons.

If a random three-coordinate network is created by agglomeration at given constant temperature T, the probability distribution of five-sided, six-sided and seven-sided polygons will be affected by the energy barriers related to the local curvature effect they produce. Let us suppose that the probability of creation of a single polygon of given type is determined by chemical composition, temperature, etc.; let us consider these probabilities as the input of our model of agglomeration of polygons, and let us denote them by

$$P_5^{(0)}\,,\ P_6^{(0)},\ \text{and}\ P_7^{(0)}.$$

By definition, $P_5^{(0)} + P_6^{(0)} + P_7^{(0)} = 1$, so that there are only two independent parameters, e.g., P_5 and P_6 whose domain of variation is the triangle (simplex) on the P_5, P_7 plane. The extremal points of the triangle correspond to "pure" configurations with $P_5 = 1, P_6 = P_7 = 0$, or $P_7 = 1, P_5 = P_6 = 0$, or $P_6 = 1, P_5 = P_7 = 0$. Forming pairs of edge-sharing polygons may be considered as a first step toward agglomeration; this is why we denote the resulting probability distribution of doublets by $P_{ik}^{(1)}$ with $i, k = 5, 6, 7$.

Let us incorporate the energy barriers resulting from the corresponding stresses provoked by local (positive or negative) curvatures into the probablities of the corresponding doublets. We assume that the probability of a $5-6$ couple, which represents one standard departure from flatness, should contain a Boltzmann factor $\mathrm{e}^{-\frac{\Delta E}{k_\mathrm{B}T}}$; the same should be true for the probability of the doublet $P_{67}^{(1)}$. The doublets 55 and 77 should be affected by Boltzmann factor $\mathrm{e}^{-\frac{2\Delta E}{k_\mathrm{B}T}}$, because they represent *two* standard deviations from flatness; finally, the doublets 66 and 57 should be considered as representing zero-energy cost, therefore the corresponding Boltzmann factor reduces to 1. Let us introduce the shortened notation: $\frac{\Delta E}{k_\mathrm{B}T} = \alpha$. With Boltzmann factors incorporated, the probabilities must be normalized to 1, so that now they should be computed as follows:

$$P_{55}^{(1)} = \frac{1}{Q}\,(P_5^{(0)})^2\,\mathrm{e}^{-2\alpha}\,,\quad P_{56}^{(1)} = \frac{2}{Q}\,P_5^{(0)}P_6^{(0)}\,\mathrm{e}^{-\alpha}\,,\quad P_{57}^{(1)} = \frac{2}{Q}\,P_5^{(0)}P_7^{(0)}\,,$$

$$P_{66}^{(1)} = \frac{1}{Q}\,(P_6^{(0)})^2\,,\ P_{67}^{(1)} = \frac{2}{Q}\,P_6^{(0)}P_7^{(0)}\,\mathrm{e}^{-\alpha}\,,\ P_{77}^{(1)} = \frac{1}{Q}\,(P_7^{(0)})^2\,\mathrm{e}^{-2\alpha}\,, \quad (3.19)$$

where the normalizing factor Q is given by:

$$Q = (P_5^{(0)})^2\,\mathrm{e}^{-2\alpha} + 2\,P_5^{(0)}P_6^{(0)}\,\mathrm{e}^{-\alpha} + 2\,P_5^{(0)}P_7^{(0)} + (P_6^{(0)})^2 + \\ + 2\,P_6^{(0)}P_7^{(0)}\,\mathrm{e}^{-\alpha} + (P_7^{(0)})^2\,\mathrm{e}^{-2\alpha}\,. \quad (3.20)$$

Now we can calculate the distribution of five-sided, six-sided and seven-sided polygons in all spontaneously agglomerated doublets. The corresponding set of probabilities is denoted by $P_k^{(1)}$ and is computed as follows:

$$P_5^{(1)} = \frac{1}{2}\,(2\,P_{55}^{(1)} + P_{56}^{(1)} + P_{57}^{(1)})\,; \qquad P_7^{(1)} = \frac{1}{2}\,(P_{57}^{(1)} + P_{67}^{(1)} + 2\,P_{77}^{(1)}); \quad (3.21)$$

and obviously,

$$P_6^{(1)} = \frac{1}{2}\,(P_{56}^{(1)} + 2\,P_{66}^{(1)} + P_{67}^{(1)} = 1 - P_5^{(1)} - P_7^{(1)})\,.$$

In a real agglomeration process, at least at its initial stage, one can observe a mixture of single polygons and freshly created doublets; later on also triplets, quadruplets of agglomerated polygons are created, too, but at the beginning, only single polygons ("singlets") and edge-sharing pairs ("doublets") dominate. Suppose that at a certain initial stage of agglomeration one has, in a unit volume, $N - m$ singlets and m doublets. The average probability of finding a k-sided polygon in such sample will be

$$P_k(s) = (1 - s)\,P_k^{(0)} + s\,P_k^{(1)}, \quad (3.22)$$

where $s = m/N$ is a natural parameter describing the progress of agglomeration process at its initial stage (reduced to doublet creation only). The derivative of $P_k(s)$ is therefore

$$\frac{\mathrm{d}P_k}{\mathrm{d}s} = P_k^{(1)} - P_k^{(0)} \quad (3.23)$$

and is independent of s in this approximation. Of course, only two out of these three differential equations are independent.

The above differential system, reduced to two ordinary differential equations, can be written explicitly as follows (we shall replace $P_k^{(0)}$ by P_k for the sake of simplicity):

$$\frac{\mathrm{d}P_5}{\mathrm{d}s} = \frac{1}{2Q}\left[2P_{55}^{(1)} + P_{56}^{(1)} + P_{57}^{(1)}\right] - P_5,$$
$$\frac{\mathrm{d}P_7}{\mathrm{d}s} = \frac{1}{2Q}\left[2P_{77}^{(1)} + P_{57}^{(1)} + P_{67}^{(1)}\right] - P_7. \quad (3.24)$$

The third equation is linearly dependent, because $P_6 = 1 - P_5 - P_7$, so that $\mathrm{d}P_6/\mathrm{d}s = -\mathrm{d}P_5/\mathrm{d}s - \mathrm{d}P_7/\mathrm{d}s$.

The differential equations can be written in a more explicit manner, with clear dependence on two independent variable P_5 and P_6:

$$\frac{\mathrm{d}P_5}{\mathrm{d}s} = \frac{1}{2Q}\left[2P_5^2\mathrm{e}^{-2\alpha} + 2P_5(1 - P_5 - P_7)\mathrm{e}^{-\alpha} + 2P_5P_7 - 2P_5Q\right],$$

$$\frac{\mathrm{d}P_7}{\mathrm{d}s} = \frac{1}{2Q}\left[2P_7^2\mathrm{e}^{-2\alpha} + 2P_7(1 - P_5 - P_7)\mathrm{e}^{-\alpha} + 2P_5P_7 - 2P_5Q\right] \quad (3.25)$$

with

$$Q = P_5^2 \mathrm{e}^{-2\alpha} + 2P_5P_6\mathrm{e}^{-\alpha} + 2P_5P_7 + P_6^2 + 2P_6P_7\mathrm{e}^{-\alpha} + P_7^2\mathrm{e}^{-2\alpha},$$

where $P_6 = 1 - P_5 - P_7$.

Finally, simplifying by 2 and putting forward common factors, we get:

$$\frac{\mathrm{d}P_5}{\mathrm{d}s} = \frac{P_5}{Q}\left[P_5\mathrm{e}^{-2\alpha} + (1 - P_5 - P_7)\mathrm{e}^{-\alpha} + P_7 - Q\right],$$

$$\frac{\mathrm{d}P_7}{\mathrm{d}s} = \frac{P_7}{Q}\left[P_7\mathrm{e}^{-2\alpha} + (1 - P_5 - P_7)\mathrm{e}^{-\alpha} + P_5 - Q\right]. \tag{3.26}$$

Now it is quite easy to perform the analysis of phase trajectories of this differential system displayed in Fig. 3.4.

First of all, we find out the position of *singular points*, corresponding to constant solutions at which the two derivatives vanish simultaneously. There are five such solutions, three of which are found at the summits of the simplex of probabilities:

$$\mathrm{A}:\ P_7 = 1, \quad P_5 = 0,\ (P_6 = 0); \quad \mathrm{B}:\ P_6 = 1 \quad (P_5 = P_7 = 0);$$
$$\mathrm{C}:\ P_5 = 1, \quad P_7 = 0, \quad (P_6 = 0);$$

while the two remaining ones are:

$$\mathrm{D}: \quad P_5 = P_7 = \frac{1}{2} \quad (P_6 = 0),$$

and the fifth one inside the triangle:

$$\mathrm{E}: \quad P_5 = \frac{1}{3 - \mathrm{e}^{-\alpha}}, \quad P_7 = \frac{1}{3 - \mathrm{e}^{-\alpha}}, \quad P_6 = 1 - P_5 - P_7 = \frac{1 - \mathrm{e}^{-\alpha}}{3 - \mathrm{e}^{-\alpha}}\,.$$

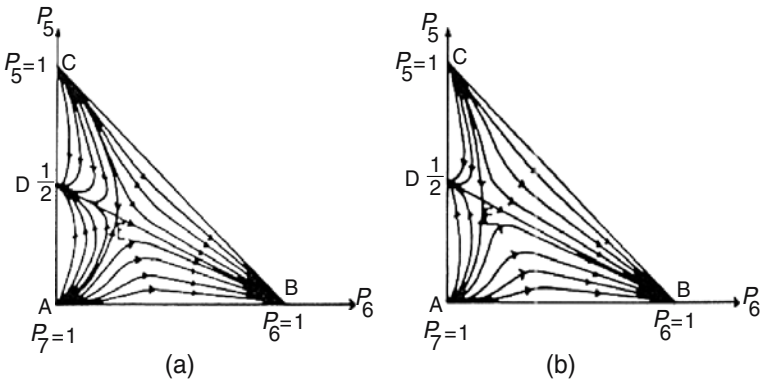

Fig. 3.4. Two phase portaits of probability trajectories describing the agglomeration process of five-sided, six-sided, and seven-sided polygons

It is easy to check by linearization of (3.26) that A and C are *repulsive* singular points, B and D are *attractive* singular points, while E is a *saddle point.* (There is another saddle point at infinity, the fact resulting also from Euler's formula for a sphere, but for the probabilities it has no physical meaning, of course.)

The points A, B, C, D keep their position steady independent of the value of parameter α (therefore, independent of temperature, too), whereas the position of the saddle point E depends on the value of the parameter α, falling on the attractive point D when $\alpha \to 0$. Typical phase trajectories are displayed in Fig. 3.4. It is worthwhile to note that the two attractive points and the saddle point are found on the line $P_5 = P_7$, which satisfies Euler's constraint for three-coordinate lattices (3.1). The separatrix curve AEC divides the simplex into two regions: on the right, the system is driven toward crystallization ($P_6 = 1$), whereas if the initial conditions happen to be on the left, the system will prefer another attractive singular point corresponding to an amorphous mixture of only pentagons and heptagons. Besides, in a more realistic model taking into account second-order effects, the system can remain infinite time in the vicinity of the metastable saddle point E.

This method has been developed and generalized to three dimensions in [6–8]. It describes quite well the process of glass formation and enables to compute certain interesting parameters of covalent network glasses, mostly the chalcogenide glasses like $Ge_xSe_{(1-x)}$, $As_xSe_{(1-x)}$, or alkali–boroxol glasses $(B_2O_3)_{(1-x)}(Na_2O)_x$ [9, 10]. in particular, the dependence of glass transition temperature T_g on the average coordination number of the network, and indirectly, on its chemical composition [11, 12]. Here we demonstrate an application of this stochastic method for modeling of nucleation and growth of fullerenes.

3.4 Application: How the Fullerene Molecules are Formed

Fullerenes are a new form of pure carbon, first predicted, then discovered in early 1980s of the last century (cf. [13,14]. The fullerene molecule is composed of 60 carbon atoms, forming a perfect Archimedean polyhedron composed of 12 pentagon and 20 hexagon rings. Each pentagon is surrounded by hexagons, while each hexagon shares three of its sides with pentagons, alternating with three other sides adjacent to hexagons. It is easy to check how Euler's formula is satisfied in this case: we have 60 summits (i.e., carbon atoms), 30 edges (the entire network is three coordinate, so that each atom is bonded with other three, and each bond is shared by two atoms, so that the total number of bonds (edges) is $(3 \times 60)/2 = 90$, and the number of faces is $12 + 20 = 32$). Equation (3.1) then becomes $60 - 90 + 32 = 2$, as it should be. Modern soccer balls are made of leather polygons following this scheme, with pentagons often painted in black, and hexagons painted in white. We show a few examples of these fancy carbon molecules in Fig. 3.5.

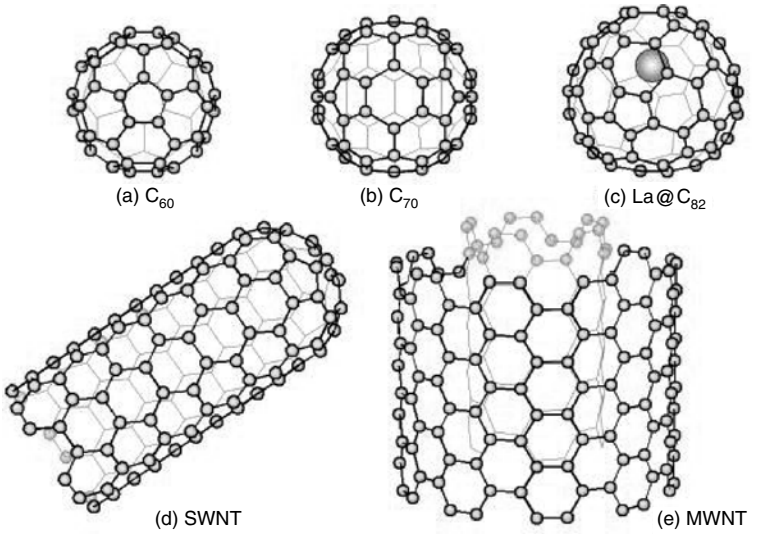

Fig. 3.5. Examples of fullerene molecules: the C_{60}, C_{70}, a nested fullerene C_{60} (with an La atom captured inside), and tubular fullerenes

Because these 32 polygons represent a simple covering of an Euclidean two-dimensional sphere embedded in three-dimensional Euclidean space, the total solid angle covered is equal to $4\pi = 720^{\circ}$. We know that the Euclidean plane can be tiled entirely by a three-coordinate hexagonal lattice, with $N_c = 3$ and $N_f = 6$. Each pentagon creates an angular deficit of $5 \times 12^{\circ} = 60^{\circ}$. This is so because in a regular pentagon the angle is equal to 108° instead of 120° in a regular hexagon. Three hexagons meeting at one vortex give exactly $3 \times 120^{\circ} = 360^{\circ}$, like it should be on the plane. If we replace one of the three hexagons by a pentagon, either the pentagon must undergo a deformation, or else, if it is supposed to keep its perfect form, the whole structure will become curved and will be possible only in three dimensions, out of the plane. The measure of this local curvature is the deficit in angle, which with one angle of the pentagon is equal to 12°; but each pentagon creates such a deficit in five different vertices.

Now, if the total curvature has to attain $4\pi = 720^{\circ}$, it is easy to see that 12 pentagons have to be inserted in a homogeneous hexagon lattice in order to form a structure topologically equivalent to a sphere. The number of hexagons is in principle arbitrary. When there are no hexagons at all, 12 pentagons form one of the Platonic polyhedra, the perfect *dodecahedron*, with 12 faces, 20 summits, and 30 edges (again, $20 - 30 + 12 = 2$!). Its *dual* polyhedron, with the same number of edges, but with $N_c^{'} = N_f$ and $N_f^{'} = N_c$ is the regular *icosahedron*, with 20 triangular faces and 12 five-coordinate summits. The next regular structure with 12 polygons and extra hexagons can be obtained from the icosahedron by slicing, in a very symmetric

manner, all the summits, which will give rise to 12 pentagons, and transform all the equilateral triangles into smaller hexagons. Such a polyhedron is called an *Archimedean* regular body, and in this case coincides with the fullerene structure.

The existence of C_{60} molecules representing a new form of pure carbon has been first conjectured by astronomers investigating the infrared spectra of red giant stars. Later on, these molecules were obtained by synthesis from carbon evaporating in a very hot electric arc between two graphite electrodes, in closed space filled with helium under the pressure of 400 mbar, thus imitating the conditions in the outer atmosphere of an average red giant. The soot that gathered at the bottom of the recipient contained as much as 10% of fullerene molecules, which could be separated from the rest of carbon precipitate because they could be easily solved in benzene, and later on recovered after its evaporation.

These ideas have been quite successfully applied to explain the growth of fullerene molecules [14–16] in a model proposed recently in the series of papers coauthored by Bennemann [17, 18]. The model of nucleation and growth of the fullerene C_{60} molecules follows strictly the ideas of self-similarity and converging algorithms exposed earlier.

We propose the following model of agglomeration processes leading to the formation of fullerene molecules. These molecules are found in great abundance in the soot falling from the electric arc discharge between two graphite electrodes, at the temperature about $2,800°C$, in helium atmosphere (acting as a moderator) under the pressure of 0.4 atm. Each fullerene C_{60}-molecule contains 60 carbon atoms disposed of in rings, with 20 hexagons and 12 pentagons arranged like in a soccer ball. Each pentagon is surrounded by hexagons only, while each hexagon is surrounded alternatively by three pentagons and three hexagons.

In the hot flame surrounding the electric arc one finds many carbon clusters, the acetylene groups C_2, the molecules C_3, C_4, etc. up to C_6 benzene rings, C_{10} naphthalene double rings, and even the C_{12} molecules built up from three rings, two hexagons, and one pentagon (see Fig. 3.6).

In an achieved fullerene molecule all the vertices are of the same type, in which two hexagons and one pentagon meet together. Two pentagons can never share an edge, nor can three hexagons share a common vertex. That is why when a new ring is formed in one of the cavities of a C_{12} molecule by the addition of a C_2 or a C_3 molecule abundant in the hot gas, out of four possible stable configurations (excluding the formation of unstable *seven-sided rings*), only *two* are appropriate for the consecutive fullerene formation, the other two containing wrong combinations of polygons.

The same is true at each consecutive step of agglomeration, consisting in the creation of a new polygon. Each time we get only *half* of configurations that are proper for the fullerene building, the other half being lost because it contains the wrong configurations, such as two neighboring pentagons, or three hexagons sharing a common vortex. An example of two new polygons

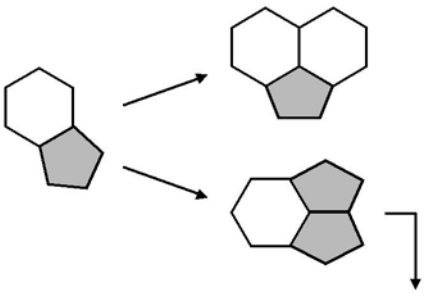

Fig. 3.6. Formation of a new polygon in a (6,5)-type cavity

formed in a cavity, one of which leading to a configuration that can lead to formation of a fullerene molecule, another one leading to a prohibited (two edge-sharing pentagons) configuration, is shown in Fig. 3.6.

Starting from the stage of 3-ring molecule C_{12}, we need still 29 more rings in order to complete a fullerene molecule with the total of 32 rings; but in order to build all the remaining 29 rings, it is sufficient to complete only about 22 or 23 polygons. It is obvious that when in such a closed structure we already have 31 or 30 *correct* polygons in place, the remaining one or two rings are also there; one can safely assume that this reasoning is still true at the level of 26 or 27 completed rings, when only no more than 5 or 6 polygons are missing.

This means that if at the beginning about 25% of carbon available in the hot gas is contained in the C_{12}, C_2, and C_3 molecules, the final yield of fullerenes is given by the geometric progression with the ratio 1:2, which gives the estimate

$$25\% \times (0.5)^{23} \sim 10^{-8}.$$

This is more than seven orders of magnitude below the observed 10% yield. Maintaining the idea that on average the yield of the fullerene-like molecules at the subsequent agglomeration stages behaves as a geometric progression, we can easily evaluate the ratio q that leads to the observed final yield:

$$25\% \times q^{23} \simeq 10\% = 0.1 \quad \Longrightarrow q \simeq 0.961. \tag{3.27}$$

In order to explain the experimental facts, we must assume that at each agglomeration step the "proper," i.e., fullerene-like configurations, are highly preferred to the "wrong" ones. Their pure combinatorial factors being the same, the only reasonable explanation could be given by the difference of energies related to the respective polygon construction; these energies should be contained in the corresponding Boltzmann factors as follows:

$e^{-\epsilon}$: for a pentagon created between two hexagons;

$e^{-\beta}$: for a pentagon created between a pentagon and a hexagon;

$e^{-\alpha}$: for a hexagon created between two hexagons;

$e^{-\eta}$: for a hexagon created between a pentagon and a hexagon.

It is easy to compute the normalized probabilities of producing the "good" clusters that are proper for the subsequent construction of C_{60} after each agglomeration step.

For example, at the stage of construction when the four-polygon molecules are being formed, out of which only two are proper for fullerene formation, their total probability is easily found to be

$$\frac{e^{-\epsilon} + 2\,e^{-\eta}}{e^{-\epsilon} + 2\,e^{-\eta} + e^{-\alpha} + 2\,e^{-\beta}}.$$

This rational expression depends only on the ratios of the Boltzmann factors involved; let us denote $y = e^{-(\eta-\epsilon)}$, $z = e^{-(\alpha-\epsilon)}$, and $t = e^{-(\beta-\epsilon)}$; then the above expression can be written as

$$F_4 = \frac{1+2y}{1+2y+2t+z}.$$

Another interesting feature related to the self-similarity is the average ratio of pentagons among all rings in the fullerene-like clusters. At the stage of four-polygon clusters it is computed as follows:

$$G_4 = \frac{1}{2}\,\frac{e^{-\epsilon}}{e^{-\epsilon}+2\,e^{-\eta}} + \frac{1}{4}\,\frac{2\,e^{-\eta}}{e^{-\epsilon}+2\,e^{-\eta}} = \frac{e^{-\epsilon}+e^{-\eta}}{e^{-\epsilon}+2\,e^{-\eta}} = \frac{1+y}{1+2\,y}.$$

We were able to compute these characteristic expressions up to the tenth step of agglomeration, when the clusters made up from 11 polygons containing 28–30 carbon atoms, by obtaining two series of functions of three variables, y, z, and t, denoted by F_n and G_n.

We need three independent equations in order to solve for (y, z, t). These equations can be easily produced if we suppose that on average the yield at each step is constant, and close to the evaluation we made shortly before, i.e., if we set

$$F_{n+1}/F_n \simeq 0.961.$$

But with the help of the second set of functions we can be more ambitious and try to find the numerical value from the first principles. The functions F_n should behave as a geometrical sequence; so we require that

$$F_{n+2}/F_{n+1} \;=\; F_{n+1}/F_n.$$

The functions G_n give the average ratio of pentagons in all fullerene-like clusters containing n polygons. It is reasonable to suppose that at the very early stages of agglomeration this ratio is very close to the ultimate limit, and that the convergence is also very rapid, i.e., exponential.

In the final product — a C_{60} molecule — this ratio is equal to $3/8 = 0.375$. It seems therefore reasonable to suppose that functions G_n obey the exponential law:

$$G_n = G_{\text{end}}\,(\,1\,-\,\mathrm{e}^{\lambda\, n}).$$

Here too, we can eliminate the unknown characteristic exponent λ and the final value $G_{\text{end}} = 0.375$ by comparing several expressions with different values of n, arriving at the following law of self-similarity:

$$\frac{G_{n+2} - G_{n+1}}{G_{n+1} - G_n} \;=\; \frac{G_{n+1} - G_n}{G_n - G_{n-1}}.$$

With two equations for F_n and one for G_n constructed with $n = 8, 9, 10$, and 11 we have solved for y, z, and t:

$$y = 0.691, \quad z = 0.122, \quad \text{and} \quad t = 0.032$$

This leads to the constant average ratio $F_{n+1}/F_n = 0.957$, which is very close to what has been anticipated, and gives the final yield of about 9.5%, which is also satisfactory.

For the obtained values of the parameters y, z, and t the characteristic exponent λ is equal to 0.635, and the rates defined by G_ns are:

$$G_7 = 0.370, \quad G_8 = 0.372, \quad G_9 = 0.373, \quad G_{10} = 0.3736,$$

which is fairly rapid, taking into account that without discriminating Boltzmann factors the same ratios would behave as:

$$G_6 = 0.3417, \quad G_7 = 0.3444, \quad G_8 = 0.3467, \quad G_9 = 0.3492, \quad G_{10} = 0.3512, \quad \text{etc.}$$

It is interesting to note that all the numerical values (the final yield of C_{60} molecules, the ratio of pentagons in the final product, and the values of the Boltzmann factors) are obtained here by applying *exclusively* the self-similarity principle — quite a remarkable result.

Finally, knowing that the temperature around the arc at which the process takes place is about $3,000\,\mathrm{K}$, we find the energy differences:

$$E_2 - E_1 = 0.104\,\mathrm{eV}, \quad E_3 - E_1 = 0.588\,\mathrm{eV}, \quad E_4 - E_1 = 0.965\,\mathrm{eV},$$

which suits reasonably well our ideas about the forces needed to bend the graphite network creating local curvature around one of the carbon atoms.

3.5 Onion Fullerenes and Carbon Tubes

Since the discovery of fullerenes, an impressive amount of new structures made of pure carbon have been put into evidence and studied by many experimentalists [19, 20] and theorists [20, 21]. These structures are supposed

to give rise to many new materials with exciting physical properties that are now intensely studied. Apparently, the most common structures besides the fullerenes proper are *carbon nanotubules* and the so-called *onion fullerenes*, which are composed of multiple layers forming most probably around a C_{60} fullerene nucleus.

Whatever the size of the subsequent layer, it must be composed of carbon polygons, as the smallest C_{60} fullerene molecule. If it contained exclusively hexagons and pentagons, the number of the latter must always be 12, as it follows from Euler's theorem. It is quite easy to form bigger and bigger symmetric molecules containing 12 pentagons and an appropriate number of hexagons. These giant molecules will possess an icosahedral form, because they can be formed out of 20 triangular flat structures made of certain number of hexagons and three pentagons at the three summits. Some of them are shown in Fig. 3.7.

When the triangles containing pieces of pure hexagonal lattice grow bigger, they are closer and closer to flat surface. Therefore, big shells built with a great number of hexagons and 12 pentagons, become more and more like perfect icosahedrons. However, pictures obtained with electronic microscopy show clearly that great onion fullerenes look like almost perfect spheres. This means that their local curvature is almost constant, and close to the value $1/R$, with R denoting the radius of spherical fullerene shell. The same is true for the spherical caps of giant carbon tubes, which apparently do not display any tendency to become like halves of corresponding icosahedrons [19].

But this is possible only if the structures contain not only hexagons and pentagons, but heptagons as well. As we have shown earlier, even a flat tiling can be achieved exclusively with a mix of pentagons, and heptagons, provided they are in equal number, or with a mix of hexagons, pentagons and

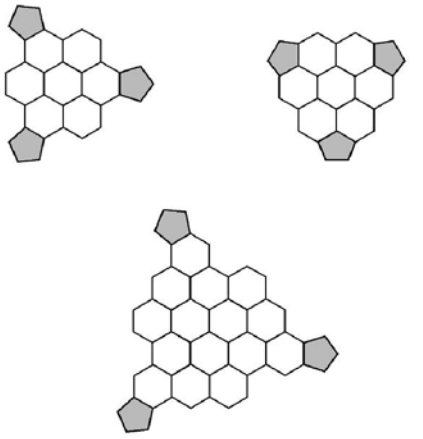

Fig. 3.7. Some possible triangular configurations containing three pentagons

heptagons, provided that $P_5 = P_7$. Topological and geometrical constraints provide equations that enable us to determine the structure of giant molecules topologically equivalent to a sphere. For example, it is easy to determine the total number of atoms if we assume that only vertices at which two hexagons and one pentagon meet are present. Let the total number of vertices (i.e., carbon atoms) be N_A. At each vertex, the angular deficit is equal to $(3\pi/5 - 4\pi/6)$. As the total angular deficit to cover a sphere must be -4π, we must have

$$N_A\left(\frac{3\pi}{5} - \frac{4\pi}{6}\right) = -4\pi$$

and the solution is, of course, $N_A = 60$, which corresponds to the fullerene molecule.

It is also quite easy to evaluate the number of atoms and polygons in the icosahedral structures made of 20 triangles represented in Fig. 3.7. Any such triangular face can be parametrised, following Coxeter [1], by two integers (p, q). The prescription is then as follows: in a perfect planar hexagonal lattice choose a hexagon and replace it by a pentagon. Go p steps (i.e., skip p hexagons) along a straight row; then turn at the angle 120− and go q steps along this new direction. Place a new pentagon; start over again, but at the direction, which makes 120− with respect to the last one; again, p steps, and q steps along the direction at the angle of 120−. The third operation will lead back to the place we started from, and the construction is complete.The triangles displayed in Fig. 3.7 correspond to the labels $(2, 2), (3, 0)$, and $(4, 2)$.

The number of carbon atoms in the icosahedral molecule made of triangles of the type (p, q) is equal to

$$N = 20\,(p^2 + pq + q^2).$$

With number of pentagons always equal to $N_5 = 12$, it is easy to evaluate the number of remaining hexagonal rings: it is equal to

$$N_6 = 10\,(p^2 + pq + q^2 - 1).$$

The first few icosahedral structures following after the fullerene C_{60} are the: C_{80}, corresponding to $p = 2, q = 0$; C_{140}, corresponding to $p = 2, q = 1$; C_{180}, with $p = 3, q = 0$; C_{240}, with $p = 2, q = 2$, and so on [22, 23].

It is also easy to evaluate the sizes of pure carbon icosahedrons that could be built as a next layer on a smaller one, so that the distance between layers is no lesser than the standard distance between layers in the graphite structure, which is of the order of $d = 3.2 \times 10^{-8}$ cm, about twice more than the standard distance between the closest neighbors in the same layer, i.e., the side of a hexagon, which is $a = 1.42 \times 10^{-8}$ cm. For example, the average radius of a C_{60} fullerene ball is about 3.45×10^{-8} cm; the average radius of the next

Table 3.1. Icosahedral structures of growing size and their radii (in Ångstroms)

Type (p, q)	Number of C atoms	$\langle r \rangle$	$r_{\min}$
(1,1)	C_{60}	3.45	3.42
(2,0)	C_{80}	3.85	3.79
(2,1)	C_{140}	5.44	5.32
(3,0)	C_{180}	5.84	5.67
(2,2)	C_{240}	6.89	6.67
(3,1)	C_{260}	7.39	7.20
(4,0)	C_{320}	7.83	7.54
(4,1)	C_{420}	9.24	8.94
(5,0)	C_{500}	9.82	9.40
(3,3)	C_{540}	10.35	9.94

regular ball, C_{80}, is equal to 3.85×10^{-8} cm, clearly too close in order to serve as a next layer.

It follows from the Table 3.1 that if multilayer icosahedral structures made of pure carbon could be observed, the first structure that could grow on a C_{60} fullerene ball would be a C_{240} or a C_{260} icosahedron, and the third could then be a C_{540} structure, and so on. Similarly, if the first structure serving as a "core" were a C_{80} ball, then the second layer should be a C_{320}, the third one a C_{620}, etc.

But the electron microscope pictures of onion fullerenes show almost perfect spherical shapes, which could not be observed with the above icosahedral structures. In spite of this theoretical possibility, Nature apparently chooses another way to create closed bottoms of nanotubes: they are apparently as close as possible to hemispheres [19].

Also the nanotubes of various sizes can close with similar icosahedral structures, which can be regarded as halves of carbon icosahedrons, centered around one of the pentagons, and continued with five-fold symmetry as five almost flat triangular sides made exclusively with hexagons, up to a layer when five new pentagons are added, thus creating solid angle of 2π; then a locally flat tubular structure made of hexagons may continue as a regular cylindric nanotube.

As a matter of fact it is not difficult to find out how such spheroidal structures can be produced with a three-coordinate lattice composed only of five-sided, six-sided and seven-sided carbon rings. In order to do this, let us analyze local curvature around each carbon atom surrounded by three

Table 3.2. Ten possible vertices and corresponding angular deficits (in radians and degrees)

Vertex type	Angular deficit (rad)	Angular deficit (deg)
(555)	$-\frac{\pi}{5}$	-36
(556)	$-\frac{2\pi}{15}$	-24
(557)	$-\frac{3\pi}{35}$	-15.43
(566)	$-\frac{\pi}{15}$	-12
(567)	$-\frac{2\pi}{105}$	-3.43
(666)	0	0
(577)	$+\frac{\pi}{35}$	$+5.14$
(667)	$+\frac{\pi}{21}$	$+8.57$
(677)	$+\frac{2\pi}{21}$	$+17.14$
(777)	$+\frac{3\pi}{21}$	$+25.71$

polygons. The angular deficit (or angular excess) of each triplet is given in Table 3.2.

We see that the first five triplets display *positive local curvature*, the triplet (666) is flat, and four remaining ones contribute to *negative local curvature*. In view of this, we shall suppose that giant onion fullerenes contain exclusively *convex* triplets. To avoid too strong local curvatures, we exclude the triplets containing two 5-fold rings at once; and to avoid the formation of locally flat surfaces, like in the icosahedral molecules discussed earlier, we also exclude the (666) triplets. This leaves us with only two kinds of vertices: (566) and (567), whose angular deficits are $-\frac{\pi}{15} = -12^\circ$ and $-\frac{2\pi}{105} = -3.43^\circ$, respectively. But it is easy to see that if we start surrounding a heptagon with (567) vertices, we must also produce at least one (667) vertex.

Now we are ready to form the equations that will enable us to determine the admissible configurations of giant onion fullerenes. The following seven numbers are to be found: N_A, the total number of atoms; N_5, N_6, and N_7, the numbers of five-sided, six-sided, seven-sided polygons; finally, the exact numbers of (566), (567), and (667) vertices. It is easy to establish *six* linear relations between these variables.

As the total angular deficit of a sphere must be -4π, taking the values of angular deficits of (566) and (567) vertices from Table 3.2, we must have

$$-\frac{\pi}{15} N_{566} - \frac{2\pi}{105} N_{567} + \frac{\pi}{21} = -4\pi, \tag{3.28}$$

from which we get

$$7\,N_{566} + 2\,N_{567} - 5\,N_{667} = 420. \tag{3.29}$$

The above equation remains valid even in the presence of (666) vertices, which are flat and do not contribute to angular deficit or excess. From Euler's theorem we also know that in a tri-coordinate lattice covering a sphere exclusively with five-sided, six-sided, and seven-sided polygons, we must have

$$N_5 = N_7 + 12, \tag{3.30}$$

the number of hexagons N_6 being arbitrary.

Next, any atom belongs to one type of vertex exclusively; therefore the total number of atoms is equal to the sum of numbers of these vertices:

$$N_{566} + N_{567} + N_{666} + N_{667} = N_A, \tag{3.31}$$

where we also counted the triplets (666) for a more general case. The total number of carbon atoms (vertices) can also be obtained by counting individual polygons weighted by their respective numbers of sides (vertices) — this will count each atom three times, because the network is three coordinate:

$$5N_5 + 6N_6 + 7N_7 = 3\,N_A. \tag{3.32}$$

Finally, supposing that only one (667) triplet can be found around a heptagon, we must have the following two relations:

$$N_{667} = N_7\,, \qquad N_{567} = 6N_7. \tag{3.33}$$

We shall choose the number of heptagons N_7 as free parameter. Then the set of equations can be reduced and represented as follows:

$$\begin{gathered} N_{566} = 60 - N_7\,, \quad N_5 = 12 + N_7, \\ 6N_6 + 12N_7 = 3N_A - 60, \quad N_{566} + N_{666} = N_A - 7N_7. \end{gathered} \tag{3.34}$$

Now we can find many solutions corresponding to different choices of N_7 and N_{666}. Not all of them are physically acceptable though. In order to maintain sphericity, we should not break the symmetry too strongly (as suggested, e.g., in [23]. For example, a good choice seems to be $N_7 = 20$, decorating all potential faces of an icosahedron; let us also choose $N_{666} = 0$. Then we get

$$\begin{gathered} N_7 = 20\,, \quad N_5 = 32\,, \qquad N_6 = 40\,, \; N_A = 180, \\ N_{566} = 40\,, \quad N_{567} = 120\,, \quad N_{667} = 20. \end{gathered}$$

In view of (3.34), N_7 should not exceed 60, in which case there are no more (566) vertices in the giant onion fullerene. If we suppose that no "flat" (666)

vertices are present, we get the following solution:

$$N_7 = 60\,, \quad N_5 = 72\,, \quad N_6 = 80\,, \quad N_A = 420, \quad N_{566} = 0\,,$$
$$N_{567} = 360\,, \quad N_{667} = 60.$$

These are only two possible solutions: the number and distribution of vertices reflects the mean curvature of the giant fullerene sphere, which is obviously higher in the first case, and lower in the second case. A rough estimate of the mean curvature can be obtained via inverse square law, $\langle R \rangle \simeq (N_A)^{-2}$; this law supposes that the average density of carbon atoms on the surfaces of subsequent fullerene onions is very much the same, as are the sides of the polygons. Therefore the average radii of the C_{180} and C_{420} onions are about 5.92 and 9.06 Å, respectively.

3.6 Rigidity and Local Structure in Covalent Glasses

Most of the glasses, especially the so-called *covalent glasses*, are formed by the elements from the third, fourth, and fifth groups of the periodic table, their oxides, or in combination with chalcogenides like sulfur or selenium. They form a very homogeneous and, at first glance, totally *random* network. However, local structures are not random at all: every atom has a well-defined valence, i.e., *coordination number*, the chemical bonds between given kinds of atoms are essentially of the same length, and even the angles between the bonds are quite rigid, with the exception of oxygen or chalcogenide bridges. This meekness of angular constraints is responsible for the breakdown of geometrical order on the medium-range (several bond lengths) scale, which in turn enhances crystalline growth. For a long time however, it remained quite unclear why certain compounds do not crystallize and prefer to keep the amorphous structure of an overcooled liquid, whereas many other covalent substances prefer to crystallize.

In 1979, Phillips [24] had formulated a very simple criterion in order to decide those chemical compounds that covalent networks would belong *a priori* to the species of good glass formers. It consisted in stating that the best glass formers, when looked upon as random networks of solid bars (symbolizing the covalent chemical bonds) and massive balls (atoms), are those, that are closest to *isostatic* networks. This means that the number of mechanical constraints per atom should be exactly equal to 3, whenever possible, because 3 is the natural number of degrees of freedom of a point-like mass in three dimensions. The theory has been developed further by Thorpe [25] and applied very successfully to the experimental analysis of many glasses by Boolchand [26–28].

A single atom is maintained in its position by bonds stretching towards its closest neighbors (from now on, "a neighbor" is synonymous with another atom with which a given atom is linked via real chemical bond, and not

just geometrically close). As seen from a given point-like atom, the bonds stretching toward its neighbors form a star-like figure, with the number of rays corresponding to the central atom's valence. Let us call this number N_c, the coordination number. The mechanical constraints that these N_c bonds impose on the system are of two types. First, their constant length represents clearly a mathematical (thus also mechanical) constraint, one equation corresponding to one bond; but each bond contributing to *two* atoms found at its extremities, the number of constraints (called *the bond-stretching constraints*) per atom is clearly $N_c/2$. But the star-like structure is still not rigid if only the lengths of bonds are kept constant. In order to ensure the rigidity of the structure, the angles between the bonds should also be fixed at given values. The number of these *angular constraints* is very easy to compute: once a specific direction is given to one of the bonds, which corresponds to a choice of the basis in vector space $\mathbf{E}^3$, the direction of a second bond stemming from the same atom is fixed with respect to the first vector by the choice of just one angle between them. The second angle corresponds just to the choice of the reference frame in $\mathbf{E}^3$. If we wish to fix a third unit vector with respect to the former two, we have to fix two new angles with respect to the former construction serving as the reference frame, as for any unit vector in three dimensions, and so forth. The simple formula for the total number of angular constraints belonging to an atom is then $2\,N_c - 3$, so that the total number of mechanical constraints per atom with coordination number N_c is

$$r_c = \frac{N_c}{2} + 2\,N_c - 3. \tag{3.35}$$

If the average number of constraints per atom is higher than 3, the corresponding network will be overconstrained; if it is lower than 3, the network will be underconstrained, thus making possible local motions characteristic for liquid state; finally, when $r_c = 3$, the network is *isostatic*, which has been proposed as an important criterion for glass-forming tendency.

When generalized to an arbitrary random network, i.e., replacing N_c by its average value $\langle N_c \rangle$, it leads to an obvious condition

$$\langle r_c \rangle = \frac{\langle N_c \rangle}{2} + 2\,\langle N_c \rangle - 3 = 3$$
$$\text{wherefrom} \quad \langle N_c \rangle = \frac{12}{5} = 2.4. \tag{3.36}$$

In many glasses (mainly the so-called chalcogenide glasses, $\mathrm{Ge}_x\mathrm{Se}_{(1-x)}$, $\mathrm{As}_x\mathrm{Se}_{(1-x)}$, $\mathrm{Si}_x\mathrm{Se}_{(1-x)}$, one can observe the change of behavior of many parameters, in particular the *glass transition temperature* and the specific heat C_p, near the threshold of $r_c = 2.4$. The best results have been obtained for simple two-component chalcogenide glasses, like $\mathrm{Ge}_x\mathrm{Se}_{(1-x)}$, $\mathrm{As}_x\mathrm{Se}_{(1-x)}$, $\mathrm{Ge}_x\mathrm{S}_{(1-x)}$, in which single atoms of selenium, germanium, or arsenic can be viewed as elementary building blocks with well-defined coordination number

equal to their chemical valence: 2 for selenium and sulfur, 4 for germanium, and 3 for arsenic. Then the average coordination number $\langle r_c \rangle$ is given by simple formula, $\langle r_c \rangle = 4\,x + 2\,(1-x)$ for $Ge_xSe_{(1-x)}$, or $\langle r_c \rangle = 3\,x + 2\,(1-x)$ for $As_xSe_{(1-x)}$, and so on.

Recently in [29] this "magic" rule has also been successfully applied to more common oxide glasses, and in the first place, to the silicate-based window glass, whose local geometry is much more complicated than that of binary chalcogenide glasses.

The composition of window glass, also used for the production of bottles and light bulbs, does not vary very much since its introduction by Italian masters in the fourteenth century in Venice. Besides some small amount of special modifiers (usually below 1.5%) intended to give the glass the desired color or to improve its mechanical resistance, the basic composition is as follows: 75% of SiO_2, 15% of Na_2O, and 10% of CaO. In a more compact notation, one can express it in a formula,

$$(SiO_2)_{1-x-y}\,(Na_2O)_x\,(CaO)_y \qquad \text{with} \qquad x = 10\%, \qquad y = 15\%\,.$$

It is clear that in order to fix the values of x and y one needs two independent equations. One of them is readily provided by the rigidity criterion, which indicates that the best glass-forming tendency will be observed when the average coordination number $\langle r_c \rangle$ is equal to 2.4. Let us express this condition as an algebraic equation. In order to compute $\langle r_c \rangle$, let us take one mole of window glass, then sum up all coordination numbers multiplied by their relative weights, and divide the result by the total number of atoms in the sample. We have only 1 (Na), 2 (Ca, O) and 4 (Si) valenced atoms in our network. The total number of bonds stemming from these atoms (per unit of mass) is then: $4\,(1-x-y)$ taking into account the 4-valenced Si atoms, plus $2 \times 2\,(1-x-y)$ from the O-atoms in the SiO_2 molecules, plus $2 \times 2\,y$ from the Ca and O atoms in the CaO molecules, plus $2\,x$ from the O atoms in Na_2O molecules, finally $1 \times 2\,x$ from the one-valenced Na atoms in the Na_2O molecules. The total number of bonds per unit of mass is therefore

$$4\,x + 4\,y + 8\,(1-x-y) = 8 - 4x - 4y.$$

The total number of atoms per unit mass is also easily evaluated: there are three atoms in a SiO_2 molecule, three atoms in a Na_2O molecule, and two atoms in a CaO molecule, which leads to the simple formula

$$3\,(1-x-y) + 3\,x + 2\,y = 3 - y.$$

The equation defining the rigidity threshold now becomes:

$$\frac{8-4x-4y}{3-y} = \frac{12}{5} = 2.4. \tag{3.37}$$

We need another independent relation between x and y in order to determine their values. Such a relation should be imposed by an independent physical

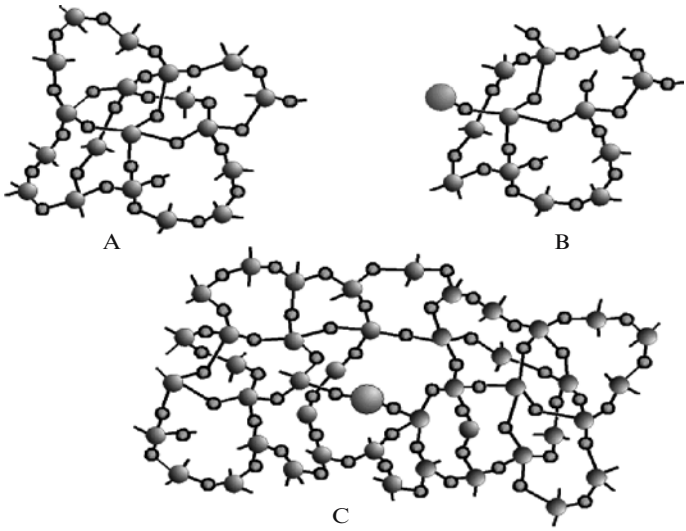

Fig. 3.8. The effect of Na^+ and Ca^{++} cations on local ring structure in the SiO_2 random lattice

or geometrical principle. Hopefully, another simple condition may be imposed if we want the network to be as homogeneous as possible, from a structural point of view. To this end, let us consider the effect of Na_2O and CaO molecules on the structure of initial amorphous SiO_2 network. If all the bonds are saturated, there are always *six minimal rings* starting with pairs of oxygen bonds stemming from each silicon atom, as shown in Fig. 3.8C. A minimal ring is defined as a circular loop made of oxygen bonds, such that there is no shorter way to start from the given bond and come back through another chosen bond. There are six different couples of bonds that can be chosen from four different bonds coming out from each Si atom, as simple combinatorics shows, the number of choices being $C_2^4 = \frac{4!(4-2)!}{2!}$.

When one Na_2O molecule is added to the network, each of its sodium atoms breaks one of the oxygen bonds of the SiO_2 network in halves, the extra O atom completing one of the two broken oxygen bonds. This creates two local structures with *three minimal rings* only, as shown in Fig. 3.8. Therefore, each Na_2O molecule removes *six* rings from the network, creating voids that enhance homogeneity of the network.

On the contrary, when one CaO molecule is added to the network, it also breaks an oxygen bond, but being two valent, it zips together two neighboring Si-centered tetrahedra, thus creating a compact local unit with as many as *six* oxygen bonds pointing out, as shown in Fig. 3.8. The corresponding number of minimal rings surrounding this new building block is 15, which is nine more rings than before. Summarizing up we see that whereas each Na_2O molecule suppresses six rings in the network, each CaO molecule creates nine new rings.

To keep the balance, i.e., to compensate the loss of rings due to Na_2O, one must add two CaO molecules for three Na_2O molecules, which gives us the new relationship we were looking for:

$$2\,x = 3\,y.$$

With these two equations the solution is immediately found to be:

$$x = \frac{3}{19} = 15.79\,\%\,, \quad y = \frac{2}{19} = 10.52\,\%\,, \quad 1 - x - y = 73.69\,\%\,,00, \quad (3.38)$$

which is almost exactly the average usual composition of window and bottle silicate glass.

Some amount of Al_2O_3 is present in most of industrial glasses, except for those whose chemical purity is specially supervised. The amount of Al_2O_3 is usually not higher than 1.5%. Our simple method enables us to evaluate the "ideal" composition also in presence of the aluminum oxide. A new variable must be introduced, denoting the relative amount of Al_2O_3; let us name it z. Then the new glass composition will be

$$(SiO_2)_{1-x-y-z}\,(Na_2O)_x\,(CaO)_y\,(Al_2O_3)_z\,.$$

We have only two basic equations at our disposal, and we would not like to introduce a new principle each time a small amount of new modifier is added. Supposing that the amount of Al_2O_3 is rather small as compared with other components, we can solve the problem by successive approximation method. This means that we take the highest rate obtained in the previous example of ternary $(SiO_2)_{1-x-y}\,(Na_2O)_x\,(CaO)_y$ glass as a starting point and solve the remaining two equations with two unknowns only. Therefore, we shall suppose that the amount of SiO_2 remains close to 75%, which enables us to write an extra equation

$$1 - x - y - z = 0.75 \quad \text{or} \quad 4x + 4y + 4z = 1. \quad (3.39)$$

The average coordination number is computed as the ratio of all valencies per mole to all atoms per mole, which gives readily the next equation:

$$\frac{8(1 - x - y - z) + 4x + 4y + 12z}{3(1 - x - y - z) + 3x + 2y + 5z} = \frac{12}{5},$$

leading to the following linear equation:

$$5x + 2y + z = 1. \quad (3.40)$$

Finally, we follow the principle of maximal homogeneity of ring structure. We already know that each Na_2O molecule suppresses six rings in the SiO_2 network; the CaO molecules, zipping together silicon-centered tetrahedra add nine extra rings to the structure. The Al_2O_3 molecules also act as "zipping"

agents, but each Al atom is three valenced, so it creates a cluster with three silicon-centered tetrahedra attached together, which leaves *nine* free valencies on which the rings can be constructed. Simple combinatorics leads to the conclusion that the number of new rings around each such cluster will be $(9 \times 8)/2 = 36$, which means an excess of three rings per one Al atom, or 60 rings per one Al_2O_3 molecule. In order for all these rings to compensate to the average, we must have

$$-6x + 9y + 60z = 0, \tag{3.41}$$

which completes the set of equations to solve. The result is as follows:

$$x = \frac{31}{184} \simeq 16,33\%\,, \quad y = \frac{14}{184} \simeq 7,61\%\,, \quad z = \frac{1}{184} \simeq 0.6\%\,.$$

and of course, $1 - x - y - z = 75\%$.

This method has its limitations too. For example, it is difficult to explain the composition of Pyrex glass (80.6%SiO_2, 13.0%B_2O_3, 4.1% Na_2O, 2.3%Al_2O_3), partly because pure B_2O_3 itself ideally satisfies the isostatic criterion: $3 \times \frac{2}{5} + 2 \times \frac{3}{5} = \frac{12}{5}$. In quaternary composition $(SiO_2)_{1-x-y-z}\,(Na_2O)_x$ $(B_2\,O_3)_y, (Al_20_3)_z$ the average coordination number is readily found as

$$\langle r_c \rangle = \frac{8(1-x-y-z) + 4x + 12y + 12z}{3(1-x-y-z) + 3x + 5y + 5z} = \frac{8 - 4x + 4y + 4z}{3 + 2y + 2z}.$$

When we fix the value of $\langle r_c \rangle$ at $12/5 = 2.4$, we get the equation $y = 1 - 5x - z$, which is far from being satisfied by the pyrex glass composition. The ring structure is dominated by silicate rings, and it is not at all clear if borons can form typical flat boroxol rings, their percentage being too low to ensure frequent clustering of three borons at once. This means that the rigidity counting must be more sophisticated, and taking into account certain local rigid structures, and others with some of the angular constraints relaxed. Also, other parameters may turn out to be more important: chemical and mechanical resistance, low thermal expansion coefficient, and others as well. Since time immemorial, good glass has always been a compromise among these multiple factors [29].

References

1. Coxeter, M.C.M.: "Regular Polytopes", Methuen and Co., London (1948)
2. Hamermesh, M.: Group Theory and Its Application to Physical Problems. Pergamon, London (1962)
3. Steinhaus, H.: Mathematical Kaleidoscope. Plenum Press, London-New York (1960)
4. Kerner, R.: Phenomenological Lagrangian for the amorphous solid state. Phys. Rev. B, **28** (10), 5756 (1983)

5. Kerner, R., dos Santos, D.M.: Nucleation and amorphous and crystalline growth: a dynamical model in two dimensions. Phys. Rev. B. **37** (8), 3881 (1988)
6. Kerner, R.: J. Non-Cryst. Solids **135**, 155 (1991)
7. Kerner, R.: Physica B **215**, 267 (1996)
8. Kerner, R.: Stochastic description of agglomeration and growth processes in Glasses. Proceedings of the Conference "Nankai Symposium on Lattice Statistics and Mathematical Physics" Int. J. Modern Phys. B **16**, 1987 (2002)
9. dos Santos-Loff, D.M., Kerner, R., Micoulaut, M.: Europhys. Lett. **28**, 573 (1994)
10. Barrio, R.A., Duruisseau, J.-P., Kerner, R.: Structural properties of alkaliborate glasses derived from a theoretical model. Philos. Mag. B, **72**, 535 (1995)
11. Barrio, R.A., Kerner, R., Micoulaut, M., Naumis, G.G.: Evaluation of the concentration of boroxol rings in vitreous B2O3 by the stochastic matrix method. J. Phys.: Cond. Matter **9**, 9219 (1997)
12. Kerner, R., Naumis, G.G.: Stochastic matrix description of the glass transition. J. Phys.: Cond. Matter **12**, 1641 (2000)
13. Smalley, R.: Probing C_{60}. Science **242**, 1017 (1988)
14. Kroto, H., Heath, J.R., O'Brien, S.C., Curl, R.F., Smalley, R.E.: Nature **318**, 162 (1985)
15. Krätschmer, W., Fostiropulos, K., Huffmann, D.R.: The infrared and ultraviolet absorbtion spectra of laboratory-produced carbon dust: evidence for presence of the C_{60} molecule. Chem. Phys. Lett. **170**, 162 (1990)
16. Schmalz, T.G., Seitz, W.A., Klein, D., Hiley, C.E.: J. Am. Chem. Soc. **110**, 1113 (1991)
17. Kerner, R., Bennemann, K.H., Penson, K.: Model for the growth of fullerenes (C60, C70) from carbon vapour. Europhys. Lett. **19**, (5), 363 (1992)
18. Kerner, R.: Nucleation and growth of fullerenes. Comput. Mater. Sci. **2**, 500–508 (1994)
19. Terrones, H., Terrones, M., Moran-Lopez, J.L.: Curved nanomaterials. Curr. Sci. **81**, 1011 (2001)
20. Tomanek, D.: Proceedings of the Tsukuba Symposium on Carbon Nanotube. Physica B **323**, 86 (2002)
21. Bennemann, K.H., Kerner, R.: Theory for the growth of fullerenes. Z. Phys. Chem. **195**, 89 (1996)
22. Kerner, R.: New forms of pure carbon and possibility of icosahedral quasicrystalline growth, in "Theories of Matter". World Scientific 119 (1994) (a festschrift for Joseph L. Birman, ed. Solomon, A.)
23. Twarock, R.: New group structures for carbon onions and carbon nanotubes via affine extensions of noncrystallographic Coxeter groups. Phys. Lett. A, **300**, 437 (2002)
24. Phillips, J.C.: J. Non-Cryst. Solids **34**, 153
25. Thorpe, M.F.: J. Non-Cryst. Solids **57**, 355 (1983)
26. Boolchand, P., Thorpe, M.F.: Phys. Rev. B **50**, 10366 (1994)
27. Zhang, M., Boolchand, P.: The Central Role of Broken Bond-Bending Constraints in Promoting Glass Formation in the Oxides. Science **266**, 1355 (1994)
28. Boolchand, P.: ed. Insulating and Semiconducting Glasses. World Scientific Publishing Co. , Singapore-London-New-York (2001)
29. Kerner, R., Phillips, J.C.: Quantitative principles of silicate glass chemistry. Solid State Commun. **117**, 47 (2000)

4

Topological Defects in Carbon Nanocrystals

V.A. Osipov

Summary. The modern status of the problem of topological defects in graphitic nanocrystals is discussed. The gauge theory and topology are proved to be powerful methods in analyzing the electronic structure of variously shaped carbon nanoparticles. Both the eigenfunctions and the local density of states (DOS) near the pentagonal defects are calculated for three geometries: sphere, cone, and hyperboloid. It is found that the low-energy DOS has a cusp, which drops to 0 at the Fermi energy for any number of pentagons at the tip except 3. For three pentagons, the nonzero DOS across the Fermi level is formed.

4.1 Introduction

Topology and geometry have many applications in modern condensed matter physics (see, e.g., the books [1, 2]). The purpose of this brief review is to present some bright examples of using topology and geometry in a study of a new interesting class of carbon materials—carbon nanoparticles. The discovery of these cage-like molecules has attracted considerable attention of both experimentalists and theorists due to unique physical properties that are directly related to their exotic geometry. Moreover, there is reason to believe that an infinite variety of both carbon-based and some other materials with particular nanoscale shapes and forms can be produced, therefore increasing the significance of geometrical methods [3] in theoretical studies.

An additional interest in carbon nanoparticles originates from the fact that the exotic geometry is accompanied by topological defects. Note that topologically nontrivial objects play an important role in various physically interesting systems. It will suffice to mention the 't Hooft–Polyakov monopole in the non-Abelian Higgs model, instantons in quantum chromodynamics, solitons in the Skyrme model, Nielsen–Olesen magnetic vortices in the Abelian Higgs model, etc. (see, e.g., [4]). Note that similar objects are known in condensed matter physics as well. For instance, vortices in liquids and liquid crystals, solitons in low-dimensional systems (e.g., in magnetics, linear polymers, and organic molecules), as well as the famous Abrikosov magnetic

vortices in superconductors are a matter of common knowledge. Mathematically, all these objects appear in the framework of nonlinear models as partial solutions of strongly nonlinear equations. An important point is that all the solutions are topologically stable and belong to nontrivial homotopic sectors.

It should be noted that elastic media also leave room for topological defects known as dislocations and disclinations. Disclinations in liquid crystals are one of the best-studied cases. In particular, the known exact "hedgehog" solution has been obtained within the continuum model of nematics. It is interesting that a hedgehog-like solution was also found for a point 4π disclination within the framework of the gauge model [5]. An important advantage of the gauge model follows from the fact that it is similar to the known field theory models, first of all to the non-Abelian and Abelian Higgs models, where topological objects are studied well. Taking into account this similarity, two exact static solutions for linear disclinations have been found [6–8].

It is now well understood that the modern problems of condensed matter physics call for using new theoretical methods. As we show here, a theoretical description of variously shaped carbon nanocrystals requires involving differential geometry, topology, and gauge theory. These methods are not typical for condensed matter theory though they are widely used in the field theory and gravity.

4.2 Geometry and Topology of Carbon Nanoparticles

The high flexibility of carbon allows producing variously shaped carbon nanostructures: fullerenes, nanotubes, nanohorns, cones, toroids, graphitic onions, etc. In some sense, the carbon nanoparticles mediate between the molecular and bulk phases and can be considered as a third form of carbon along with diamond and graphite. Historically, fullerenes C_{60} (nicknamed also as Buckminsterfullerene or "bucky ball") were first discovered in 1985 [9]. They are tiny molecular cages of carbon having 60 atoms and making up the mathematical shape called truncated icosahedron (12 pentagons and 20 hexagons). Although the amount of C_{60} actually produced in the experiment was very small, these curious molecules right away attracted the attention of theorists. In 1990, the adaptation of arc technique for carbon rods gave a possibility of making C_{60} in gram quantities [10]. Since then, in the process of graphite vaporization variously shaped fullerene molecules have been produced. The more spherical of them are the C_{60} molecule and its generalizations like C_{240} and C_{540} molecules. Others are either slightly (like the C_{70} (see Fig. 4.1)) or remarkably deformed.

Soon after the fullerenes, other interesting carbon structures were discovered. First of all, carbon nanotubes of different diameters and helicity [11] were produced. It turns out that single-walled carbon nanotubes can be twisted, flattened, or bent around to form sharp corners. These distortions do not cause them to break (see Fig. 4.2). The mechanical, magnetic, and especially electronic properties of carbon nanotubes are found to be very specific (see,

Fig. 4.1. The fullerene C_{60} (*left*) and C_{70} molecules (*right*)

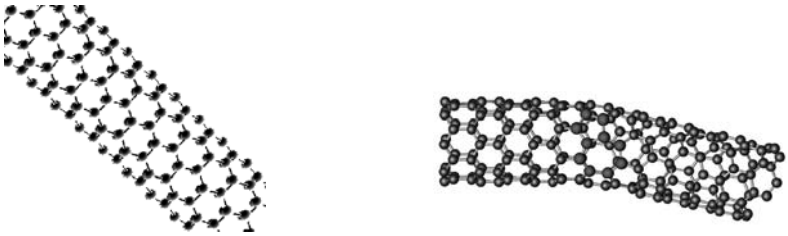

Fig. 4.2. Carbon nanotubes

e.g., [12]). For example, the nanotube can be either metallic or semiconducting depending on its diameter and helicity (see discussion later).

Carbon "onions" have also been found and they can be considered as carbon cages one inside the other [13]. The tubes and onions are likely to be composed of hexagonal and pentagonal carbon rings just like the fullerenes. However, structures having heptagonal rings are also possible. There has been much progress in recent years in producing toroids [14], cones [15, 16] (see Figs. 4.3,4.4), nanohorns [17], boxes [18], and helically coiled graphite [19].

One can expect that even more exotic configurations can be produced in experiments (see, e.g., Fig. 4.5). Indeed, theoretically the closed (without dangling bonds) fullerenes and nanotubes exhibiting high topologies (from genus 5 to genus 21) were suggested in [20]. This follows from the known Euler's theorem that relates the number of vertices, edges, and faces of an

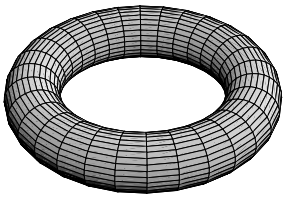

Fig. 4.3. Torus

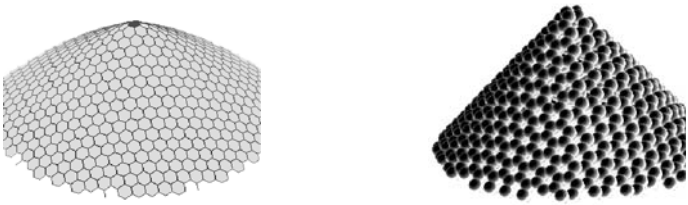

Fig. 4.4. Nanocones containing one (*left*) and two (*right*) pentagons at the apex

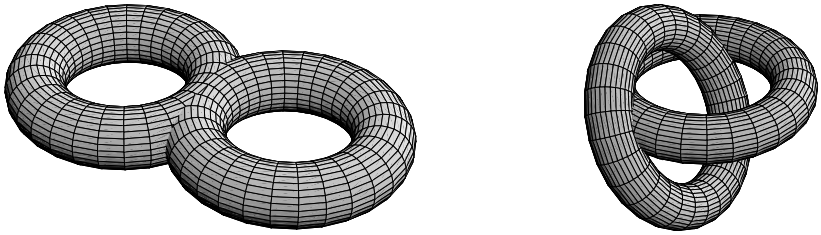

Fig. 4.5. Exotic configurations

object. For the hexagonal carbon lattice it can be written in the form [20]

$$\cdots 2n_4 + n_5 - n_7 - 2n_8 \cdots = \sum (6 - x) n_x = \chi = 12(1 - g), \qquad (4.1)$$

where n_x is the number of polygons having x sides, χ is the Euler characteristic, which is a geometrical invariant related to the topology of the structure, and g is the genus or a number of handles of an arrangement. So, for a sphere $g = 0$, a torus has $g = 1$ while for two "sticked" torii in Fig. 4.5 one has $g = 2$. According to (4.1) there is no contribution to the Gaussian curvature for $x = 6$. This means that 2D carbon lattice consisting only of hexagons is flat. On the contrary, to obtain a nontrivial shape one has to introduce some additional polygons. For example, in order to make a fullerene with genus 0 we need additionally 12 pentagons. In general, Euler's theorem allows to determine all the possible graphitic structures. As mentioned in [20], in accordance with (4.1) the complex structures with no pentagons (no positive Gaussian curvature) can be constructed if the genus is increased. In particular, an existence of the new stable family of fullerene-like structures (holey-balls and holey-tubes), which have high genus and no pentagonal rings, was predicted in [20].

By their nature, pentagons (as well as other polygons with $x \neq 6$) in a graphite sheet are topological defects. In particular, fivefold coordinated particles are orientational 60° disclination defects in the otherwise sixfold coordinated triangular lattice. This can be understood by realizing that a pentagon can be inserted in the hexagonal lattice by a cut-and-glue procedure typical for disclination defects. Namely, one has to cut out a 60° sector from a

graphene (a single layer of graphite) sheet and then glue together the two cut sides of the sheet. Moreover, if the departure from the flat surface is allowed, a cone whose apex angle is directly related to the disclination angle will be generated. Pentagonal defects in cones can therefore be considered as apical disclinations, and the opening angle is directly connected to the Frank index of the disclination. A cone's apex may consist of a combination of ring defects. Because of the symmetry of the graphite sheet, only five types of cones can be created from a continuous sheet of graphite. The total disclinations of all these cones are multiples of 60°, corresponding to the presence of a given number (n) of pentagons at the apices. It is important to mention that carbon nanocones with the cone angles of 19°, 39°, 60°, 85°, and 113° have been observed in a carbon sample [16]. Note that these angles might correspond to 300°, 240°, 180°, 120°, and 60° disclinations in graphite, respectively. Disks ($n = 0$) and one-open-end nanotubes ($n = 6$) have also been observed in the same sample [16]. This case was theoretically studied in [21–23]. At the same time, cones with apex angles of 30°, 50°, and 70° have also been found [20,24]. These angles are forbidden within this scenario. In [24, 25] the appearance of such cones was explained in terms of an open cone model. Another possibility gives a creation of partial disclinations. As is known, a finite graphite sheet with disclinations will be buckled to screen its energy [26]. In this case, one of the allowed geometries is the hyperboloid.

It should be noted that the presence of topological defects in the elastic medium changes the topology of space, a simple connected region becomes multiply connected whenever there are defects. As a result, the physical characteristics of quantum particles moving in defect medium can be modified in comparison with the defect-free case. Indeed, the Aharonov–Bohm-like (AB-like) effect in dislocated crystals (called "phase-dismatching") was predicted in [27]. It was found that the Schrödinger equation for a tight-binding electron is reduced to the AB-like equation in the presence of a screw dislocation (see also [28, 29]). In experiments, the effects of Berry's geometrical phase were established by analyzing the high-energy electron diffraction from a screw dislocation [30].

Among other effects it is necessary to note the prediction of the AB-like electron scattering due to disclinations [6, 31], an electron localization near topological defects [32, 33] as well as a formation of the polaron-type states near dislocations [34]. Note that a possibility of the solid state realization of the AB effect was suggested earlier in metals [35, 36] and in dielectrics [37]. It has been shown that the AB effect results in oscillations of physical characteristics (transport properties, magnetic susceptibility) with a certain fundamental period $\Phi_0 = hc/ne$, where $n = 1$ for pure metals and $n = 2$ for disordered metals and dielectrics.

One would expect some new physical phenomena arising from nontrivial topology of carbon nanoparticles. It is interesting to note in this connection that an important role of topology has been recently discovered in experiments with niobium and selenium. In particular, a Möbius strip (see Fig. 4.6) of

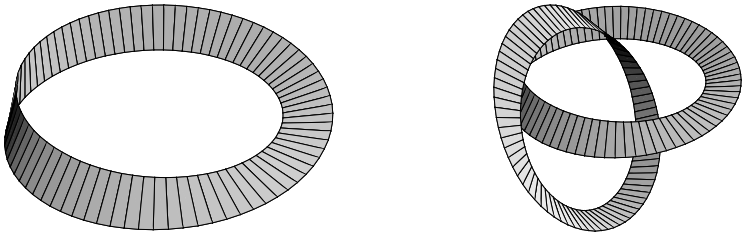

Fig. 4.6. Möbius stripe (*left*) and a more exotic configuration (*right*)

single microcrystals $NbSe_3$ has been produced by twisting a ribbon of material through 180° and joining its two ends, resulting in a distinct one-sided topology [38]. In a sense, these crystals can be considered as global disclinations. It was established that the electronic properties of the Möbius crystals are modified in comparison with the ring configuration. Namely, the temperature of charge-density-wave phase transition was observed to be 4 K lower than this in the ring. There is reason to believe that this effect is pure topological in origin. Evidently, topologically nontrivial crystal forms offer a new route to study topological effects in solid state physics.

4.3 Electronic Properties

Among the most unique features of carbon nanoparticles are their electronic properties. Electronic states in nanotubes, fullerenes, nanocones, nanohorns, as well as in other carbon configurations are the subject of an increasing number of experimental and theoretical studies. They already find use in the development of modern nanoscale electronic devices: flat panel displays, nanoswitches, molecular memory devices, transistors, electron field emitters, etc. It has been predicted and later observed in experiments that bending or stretching a nanotube changes its band structure, therefore changing the electrical properties: stretched nanotubes become either more or less conductive. Moreover, a nanotube's chiral angle (the angle between the axis of its hexagonal pattern and that of the tube) determines whether the tube is metallic or semiconducting (see, e.g., [12]). This finding could allow to build nanotube-based transducers sensitive to tiny forces.

Interesting changes in the electronic properties arise from topological defects. The peculiar electronic states due to topological defects have been observed in different kinds of carbon nanoparticles by scanning tunneling microscopy (STM). For example, STM images with five-fold symmetry (due to pentagons in the hexagonal graphitic network) have been obtained in the C_{60} fullerene molecule [39]. The peculiar electronic properties at the ends of carbon nanotubes (which include several pentagons) have been probed experimentally in [40, 41]. Recently, the electronic structure of a single disclination has been revealed on an atomic scale by STM [42], where the enhanced charge

density at the disclination, which was located at the apex of the conical protuberance of the graphitic particle, has been experimentally clarified.

The problem of peculiar electronic states near the pentagons in curved graphite nanoparticles was the subject of intensive theoretical studies in fullerenes [43, 44], nanotubes [45], nanohorns [46], and cones [21, 47]. In particular, analysis within the effective-mass theory has shown that a specific $\sqrt{3} \times \sqrt{3}$ superstructure induced by pentagon defects can appear in nanocones [48]. This prediction has been experimentally verified in [42]. A recent study [23] within both tight-binding and ab initio calculations shows the presence of sharp resonant states in the region close to the Fermi energy. The strength and the position of these states with respect to the Fermi level were found to depend sensitively on the number and the relative positions of the pentagons constituting the conical tip. In particular, a prominent peak that appears just above the Fermi level was found for the nanocone with three symmetrical pentagons (which corresponds to a 60° opening angle or, equivalently, to 180° disclination). A similar result has been recently obtained in the framework of the gauge-theory approach [47]. Note also that localized cap states in nanotubes have been recently studied in [49].

It is interesting to note that the problem of specific electronic states at the Fermi level due to disclinations is similar to that of the fermion zero modes for planar systems in a magnetic field. Generally, zero modes for fermions in topologically nontrivial manifolds have been of current interest both in the field theory and condensed matter physics. As was revealed, they play a major role in getting some insight into understanding anomalies [50] and charge fractionalization that results in unconventional charge–spin relations (e.g., the paramagnetism of charged fermions) [51] with some important implications for physics of superfluid helium (see, e.g., review [52]). Three-dimensional space-time Dirac equation for massless fermions in the presence of the magnetic field was found to yield $N - 1$ zero modes in the N-vortex background field [53]. As shown in [44], the problem of the local electronic structure of fullerene is closely related to Jackiw's analysis [53]. The importance of the fermion zero modes was also discussed in the context of the high-temperature chiral superconductors [54–56].

4.3.1 Theory: Basic Assumptions

Investigation of the electronic structure requires formulating a theoretical model describing electrons on arbitrary curved surfaces with disclinations taken into account. An important ingredient of this model can be provided by the self-consistent effective-mass theory describing the electron dynamics in the vicinity of an impurity in graphite intercalation compounds [57]. The most important fact found in [57] is that the electronic spectrum of a single graphite plane linearized around the corners of the hexagonal Brillouin zone coincides with that of the Dirac equation in (2+1) dimensions. This finding stimulated formulation of some field-theory models for Dirac fermions on

hexatic surfaces to describe electronic structure of variously shaped carbon materials: fullerenes [44], nanotubes [45, 49], and cones [21, 22].

The effective-mass theory for a 2D graphite lattice is equivalent to the $\boldsymbol{k}\cdot\boldsymbol{p}$ expansion of the graphite energy bands about the $\boldsymbol{K}$ point in the Brillouin zone when the intercalant potential is equal to 0. In fact, there are two kinds of sublattice points in a unit cell (two degenerate Bloch eigenstates at $\boldsymbol{K}$) and the electron wave function can therefore be approximated by

$$\Psi(\boldsymbol{k},\boldsymbol{r}) = f_1(\boldsymbol{\kappa})\mathrm{e}^{\mathrm{i}\boldsymbol{\kappa}\boldsymbol{r}}\Psi_1^S(\boldsymbol{K},\boldsymbol{r}) + f_2(\boldsymbol{\kappa})\mathrm{e}^{\mathrm{i}\boldsymbol{\kappa}\boldsymbol{r}}\Psi_2^S(\boldsymbol{K},\boldsymbol{r}),$$

where $\boldsymbol{k} = \boldsymbol{K} + \boldsymbol{\kappa}$. Keeping the terms of the order of $\boldsymbol{\kappa}$ in the Schrödinger equation results in a secular equation for the amplitudes $f_{1,2}(\boldsymbol{\kappa})$, which after diagonalization finally yields the 2D Dirac equation (see, for details, Ref. [57])

$$\mathrm{i}\gamma^\mu\partial_\mu\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \tag{4.2}$$

Here γ^μ are the Dirac matrices that in 2D reduce to the conventional Pauli matrices, the energy E is measured relative to the Fermi energy, and the two-component wave function ψ represents two graphite sublattices. As mentioned in [57], the $\boldsymbol{k}\cdot\boldsymbol{p}$ approximation essentially amounts to replacing the graphite bands by conical dispersions at the Fermi energy.

For our purpose, we need a generalization of (4.2) incorporating both a disclination field and a nontrivial background geometry. A possible description of disclinations on arbitrary 2D elastic surfaces is offered by the gauge approach [58]. In accordance with the basic assumption of this approach, disclinations can be incorporated in the elasticity theory Lagrangian by introducing a compensating $U(1)$ gauge fields W_μ. It is important that the gauge model admits exact vortex-like solutions for wedge disclinations [58], thus representing a disclination as a vortex of elastic medium. The physical meaning of the gauge field is that the elastic flux due to rotational defect (which is directly connected with the Frank vector (see Sect. 4.1)) is completely determined by the circulation of the W_μ field around the disclination line. In the gauge theory context, the disclination field can be straightforwardly incorporated in (4.2) by the standard substitution $\partial_\mu = \partial_\mu - \mathrm{i}W_\mu$.

Within the linear approximation to gauge theory of disclinations (which amounts to the conventional elasticity theory with linear defects), the basic field equation that describes the $U(1)$ gauge field in a curved background is given by

$$D_\mu F^{\mu k} = 0, \quad F^{\mu k} = \partial^\mu W^k - \partial^k W^\mu, \tag{4.3}$$

where covariant derivative $D_\mu := \partial_\mu + \Gamma_\mu$ involves the Levi-Civita (torsion-free, metric compatible) connection

$$\Gamma^k_{\mu\lambda} := (\Gamma_\mu)^k_\lambda = \frac{1}{2}g^{kl}\left(\frac{\partial g_{l\lambda}}{\partial x^\mu} + \frac{\partial g_{\mu l}}{\partial x^\lambda} - \frac{\partial g_{\mu\lambda}}{\partial x^l}\right) \tag{4.4}$$

with $g_{\mu k}$ being the metric tensor on a Riemannian surface Σ with local coordinates $x^\mu = (x^1, x^2)$. For a single disclination on an arbitrary elastic surface,

a singular solution to (4.3) is found to be [58]

$$W^k = -\nu\varepsilon^{k\lambda}D_\lambda G(x,y), \tag{4.5}$$

where

$$D_\mu D^\mu G(x^1,x^2) = 2\pi\delta^2(x^1,x^2)/\sqrt{g}, \tag{4.6}$$

with $\varepsilon_{\mu k} = \sqrt{g}\epsilon_{\mu k}$ being the fully antisymmetric tensor on $\Sigma, \epsilon_{12} = -\epsilon_{21} = 1$. It should be mentioned that (4.3)–(4.6) self-consistently describe a defect located on an arbitrary surface [58].

To describe fermions in a curved background, we need a set of orthonormal frames $\{e_\alpha\}$, which yield the same metric, $g_{\mu\nu}$, related to each other by the local SO(2) rotation,

$$e_\alpha \to e'_\alpha = \Lambda^\beta_\alpha e_\beta, \quad \Lambda^\beta_\alpha \in \mathrm{SO}(2).$$

It then follows that $g_{\mu\nu} = e^\alpha_\mu e^\beta_\nu \delta_{\alpha\beta}$, where e^μ_α is the zweibein, with the orthonormal frame indices being $\alpha,\beta = \{1,2\}$, and coordinate indices $\mu,\nu = \{1,2\}$. As usual, to ensure that physical observables are independent of a particular choice of the zweinbein fields, a local SO(2) valued gauge field ω_μ must be introduced. The gauge field of the local Lorentz group is known as the spin connection. For the theory to be self-consistent, the zweibein fields must be chosen to be covariantly constant [59]:

$$\mathcal{D}_\mu e^\alpha_\nu := \partial_\mu e^\alpha_\nu - \Gamma^\lambda_{\mu\nu} e^\alpha_\lambda + (\omega_\mu)^\alpha_\beta e^\beta_\nu = 0,$$

which determines the spin connection coefficients explicitly

$$(\omega_\mu)^{\alpha\beta} = e^\alpha_\nu D_\mu e^{\beta\nu}. \tag{4.7}$$

Finally, the Dirac equation (4.2) on a surface Σ in the presence of the $U(1)$ external gauge field W_μ is written as

$$\mathrm{i}\gamma^\alpha e^\mu_\alpha(\nabla_\mu - \mathrm{i}W_\mu)\psi = E\psi, \tag{4.8}$$

where $\nabla_\mu = \partial_\mu + \Omega_\mu$ with

$$\Omega_\mu = \frac{1}{8}\omega^\alpha{}_\mu{}^\beta[\gamma_\alpha,\gamma_\beta] \tag{4.9}$$

being the spin connection term in the spinor representation.

Note that the general analytical solution to (4.8) is known only for chosen geometries. One of them is the cone [21, 22]. For the sphere and the hyperboloid, which are of interest here, some approximations were used. In particular, asymptotic solutions at small r (which allow us to study electronic states near the disclination line) were considered in [47]. For this reason, the numerical calculations for all three geometries were performed in [60]. The results of both analytical and numerical studies are presented in Sects. 4.2 and 4.3.

4.4 Spherical Molecules

4.4.1 The Model

To describe a sphere, we employ the polar projective coordinates $x^1 = r$, $x^2 = \varphi$; $0 \le r < \infty, 0 \le \varphi < 2\pi$ with R being the radius of the sphere. In these coordinates, the metric tensor becomes

$$g_{rr} = 4R^4/(R^2 + r^2)^2, \quad g_{\varphi\varphi} = 4R^4 r^2/(R^2 + r^2)^2, \quad g_{r\varphi} = g_{\varphi r} = 0, \quad (4.10)$$

so that

$$\sqrt{g} := \sqrt{\det ||g_{\mu\nu}||} = 4R^4 r/(R^2 + r^2)^2.$$

Nonvanishing connection coefficients (4.4) take the form

$$\Gamma^r_{rr} = -\frac{2r}{R^2 + r^2}, \quad \Gamma^r_{\varphi\varphi} = -r\frac{R^2 - r^2}{R^2 + r^2}, \quad \Gamma^\varphi_{r\varphi} = \frac{1}{r}\frac{R^2 - r^2}{R^2 + r^2},$$

and the general representation for the zweibeins is found to be

$$e^1_r = e^2_\varphi = 2R^2 \cos\varphi/(R^2 + r^2), \quad e^1_\varphi = -e^2_r = -2R^2 \sin\varphi/(R^2 + r^2),$$

which in view of (4.7) gives

$$\omega^{12}_r = \omega^{21}_r = 0, \quad \omega^{12}_\varphi = -\omega^{21}_\varphi = 2r^2/(R^2 + r^2) =: 2\omega. \quad (4.11)$$

The following solution to (4.5) and (4.6) can be easily found

$$G = \log r; \quad W_r = 0, \quad W_\varphi = \nu, \quad r \neq 0.$$

Locally, it describes a topological vortex on the Euclidean plane, which confirms the observation that disclinations can be viewed as vortices in elastic media. Note that the elastic flux is actually characterized by the Frank vector $\boldsymbol{\omega}$, $|\boldsymbol{\omega}| = 2\pi\nu$ with ν being the Frank index. The elastic flow through a surface on the sphere is given by the circular integral

$$\frac{1}{2\pi}\oint \boldsymbol{W}\mathrm{d}\boldsymbol{r} = \nu.$$

Generally, there are no restrictions on the value of the winding number ν apart from $\nu > -1$ for topological reasons. This means that the elastic flux is "classical" in its origin; i.e., there is no quantization (in contrast to the magnetic vortex). However, if we take into account the symmetry group of the underlying crystal lattice, the possible values of ν become "quantized" in accordance with the group structure (e.g., $\nu = 1/6, 1/3, 1/2, ...$ for the hexagonal lattice). It is interesting to note that in some physically interesting applications vortices with the fractional winding number have already been considered (see, e.g., the discussion in [54]). Note also that a detailed theory of magnetic vortices on a sphere has been presented in [61].

In 2D, the Dirac matrices can be chosen as the Pauli matrices: $\gamma^1 = -\sigma^2$, $\gamma^2 = \sigma^1$ and (4.9) reduces to

$$\Omega_\varphi = \mathrm{i}\omega\sigma^3. \tag{4.12}$$

As a result, the Dirac operator $\widehat{\mathcal{D}} := \mathrm{i}\gamma^\alpha e_\alpha^{\ \mu}(\nabla_\mu + \mathrm{i}W_\mu)$ on the two-sphere becomes

$$\widehat{\mathcal{D}} = \widehat{\mathcal{D}}^\dagger = \frac{r^2+R^2}{2R^2}\begin{bmatrix} 0 & \mathrm{e}^{-\mathrm{i}\varphi}(-\partial_r + \frac{\mathrm{i}\partial_\varphi+\nu}{r} + \frac{\omega}{r}) \\ \mathrm{e}^{\mathrm{i}\varphi}(\partial_r + \frac{\mathrm{i}\partial_\varphi+\nu}{r} - \frac{\omega}{r}) & 0 \end{bmatrix}. \tag{4.13}$$

For massless fermions σ^3 serves as a conjugation matrix, and the energy eigenmodes are symmetric with respect to $E = 0$ ($\sigma^3\psi_E = \psi_{-E}$). The generator of the local Lorentz transformations $\Lambda_\alpha^\beta \in SO(2)$ takes the form $-\mathrm{i}\partial_\varphi$, whereas that of the Dirac spinor transformations $\rho(\Lambda)$ is

$$\Sigma_{12} = \frac{\mathrm{i}}{4}[\gamma_1, \gamma_2] = \frac{1}{2}\sigma^3.$$

The total angular momentum of the 2D Dirac system is therefore given by

$$L_z = -\mathrm{i}\partial_\varphi + \frac{1}{2}\sigma^3,$$

which commutes with the operator (4.13). Consequently, the eigenfunctions are classified with respect to the eigenvalues of $J_z = j+1/2,\ j = 0, \pm1, \pm2, ...,$ and are to be taken in the form

$$\psi = \begin{pmatrix} u(r)\mathrm{e}^{\mathrm{i}\varphi j} \\ v(r)\mathrm{e}^{\mathrm{i}\varphi(j+1)} \end{pmatrix}. \tag{4.14}$$

As follows from (4.13) the spin connection term can be taken into account by redefining the wave function as

$$\psi = \tilde{\psi}\sqrt{R^2+r^2}, \tag{4.15}$$

which reduces eigenvalue problem (4.8) to

$$\begin{aligned} \partial_r\tilde{u} - \frac{(j-\nu)}{r}\tilde{u} &= \tilde{E}\tilde{v}, \\ -\partial_r\tilde{v} - \frac{(j+1-\nu)}{r}\tilde{v} &= \tilde{E}\tilde{u}, \end{aligned} \tag{4.16}$$

where $\tilde{E} = 2R^2E/(R^2+r^2)$.

4.4.2 Extended Electron States

Let us consider an approximate solution to (4.16). The point is that, because we are mainly interested in electronic states near the disclination line, we can restrict our consideration to the case of small r. In this case, a solution to (4.16) (with (4.15) taken into account) is found to be

$$\begin{pmatrix} u \\ v \end{pmatrix} = A \begin{pmatrix} J_\eta(2Er) \\ \pm J_{\bar\eta}(2Er) \end{pmatrix}, \tag{4.17}$$

where $\eta = \pm(j-\nu)$, $\bar\eta = \pm(j-\nu+1)$, and A is a normalization factor. Therefore, there are two independent solutions with $\eta(\bar\eta) > 0$ and $\eta(\bar\eta) < 0$. Note that respective signs $\pm$ in (4.17) correspond to states with $E > 0$ and $E < 0$. As noted already, σ^3 serves as the conjugation matrix for massless fermions, and the energy eigenmodes are symmetric with respect to $E = 0$. One can therefore consider either case, for instance, $E > 0$.

The important restrictions come from the normalization condition

$$\int (|u|^2 + |v|^2)\sqrt{g}\,\mathrm{d}x^1\,\mathrm{d}x^2 = 1. \tag{4.18}$$

From (4.17), it follows that $A^2 \sim E$. On the other hand, the integrand in (4.18) must be nonsingular at small Er. This imposes a restriction on possible values of j. Namely, for $\eta, \bar\eta > 0$ one obtains $j - \nu > -1/2$, and for $\eta, \bar\eta < 0$ one has $j - \nu < -1/2$. As seen here, possible values of j do not overlap at any ν.

In the vicinity of a pentagon, the electron wave function reads

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \begin{pmatrix} E^{1/2+\eta} r^\eta, \\ E^{1/2+\bar\eta} r^{\bar\eta} \end{pmatrix}. \tag{4.19}$$

In particular, in the leading order, one obtains $\Psi \sim \sqrt{E}$, $\Psi \sim E^{1/3} r^{-1/6}$, and $\Psi \sim E^{1/6} r^{-1/3}$ for $\nu = 0{,}1/6{,}1/3$, respectively. Because the local density of states diverges as $r \to 0$, it is more appropriate to consider the total density of states (DOS) on a patch $0 < r \le \delta$ for small δ, rather than the local quantities. For this, the electron density should be integrated over a small disk $|r| < \delta$ (recall that r, φ are stereographically projected coordinates on the sphere). The result is

$$D(E,\delta) \propto \begin{cases} (E\delta)\delta, & \nu = 0; \\ (E\delta)^{2/3}\delta, & \nu = 1/6, 5/6; \\ (E\delta)^{1/3}\delta, & \nu = 1/3, 2/3; \\ \delta, & \nu = 1/2. \end{cases} \tag{4.20}$$

For the defect-free case ($\nu = 0$) we obtain the well-known behavior of the DOS in the δ disk given by $D(E,\delta) \sim E\delta^2$ (in accordance with the previous

analysis [57]). For $\nu = 1/6, 1/3, 2/3, 5/6$, the low-energy total DOS has a cusp that drops to 0 at the Fermi energy. Most intriguing is the case where $\nu = 1/2$ and a region of a nonzero DOS across the Fermi level is formed. This implies local metallization of graphite in the presence of 180° disclination. In the fullerene molecule, however, there are twelve 60° disclinations, and therefore, the case $\nu = 1/6$ is actually realized.

4.4.3 Numerical Results

The numerical calculations for the case of sphere a are presented in a recent paper [60]. As a starting point, the analytical asymptotic solutions found in Sect. 4.2 are considered. The initial value of the parameter r is defined as $r = 10^{-4}$. It is worth noting that the choice of the boundary conditions does not influence the behavior of the calculated wave functions and only the starting point depends on it. A dimensionless substitution $x = Er$ is used. The normalized numerical solutions to (4.16) are given in Fig. 4.7. The parameters are chosen to be $E = 0.01$ and $R = 1$. Note that here we present the solutions for dotted values $v'(= \tilde{v})$ and $u''(= \tilde{u})$. The local DOS is shown schematically in Figs. 4.8 and 4.9 for different n. Note that Fig. 4.9 also describes the dependence of the local DOS on a position of the maximum value of integrand in (4.18)(which actually characterizes the numerically calculated localization point of an electron).

Here and below $\delta = 0.1$. Note that in fact the choice of the value of δ does not influence the characteristic behavior of LDOS. As seen earlier, the DOS has a cusp that drops to 0 at the Fermi energy. The case $n = 3$ becomes

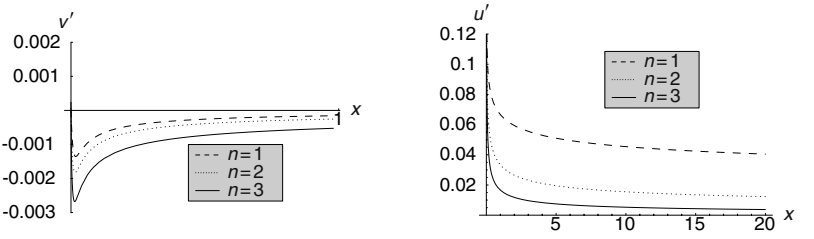

Fig. 4.7. The solutions $v'(x), u'(x)$ for different n

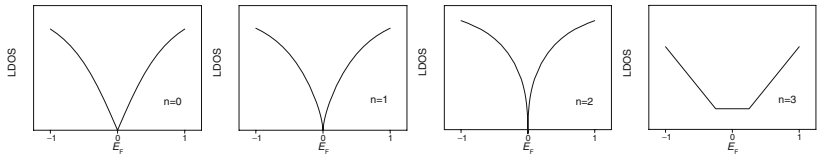

Fig. 4.8. Schematic densities of states near the Fermi energy in the case of sphere

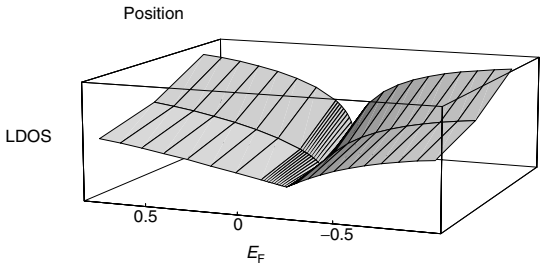

Fig. 4.9. 3D schematic plotting of the DOS near the Fermi energy for $n = 0, 1, 2$ (going from the front side to the back side)

distinguished. Let us emphasize once more that in the fullerene molecule there are twelve 60° disclinations, so that the case $n = 1$ is actually realized.

4.4.4 Zero-Energy Modes

An interesting issue to be addressed is the existence of zero-energy modes. For the two sphere, this problem can be solved exactly (see [44, 53]). Namely, for $E = 0$, (4.16) reduces to

$$\begin{aligned} \partial_r \tilde{u}_0 - \frac{(j-\nu)}{r}\tilde{u}_0 &= 0, \\ -\partial_r \tilde{v}_0 - \frac{(j+1-\nu)}{r}\tilde{v}_0 &= 0. \end{aligned} \tag{4.21}$$

One can construct self-conjugate solutions $\binom{\tilde{u}_0}{0}$ and $\binom{0}{\tilde{v}_0}$ where

$$\tilde{u}_0 = Ar^{j-\nu}, \qquad \tilde{v}_0 = Ar^{-(j-\nu+1)}. \tag{4.22}$$

The normalization condition

$$\int |\psi_0|^2 \sqrt{g}\, \mathrm{d}r\, \mathrm{d}\varphi = 1 \tag{4.23}$$

yields

$$2\pi A^2 \int_0^\infty \frac{4R^4 r^{2l}}{R^2 + r^2} r\, \mathrm{d}r = 1, \tag{4.24}$$

where $l = j - \nu$ for u_0 and $l = -(j - \nu + 1)$ for v_0. Finally, $A^2 = \sin \pi\bar{\eta}/4\pi^2 R^{2(1+\bar{\eta})}$ for u_0 and $A^2 = -\sin \pi\eta/4\pi^2 R^{2(1-\eta)}$ for v_0, respectively. Note that the restriction $-1 < j - \nu < 0$ serves to avoid divergence in (4.23). In the defect-free case ($\nu = 0$), this yields no zero modes on a sphere. Note that this agrees with a general observation that the Dirac operator can have no zero modes on a manifold with an everywhere positive Ricci scalar curvature $\mathcal{R}$. Indeed, one easily obtains $\widehat{\mathcal{D}}^2 = \Delta + \mathcal{R}/4$, where the Laplace–Beltrami

operator Δ has non-negative eigenvalues. For the two-sphere $\mathcal{R} = 1/R^2$, and therefore, $\widehat{\mathcal{D}}^2 > 0$.

For $\nu = 1/6$, which is of our interest here, the only possible value of j is $j = 0$, so that $u_0 \sim r^{-1/6}$ and $v_0 \sim r^{-5/6}$ near the disclination line. Thus, our analysis shows that two normalizable zero modes can exist on a sphere in the presence of a disclination vortex. Let us note that this conclusion agrees with [43] (where different continuum model was formulated) and differs from [44, 53] where either u_0 or v_0 was found to be normalizable. The reason is that in [44, 53] the external gauge field was assumed to be well behaved at the origin whereas here singular solutions are also admitted.

The total DOS on a patch $0 < r \leq \delta$ becomes

$$D(\delta) \propto \begin{cases} \delta^{1/3}, & \nu = 1/6, 5/6; \\ \delta^{2/3}, & \nu = 1/3, 2/3; \\ \delta, & \nu = 1/2. \end{cases} \tag{4.25}$$

As seen here, this behavior differs from (4.20) thus allowing to recognize the zero-eigenvalue states in experiments.

4.5 Nanocones

4.5.1 The Model

In the polar coordinates $(r, \varphi) \in R^2$ a cone can be regarded as an embedding

$$(r, \varphi) \to (ar\cos\varphi,\ ar\sin\varphi,\ cr), \quad 0 < r < 1,\ 0 \leq \varphi < 2\pi,$$

with a and c being the cone parameters. From this, the components of the induced metric can be easily read:

$$g_{rr} = a^2 + c^2, \qquad g_{\varphi\varphi} = a^2 r^2, \qquad g_{r\varphi} = g_{\varphi r} = 0. \tag{4.26}$$

The opening angle of a cone, θ, is determined by $\sin(\theta/2) = a/\sqrt{a^2 + c^2}$. Because the cone itself appears when one or more sectors are removed from graphene, all possible angles are divisible by $\pi/3$. Therefore, the Frank index of the apical disclination can be specified by $\nu = 1 - \sin(\theta/2)$. At $\nu = 0$ one gets a flat graphene sheet ($\theta = \pi$). For convenience, we introduce a parameter $\xi = 1 + c^2/a^2$, so that $\sin(\theta/2) = 1/\sqrt{\xi}$ and $1/\sqrt{\xi} = 1 - \nu$.

Nonvanishing connection coefficients (4.4) are now given by

$$\Gamma^r_{\varphi\varphi} = -r/\xi, \quad \Gamma^\varphi_{r\varphi} = \Gamma^\varphi_{\varphi r} = 1/r.$$

The general representation for the zweibeins is found to be

$$e^1_r = a\sqrt{\xi}\cos\varphi, \quad e^1_\varphi = -ar\sin\varphi, \quad e^2_r = a\sqrt{\xi}\sin\varphi, \quad e^2_\varphi = ar\cos\varphi,$$

which in view of (4.7) gives

$$\omega_r^{12} = \omega_r^{21} = 0, \quad \omega_\varphi^{12} = -\omega_\varphi^{21} = 1 - 1/\sqrt{\xi} =: 2\omega. \tag{4.27}$$

The external gauge potential is then found to be $W_r = 0$, $W_\varphi = \nu$, and the Dirac operator on the cone takes the form

$$\widehat{\mathcal{D}} = \begin{bmatrix} 0 & \mathrm{e}^{-\mathrm{i}\varphi}\left(-\frac{\partial_r}{\sqrt{a^2+c^2}} + \frac{1}{ar}(\mathrm{i}\partial_\varphi + \nu + \omega)\right) \\ \mathrm{e}^{\mathrm{i}\varphi}\left(\frac{\partial_r}{\sqrt{a^2+c^2}} + \frac{1}{ar}(\mathrm{i}\partial_\varphi + \nu - \omega)\right) & 0 \end{bmatrix}.$$

Making the substitution

$$\psi = \tilde{\psi} r^\alpha, \quad \alpha = \sqrt{\xi}\omega,$$

one reduces the eigenvalue problem (4.8) to

$$\begin{aligned} \partial_r \tilde{u} - \frac{\sqrt{\xi}}{r}(j-\nu)\tilde{u} &= \tilde{E}\tilde{v}, \\ -\partial_r \tilde{v} - \frac{\sqrt{\xi}}{r}(j+1-\nu)\tilde{v} &= \tilde{E}\tilde{u}, \end{aligned} \tag{4.28}$$

where $\tilde{E} = \sqrt{\xi} a\, E$.

4.5.2 Electron States

Unlike the previous case of the two sphere, the cone is essentially a flat manifold (the scalar curvature $\mathcal{R} = 0$ everywhere on the cone, except for the origin), and as a result, (4.28) allows an exact solution. Namely, the general solution to (4.28) is found to be [22]

$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = A r^{-\alpha} \begin{pmatrix} J_\eta(\tilde{E}r) \\ \pm J_{\bar{\eta}}(\tilde{E}r) \end{pmatrix}, \tag{4.29}$$

where $\eta = \pm(\sqrt{\xi}(j-\nu+1/2) - 1/2)$, and $\bar{\eta} = \pm(\sqrt{\xi}(j-\nu+1/2) + 1/2)$. As earlier, we consider the case where $E > 0$. Normalization condition (4.18) takes the form

$$2\pi\sqrt{\xi} a^2 A^2 \int_0^1 (J_\eta^2(\tilde{E}r) + J_{\bar{\eta}}^2(\tilde{E}r)) r \,\mathrm{d}r = 1. \tag{4.30}$$

The normalization factor can be extracted from the asymptotic formula for Bessel functions at large arguments. Indeed, in our case, $\bar{\eta} - \eta = 1$ so that $J_\eta^2 + J_{\bar{\eta}}^2 \to 2/\pi\tilde{E}r$ for $\tilde{E}r \gg 1$. Substituting this in (4.30) yields $A^2 = E/4a$. Clearly, (4.30) must be nonsingular at small r. This imposes a restriction on possible values of j. Namely, for $\eta, \bar{\eta} > 0$ one gets $j > -1$ (i.e., $j = 0, 1, 2, ...$) while for $\eta, \bar{\eta} < 0$ one has $j < -2\nu$ ($j = -1, -2, ...$ at $\nu < 1/2$).

We are interested in the electron states near the apex of a cone. As it follows directly from (4.29), for small r the wave functions behave as

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \begin{pmatrix} E^{1/2+\eta} r^{\eta} \\ E^{1/2+\bar{\eta}} r^{\bar{\eta}} \end{pmatrix}. \tag{4.31}$$

In the leading order, one obtains

$$\tilde{\Psi} \sim E^{(1-2\nu)/2(1-\nu)} r^{-\nu/(1-\nu)}.$$

In particular, we obtain $\tilde{\Psi} \sim \sqrt{E}$, $\tilde{\Psi} \sim E^{2/5} r^{-1/5}$, and $\tilde{\Psi} \sim E^{1/4} r^{-1/2}$ for $\nu = 0,\ 1/6,\ 1/3$, respectively.

Finally, the total DOS on the patch $0 < r \le \delta$ is found to be

$$D(E,\delta) \propto \begin{cases} E^{(1+2\nu)/(1-\nu)} \delta^{(\nu+2)/(1-\nu)}, & \eta, \bar{\eta} > 0; \\ E^{(1-2\nu)/(1-\nu)} \delta^{(2-3\nu)/(1-\nu)}, & \eta, \bar{\eta} < 0. \end{cases} \tag{4.32}$$

It should be stressed that, according to (4.32), a specific behavior of $D(E,\delta)$ occurs only for $\nu = 1/2$ where $D \sim E^0 \delta$. This agrees with the finding in [23], where the prominent peak just above the Fermi level was found for the nanocone with three symmetric pentagons (180° disclination). In the leading order, it follows from (4.32) that

$$D(E,\delta) \propto \begin{cases} E\delta^2, & \nu = 0; \\ E^{4/5}\delta^{9/5}, & \nu = 1/6; \\ E^{1/2}\delta^{3/2}, & \nu = 1/3; \\ \delta, & \nu = 1/2. \end{cases} \tag{4.33}$$

As seen here, the extended states with a nonzero DOS at E_{F} appear only at $\nu = 1/2$.

To examine the electron states at the Fermi energy, one has to return to (4.28) and set $E = 0$. The solution reads

$$u_0 = A r^{-\frac{1}{2}+\widetilde{j}} \sqrt{\xi}, \qquad v_0 = B r^{-\frac{1}{2}-\widetilde{j}} \sqrt{\xi}, \tag{4.34}$$

where $\widetilde{j} = j - \nu + 1/2$. A simple analysis shows that for $j = 0$ both u_0 and v_0 are normalizable on the cone of a finite size. Both solutions are singular. For $\nu = 1/6$ one gets $|u_0|^2 \sim r^{-1/5}/a^2$ and $|v_0|^2 \sim r^{-9/5}/a^2$. For any other j, either u_0 or v_0 is found to be normalizable and the solutions become nonsingular. As before, for singular states one can consider the total DOS. It is easy to find that $D \sim \delta^{1/5}$ for u_0 and $D \sim \delta^{9/5}$ for v_0. This result differs from [48] where, although in a different framework, the states on a finite cone with a single-pentagon defect have been found at the Fermi energy (these states decay away from the apex as $|\psi|^2 \sim r^{-2/5}$). At the same time, our study confirms the principal conclusion in [48,62] that the states contributing to the

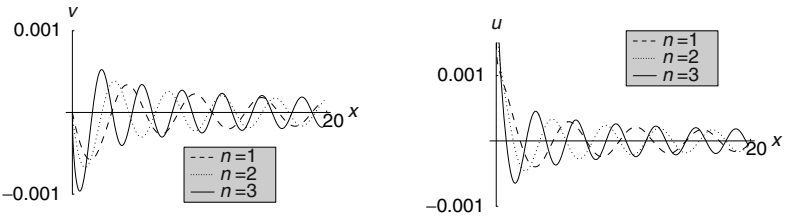

Fig. 4.10. The solutions $v(x), u(x)$ for different n

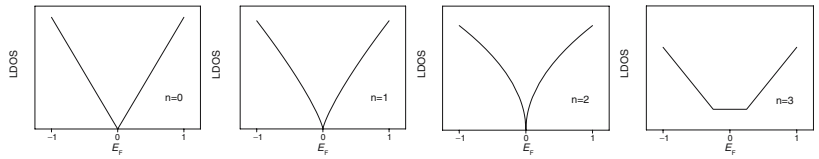

Fig. 4.11. Schematic densities of states near the Fermi energy in the case of cone

nonzero DOS at the Fermi energy exhibit a power-law behavior for a single-pentagon defect. Note also that in a monolayer graphite of infinite length ($a \to \infty$) there are no zero-energy electronic states on a single disclination. It should be emphasized that this conclusion agrees with the results of numerical calculations [62] where the local DOS at the Fermi level was found to be 0 for five-membered rings (pentagons). Note also that for $\nu = 1/2$, $D \sim \delta$ for both u_0 and v_0.

4.5.3 Numerical Results

It is interesting to present the results of numerical calculations [60]. The normalized numerical solutions to (4.28) for different n are shown in Fig. 4.10. The parameters are chosen to be $E = 0.01$, $a = 1$, and $c = 1$. The "total" DOS near the Fermi energy for the case of the cone is illustrated schematically in Fig. 4.11.

One can see that the "total" DOS has a cusp that drops to 0 at the Fermi energy. It should be stressed that a specific behavior takes place only for $n = 3$ where a nonzero DOS near the Fermi energy is found.

4.6 The Geometry of Hyperboloid

4.6.1 The Model

The upper half of a hyperboloid can be regarded as the embedding

$$(\chi, \varphi) \to (a \sinh\chi \cos\varphi, a \sinh\chi \sin\varphi, c \cosh\chi), \quad 0 \le \chi < \infty, 0 \le \varphi < 2\pi,$$

From these, the components of the induced metric can be obtained as

$$g_{\chi\chi} = a^2 \cosh^2 \chi + c^2 \sinh^2 \chi, \quad g_{\varphi\varphi} = a^2 \sinh^2 \chi, \quad g_{\varphi\chi} = g_{\chi\varphi} = 0, \quad (4.35)$$

which, for the nonvanishing connection coefficients, yields

$$\Gamma^{\chi}_{\chi\chi} = \frac{(a^2 + c^2) \sinh 2\chi}{2g_{\chi\chi}}, \quad \Gamma^{\chi}_{\varphi\varphi} = -\frac{a^2 \sinh 2\chi}{2g_{\chi\chi}}, \quad \Gamma^{\varphi}_{\varphi\chi} = \Gamma^{\varphi}_{\chi\varphi} = \coth \chi. \quad (4.36)$$

In a rotating SO(2) frame, the zweibeins become

$$e^1_\chi = \sqrt{g_{\chi\chi}} \cos\varphi,\ e^2_\chi = \sqrt{g_{\chi\chi}} \sin\varphi,\ e^1_\varphi = -a \sinh\chi \sin\varphi,\ e^2_\varphi = a \sinh\chi \cos\varphi, \quad (4.37)$$

which in view of (4.7) gives the spin connection coefficients

$$\omega^{12}_\chi = \omega^{21}_\chi = 0, \quad \omega^{12}_\varphi = -\omega^{21}_\varphi = \frac{1}{2}\left[1 - \frac{a \cosh\chi}{\sqrt{g_{\chi\chi}}}\right] =: \omega, \quad (4.38)$$

and therefore,

$$\Omega_\varphi = \mathrm{i}\omega\sigma^3. \quad (4.39)$$

The external gauge potential in this case is found to be $W_\chi = 0$, $W_\varphi = \nu$, and the Dirac operator on the hyperboloid takes the form

$$\widehat{\mathcal{D}} = \begin{bmatrix} 0 & \mathrm{e}^{-\mathrm{i}\varphi}\left(-\frac{\partial_\chi}{\sqrt{g_{\chi\chi}}} + \frac{1}{a\sinh\chi}(\mathrm{i}\partial_\varphi + \nu + \omega)\right) \\ \mathrm{e}^{\mathrm{i}\varphi}\left(\frac{\partial_\chi}{\sqrt{g_{\chi\chi}}} + \frac{1}{a\sinh\chi}(\mathrm{i}\partial_\varphi + \nu - \omega)\right) & 0 \end{bmatrix}.$$

It can be verified that $\widehat{\mathcal{D}} = \widehat{\mathcal{D}}^\dagger$.

The substitution

$$\tilde{\psi} = \psi\sqrt{\sinh\chi}$$

reduces the eigenvalue problem (4.8) to

$$\begin{aligned} \partial_\chi \tilde{u} - \sqrt{\coth^2\chi + b^2}\,\tilde{j}\tilde{u} &= \tilde{E}\tilde{v}, \\ -\partial_\chi \tilde{v} - \sqrt{\coth^2\chi + b^2}\,\tilde{j}\tilde{v} &= \tilde{E}\tilde{u}, \end{aligned} \quad (4.40)$$

where $\tilde{E} = \sqrt{g_{\chi\chi}}\,E$, $b = c/a$, and $\tilde{j} = j - \nu + 1/2$.

4.6.2 Electron States

To study electronic states on the hyperboloid one has to analyze (4.40). As for the sphere, let us consider the behavior of the electron states near the apex,

which is the case of small χ. One obtains

$$\begin{aligned} \partial_\chi \tilde{u} - \frac{\tilde{j}}{\chi}\tilde{u} &= Ea\tilde{v}, \\ -\partial_\chi \tilde{v} - \frac{\tilde{j}}{\chi}\tilde{v} &= Ea\tilde{u} \end{aligned} \tag{4.41}$$

with the obvious solution

$$\tilde{u} = A\sqrt{Ea\chi}J_{|j-\nu|}(Ea\chi), \quad \tilde{v} = A\sqrt{Ea\chi}J_{|j-\nu+1|}(Ea\chi).$$

As seen here, this is exactly the case of a sphere, which should not be surprising, because these two manifolds are locally diffeomorphic. Evidently, the "total" (on the disk $|r| < \delta$) DOS is the same as on the sphere. However, for hyperboloid the problem is more intricate due to the requirement to fulfill the normalization condition (see the numerical calculations given later).

An interesting situation arises for the zero-energy solution. Let us consider the zero-energy modes setting $E = 0$ in (4.41). The general solution is found to be

$$\begin{aligned} \tilde{u}(\chi) &= A\left[(k\cosh\chi + \Delta)^{2k}\,\frac{\Delta - \cosh\chi}{\Delta + \cosh\chi}\right]^{\frac{\tilde{j}}{2}}, \\ \tilde{v}(\chi) &= A\left[(k\cosh\chi + \Delta)^{2k}\,\frac{\Delta - \cosh\chi}{\Delta + \cosh\chi}\right]^{-\frac{\tilde{j}}{2}}, \end{aligned} \tag{4.42}$$

where $k = \sqrt{1+b^2}$, $\Delta = \sqrt{1 + k^2\sinh^2\chi}$. An important restriction comes from the normalization condition, which on a finite hyperboloid yields $\tilde{j} > -1/2$ for $u(\chi)$ and $\tilde{j} < 1/2$ for $v(\chi)$. One can see that for $-1/2 < \tilde{j} < 1/2$ both $u(\chi)$ and $v(\chi)$ are normalizable simultaneously. For the zero-energy mode, the total DOS on a finite hyperboloid is found to be the same as on the sphere (see (4.25)).

Although the local electronic structures are similar on the hyperboloid and the sphere, there is a principal global distinction. In proving this, let us consider an unbounded hyperboloid (full locus). In this case, one has to take into account additional restrictions at the upper limit of the integral in (4.18). One obtains $-1/2 < \tilde{j} < -1/2k$ for $u(\chi)$ and $1/2k < \tilde{j} < 1/2$ for $v(\chi)$. Thus, either $u(\chi)$ or $v(\chi)$ becomes normalizable on the hyperboloid of infinite volume. One can see that as $(c/a) \to 0$ a normalizable solution does not exist. In fact, under this condition the hyperboloid changes over to a plane. Consequently, our results are in accordance with the planar case. The total DOS on an infinite hyperboloid for a variety of defects is as follows:

$$D(\delta) \propto \begin{cases} \delta^{1/3}, & \nu = 1/6, 5/6; \quad c/a > \sqrt{5}/2, \\ \delta^{2/3}, & \nu = 1/3, 2/3; \quad c/a > 2\sqrt{2}. \end{cases} \tag{4.43}$$

Note that normalizable zero-energy states do not exist for the defect with $\nu = 1/2$ or for the defect-free case $\nu = 0$. The most important conclusion from our consideration is that there is a possibility for the true zero-mode fermion state on the hyperboloid. As we have shown, the normalized zero-mode states on both the sphere and the cone exist only for a finite system size and disappear in the infinite-size limit. For an infinite hyperboloid, a normalized zero-energy electron state can exist in the presence of a disclination flux.

4.6.3 Numerical Results

A more clear difference comes from the numerical study. The results of the numerical calculations are shown in Fig. 4.12 where the parameters are chosen to be $E = 0.01$, $a = 1$, and $c = 1$. Note that the starting point in this case was chosen to be $y = 0.01$. As can be seen, for the hyperboloid the electron eigenfunctions behave similar to the sphere near the disclination line and differ remarkably at large distances. In addition, there is a problem with the normalization of the solution for hyperboloid. Actually, the integrand is found to be constantly growing with increasing parameter y. Due to this problem (coming from the hyperboloid geometry itself) it is impossible to perform numerical calculations of the DOS.

To compare the behavior of the solutions $u(u')$ for every kind of the geometries the combined pictures are shown in Fig. 4.13 for $n = 1, 2$. It can be seen that the solutions for the sphere and the hyperboloid have a similar behavior

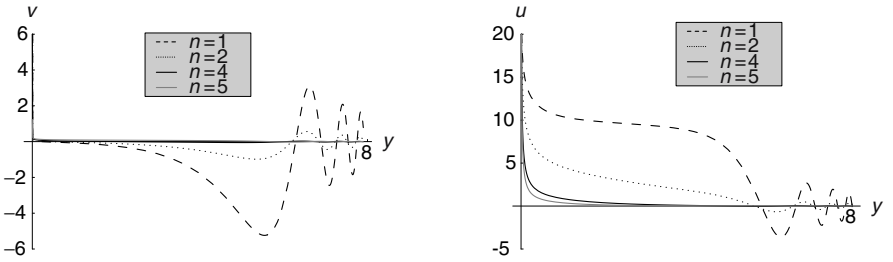

Fig. 4.12. Solutions $v(y), u(y)$ for different n

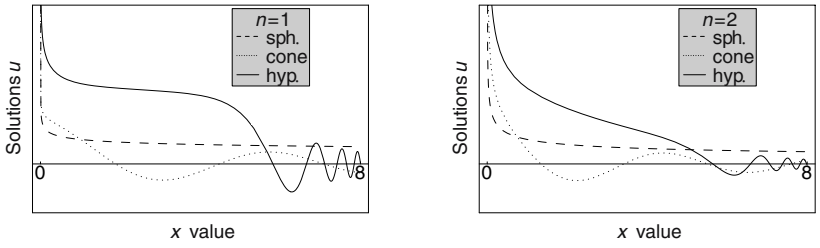

Fig. 4.13. The schematic pictures for $u(x)$ for three geometries in the case of $n = 1, 2$

near the disclination line at small $x(y)$, as already discussed. Let us note that the solution for $u(u')$ is found to be of the decisive importance in the final results for all three geometries, which is consistent with the previous analytical results. The choice of the parameters (R, c, a) does not influence the main characteristics of the calculated wave functions.

In summary, the numerical calculations confirm the finding that the pentagonal defects in graphite nanoparticles markedly modify the low-energy electronic structure. This is evident from both the exact form of wave functions and the local density of electron states. As seen from Fig. 4.9, in the case of the sphere the local DOS increases with a distance from the disclination line for defects with $n = 1, 2$. The low-energy total DOS has a characteristic cusp at the Fermi energy for any number of pentagons except $n = 3$, where the enhanced charge density at the Fermi energy is found.

4.7 Conclusions

There are many interesting applications of geometrical and topological methods to actual problems of modern condensed matter physics. As shown earlier, the physics of carbon nanoparticles is one of the striking examples. The geometry and topology is found to influence the main physical characteristics of graphite nanoparticles, first of all, their electronic properties. The topological defects (disclinations) appear as generic defects in closed carbon structures. For 180° disclination (three pentagons), the electronic DOS is found to be remarkably increased. Physically this means local metallization, thus suggesting some important applications of nanocone-based structures in microelectronic devices. First of all, such a remarkable increase of the DOS must provoke the corresponding enhancement of the field emission current, thereby decreasing the threshold voltage for emitted electrons. It should be noted that this conclusion agrees well with the results in [23], where the prominent peak appearing just above the Fermi level was established in a nanocone with three pentagons at the apex. It was proposed that such peculiar nanocones are good candidates for nanoprobes in scanning probe microscopy and excellent candidates for field-emission devices. As mentioned in [23], the nanocones with free pentagons at the tip have the highest probability of nucleation and are frequently observed [16]. It is expected that localized states at the Fermi level may give rise to materials with novel electronic and magnetic properties.

It should be emphasized that a large variety of closed graphitic structures is generally expected to be produced. Therefore, the theoretical study of various topologically nontrivial objects as well as of topological defects in graphite and other materials is of great importance. There is reason to believe that application of geometrical and topological methods to condensed matter physics will result in considerable progress in the near future.

Acknowledgments. I would like to thank Profs. A.M. Kosevich, M.I. Monastyrskii, and B. Zhilinskii for fruitful discussions during the workshop. The main results presented in this short review have been obtained in coauthorship with Drs E.A. Kochetov, M. Pudlak, and R. Pincak to whom I am very grateful. This work has been supported by the Science and Technology Department of Moscow region and the Russian Foundation for Basic Research under grant no. 01-02-97021.

References

1. Monastyrskii, M.I.: Topology of Gauge Fields and Condensed Matter. Plenum, New York (1993)
2. Sadoc, J.F. ed: Geometry in Condensed Matter Physics. World Scientific (1990)
3. Dubrovin, B.A., Novikov, S.P., Fomenko, A.T.: Modern Geometry—Methods and Applications: The Geometry of Surfaces, Transformation Groups and Fields. Springer-Verlag, New York (1993)
4. Rajaraman, R.: Solitons and Instantons. North-Holland, Amsterdam (1982)
5. Osipov, V.A.: Phys. Lett. A **146**, 67 (1990)
6. Osipov, V.A.: Phys. Lett. A **164**, 327 (1992)
7. Osipov, V.A.: J. Phys. A: Math. Gen. **26**, 1375 (1993)
8. Osipov, V.A.: Phys. Lett. A **193**, 97 (1994)
9. Kroto, H.W., Heath, J.R., O'Brien, S.C. et al.: Nature **318**, 162 (1985)
10. Kratschmer, W., Lamb, L.D., Fostiropoulos, K. et al.: Nature **347**, 354 (1990)
11. Iijima, S.: Nature **354**, 56 (1991)
12. Ebbesen, T.W.: Phys. Today, (June) **26** (1996)
13. Ugarte, D.: Nature **359**, 707 (1992)
14. Liu, J., Dai, H.J., Hafner, J.H. et al.: Nature **385**, 780 (1997)
15. Ge, M., Sattler, K.: Chem. Phys. Lett. **220**, 192 (1994)
16. Krishnan, A., Djuardin, E., Treacy, M.M.J. et al.: Nature **388**, 451 (1997)
17. Iijima, S., Yudasaka, M., Yamada, R. et al.: Chem. Phys. Lett. **309**, 165 (1999)
18. Saito, Y., Matsumoto, T.: Nature **392**, 237 (1998)
19. Amenlinckx, S., Zhang, X.B., Berbaerts, D.: Science **265**, 635 (1994)
20. Terrones, H., Terrones, M.: Carbon **36**, 725 (1998)
21. Lammert, P.E., Crespi, V.H.: Phys. Rev. Lett. **85**, 5190 (2000)
22. Osipov, V.A., Kochetov, E.A.: JETP Lett. **73**, 631 (2001)
23. Charlier, J.-C., Rignanese, G.-M.: Phys. Rev. Lett. **86**, 5970 (2001)
24. Kiselev, N.A., Sloan, J., Zakharov, D.N. et al.: Carbon **36**, 1149 (1998)
25. Terrones, H., Hayashi, T., Munoz-Navia, M. et al.: Chem. Phys. Lett. **343**, 241 (2001)
26. Nelson, D.R., Peliti, L.: J. Phys. (Paris) **48**, 1085 (1987)
27. Kawamura, K.: Z. Physik B. **29**, 101 (1978); ibid. **30**, 1 (1978)
28. Bausch, R., Schmitz, R., Turski, L.A.: Phys. Rev. B **59**, 13491 (1999)
29. Furtado, C., Bezerra, V.B., Moraes, F.: Europhys. Lett. **52**, 1 (2000)
30. Bird, D.M., Preston, A.R.: Phys. Rev. Lett. **61**, 2863 (1988)
31. Azevedo, S., Moraes, F.: Phys. Lett. A **246**, 374 (1998)
32. Osipov, V.A.: J. Phys. A: Math. Gen. **24**, 3237 (1991)

33. Furtado, C., Moraes, F.: Phys. Lett. A **188**, 394 (1994)
34. Osipov, V.A.: Phys. Rev. B **51**, 8614 (1995)
35. Kulik, I.O.: JETP Lett. **11**, 275 (1970)
36. Altshuler, B.L., Aronov, A.G., Spivak, B.Z.: JETP Lett. **33**, 94 (1981)
37. Bogachek, E.N., Krive, I.V., Kulik, I.O. et al.: JETP **97**, 603 (1990)
38. Tanda, S., Tsuneta, T., Okajima, Y. et al.: Nature (London) **417**, 397 (2002)
39. Hou, J.G., Yang, J., Wang, H. et al.: Phys. Rev. Lett. **83**, 3001 (1999)
40. Carroll, D.L., Redlich, P., Ajayan, P.M. et al.: Phys. Rev. Lett. **78**, 2811 (1997)
41. Kim, P., Odom, T.W., Huang, J.-L. et al.: Phys. Rev. Lett. **82**, 1225 (1999)
42. An, B., Fukuyama, S., Yokogawa, K. et al.: Appl. Phys. Lett. **78**, 3696 (2001)
43. González, J., Guinea, F., Vozmediano, M.A.H.: Phys. Rev. Lett. **69**, 172 (1992); Nucl. Phys. B **406**, 771 (1993)
44. Osipov, V.A., Kochetov, E.A.: JETP Lett. **72**, 199 (2000)
45. Kane, C.L., Mele, E.J.: Phys. Rev. Lett. **78**, 1932 (1997)
46. Berber, S., Kwon, Y.-K., Tománek, D.: Phys. Rev. B **62**, R2291 (2000)
47. Osipov, V.A., Kochetov, E.A., Pudlak, M.: JETP **96**, 140 (2003)
48. Kobayashi, K.: Phys. Rev. B **61**, 8496 (2000)
49. Yaguchi, T., Ando, T.: J. Phys. Soc. Jpn. **71**, 2224 (2002)
50. Jackiw, R., Rebbi, C.: Phys. Rev. D **16**, 1052 (1977)
51. Jackiw, R., Schrieffer, J.R.: Nucl. Phys. B **190**, 253 (1981)
52. Salomaa, M.M., Volovik, G.E.: Rev. Mod. Phys. **59**, 533 (1987)
53. Jackiw, R.: Phys. Rev. D **29**, 2375 (1984)
54. Volovik, G.E.: JETP Lett. **63**, 763 (1996)
55. Volovik, G.E.: JETP Lett. **70**, 609 (1999)
56. Volovik, G.E.: JETP Lett. **70**, 792 (1999)
57. DiVincenzo, D.P., Mele, E.J.: Phys. Rev. B **29**, 1685 (1984)
58. Kochetov, E.A., Osipov, V.A.: J. Phys. A: Math. Gen. **32**, 1961 (1999)
59. Green, M.B., Schwartz, J.H., Witten, E.: Superstring Theory Vol. 2 Cambridge (1988)
60. Pincak, R., Osipov, V.A.: Phys. A **314**, 315 (2003)
61. Ovrut, B.A., Tomas, S.: Phys. Rev. D **43**, 1314 (1991)
62. Tamura, R., Tsukada, M.: Phys. Rev. B **49**, 7697 (1994)

5

Physics from Topology and Structures

J. Yi

Summary. This chapter reviews physical properties resulting from topology, taking representative systems whose space of topological origin is presented. In addition, we review Josephson junction and one-dimensional conductors described by the collective degrees of freedom, that is, phase variables defined on a compact interval. Then the Shapiro steps in Josephson junctions and the Aharonov–Bohm effects by instanton tunneling are discussed. Various quantum phases defined on a closed path in real space are introduced, and as their well-known demonstration, persistent currents are discussed. Finally, we introduce carbon nanotubes where genuine lattice structures together with the topology define their electric properties.

5.1 Introduction

In condensed matter physics, one can easily note that the main stream runs with, so to say, a *dirty* spinnaker—randomness, disorder, frustration, and chaos. Whereas one may think of high-energy terminology, unification, charm, beauty, and supersymmetry (we are often interested even in symmetry breaking!), dirty systems could rather discourage a freshman in graduate school from choosing her or his major in condensed matter physics. Yet, here is a fact that we should not overlook : the best precision comes from dirty systems. In the presence of strong magnetic fields, Hall resistance has been observed to be quantized in units of $h/2e^2$ with an accuracy in a few parts per million. The accuracy of the quantum Hall effect (QHE) is indeed so impressive that it is used as a resistance standard. One may ask "why so accurate?" Instead of a detailed discussion (that is presented in Sect. 5.2), we merely answer that it is a consequence of *nontrivial topology* robust to any external perturbations, unlike symmetry that is easily broken.

To appreciate what nontrivial topology stands for, we draw attention to the question: how many Ts in TOPOLOGY? At first glance, one can answer 1, which is, of course, not the right answer in the view of topology—often called *rubber geometry*. Allowing deformation of the characters, we have four Ts by spreading the two upper branches of Y to form a π-angle, unfolding the

arc of G, and stretching the tiny head decoration of L, for example. On the other hand, O or P cannot fall into the "T" group, and nor, be deformed into a single point via continuous transformation. Such noncontractible geometries are referred to as nontrivial (or nonsimply connected) topology.

This chapter addresses representative systems having characteristic properties resulting from topology. We organize this chapter according to the space embedding the topology; (i) momentum space; (ii) phase space; and (iii) real space. For (i), QHE and its topological origin are presented in Sect. 5.2. For (ii), Sects. 5.3 and 5.4 review Josephson junction (JJ) and one-dimensional conductors described by the collective degrees of freedom, respectively, that is, phase variables defined on a compact interval. For an introduction to topological aspects, the Shapiro steps in JJs and the Aharonov–Bohm (AB) effects by instanton tunneling are discussed. Section 5.5 is devoted to introducing various quantum phases defined on a closed path in real space. Quantum phases that can alter the ground state properties of small-scaled conductors, whereby currents can flow eternally (if without dissipation), namely the persistent currents, in the system are also discussed. In Sect. 5.6, we introduce carbon nanotubes where a genuine lattice structure together with the topology defines the electric properties. We also study how a tiny change in the windings of hexagons along the tube can transform the tube metallic into a large-gap semiconductor. Furthermore, an even more transparent view of the role of topology can be made in transport properties through a carbon nanotube torus.

5.2 Quantum Hall Effect

One of the most significant discoveries in the 1980s was QHE [1]. Normally in solid state experiments, scattering processes introduce enough uncertainty, so that most results have error bars of several percentages. For example, the conductance of a ballistic conductor has been shown to be quantized in units of $h/2e^2$. However, this is true as long as we are not bothered by deviations of a few percentages since real conductance is usually not truly ballistic. On the other hand, in the presence of strong magnetic fields, the Hall resistance has been observed to be quantized in units of $h/2e^2$ with an accuracy in a few parts per million.

This impressive accuracy arises from the nearly complete suppression of momentum relaxation processes in the quantum Hall regime, resulting in a truly ballistic conductor. This is achieved because at high magnetic fields the electronic states carrying current in one direction are localized on one side of the sample, while those carrying current in other direction are localized on the other side. Due to the formation of this divided highway, there is hardly any overlap between the two groups of states and hence back scattering cannot take place even though impurities are present in the system. In this section we give a general discussion of the two-dimensional electron gas (2DEG) in

a magnetic field, including the integer quantum Hall effect (IQHE) and its topological origin.

We consider noninteracting electrons moving on a $(x-y)$ plane in the presence of magnetic field $\mathbf{B} = B\hat{z}$. We take the vector potential of the form $\mathbf{A} = By\widehat{x}$ so that the time-independent Schrödinger equation can be written as

$$\left[\frac{1}{2m}\left(p_x + \frac{eB}{c}y\right)^2 + \frac{p_y^2}{2m}\right]\Psi(x,y) = E\Psi(x,y). \tag{5.1}$$

Note that the choice of the vector potential is not unique for the given magnetic field: for instance, we could choose $A_x = 0$ and $A_y = -Bx$. The solution would then look very different although the physics must remain the same. With the translational invariance along x-direction in (5.1), we can express the solution in the form of plane waves, $\Psi(x,y) = \mathrm{e}^{-\mathrm{i}kx}\chi(y)$, where the transverse function satisfies the equation

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + \frac{1}{2}m\omega_{\mathrm{c}}^2(y-y_k)^2\right]\chi(y) = E\chi(y). \tag{5.2}$$

Here we have defined the cyclotron frequency $\omega_{\mathrm{c}} \equiv |e|B/mc$ and $y_k \equiv c\hbar k/|e|B$. In fact, (5.2) is simply a one-dimensional Schrödinger equation of a harmonic oscillator centered at y_k. So, it is easy to get the well-known eigenenergies and eigenfunctions: $E_n = (n+1/2)\hbar\omega_{\mathrm{c}}$ and $\chi_{n,k} = e^{(q-q_k)^2/2}H_n(q-q_k)$, with $H_n(q)$ being the nth Hermite polynomial, and $q \equiv \sqrt{m\omega_{\mathrm{c}}/\hbar}y$. These energy levels E_n with different values of n are referred to as the Landau levels. Although the eigenfunctions have the form of plane waves, these waves have no group velocity because the energy is independent of k. If we were to construct a wavepacket out of these localized states, it would not move. This keeps parallel with what we would expect from classical dynamics, which predicts that an electron in a magnetic field is described by the closed orbit not moving in any particular direction. Further, the spatial extent of each eigenfunction in the y-direction is approximately $\sqrt{\hbar/m\omega_{\mathrm{c}}}$, which is equal to the radius of the classical orbit if the energy of the electron is $\hbar\omega_{\mathrm{c}}/2$.

One question that often comes up is: how many electrons can fit into one Landau level? We can obtain the answer to the question by noting that the allowed values of k are spaced by $2\pi/L_x$ with L_x being the length of the plane in the x-direction, which means that the corresponding wavefunctions are spaced by $\Delta y_k = 2\pi\hbar c/|e|BL_x$ along the y-direction with the spatial extent L_y. Hence the total number of states is given by $N = 2L_y/\Delta y_k = |e|BS/\pi\hbar c$, where factor 2 accounts for the spin states and S is the sample area ($S = L_xL_y$).

After these preliminaries we now turn to the Hall effect by starting with a brief account of its conventional (classical) theory. Consider a slab of conducting material in crossed electric and magnetic fields, $\mathbf{E} = E\widehat{y}$ and $\mathbf{B} = B\widehat{z}$,

where voltage is induced in the direction orthogonal to the crossed fields, as manifested by an induced current flowing in that direction, i.e., the Hall current. If $\mathbf{E} = 0$, the electrons move in circular orbits. The effect of the non-vanishing $\mathbf{E}$ can be boosted by performing a Lorentz boost along the x-axis with the boost velocity $v_{\mathrm{D}} = cE/B$ (as $E \ll B$ in typical experiments, so $v_{\mathrm{D}} \ll c$). Therefore the resulting trajectories become the superpositions of the circular and the uniform drift motions. The net current (with n being the 2D carrier density)

$$J_{\mathrm{H}} = env_{\mathrm{D}} = \left(\frac{enc}{B}\right) E,$$

defines the Hall conductivity $\sigma_{\mathrm{H}} = enc/B$. In two dimensions, the ratio e^2/h has the dimension of the conductivity, and $nhc/eB = \nu$ is defined as the filling factor of the system. We thus obtain $(h/e^2)\sigma_{\mathrm{H}} = \nu$. The main features of IQHE are the plateaus of integer values (in units of e^2/h) in the Hall conductivity, centered around the integer fillings.

The system, in which QHE shows up, may be disordered. In a perfectly pure sample, the density of states is composed of a series of Δ-functions centered at the position of the Landau levels. On the other hand, in the presence of impurities each Landau level is broadened into a band. In general, the localized states exist near the band edge. As long as the Fermi level lies within the tail of the localized states, there are no electron states available for contributing to conduction. On the contrary, conductivity is expected to rise sharply when the Fermi level sweeps the new set of the extended states. This picture for the density of states explains the plateau, but it does not explain the integer values of the Hall conductance.

Let us now have a rather detailed look at the Hall conductance obtained through the use of the Kubo formula, which is given for an individual eigenstates $|n\rangle$

$$\sigma_{H,n} = \frac{\mathrm{i}\hbar e^2}{A} \sum_{m \neq n} \frac{(v_x)_{nm}(v_y)_{mn} - (v_y)_{nm}(v_x)_{mn}}{(E_n - E_m)^2},$$

where A is the sample area and $\mathbf{v}_{nm}$ is defined as

$$\begin{aligned} \mathbf{v}_{nm} &\equiv \langle n|\mathbf{v}|m\rangle \\ &= \left\langle u_n \middle| \frac{1}{m}\left(\mathbf{p} + \hbar\mathbf{k} + \frac{e}{c}\mathbf{A}\right) \middle| u_m \right\rangle. \end{aligned}$$

Here the Bloch function $u_n(x)$ has been defined according to $\psi_n(\mathbf{x}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} u_n(\mathbf{x})$, which leads to the expression for the x-component of $\mathbf{v}_{nm}$

$$(v_x)_{nm} = \frac{E_n - E_m}{\hbar} \left\langle \frac{\partial u_n}{\partial k_x} \middle| u_m \right\rangle.$$

We thus obtain the Hall conductivity in the compact form:

$$\sigma_{H,n} = \frac{\mathrm{i}e^2}{\hbar} \int \frac{d^2}{(2\pi)^2} \left[\left\langle \frac{\partial u_n}{\partial k_x} \middle| \frac{\partial u_n}{\partial k_y} \right\rangle - \left\langle \frac{\partial u_n}{\partial k_y} \middle| \frac{\partial u_n}{\partial k_x} \right\rangle \right],$$

where the integration is taken over by the Brillouin zone $(k_x, k_y) = (-\pi, \pi)$. With the definition of $\sigma_{\mathrm{H},n} \equiv (e^2/h)C_1$, one can write

$$C_1 \equiv \int \frac{\mathrm{d}^2 k}{(2\pi)^2} \left(\frac{\partial \mathcal{A}_y^n}{\partial k_x} - \frac{\partial \mathcal{A}_x^n}{\partial k_y} \right) = \frac{1}{2\pi} \oint \mathrm{d}\mathbf{k} \cdot \mathcal{A}^n, \tag{5.3}$$

where C_1 is the first Chern number and $\mathcal{A}_{x(y)}^n \equiv \mathrm{i}\langle u_n | \partial u_n / \partial k_{x(y)} \rangle$ is called Berry's connection. The equation appears suggestive for nonsimply connected geometry in momentum space to give nonzero values of C_1: when $\mathcal{A}$ is a single-valued function, $C_1 = 0$ and thus $\sigma_{\mathrm{H}} = 0$. Recalling that the momentum space is periodically defined in the Brillouin zone, we can easily notice that the space spanned by $\mathbf{k}$ is defined on a torus. In this case, C_1 measures winding numbers of the $\mathcal{A}$-trajectory, where the paths can start and end at the same point but have different windings. For a better understanding, let us lift the (nonsimply-connected) torus to its (simply-connected) universal covering space. Any curve on the universal covering space can be characterized by an element (n_x, n_y) of the fundamental group of the torus. Here $n_{x(y)}$ measures the total number of boundary crossings between the squares in the horizontal and vertical directions, respectively.

We now discuss the effect of lattice potential, which is periodic in x and y and takes the form $V(x, y) = V_1 \cos(2\pi x/a) + V_2 \cos(2\pi y/a)$. In the weak-potential limit ($V \ll \hbar\omega_{\mathrm{c}}$), the band structure deformed by the lattice potential can be obtained within a perturbative scheme. The Schrödinger equation can be expanded in the unperturbed basis as $(E - E_n)\Psi_k = \sum_{k'} \langle n, k | V | n, k' \rangle \Psi_{k'}$, where Ψ_k is the amplitude corresponding to the state $|n, k\rangle$ of the eigenstates of the perturbed system: $\Psi\rangle = \sum_k \Psi_k |n, k\rangle$. For simplicity, we consider the lowest Landau level ($n = 0$), where the matrix elements of the periodic potential can be computed easily by using $H_0(x) = 1$:

$$\langle k | V(x) | k' \rangle = \frac{V_1}{2} \mathrm{e}^{-\pi/2f} \Delta_{k,k'}$$

$$\langle k | V(y) | k' \rangle = \frac{V_2}{2} \mathrm{e}^{-\pi/2f} (\Delta_{k,k'+2\pi/a} + \Delta_{k,k'-2\pi/a}),$$

with f being the magnetic flux per lattice cell in units of flux quantum. Thus we obtain the Harper equation:

$$\Psi_{k-2\pi/a} + \Psi_{k+2\pi/a} + \mathcal{V} \cos(2\pi f^{-1} m + \theta) \Psi_k = \varepsilon \Psi_k,$$

where we set $k - 2\pi m/a \equiv f\theta/a$, $2V_1/V_2 \equiv \mathcal{V}$, and $\varepsilon \equiv E - E_0$. The Harper equation has been studied by a number of authors [2]. When f^{-1} is given by a rational number p/q, with p and q being the relative prime integer, the energy spectra are known to consist of q bands together with $q - 1$ gaps in between. Note here that the original Brillouin zone has been reduced such that $-\pi/q \leq k \leq \pi/q$, which reflects the formation of the superlattice consisting of q original cells. This reduction of the Brillouin zone can be viewed as the

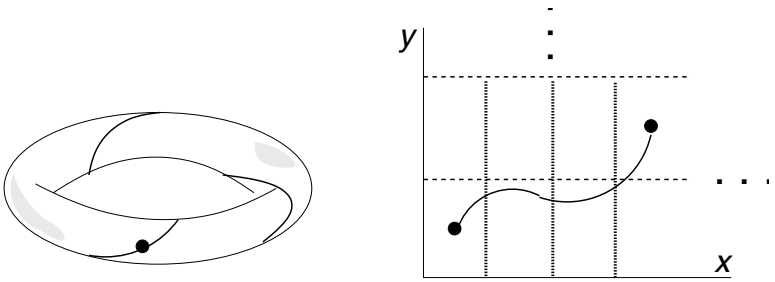

Fig. 5.1. A trajectory defined on a torus, having three turns along the small radius and a single turn along the large radius. The trajectory is equivalently mapped onto the universal covering space, whereby there are three crossings in the horizontal direction and one crossing in the vertical direction

torus perimeter along k-direction becoming π/q. When the path in (5.3) is completed over the original Brillouin zone, the winding number becomes q (see Fig. 5.1). This provides a topological picture of the quantized conductivity steps whose heights vary according to the magnetic fields.

5.3 Shapiro Steps in Josephson Junctions

In this section, we give an introduction to the Josephson effect [3], a representative macroscopic quantum system [4]. Traditionally, macroscopic systems have been considered to behave classically. In general, a macroscopic system is described by a few collective degrees of freedom, where a number of remaining microscopic degrees of freedom may be regarded as the environment. Here the coupling between the collective degree of freedom and the environmental degrees of freedom leads to dissipation in the system [5]. Accordingly, quantum coherence is not usually maintained in the macroscopic system, which makes it difficult to observe macroscopic quantum effects. Nevertheless, there do exist macroscopic quantum systems, where the coupling is too weak to destroy quantum coherence; one of the pronounced examples being the JJ system.

JJ is a heterostructure in which a normal metal or an insulator is sandwiched between superconductors (SCs). As is well known, many-particle wavefunction in the SC can be written as $\psi(\mathbf{x}) = |\psi(\mathbf{x})|\mathrm{e}^{\mathrm{i}\theta(\mathbf{x})}$, where $|\psi(\mathbf{x})|^2$ corresponds to the number density of Cooper pairs at $\mathbf{x}$, and the phase coherence over the whole sample guarantees superconductivity. This makes it possible to take $\psi(\mathbf{x})$ as the superconducting order parameter in the Ginzburg–Landau (GL) theory of superconductivity [6]. The postulate of the GL theory is that for ψ small and varying slowly in space the free energy density $\mathcal{F}$ can be expanded in the form

$$\mathcal{F} = \alpha|\psi|^2 + \frac{\beta}{2}|\psi|^4 + \frac{1}{2m^*}\left|\left(\frac{\hbar}{\mathrm{i}} - \frac{e^*}{c}\mathbf{A}\right)\psi\right|^2 + \frac{h^2}{8\pi^2} \tag{5.4}$$

with appropriate constants α and β, where $\mathbf{A}$ is the vector potential corresponding to the magnetic field h. The order parameter adjusts itself to minimize the overall free energy given by the volume integral of (5.4), which leads to the GL differential equation,

$$\beta|\psi|^2\psi + \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}} - \frac{e^*}{c}\mathbf{A}\right)^2\psi = -\alpha\psi. \tag{5.5}$$

Note that this is analogous to the Schrödinger equation for a free particle, but with the nonlinear term. The corresponding equation for the supercurrent of the system is given as

$$\mathbf{J}_s = \frac{c}{2\pi}\nabla\times\mathbf{h} = \frac{e^*\hbar}{2m^*\mathrm{i}}(\psi^*\nabla\psi - \psi\nabla\psi^*) - \frac{e^{*2}}{m^*c}\psi^*\psi\mathbf{A}, \tag{5.6}$$

which is also the same as the usual quantum mechanical current expression for particle of charge e^* and mass m^*. In the theory of Bardeen, Cooper, and Shrieffer (BCS) [7], and Gor'kov [8], it has been shown that $m^* = 2m$ and $e^* = -2e$, where m and $-e$ are the mass and the charge of an electron.

We first simplify the situation by considering the case where the magnetic fields are absent. Then we can take ψ to be real since the differential equation (5.5) has only real coefficients. If a normalized wavefunction $\widetilde{\psi} \equiv \psi/\psi_\infty$ is introduced, where $\psi_\infty^2 = -\alpha/\beta$ with negative α, the equation becomes, in 1D,

$$\frac{\hbar^2}{2m^*|\alpha|}\frac{\partial^2\widetilde{\psi}}{\partial x^2} + \widetilde{\psi} - \widetilde{\psi}^2 = 0, \tag{5.7}$$

which can yield a zero-voltage supercurrent flowing between two superconducting electrodes separated by an insulating barrier. We assume two massive electrodes for which $|\widetilde{\psi}| = 1$, with different phases allowed. Since the absolute phase is undefined, without loss of generality we can take the phase at each electrode to be 0 and $\Delta\theta$, respectively; here $\Delta\theta$ can be regarded as the phase difference between the two electrodes. The appropriate solution of (5.7) in the insulating barrier matches the boundary conditions $\widetilde{\psi} = 1$ at $x = 0$ and $\widetilde{\psi} = \mathrm{e}^{\mathrm{i}\Delta\theta}$ at $x = L$ with L being the barrier width. As long as $L \ll \xi(\equiv\hbar^2/2m^*\alpha)$, the first term in (5.7) dominates the other two terms for any nonzero $\Delta\theta$ [9]. In this limit, applying the boundary conditions leads to the solution

$$\widetilde{\psi} = (1 - x/L) + (x/L)\mathrm{e}^{\mathrm{i}\Delta\theta}, \tag{5.8}$$

where the first term represents the spread of the order parameter from the left superconducting electrode with phase 0, and the second represents the spread from the right one with phase $\Delta\theta$. Inserting $\widetilde{\psi}$ given by (5.8) into the supercurrent expression, we obtain

$$I_\mathrm{s} = I_\mathrm{c}\,\sin\Delta\theta,$$

where the critical current is given by $I_c = (2e\hbar\psi_\infty^2/m^*)(A/L)$ with A being the cross-sectional area of the insulating barrier.

The previous discussion had been carried out without taking magnetic flux into consideration. However, it is important to work out the effects of an applied field on the supercurrent. Since $\Delta\theta$ is not a gauge-invariant quantity, it is appropriate to introduce the gauge-invariant phase difference ϕ defined by

$$\phi = \Delta\theta - (2\pi/\Phi_0)\int \mathbf{A}\cdot \mathrm{d}\mathbf{l},$$

leading to the general form of the supercurrent $I_\mathrm{s} = I_\mathrm{c}\sin\phi$, which can also be obtained had we used the full gauge-invariant term $[-\mathrm{i}\hbar\nabla - e^*\mathbf{A}/c]$ throughout the preceding evaluation of the current.

Armed with this gauge-invariant form of the current, we now evaluate the maximum current of dc-SQUID (superconducting quantum interference devices). To avoid sensitivity to the gauge choice of the vector potential, results would rather be obtained in terms of the magnetic flux Φ through a specified contour: since $\mathbf{B} = \nabla\times\mathbf{A}$, the line integral of $\mathbf{A}$ along the path passing through both links (denoted 1,2) and the supercurrent electrodes (denoted A,B) in Fig. 5.2 gives the enclosed flux Φ. The integration contour, taken in the interior of the electrodes, is assumed to be thick enough where the supercurrent density vanishes. Therefore, the enclosed flux is given by

$$\Phi = \oint \mathbf{A}\cdot\mathrm{d}\mathbf{l} = (\Phi_0/2\pi)\int_{\text{electrodes}} \nabla\theta\cdot\mathrm{d}\mathbf{l} + \int_{\text{links}} \mathbf{A}\cdot\mathrm{d}\mathbf{l}.$$

The single-valuedness of the phase θ allows us to replace the integration taken in the region of the electrodes by the sum of finite phase difference ϕ_i at each link related through

$$\phi_1 - \phi_2 = 2\pi\frac{\Phi}{\Phi_0}(\mathrm{mod}2\pi).$$

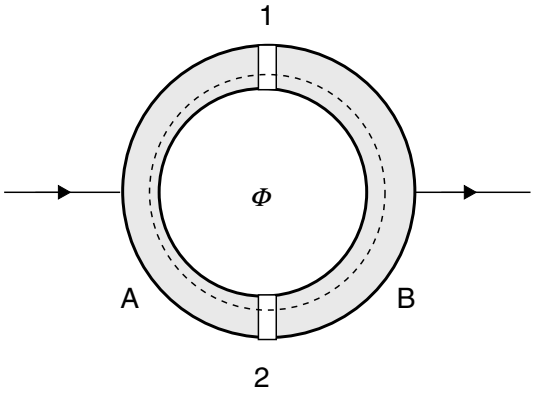

Fig. 5.2. Schematic diagram showing geometry for quantum interference of Josephson tunneling in the presence of the magnetic flux Φ

This relation implies that ϕ_1 and ϕ_2 cannot simultaneously take the value $\pi/2$, as it would be regarded to give the greatest total supercurrent, unless Φ is an integral multiple of Φ_0. In fact, if $I_c^{(1)} = I_c^{(2)} = I_c$, then the maximum supercurrent can be shown by a simple trigonometric argument to be

$$I_{\max} = 2I_c|\cos(\pi\Phi/\Phi_0)|,$$

which is the basis of the dc-SQUID magnetometer

Here we have a simple and interesting feature revealed in a junction with combined dc and ac voltage bias $V = V_{dc} + V_{ac}\cos\Omega t$. According to the Josephson relation $\dot{\phi} = 2eV/\hbar$, the phase difference across the junction is given by

$$\phi(t) = \phi_0 + \omega t + (2eV_{ac}/\hbar\Omega)\sin\Omega t$$

with an integration constant ϕ_0 and $\omega \equiv 2eV_{dc}/\hbar$. The resulting supercurrent can be expressed in terms of the Bessel functions

$$I_s = I_c \sum_{n=-\infty}^{\infty} (-1)^n J_n(2eV_{ac}/\hbar\Omega)\sin(\phi_0 + \omega t - n\Omega t),$$

which displays that the dc voltage satisfying $\omega = n\Omega$ gives a contribution toward the nonvanishing time-averaged current. In consequence, when $V_{dc} = n\hbar\Omega/2e$, the total direct current including the normal current can take any value in the range

$$V_n/R - I_c J_n(2eV_{ac}/\hbar\Omega) \le I \le V_n/R + I_c J_n(2eV_{ac}/\hbar\Omega)$$

with R shunting resistance, which gives rise to the Shapiro steps in the current–voltage characteristics [10].

This discussion has been made possible by solving the equation of motion for phase. A more complete description based on topological argument is made in the following. The Hamiltonian of JJ is given by

$$\mathcal{H} = \frac{(Q - Q_{ext})^2}{2C} - E_J \cos\phi.$$

The first term is responsible for the charging energy, where external current (or voltage) bias is incorporated with the gauge charge $Q_{ext} = \dot{I}_{ext} = CV_{ext}$, and the second term is Josephson energy with $E_J = (\hbar/2e)I_c$. Here the charge variable Q and the phase variable ϕ are conjugate to each other via $[\hbar\phi/2e, Q] = i\hbar$. Thus, we can consider $\hbar\theta/2e$ as a coordinate x and Q as a momentum $p = (-i\hbar)\partial/\partial x$. It is of interest to note that the problem can also be mapped onto one-dimensional electron moving on a periodic potential. Writing the Schrödinger equation $\mathcal{H}\psi = E\psi$, we have the external charge part gauged away via transformation $\psi'(x) = e^{i\theta_0 x}\psi(x)$ with $\hbar\theta_0 = Q_{ext}$.

Now we recall that the coordinate x is, in fact, a phase variable defined on a compact interval $2ex/\hbar \equiv \phi = (0, 2\pi]$, and we have a definite closed

path defined in the phase space. Let us now examine what the physical consequence is. For example, when $Q_{\rm ext} = q_{\rm d} + q_{\rm a}\cos(\Omega t)$ and the system has no degeneracy, the wavefunction should be periodic after one frequency evolution of the periodic driving: $\psi(x, t+\tau) = \psi(x, t)$ with $\tau \equiv 2\pi/\Omega$. Such boundary conditions can then be guaranteed only by

$$\phi(t+\tau) - \phi(t) = 2\pi n_w$$

with integer n_w. Using the Josephson relation, the time-averaged voltage across the junction is given by

$$\langle V \rangle = \frac{1}{\tau}\int_0^\tau \frac{\hbar\dot{\phi}}{2e} = n_w \frac{\hbar\Omega}{2e},$$

which is precisely the same as the voltage quantization obtained previously. Interestingly, we can also predict *fractional* quantization resulting from the system degeneracy. For instance, let us consider the ground state having q-fold degeneracy whose the wavefunction can be written as $\psi^{(\alpha)}$ $(\alpha = 1, 2, \ldots, q)$. Thus the average over the q degenerate ground states should be taken in calculating the properties. The αth ground states exhibits its phase change $\phi^{(\alpha)}(t+\tau) - \phi^{(\alpha)}(t) = 2\pi n^{(\alpha)}$, and the average change rate of phase difference is given by

$$\langle \dot{\phi} \rangle = \frac{1}{q}\sum_{\alpha=1}^{q} \dot{\phi}^{(\alpha)} = \frac{n}{q}$$

with $n \equiv \sum_\alpha n^{(\alpha)}$, leading to the fractional Shapiro steps.

5.4 Charge Density Waves

A characteristic property of a one-dimensional electron system is its instability against the potential having the wave number $2k_{\rm F}$ with $k_{\rm F}$ being the Fermi momentum: the amplitude of the electron density wave coupled to a periodic potential increases divergently when the wave number approaches $2k_{\rm F}$. This Peierls instability gives rise to a collective state of electrons called the charge density wave (CDW) and causes a metal-insulator transition known as the Peierls transition [11].

Let us have a brief look at the origin of the instability. The potential V_q oscillating spatially with the wave number q causes the spatial modulation ρ_q of the electron density. For sufficiently small V_q, the linear response, i.e., ρ_q proportional to V_q is dominant, and we thus introduce the response function χ_q, which relates ρ_q and V_q according to $\rho_q = \chi_q V_q$ [12]. The standard linear response theory gives the expression

$$\chi_q = \frac{1}{N_e}\sum_k \frac{f_{k+q} - f_k}{E_k - E_{k+q}}, \tag{5.9}$$

known as the Lindhard function, where f_k is the Fermi distribution function. For simplicity, let us consider the zero-temperature case. Each term in (5.9) has a nonvanishing contribution only when either $E_k < E_\mathrm{F}$ and $E_{k+q} > E_\mathrm{F}$ or $E_{k+q} < E_\mathrm{F}$ and $E_k > E_\mathrm{F}$. In the former case we have $f_k = 1$ and $f_{k+q} = 0$, whereas in the latter $f_k = 0$ and $f_{k+q} = 1$. Replacing the summation by integration, we obtain

$$\chi_q = \frac{2m}{\pi\hbar^2 nq} \ln \left| \frac{q + 2k_\mathrm{F}}{q - 2k_\mathrm{F}} \right| ,$$

where the free-electron energy $E_k = \hbar^2 k^2/2m$ has been used, and n is the number density of electrons.

The Lindhard function in (5.9) can also be evaluated in two dimensions and three dimensions. While in 1D the response function diverges at $q = 2k_\mathrm{F}$, in higher dimensions it does not diverge, and instead its derivative has jump discontinuity or divergence at $2k_\mathrm{F}$. The origin of the singularity in χ_q at $2k_\mathrm{F}$ can be understood in the following way. In 1D only the term with $k \pm k_\mathrm{F}$ of 5.9 can give a divergent contribution at $q = \pm 2k_\mathrm{F}$. Other terms do not contribute to the divergence even when $E_k = E_{k+q}$, because numerator $f_{k+q} - f_k$ also becomes exactly equal to 0. In higher dimensions, the situation is not different with regard to the singularity of the terms $|k| = k_\mathrm{F}$. However, from the 1D and 2D Fermi surfaces shown in Fig. 5.3, it can be seen that the number of diverging terms for given q is of the order $1/k_\mathrm{F}$ in 1D but only of the order $1/k_\mathrm{F}^2$ in 2D. Similar conclusions can be reached for the 3D case. Thus we can qualitatively understand why the singularity at $q = 2k_\mathrm{F}$ weakens with increasing dimensionality, and in higher dimensions the singularity appears only in the derivative of χ_q at $q = 2k_\mathrm{F}$.

This instability gives rise to the CDW displaying periodic variations of the charge density with the wave number $2k_\mathrm{F}$ [13]. In general, a CDW is described by the modulated density of electrons $\rho = \rho_0 \cos(2k_\mathrm{F} x + \phi)$, where

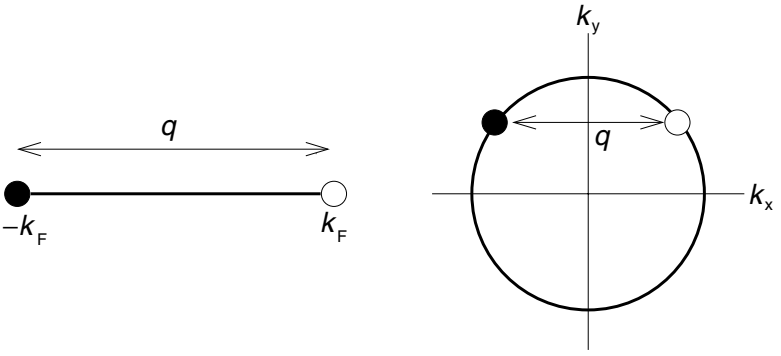

Fig. 5.3. Fermi surfaces of noninteracting electrons in one dimension (*left*) and in two dimensions (*right*)

the phase ϕ represents the location of the CDW. Here the temporal evolution of ϕ corresponds to the sliding motion of the CDW, and hence, the kinetic energy of the CDW is proportional to $\dot{\phi}^2$. In the case of the Mth order commensurability ($2k_{\mathrm{F}}/G = N/M$ with relatively prime integers N and M, and G is the reciprocal lattice vector), the effective potential takes the form of $V = V_0(1 - \cos M\phi)$. Here, the coherent many-particle AB effect can be expected through the collective motion of electrons described by the phase variable ϕ. In the presence of the AB flux f_{AB}, the effective Lagrangian can be written in terms of the collective degree of freedom ϕ [14]:

$$\mathcal{L} = \frac{1}{2}M_{\mathrm{c}}\left(\frac{\partial\phi}{\partial t}\right)^2 - f_{\mathrm{AB}}\left(\frac{\partial\phi}{\partial t}\right) - V_0(1 - \cos M\phi)$$

from which we can calculate the density matrix and the partition function with the help of the path integral method [15]. For that purpose, let us introduce the imaginary time via $\tau \equiv \mathrm{i}t$ and the Euclidean action $S_{\mathrm{E}} = \int \mathcal{L}_{\mathrm{E}}\,\mathrm{d}\tau$, where the corresponding Euclidean Lagrangian takes the form

$$\mathcal{L}_{\mathrm{E}} = \frac{1}{2}M_{\mathrm{c}}\left(\frac{\partial\phi}{\partial\tau}\right)^2 - \mathrm{i}f_{\mathrm{AB}}\left(\frac{\partial\phi}{\partial\tau}\right) - V_0(1 - \cos M\phi).$$

The classical solution satisfying $\Delta S_{\mathrm{E}} = 0$ can be obtained from the following equation of motion

$$M_{\mathrm{c}}\frac{\partial^2\phi}{\partial\tau^2} - MV_0\,\sin\,M\phi = 0,$$

which of this form is called the sine-Gordon equation. Trajectories on a circle, which is nonsimply connected, can be classified into homotopically nonequivalent classes labeled by the winding number n. It represents the number of rotations of the trajectories around the circle in the anticlockwise direction

$$\phi(\tau + \beta) - \phi(\tau) = \frac{2\pi n}{M},$$

where $\beta \equiv 1/k_{\mathrm{B}}T$ is the inverse temperature. Assuming that the temperature is very low, one can get the solution corresponding to $n = 1$ given by [16]

$$\phi(\tau) = \frac{4}{M}\tan^{-1}[\exp(\pm\omega\tau)],$$

where $\omega^2 = MV_0/M_c$, and the positive (negative) sign represents an instanton (anti-instanton). We can compute single-instanton contribution to the partition function, and then straightaway generalize it to contain the contributions of n_1 instantons and n_2 anti-instantons by neglecting the interaction among them, yielding

$$\mathcal{Z} = \mathrm{e}^{-\omega\beta/2}\sqrt{\omega/\pi}\sum_{n=-\infty}^{\infty}\mathrm{e}^{-2\pi f_{\mathrm{AB}}n}I_n(\beta\Delta E),$$

where $\Delta E \equiv 8\sqrt{V_0\omega/\pi}\exp(-8V_0/\omega)$, and $I_n(x)$ is the nth modified Bessel function. In the limit $\beta\Delta E \ll 1$, we obtain

$$\mathcal{Z} \propto 1 + \beta\Delta E \cos\left(\frac{2\pi f_{\mathrm{AB}}}{M}\right).$$

The partition function is therefore periodic in the AB flux with the period M, resulting in the persistent current given by the derivative of the free energy with respect to the flux. The periodicity of the current is different from that in the free electron case, which reflects the effects of the commensurate potential. Moreover, the amplitude of the current is finite but small due to the exponential factor in ΔE. It is thus concluded that the (nonconventional) AB effect in a Peierls insulator is still possible due to the collective states, which provides a good example of a macroscopic quantum coherence.

5.5 Quantum Phases

It has been realized that electromagnetic fields affect the state of matter even in the spatial regions where they do not exert any forces. The most prominent example is the AB effect in the configuration, where charged particles move on a field-free plane, which is pierced by a tube of magnetic flux [17]. Although the classical motion of a particle in such a configuration is indistinguishable from that of a free particle in quantum mechanics, the presence of the flux tube gives a phase shift in the wave packet of the particle and changes the interference pattern.

The Hamiltonian of the particle, for example, an electron with charge e, and mass m subject to such a magnetic flux leads to the Schrödinger equation for the wavefunction

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\nabla + \frac{\mathrm{i}e}{\hbar c}\mathbf{A}\right)^2\psi.$$

Here the vector potential is associated with the magnetic flux in such a way that $\nabla \times \mathbf{B} = \mathbf{A}$. If we impose invariance of quantum mechanics under the gauge transformation $\mathbf{A}' = \mathbf{A} + \nabla\Lambda$, the corresponding transformed wavefunction acquires the quantum phase according to $\psi' = \exp(\mathrm{i}e\Lambda/\hbar c)\psi$. As a result, the phase difference between the two paths due to the localized magnetic flux is given by

$$\frac{e}{\hbar c}\int_{\mathrm{path1}} \mathbf{A}\cdot\mathrm{d}\mathbf{l} - \frac{e}{\hbar c}\int_{\mathrm{path2}} \mathbf{A}\cdot\mathrm{d}\mathbf{l} = \frac{e}{\hbar c}\oint \mathbf{A}\cdot\mathrm{d}\mathbf{l} \equiv 2\pi\frac{\Phi_{\mathrm{AB}}}{\Phi_0},$$

where $\Phi_{\mathrm{AB}} \equiv \oint \mathbf{A}\cdot\mathrm{d}\mathbf{l} = \int \mathbf{B}\cdot\mathrm{d}\mathbf{a}$ is the total magnetic flux enclosed by the paths, and Φ_0 is the flux quantum, $\Phi_0 = hc/e = 4.135\times10^{-7}$ Gauss cm^2. The probability for finding electrons in the screen depends on the phase difference,

which results in the interference pattern oscillating with the period Φ_0. This genuine quantum effect has been observed in a series of experiments [18].

Subsequent developments have revealed that the appearance of the phase shift of the AB-type is in fact rather ubiquitous in quantum mechanics. By exchanging the roles of the flux tube and of the electric charge, for example, Aharonov and Casher found that a neutral particle with a magnetic moment $\boldsymbol{\mu} = \mu\widehat{z}$ moving around a line of electric charges experiences a similar phase shift $2\pi\Phi_{\mathrm{AC}}/\Phi_0$, where $\Phi_{\mathrm{AC}} \equiv (\mu/e)\int \mathbf{E}\times \mathrm{d}\mathbf{l}\cdot\widehat{z}$ is the Aharonov–Casher (AC) flux [19]. Although in this case the particles move in the presence of an electric field $\mathbf{E}$, the resulting AC phase is indeed of the AB type, if the line charge and the magnetic moment are aligned to be parallel so that the force acting on the particle vanishes. The AC phase has been detected in an experiment using a Bonse–Hart single-crystal neutron interferometer [20].

There also exists another quantum phase, which is dual to the AC phase: a neutral particle with an electric dipole moment $\mathbf{d}$ moving in a magnetic field $\mathbf{B}$ obtains a nontrivial phase $\phi_{\mathrm{d}} = (1/\hbar c)\int \mathbf{B}\times\mathbf{d}\cdot\mathrm{d}\mathbf{l}$, again suggesting the possible interference effects [21]. Here we would point out that the dual phase has a relativistic theory origin in the following. Consider spin-1/2 (as an example of a fermion) and spin-1 (as an example of a boson) particles. The Dirac equation for a massive neutral spin-1/2 particle carrying an electric dipole moment d is

$$\left(\mathrm{i}\hbar\gamma^{\mu}\partial_{\mu} - mc + \frac{d}{4c}\epsilon^{\mu\nu\alpha\beta}\sigma_{\mu\nu}F_{\alpha\beta}\right)\psi = 0,$$

where ψ is the Dirac spinor and the convention for the metric, γ_μ and $\epsilon^{\mu\nu\alpha\beta}$ follows [22]. Note that the interaction of the electric moment with the external electromagnetic field takes the dual form of the interaction of a magnetic moment. This terms breaks the CP invariance such that CP violation is generally expected in the standard model of fundamental interactions. Now suppose that the external field is generated by a line of magnetic monopoles with a uniform density. The corresponding magnetic field is two-dimensional such as $\mathbf{B} = B^1(x,y)\widehat{x} + B^2(x,y)\widehat{y}$. By denoting $\sqrt{c}\phi\exp(-\mathrm{i}mc^2t/\hbar)$ to be the upper component of the spinor ψ, we can deduce the Dirac equation into the Schrödinger equation

$$\mathrm{i}\hbar\partial_t\phi = -\frac{\hbar^2}{2m}\left(\left(\nabla_x - \mathrm{i}s\frac{d}{\hbar c}B^2\right)^2 + \left(\nabla_y + \mathrm{i}s\frac{d}{\hbar c}B^1\right)^2 + \nabla_z^2 - \frac{d}{\hbar c}\nabla\cdot\mathbf{B}\right)\phi, \tag{5.10}$$

in the nonrelativistic limit. Here s is twice the spin value of ϕ and thus takes the value $+1$ for spin up and -1 and for spin down.

We now turn to the case of a spin-1 particle and consider a complex vector field $\phi_{/mu}$ satisfying the following equation:

$$\hbar^2\partial_\mu(\partial^\mu\phi^\nu - \partial^\nu\phi^\mu) + m^2c^2\phi^\nu + \mathrm{i}md\epsilon^{\nu\mu\alpha\beta}\phi_\mu F_{\alpha\beta} = 0, \tag{5.11}$$

where d again represents the magnitude of the electric dipole moment. In this equation, ϕ_0 is not a dynamical variable and can be solved formally as

$$\phi_0 = \frac{\hbar^2}{m^2c^2}(\nabla^2\phi^0 + \partial^0\nabla_i\phi^i - \mathrm{i}d\frac{m}{\hbar}^2\epsilon^{0ijk}\phi_i F_{jk}).$$

To take the nonrelativistic limit, we set $\phi_i = \exp(-\mathrm{i}mc^2t/\hbar)\Phi_i/\sqrt{2m}$ and obtain

$$\phi_0 \sim -\mathrm{i}\hbar\exp(-\mathrm{i}mc^2t/\hbar)(\nabla_i\Phi_i - \frac{d}{\hbar c}\epsilon^{ijk}\Phi_i F_{jk})/mc\sqrt{2m},$$

which allows to reduce (5.11) into the Schrödinger equation

$$\mathrm{i}\hbar\partial_t\Phi_\pm = -\frac{\hbar^2}{2m}\left(\nabla_x - \mathrm{i}s\frac{d}{\hbar c}B^2)^2 + (\nabla_y + \mathrm{i}s\frac{d}{\hbar c}B^1)^2 + \nabla_z^2\right)\Phi_\pm + \cdots \quad (5.12)$$

and $\Phi_\pm = (\Phi_1 \pm \Phi_2)/\sqrt{2}$. The subscript $\pm$ for $\Phi_\pm$ corresponds to the spin of the particle (± 1) and the parameter s still takes the value 1 or -1 depending on the spin of $\Phi_\pm$. The external magnetic field is given as in the case of the spin-1/2 system considered earlier, and "$\cdots$" stands for interactions between Φ_+ and Φ_-, which are not important in the consideration of the phase. In both cases ((5.10) and (5.12)), the Hamiltonian contains the term

$$-\frac{\hbar^2}{2m}(\nabla + s\frac{\mathrm{i}}{\hbar c}\mathbf{d}\times\mathbf{B})^2$$

with $\mathbf{d} = d\hat{z}$. It is thus obvious that the above Hamiltonian yields the dual-AC phase given by $\Phi_\mathrm{d} = (1/e)\oint(\mathbf{B}\times\mathbf{d})\cdot\mathrm{d}\mathbf{l}$.

In the following, we will see such quantum phases are a special case of geometric phases [23]. Let us consider the dual AC phase where an electric dipole $\mathbf{d}$ in a small box is located at $\mathbf{R}$. In the absence of a magnetic field, the (nonrelativistic) Hamiltonian of the electric dipole can be written in the form: $H_0 = H(\mathbf{p}, \mathbf{r}-\mathbf{R})$, and the wavefunction takes the form $\psi_n(\mathbf{r}-\mathbf{R})$ with the corresponding energy E_n independent of $\mathbf{R}$. Now suppose that there exists a magnetic field produced by a line of magnetic monopoles. The energy eigenstates $|n(\mathbf{R})\rangle$, which satisfy the Schrödinger equation

$$H\left(\mathbf{p} - \frac{1}{c}\mathbf{B}\times\mathbf{d}, \mathbf{r}-\mathbf{R}\right)|n(\mathbf{R})\rangle = E_n|n(\mathbf{R})\rangle,$$

can be written in the form

$$\langle\mathbf{r}|n(\mathbf{R})\rangle = \exp\left\{\frac{\mathrm{i}}{\hbar c}\int_\mathbf{R}^\mathbf{r}\mathrm{d}\mathbf{r}'\cdot\mathbf{B}(\mathbf{r}')\times\mathbf{d}\right\}\psi_n(\mathbf{r}-\mathbf{R}).$$

We then transport the box around a *closed* loop Γ encircling the monopole line. The geometric phase acquired by the wavefunction during this transport

is given by [23]

$$\gamma_n(\Gamma) = \mathrm{i}\oint_\Gamma \langle n(\mathbf{R})|\nabla_{\mathbf{R}} n(\mathbf{R})\rangle \cdot \mathrm{d}\mathbf{R} = \frac{1}{\hbar c}\oint_\Gamma \mathbf{B}\times\mathbf{d}\cdot\mathrm{d}\mathbf{R}.$$

One of the most well-known demonstrations of quantum phase effects is the persistent current in a mesoscopic ring threaded by AB flux. For simplicity, let us consider free electrons in an ideal ring of radius R at zero temperature. In the presence of the AB flux, the kinematic momentum of an electron, shifted by the vector potential, can be chosen by $\mathbf{A} = (\Phi_{\mathrm{AB}}/2\pi R)\widehat{\phi}$, which leads to the Schrödinger equation

$$\frac{\hbar^2}{2mR^2}\left(-\mathrm{i}\frac{\partial}{\partial\phi} + f\right)^2 \psi_n = E_n\psi_n$$

with $f \equiv \Phi_{\mathrm{AB}}/\Phi_0$. This yields the simultaneous eigenfunction ψ_n of the angular momentum $L_z = -\mathrm{i}\hbar\partial/\partial\phi + f\hbar$ and the energy $\psi_n = \mathrm{e}^{\mathrm{i}n\phi}/\sqrt{2\pi}$. The corresponding energy levels of the system are given by

$$E_n = \frac{\hbar^2}{2mR^2}(n+f)^2.$$

These flux-dependent energy levels yield interesting ground-state properties. The current density of a quantum mechanical system is given by $\mathbf{J} = \mathrm{Re}[\psi^*\mathbf{v}\psi]$, where $\mathbf{v}$ is the velocity operator. In the system of interest here, the corresponding current density can be written in the form

$$\mathbf{J}_n = \frac{1}{m}\mathrm{Re}\left[\psi_n^*\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)\psi_n\right] = \frac{R}{\hbar}\mathrm{Re}\left[\psi_n^*\frac{\partial H}{\partial f}\psi_n\right]\widehat{\phi}.$$

This yields the charge current carried in the nth level.

$$I_n = -e\int \mathrm{d}x J_n = -\frac{e}{2\pi\hbar}\frac{\partial E_n}{\partial f}.$$

The total current of the system is then given by the summation over all the occupied levels. Therefore, the flux dependence of the ground-state energy implies that the current flows persistently around the ring, thus bearing the name *persistent current*. Although only the persistent current driven by the AB flux is mentioned, one can easily infer that persistent *spin* current [24] and *dipole* current [25] are also present via AC phase and dual AC phase, respectively.

5.6 Carbon Nanotubes

A carbon nanotube can be thought of as a layer of graphite sheet folded into a cylinder [26]. Hence, understanding the basic properties can be made by

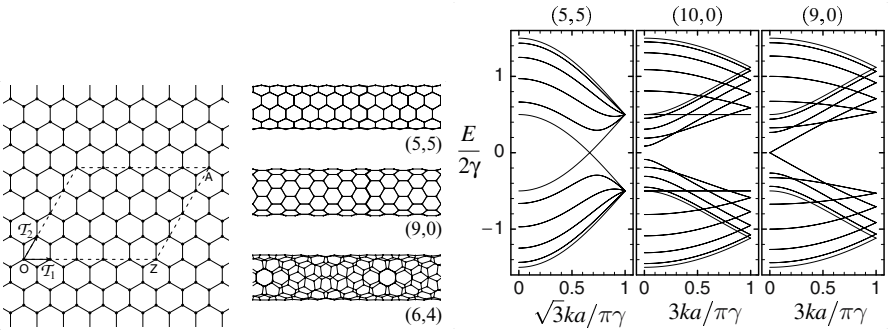

Fig. 5.4. The lattice structure of graphite and the translational vectors $\mathcal{T}_1$ and $\mathcal{T}_2$ are shown in the left panel. The so-called armchair, zig-zag, and a general chiral tube are shown in the center, and band structures are in the right panel

introducing lattice structures and energy bands of graphite. A graphite sheet consists of many hexagons whose vertices are occupied by the carbon atoms, each of which supplies one conduction electron. The electron can transfer from site to site through tunneling. In this case, the following Hamiltonian can describe the system:

$$\mathcal{H} = \epsilon_0 \sum_i c_i^\dagger c_i - \sum_{ij} \gamma_{ij} c_i^\dagger c_j.$$

In fact, the simple tight-binding Hamiltonian has been revealed to yield remarkably similar density of states to the measured values by using the tip of a scanning tunneling microscope [27]. All the lattice sites shown in Fig. 5.4 can be defined through linear combinations of the translational vectors,

$$\mathcal{T}_1 = \sqrt{3}a\widehat{x}, \quad \mathcal{T}_2 = \frac{\sqrt{3}a}{2}\widehat{x} + \frac{3}{2}a\widehat{y}.$$

Noting the translation symmetry, we can choose the Hilbert space spanned by the Bloch basis. Further a bipartite lattice, which can be defined for neglecting next-nearest-neighbor hopping, that is, to restrict γ_{ij} with the nearest neighboring i and j, can be separated into two disjoint sublattices $\mathcal{A}$ and $\mathcal{B}$. Thus, we have $|\psi^k_{\mathcal{A}(\mathcal{B})}\rangle = \sum_{i\in\mathcal{A}(\mathcal{B})} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_i} c_i^\dagger |0\rangle$. We construct a state vector that is an eigenstate of the Hamiltonian as $|\psi^k\rangle = \mathcal{A}^k|\psi^k_{\mathcal{A}}\rangle + \mathcal{B}^k|\psi^k_{\mathcal{B}}\rangle$. In the matrix representation, the energy eigenvalue equation reads

$$\begin{pmatrix} h_{\mathcal{A}\mathcal{A}} & h_{\mathcal{A}\mathcal{B}} \\ h_{\mathcal{B}\mathcal{A}} & h_{\mathcal{B}\mathcal{B}} \end{pmatrix} \begin{pmatrix} \mathcal{A}^k \\ \mathcal{B}^k \end{pmatrix} = E_k \begin{pmatrix} \mathcal{A}^k \\ \mathcal{B}^k \end{pmatrix},$$

where $h_{\mathcal{A}\mathcal{A}} = h_{\mathcal{B}\mathcal{B}} \equiv \langle\psi^k_{\mathcal{A}}|\mathcal{H}|\psi^k_{\mathcal{A}}\rangle = \epsilon_0$, and

$$h_{\mathcal{A}\mathcal{B}} \equiv \langle\psi^k_{\mathcal{B}}|\mathcal{H}|\psi^k_{\mathcal{A}}\rangle = -\gamma \mathrm{e}^{\mathrm{i}k_y a} - 2\gamma \mathrm{e}^{-\mathrm{i}k_y a/2} \cos\left(\frac{\sqrt{3}a}{2}k_x\right).$$

Thus, the energy bands of a graphite are given by

$$E_{\pm}(k_x, k_y) = \epsilon_0 \pm \gamma \sqrt{1 + 4\cos\left(\frac{3a}{2}k_y\right)\cos\left(\frac{\sqrt{3}a}{2}k_x\right) + 4\cos^2\left(\frac{\sqrt{3}a}{2}k_x\right)}.$$

How to roll up the graphite sheet into a cylinder is specified by a wrapping vector

$$\mathcal{W}_{\mathcal{N},\mathcal{M}} = n\mathcal{T}_1 + m\mathcal{T}_2$$

with the two integers defining its electrical properties [27, 28], which we show in the following. Two often-cited examples among nanotubes are armchair and zigzag types, characterized by (ℓ, ℓ) and $(\ell, 0)$, respectively.

We first consider an (ℓ, ℓ)-armchair case. As shown in Fig. 5.4, starting from the point 0, we make ℓ-step in the vector direction $\mathcal{T}_1$ and again, another in the direction $\mathcal{T}_2$. We then reach the point A. Cutting the sheet along the line perpendicular to $\overline{OA}$, and folding up the cut-out so that the two ends coincide, we then have an ℓ, ℓ-armchair tube. Now, it is natural to impose periodic boundary conditions according to the tube diameter L: $k_y L = 2\pi n$. Here the tube diameter is given by $L = 3a\ell$: there are 2ℓ lattice spacings and half of them are given by $2a$ and the other by a. Thus, feeding the boundary condition into the graphite energy bands, we obtain

$$E^a_{\pm}(k, n) = \epsilon_0 \pm \gamma \sqrt{1 + 4\cos\left(\frac{n\pi}{\ell}\right)\cos\left(\frac{\sqrt{3}a}{2}k_x\right) + 4\cos^2\left(\frac{\sqrt{3}a}{2}k_x\right)}.$$

Here we have the first Brillouin zone defined by

$$-\frac{\pi}{\sqrt{3}} \le k_x a \le -\frac{\pi}{\sqrt{3}} \quad -\pi \le \frac{\sqrt{3}}{2}k_x a + \frac{3}{2}k_y a \le \pi,$$

leading to $n = 1, 2, \ldots, 2\ell$ and $|ka| \le \pi/\sqrt{3}$. This resulting one-dimensional dispersion has a large degeneracy at the zone boundary where $|ka| = \pi/\sqrt{3}$, in particular, for half-filling case where the Fermi energy lies in $E = \epsilon_0$. The two subbands E_+ and E_- have degeneracy, indicating the absence of the band gap, and thus, the armchair nanotube is metallic, as reflected in Fig. 5.4.

Similarly for $(\ell, 0)$-zigzag tube, being at the same point A, we make ℓ-step along $\mathcal{T}_1$ direction and then reach the point Z. The cut lines perpendicular to $\overline{AZ}$ are attached, leading to the boundary conditions: $k_x\sqrt{3}a\ell = 2\pi n$, where every ℓ lattice spacing is given by $\sqrt{3}a$ for the configuration. Then, we have the energy dispersion

$$E^z_{\pm}(k, n) = \epsilon_0 \pm \gamma \sqrt{1 + 4\cos\left(\frac{3a}{2}k_y\right)\cos\left(\frac{n\pi}{\ell}\right) + 4\cos^2\left(\frac{n\pi}{\ell}\right)}.$$

Here the Brillouin zone is defined as $|k|a \le \pi/3$ and $n = 1, 2, \ldots, 2\ell$. Unlike the armchair case, $E^z_{\pm}(k, n)$ can have gapful bands, depending on ℓ: when ℓ

is given by integer multiple of 3 there is no gap at the Fermi level, thus the system becomes metallic. On the other hand, for example, zigzag tube with $\ell = 10$ bears the band gap. Especially in the case where ℓ is given by an even number dispersionless bands exist for $n = \ell/2$ (see Fig. 5.4).

We can also manipulate the tubes according to arbitrary wrapping vectors, which are called chiral tubes. The boundary condition can be obtained by a symmetry translation operator $\widehat{T}$:

$$\widehat{T}(\mathcal{N}\mathcal{T}_2 + \mathcal{M}\mathcal{T}_2)|\psi^k\rangle = |\psi^k\rangle,$$

leading to

$$\frac{\sqrt{3}}{2}(2\mathcal{N} + \mathcal{M})k_x a + \frac{3}{2}\mathcal{M}k_y a = 2\pi n.$$

A detailed analysis of the energy dispersion given by the above boundary condition leads to the classification, depending on $\mathcal{N} - \mathcal{M}$ (i) a multiple of 3, (ii) a multiple of $3C$ with C being the highest common divisor of $\mathcal{M}$ and $\mathcal{N}$. Carbon nanotubes can also be in the form of multiwalled cylinder where a few single-walled tubes are concentrically arranged. Since the distance between adjacent walls is larger than atomic spacing, the properties of multiwalled tubes bear overall similarity to single-walled tubes.

Among other physical quantities, electron transport can be regarded as a direct reflection of the system. With a ballistic transport to be kept in mind, electrons appreciate mainly the geometry of the system that they travel through. For the question on how to formulate electron transport, by referring the readers to the monographs and surveys in [29], we would provide a quick view on the transport through a carbon nanotube torus as *fully* topological object; besides the noncontractible path along the tube diameter, bending the cylinder into a torus introduces additional nonsimply connected path [30]. Figure 5.5 shows the linear transmission and its dependence on the position of the second lead (see also the schematic of the set-up). The density plot is also

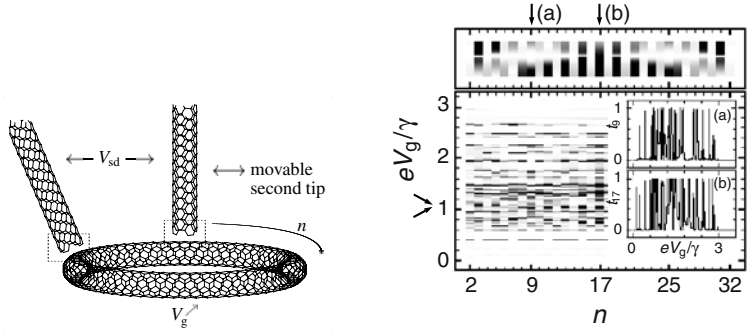

Fig. 5.5. The left panel: A pure-carbon set-up. The two semi-infinite CNT leads can scan the upper surface of the CNT ring. The right panel: Density plot of the linear transmission as a function of the position n and of the gate voltage 1. In this scale, white corresponds to 0 and black to 1

configured by varying the gate voltage, which one can consider as the Fermi level of electrons in the tube. Along the gate voltage axis (with fixed n), one can clearly see the dark-and-bright pattern of the transmission, indicating off-and-on resonance of the energy of an incident electron with the Fermi level. Even more interestingly, at a fixed gate voltage, the transmission can also vary with positional change of the second lead. For particular values of the gate voltage, zeros in the transmission occur, as shown in the upper panel blow-up of the region indicated by the arrows. If the readers were now to view topologically, a surmise that a one-dimensional ring might lead similar effects to that the nanotorus revealed can be readily made. Indeed, essential properties such as "chessboard" pattern and transmission zeros are not the exclusive possession of carbon nanotorus but are fairly shared in a community of affine topology.

5.7 Conclusions

We had a retrospective study of topology revealed in condensed matter physics. Although many important issues might be overlooked or barely touched, this chapter covers the essence.

In this age where a huge body of numerics powered by superfast machinery is regarded to be more convincing than an elegant and potent argument, a word such as, *topology* may sound ancient (for a few, even a geologic "language"). Latin is quite a suggestive example of this—once it was believed to be the closest language to God but by new most of priests do not (cannot) speak Latin. Nonetheless, our ambition is that one should fully recognize topology not only as a fascinating object in mathematics but also as a powerful tool in studying a physical system.

Acknowledgments. The author would like to acknowledge first M.Y. Choi, a wonderful lecturer as well as a scientist who provided the inspiration for this chapter through his lecture. Thanks also go to G. Cuniberti and M. Porto who have worked on many of the problems and shared stimulating discussions. The author is also indebted to all the people who helped in different ways.

References

1. von Klitzing, K., Gorba, G., Pepper, M.: Phys. Rev. Lett. **45**, 494 (1980); for a review, Prange, R.E. (ed.), The Quantum Hall Effect. Springer, New York (1990)
2. Harper, P.G.: Proc. Roy. Soc. Lond. A **68**, 874 (1955); D. Hofstadter: Phys. Rev. B **14**, 2239 (1976); J. B. Sokoloff: Phys. Rep. **126**, 189 (1985)
3. For a review see, Barone, A., Paterno, G.: Physics and Applications of the Josephson Effect. Wiley, New York (1982)
4. Caldeira, H.A., Kramer, B., Schön, G. (eds): Quantum Dynamics of Submicron Structures. Kluwer, Dordrecht (1995)

5. Weiss, U.: Quantum Dissipative Systems. World Scientific, Singapore (1993)
6. Ginzburg, V.L., Landau, L.D.: Zh. Eksp. Teor. Fiz. **20**, 1064 (1950)
7. Bardeen, J., Cooper, L.N., Schrieffer, J.R.: Phys. Rev. **108**, 1175 (1957)
8. Gor'kov, L.P.: Sov. Phys. JETP **9**, 1364 (1959)
9. Aslamazov, L.G., Larkin, A.I.: Zh. Eksp. Toer. Fiz. **9**, 150 (1968) [JETP Lett. **9**, 87 (1968)]
10. Shapiro, S.: Phys. Rev. Lett. **11**, 80 (1963)
11. For a review see, Grüner, G.: Density Waves in Solids. Addison-Wesley (1994)
12. Ashcroft, N.W., Mermin, N.D.: Solid State Physics, Ch. 17 Hole, Reinhart, and Winstron, New York (1976); Ziman, J.M.: Principles of the Theory of Solids, Ch. 5 2nd ed. Cambrige University Press, Cambridge (1972)
13. Grüner, G., Zettl, A.: Phys. Rep. **119**, 117 (1985); Grüner, G.: Rev. Mod. Phys. **60**, 1129 (1988)
14. Fukuyama, H.: J. Phys. Soc. Jpn. **41**, 513 (1976)
15. Feynmann, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)
16. Bogacheck, E.N., Krive, I.V., Kulik, I.O., Rozhavsky, A.S.: Phys. Rev. B **42**, 7614 (1990); Zh. Eksp. Teor. Fiz. **97**, 603 (1990) [Sov. Phys. JETP **70**, 336 (1990)]
17. Aharonov, Y., Bohm, D.: Phys. Rev. Lett. **115**, 485 (1959)
18. Tonomura, A., Osakabe, N., Matsuda, T., Kawasaki, T., Endo, J., Yano, S., Yamada, H.: Phys. Rev. Lett. **56**, 792 (1986)
19. Aharonov, Y., Casher, A.: Phys. Rev. Lett. **53**, 319 (1984); Hagen, C.R.: ibid. **64**,2347 (1990)
20. Cimino, A., Opat, G.I., Klein, A.G., Kaiser, H., Werner, S.A., Arif, M., Clothier, R.: Phys. Rev. Lett. **63**, 380 (1989); Balatsky, A.V., Altshuler, B.L.: Phys. Rev. Lett. **70**, 1678 (1993)
21. Wilkes, M.: Phys. Rev. Lett. **72**, 5 (1994)
22. Bjorken, J.D., Drell, S.: Relativistic Quantum Mechanics. McGraw-Hill, New York (1964)
23. Berry, M.V.: Proc. R. Soc. Lond. A **392**, 45 (1984)
24. Choi, M.Y.: Phys. Rev. Lett. **71**, 2987 (1993); Mod. Phys. Lett. B **8**, 1667 (1994)
25. Yi, J., Jeon, G.S., Choi, M.Y.: Phys. Rev. B **52**, 7838 (1995)
26. Dresselhaus, M.S., Dresselhaus, G., Eklund, P.C.: Science of Fullerences and Carbon Nanotubes. Academic, New York (1995)
27. Wildöer, J.W.G. et al.: Nature **386**, 474 (1997); Odom, T.W. et al.: Nature **391**, 62 (1998)
28. Mintmire, J.W., Dunlap, B.I., White, C.T.: Phys. Rev. Lett. **68**, 631 (1992); Hamada, N., Sawada, S.-I., Oshiyama, A.: ibid. **68**, 579 (1992)
29. Fisher, D.S., Lee, P.A., Phys. Rev. B, **23**, 6851 (1981); for a review, see for example, Datta, S.: Electronic Transport in Mesoscopic Systems. Cambridge University Press, Cambridge, UK (1999)
30. Cuniberti, G., Yi, J., Porto, M.: Appl. Phys. Lett. **81**, 850 (2002); see also, Yi, J., Cuniberti, G., Porto, M.: Euro. Phys. J B **33**, 221 (2003)

6

Phason Dynamics in Aperiodic Crystals

T. Jannsen

Summary. Aperiodic crystals have a dynamical behavior that is different from that of lattice periodic crystals. The properties are related to the possibility of describing aperiodic crystals in a higher-dimensional space, for which the physical space is a subspace. The special motions can be considered as motions in the additional space. Such motions may have low or zero frequencies due to the existence of symmetries that are special for aperiodic crystals. The ensuing large amplitude motions require a treatment in the framework of nonlinear dynamics. Under certain conditions, the low-frequency vibrations are related to friction. The phenomena are discussed for a number of classes of aperiodic crystals.

6.1 Introduction

6.1.1 Quasiperiodic Crystals

The majority of crystal structures have lattice translation symmetry. This means the existence of a unit cell and invariance under one of the 230 space groups in three dimensions. The proof that a crystal is lattice periodic can come from diffraction experiments, or, more indirectly, from morphology. The diffraction peaks of a lattice periodic system are on a three-dimensional reciprocal lattice ($\boldsymbol{k}$-space), and they can be labeled by three integer indices. The reciprocal lattice is generated by vectors $\boldsymbol{a}_i^*$ and the system is invariant under three independent lattice translations $\boldsymbol{a}_i$ with

$$\boldsymbol{k} = \sum_{i=1}^{3} h_i \boldsymbol{a}_i^*, \quad \boldsymbol{a} = \sum_{i=1}^{3} n_i \boldsymbol{a}_i, \quad \boldsymbol{a}_i^* \boldsymbol{a}_j = 2\pi\delta_{ij} \; .$$

In the morphology facets are seen that can be labeled by the same reciprocal lattice vectors.

However, there are many examples of materials that cannot be indexed by n indices, but nevertheless show sharp Bragg peaks. These belong to what can be seen as a generalization of the reciprocal lattice, the Fourier module.

This module consists of all vectors of the set

$$\boldsymbol{k} = \sum_{i=1}^{n} h_i \boldsymbol{a}_i^* \ ,$$

where n, which may be different from 3, is called the rank of the Fourier module M^*. The material then has a density function

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{k} \in M^*} \hat{\rho}(\boldsymbol{k}) \exp(i \boldsymbol{k} \boldsymbol{r}) \ . \tag{6.1}$$

If $n \neq 3$ this function does not have three-dimensional translation symmetry, although it may be periodic in one or two directions. A function of this type is called quasiperiodic, even if $n = 3$. If $n > 3$ the structure is aperiodic.

Quasiperiodicity can also be seen macroscopically in the morphology. Crystallites in the shape of a dodecahedron, or crystals with facets for the indexing of which one needs more than three indices are usually quasiperiodic.

6.1.2 Examples of Quasiperiodic Crystals

Roughly speaking, there are three big classes of quasiperiodic (also called incommensurate if they are aperiodic) systems. First are the modulated crystal structures, second the incommensurate composites, and third the quasicrystals. This is not a strict division into classes. Some materials can be considered to belong to two classes.

Incommensurate displacively modulated phases have a structure, which can be described as a periodic deformation of a lattice periodic structure. The positions of the atoms are

$$\boldsymbol{r}_{nj} = \boldsymbol{n} + \boldsymbol{r}_j + (\boldsymbol{q}\boldsymbol{n})\, \boldsymbol{f}_j \ , \tag{6.2}$$

where n belongs to the three-dimensional lattice, $\boldsymbol{r}_j$ is the average position of the j-th atom in the unit cell, and $\boldsymbol{f}_j$ is a periodic function with period $2\boldsymbol{\pi}$. The function is called the modulation function. Because $\boldsymbol{q}$ is incommensurate (i.e., has irrational indices) two atoms of type j in the unit cells $\boldsymbol{n}$ and $\hat{\boldsymbol{n}}$ only give the same argument in $\boldsymbol{f}_j$ if their difference is perpendicular to $\boldsymbol{q}$. The diffraction pattern consists of main reflections belonging to the reciprocal lattice of the basic structure, and satellites as a consequence of the periodic modulation. If there is only one modulation wave the diffraction spots are situated at

$$\boldsymbol{k} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + \ell \boldsymbol{c}^* + m\boldsymbol{q},$$

where h, k, and ℓ are the usual indices for the main reflections.

Apart from this modulation in position, a composition wave may be present. The probability of finding an atom A at position $\boldsymbol{n} + \boldsymbol{r}_j$ is $p_j(\boldsymbol{q}\boldsymbol{n})$ and the probability of finding another type B is equal to $1 - p_j(\boldsymbol{q}\boldsymbol{n})$.

A second class of aperiodic crystals is formed by incommensurate composites [1]. Such composites consist of two or more subsystems, which are

themselves incommensurately modulated. The subsystems labeled by ν have periodic basic structures with lattice Λ_ν. The position of the j-th atom in unit cell n_ν of system ν is given by

$$\boldsymbol{x}_{\nu nj} = \boldsymbol{n}_\nu + \boldsymbol{r}_{\nu j} + \sum_{k \in M*} \boldsymbol{f}\big(\boldsymbol{k}(\boldsymbol{n}_\nu + \boldsymbol{r}_{\nu j})\big) \,, \tag{6.3}$$

where $\boldsymbol{n}_\nu$ is a lattice vector of lattice Λ_ν^*, $\boldsymbol{r}_{\nu j}$ a vector in the unit cell of that lattice, whereas the modulation function has Fourier components in the Fourier module. The modulations are caused by the interactions with the other subsystems. That is the reason why the modulation wave vectors are combinations of reciprocal lattice vectors of the other subsystems. The vector module is spanned by the basis vectors $\boldsymbol{a}_{\nu i}^*$ of the reciprocal lattice vectors. If the subsystems are mutually incommensurate, the full system is aperiodic and the positions of the Bragg peaks are given by

$$\boldsymbol{k} = \sum_{\nu i} m_{\nu i} \boldsymbol{a}_{\nu i}^* = \sum_{i=1}^{n} h_i \boldsymbol{a}_i^* \,, \tag{6.4}$$

where the vectors $\boldsymbol{a}_i^*$ are linear combinations of the vectors $\boldsymbol{a}_{\nu i}^*$ such that they are a minimal set of vectors spanning the Fourier module. In principle, additional satellites might be present due to other mechanisms, not belonging to the span of the various reciprocal lattices Λ_ν^*, but we shall neglect this possibility here. Special cases of incommensurate composites are misfit structures, intercalation compounds, and adsorbed monolayers on a crystal surface.

A third class of aperiodic crystals is the quasicrystals [2, 3]. They have a rank higher than 3. A precise definition is lacking, but most of them can be considered as tilings or have a (possibly broken) point group symmetry, which is noncrystallographic in three dimensions. An example is the alloy i-$\mathrm{Al}_x\,\mathrm{Mn}_y\,\mathrm{Pd}_z$ in a certain composition range. It has the noncrystallographic symmetry group of the icosahedron as point group.

Quasicrystals often contain Al, and most of them, but not all, are ternary or quaternary alloys. Building blocks often found are Mackay and Bergman clusters with icosahedral symmetry. In general, these clusters overlap. There are families with icosahedral, decagonal, dodecagonal, or octagonal symmetry. However, such noncrystallographic three-dimensional point groups are not essential. There are often periodic, structurally related compounds, called approximants. These have strictly speaking only one of the 32 three-dimensional crystallographic point groups as symmetry.

6.1.3 Symmetry

The diffraction pattern of a quasiperiodic crystal has intensities

$$I(\boldsymbol{k}) = \sum_{\boldsymbol{k}_\mathrm{B} \in M^*} a(\boldsymbol{k}_\mathrm{B}) \delta(\boldsymbol{k} - \boldsymbol{k}_\mathrm{B}) \,. \tag{6.5}$$

If the orthogonal transformation R leaves these intensities invariant ($I(R\boldsymbol{k}) = I(\boldsymbol{k})$ for all $\boldsymbol{k} \in M^*$), then $RM^* = M^*$ and in particular

$$R\boldsymbol{a}_i^* = \sum_{i=1}^{n} \Gamma(R)_{ji} \boldsymbol{a}_j^* \,, \tag{6.6}$$

where the integer $n \times n$ matrices $\Gamma(R)$ form a representation of the finite symmetry group K of the diffraction pattern. A well-known theorem in group-theory then states that this representation is equivalent with an orthogonal transformation. If one then chooses three basis vectors, the space spanned by them (the physical space) is left invariant. So, the n-dimensional representation is reducible. This implies that there is a basis transformation such that the operations are represented by orthogonal matrices of the form

$$\begin{pmatrix} R_{\mathrm{E}} & 0 \\ 0 & R_{\mathrm{I}} \end{pmatrix} \sim \Gamma(R) \,. \tag{6.7}$$

The orthogonal matrices R_E and R_I are three and $(n-3)$-dimensional, respectively. The $n \times n$ matrices form the symmetry group of the diffraction pattern. It is a finite point group, but not necessarily one of the 32 three-dimensional crystallographic point groups. For example, the symmetry group of the diffraction of the icosahedral Al Mn Pd quasicrystal is the icosahedral group of order 120. The three-dimensional noncrystallographic point group $\bar{5}\bar{3}$m leaving the diffraction pattern invariant has generators, which are integer matrices on the basis of the Fourier module:

$$\begin{pmatrix} 1&0&0&0&0&0 \\ 0&0&0&0&0&1 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \end{pmatrix}, \quad \begin{pmatrix} 0&0&0&0&0&1 \\ 1&0&0&0&0&0 \\ 0&0&0&0&1&0 \\ 0&0&-1&0&0&0 \\ 0&0&0&-1&0&0 \\ 0&1&0&0&0&0 \end{pmatrix}, \quad -\begin{pmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{pmatrix}.$$

Apart from this symmetry, quasiperiodic crystals may show symmetries that are important for the dynamics. We show this in one dimension. For a modulated phase the positions of the atoms are given by

$$x_n = x_0 + na + f(qna) \,, \tag{6.8}$$

where the function f has periodicity 2π. A change in position $x_n \to x'_n = x_n + \delta$ does not change the mutual distances, and leaves, for that reason, the potential energy invariant. The consequence of this symmetry is that the momentum is conserved. A second symmetry is actually only a pseudo-symmetry. A change in position

$$x_n \to x'_n = x_{n+p} - pa = x_0 + na + f\big(q(na + pa)\big) \approx x_n + \epsilon f'(qna) \tag{6.9}$$

if an integer p is chosen in such a way that $\epsilon = paq \mod (2\pi)$ is (arbitrarily) small. Then the displaced atoms with a renumbering have the same mutual distances as before with arbitrary position.

For incommensurate composites the same symmetries are present. We consider a model consisting of two chains. The positions of the atoms in the two chains are

$$\begin{aligned} x_n &= x_0 + na + f(na), \qquad f(x) = f(x+b) \\ y_m &= y_0 + mb + g(mb), \qquad g(y) = g(y+a)\ . \end{aligned}$$

The displacement of all atoms in both chains by δ leaves again the potential energy invariant. The pseudosymmetry is seen by choosing a small number $\epsilon = pa - qb$ and displacing the atoms as

$$\begin{aligned} x_n &\to x_{n+p} - pa = x_0 + na + f(na+pa) \approx x_n + \epsilon f'(na) \\ y_m &\to y_{m-q} + qb = y_0 + mb + g(mb-qb) \approx y_m + \epsilon g'(mb)\ . \end{aligned}$$

These symmetries will have consequences for the dynamics.

6.2 Embedding in Superspace

By definition a quasiperiodic structure has a density function $\rho(\boldsymbol{r})$ with Fourier decomposition

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{k}\in M^*} \hat{\rho}(\boldsymbol{k}) \exp(i\boldsymbol{k}\boldsymbol{r})\ , \tag{6.10}$$

where the Fourier module M^* is the set of vectors

$$\boldsymbol{k} = \sum_{i=1}^{n} m_i \boldsymbol{a}_i^*\ . \tag{6.11}$$

Because there are no three independent vectors $\boldsymbol{k}$ such that the inner product with any vector from the Fourier module is a multiple of 2π, there is no lattice periodicity. The function $\rho(\boldsymbol{r})$, however, is a section of a lattice periodic structure in n dimensions and a three-dimensional physical space. The crucial observation is that the basis $\boldsymbol{a}_i^*$ of the module is the projection of a (reciprocal) lattice Σ^* in n dimensions with basis $(\boldsymbol{a}_i^*, \boldsymbol{b}_i^*)$. Then each vector $(\boldsymbol{k}, \boldsymbol{k}_\mathrm{I})$ in the lattice spanned by this basis is the unique vector that projects on the vector $\boldsymbol{k}$ of the Fourier module. The atom positions in the physical space are taken modulo the n-dimensional lattice Σ, for which Σ^* is the reciprocal lattice. The periodic function on the n-dimensional space is

$$\rho_s(\boldsymbol{r}, \boldsymbol{r}_I) = \sum_{\boldsymbol{k_s}\in\Sigma^*} \hat{\rho}(\boldsymbol{k}) \exp\bigl(i(\boldsymbol{k}\boldsymbol{r} + \boldsymbol{k}_\mathrm{I}\boldsymbol{r}_\mathrm{I})\bigr)\ . \tag{6.12}$$

Typical examples of embeddings of quasiperiodic systems belonging to the various classes are given in Fig. 6.1. Modulated crystals are embedded as arrays of $(n-3)$-dimensional hypersurfaces in n dimensions. These are called the

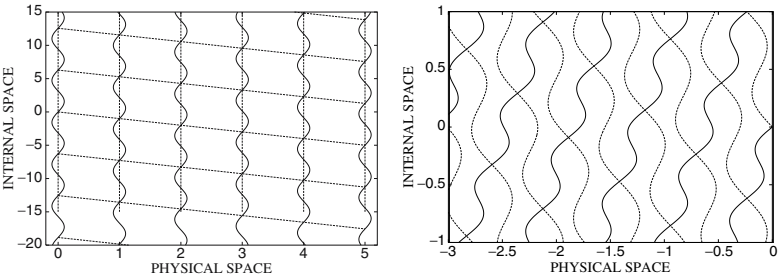

Fig. 6.1. Embeddings into higher-dimensional space of two quasiperiodic structures. *Left:* sinusoidally modulated chain (*dashed:* lattice), *right:* IC composite (*dashed:* system 1, *solid:* system 2)

atomic surfaces. When the modulation functions are continuous the atomic surfaces stretch out until infinity. For incommensurate composites there are periodic arrays of atomic surfaces for each of the subsystems. For quasicrystals, the atomic surfaces are bounded, generally, although it is possible to construct three-dimensional quasiperiodic patterns that could be considered as quasicrystals from a periodic array of unbounded atomic surfaces.

For modulated phases the additional coordinate can be seen from the embedding:

$$x_n = na + f(qna) \;\rightarrow\; (na + f(nqa + t),\; t) \qquad -\infty < t < \infty\,. \tag{6.13}$$

This is a periodic pattern with translation symmetry generated by $(a, -qa)$ and $(0,\, 2\pi)$. The variable t is just the phase of the modulation function. t is the phase variable. Certain excitations in such system may be considered to be phase oscillations, for which the term *phason* was introduced.

For incommensurate composites we consider an example with two subsystems and with rank 4. Suppose that the lattice constants in x- and y-direction are the same, but incommensurate in the z-direction. Suppose furthermore, for simplicity, that there is only atom per unit cell in each subsystem. Then the positions $\boldsymbol{n} + \boldsymbol{f}(\boldsymbol{q}_1\boldsymbol{n})$ and $\boldsymbol{m} + \boldsymbol{g}(\boldsymbol{q}_2\boldsymbol{m})$ can be embedded as

$$\begin{aligned} &(\boldsymbol{n} + \boldsymbol{f}(\boldsymbol{q}_1\boldsymbol{n} + t) + \boldsymbol{Z_1}t,\, t),\\ &(\boldsymbol{m} + \boldsymbol{g}(\boldsymbol{q}_2\boldsymbol{m} + t) - \boldsymbol{Z_2}t,\, t)\,. \end{aligned} \tag{6.14}$$

Here the vector functions $\boldsymbol{f}$ and $\boldsymbol{g}$ have periodicity b (=lattice constant in the z-direction of system 2) and a (=lattice constant in the z-direction of system 1). The vectors $\boldsymbol{q}_i$ are parallel to the z-axis. The system has a four-dimensional lattice periodicity. The internal degree of freedom here is not only the phase of the modulation function, but also the relative positions

of the centers of mass of the two subsystems. There is a dense set of shifts in the internal coordinate for which the potential energy does not change. Therefore, these are dynamic symmetries. The Goldstone mode associated with this symmetry is a (dynamical) shift in internal space, which for smooth embeddings may have zero frequency. These motions are called phasons.

The excitations in quasicrystals have certain aspects in common with this. There are jumps that can be seen as jumps in superspace, and here also the term phason was used. Very often phenomena involving the additional space (for quasicrystals usually called perpendicular space) get a name with the term phason. For example, a strain in a quasicrystal can be divided into "phonon strain" and "phason strain."

6.3 Simple Models for Incommensurate Structures

6.3.1 Displacively Modulated Phases

The structure and vibrational excitations of incommensurate modulated phases may be exemplified on simple model systems. A very simple model is one dimensional. It is a linear chain with particles with one degree of freedom, for example the deviation of its position from that in an equidistant array. The potential is a non linear function of the deviations and there is an interaction between a particle and its first and second neighbors. The Hamiltonian then is given by

$$H = \sum_n \left(\frac{p_n^2}{2} + V_1(x_n) + V_2(x_n - x_{n-1}) + V_3(x_n - x_{n-2}) \right) . \tag{6.15}$$

An example is

$$H = \sum_n \left(\frac{p_n^2}{2} + \frac{Ax_n^2}{2} + \frac{x_n^4}{4} + Bx_n x_{n-1} + Cx_n x_{n-2} \right) . \tag{6.16}$$

The on-site potential is here a fourth-order polynomial. The terms with B and C may favor different ground states, which leads to frustration. Therefore, the model is called the discrete frustrated ϕ^4 (DIFFOUR) model (Fig. 6.2) [4, 5].

The ground state of (6.16) for $T = 0$ is given by the coupled nonlinear equations

$$Ax_n + x_n^3 + B(x_{n+1} + x_{n-1}) + C(x_{n+2} + x_{n-2}) = 0 . \tag{6.17}$$

Periodic solutions with period N can be found by the solution of a finite set of coupled equations. Aperiodic solutions with wave vector $2\pi q_i$ can be found as the limit of periodic solutions with $q = L/N$ when N tends to infinity such

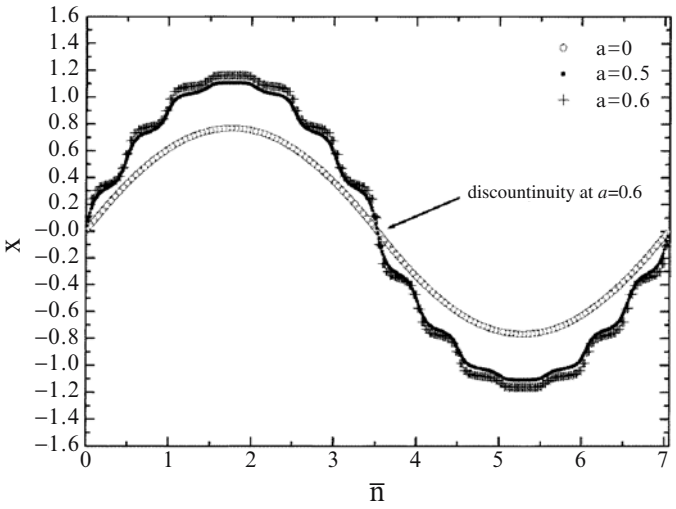

Fig. 6.2. The modulation function for the ground state of the DIFFOUR model determined for three different values of the parameters

that q tends to the irrational value q_i. The ground state is found as the lowest energy solution for all wave vectors q.

If A is sufficiently high the ground state is an undeformed, periodic chain ($x_n = 0$). For a critical value of A this state becomes unstable, and the instability is a wave with a critical wave vector q_i. Below the critical value of A the ground state is a displacively modulated chain. If q_i is irrational, the ground state is quasiperiodic. Close to A_i the modulation function is sinusoidal, but for lower values there are higher harmonics, leading to a squaring up. Generally, the modulation function remains smooth down to a value A_d. Below A_d the modulation is still incommensurate but the modulation function is discontinuous. The typical situation is that below $A_\mathrm{c} < A_\mathrm{d}$, the ground state is modulated with a modulation vector that locks in at a commensurate value. Then the chain is again periodic, generally with a larger unit cell than for $A > A_i$.

The phase diagram can be constructed from the determination of the ground state for given values of the parameters. In the A/C versus B/C plane incommensurate phases are concentrated around the origin. For high values of A/C the solution $x_n = 0$ (the paraphase) is the ground state. For large absolute values of B/C the ground state is ferroic (period 1 different from the paraphase) or antiferroic (period 2). For low values of A/C ground states are commensurate. Around the origin the wave vector is incommensurate or commensurate and the phase diagram is complicated. This means that for comparable values of A/C and B/C the ground state may be quasiperiodic and the ground state is degenerate.

If x_n is considered as the displacement of particle n, the positions of the particles are

$$r_n = na + x_n = na + f(qna).$$

The modulation function f is obtained numerically, from a solution of (6.17), which minimizes the potential energy, as $f(qna) = x_n$, where q is twice the number of nodes divided by the periodicity of the solution. With this modulation function the solution can be embedded in a two-dimensional space according to (6.13). This means that for $A_\mathrm{d} < A < A_i$ the atomic surfaces are unbounded and smooth, whereas for $A_\mathrm{c} < A < A_\mathrm{d}$ they are disjoint and bounded.

6.3.2 The Double-Chain Model for Incommensurate Composites

For aperiodic composites we study a system consisting of two one-dimensional chains, one with atoms at positions x_n, and the other with atoms at y_m. The potential energy is given by

$$V = \sum_n V_1(x_n - x_{n-1}) + \sum_m V_2(y_m - y_{m-1}) + \sum_{nm} W(x_n - y_m)\,. \qquad (6.18)$$

The intrachain couplings are either harmonic ($V_1(x) = \alpha(x-a)^2/2, V_2(y) = \beta(y-b)^2/2$) or they are Lennard-Jones potentials with minima for $x = a$ and $y = b$, respectively [6, 7]. For the interchain coupling a Lennard-Jones potential has been chosen:

$$W(r) = \lambda\left(\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right),$$

where $r^2 = (x-y)^2 + d^2$, if d is the interchain distance.

The model is in fact a generalization of a model introduced by Dehlinger, and which was studied by Frenkel and Kontorova and by Frank and Van der Merwe [8]. It consists of a linear chain on a fixed substrate, with the Hamiltonian

$$H = \sum_n \left(\frac{p_n^2}{2} + \frac{\alpha(x_n - x_{n-1} - a)^2}{2} + \lambda\cos(2\pi x_n/b + \phi)\right)\,. \qquad (6.19)$$

Usually it is called the Frenkel–Kontorova model, also when the lattice constant a is incommensurate with the periodicity b of the substrate potential. In the DCM also the substrate is deformable.

6.3.3 The Ground State of the DCM

The ground state is obtained as the configuration that minimizes the potential energy. This is done numerically for commensurate approximants. If a

and b are the lattice constants of the two chains, solutions for x_n and y_m are obtained with the condition $La = Mb$ for integers L and M such that L/M approximates the irrational value a/b for the incommensurate chain. In general, the solutions can be written as

$$x_n = x_0 + na + f(x_0 + na), \qquad f(x) = f(x + b), \tag{6.20}$$

$$y_m = y_0 + mb + g(y_0 + mb), \qquad g(y) = g(y + a). \tag{6.21}$$

The modulation functions can be determined from

$$f(na \mod b) = x_n - na, \quad g(mb \mod a) = y_m - mb,$$

when x_n, y_m are found numerically. For small interaction parameter λ the functions f and g are continuous, for larger values of λ they become simultaneously discontinuous. In the latter region the modulation functions are approximately piecewise linear. This means that locally the lattice parameter of each chain is changed. Because the density of the particles is fixed, the discontinuities provide an overall incommensurability. In this sense the transition from smooth to discontinuous can be called a discommensuration transition as well.

Using the modulation functions f and g the solution can be embedded in a higher-dimensional space according to (6.14). For values of the parameter λ smaller than the critical value the atomic surfaces extend to infinity. For values exceeding this value the atomic surfaces are bounded or form even fractal structures.

A line in the plane of the relevant parameters (λ/α and λ/β) forms the transition from smooth modulation functions to discontinuous modulation functions. This has been obtained by keeping λ/α fixed and varying λ/β monitoring the value of the discontinuity. In the real calculations, which were based on approximants, this means that the largest gap exceeds a threshold value. If one increases the size of the approximant the transition becomes more pronounced (Fig. 6.3).

6.4 Phonons and Phasons

6.4.1 Phonons in Aperiodic Crystals

Phonons are collective dynamical excitations in solids describing oscillations of the atomic positions around the equilibrium positions. When the displacements of the atoms from their equilibrium positions are denoted by u_n the potential energy may be written as $V(u_1, \ldots, u_N)$. A development in a series gives, up to second order in the displacements:

$$V(u_1, \ldots, u_N) = \frac{1}{2} \sum_{n\alpha m\beta} U(nm)_{\alpha\beta} u_{n\alpha} u_{m\beta} + \cdots .$$

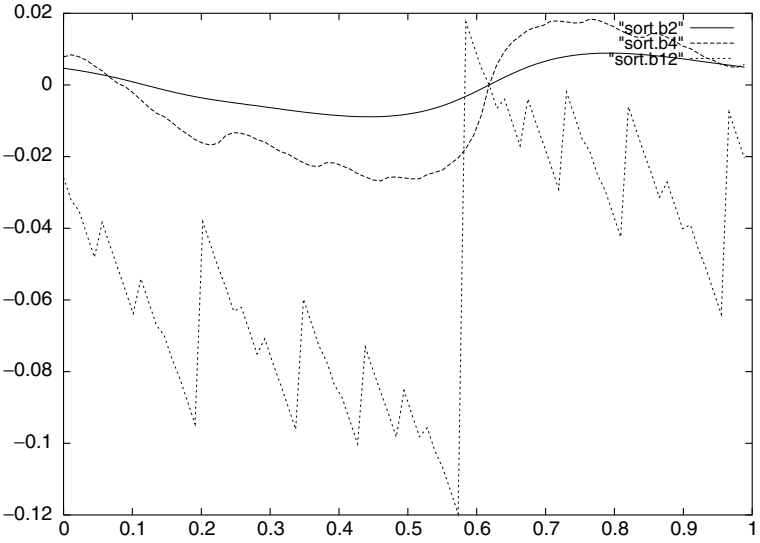

Fig. 6.3. The modulation function for one of the chains in the ground state of the DCM. There is a transition from smooth to discontinuous functions

Here $u_{n\alpha}$ is the Cartesian α-coordinate of the displacement of particle n. The equation of motions

$$m_n \ddot{u}_{n\alpha} = -\sum_{m\beta} U(nm)_{\alpha\beta} u_{m\beta}$$

can be reduced to a problem in $3s$ dimensions (s is the number of atoms per unit cell) if the crystal is periodic. Then the position of the j-th atom in the unit cell n has coordinates $x(nj)_\alpha$ and displacement $u(nj)_\alpha$. The second derivative of the potential energy with respect to the positions is a tensor

$$U(n-m, ij)_{\alpha\beta}.$$

The eigenvectors of the Fourier transform of this tensor

$$D(ij|q)_{\alpha\beta} = \sum_n U(n, ij)_{\alpha\beta} (m_i m_j)^{-1/2} \exp(-iqn)$$

are $e(q\nu|j\alpha)$, have eigenvalues $\omega^2_{q\nu}$, and specify the phonons

$$u(nj)_\alpha = \sum_{q\nu} Q_{q\nu} e(q\nu|j\alpha) \exp(iqn) + \text{c.c.}$$

The modes are characterized by a wave vector q and a branch label ν.

For aperiodic crystals there is nither such a unit cell nor a Brillouin zone. For periodic approximants the number of branches, labeled by ν, increases

with the size of the unit cell of the approximant, and becomes infinite for the aperiodic structure. Nevertheless, it is possible to consider pseudo-Brillouin zones and a function that can be measured with inelastic neutron scattering, the function $S(q,\omega)$. A Brillouin zone boundary occurs in those places in reciprocal space where two plane waves with wave vectors differing by a reciprocal lattice vector are degenerate. For example, in one dimension elastic waves with wave vectors k and $-k$ are degenerate and a gap originates if $k = -k + 2\pi/a$, at the Brillouin zone boundary $k = \pi/a$. For aperiodic crystals the wave vectors of the structure form a dense set, but the coupling between two degenerate waves k_1 and k_2 is only strong when the structure has strong Fourier components $K = k_1 - k_2$. From the diffraction pattern it is clear that the strong peaks are not dense. Therefore, there is a discrete set of surfaces where a substantial gap may occur. These are the pseudo-Brillouin zone boundaries.

The inelastic neutron scattering is described by the differential cross section, which for s atoms at positions $\boldsymbol{r}_j$ is determined by the function

$$S(\boldsymbol{q},\omega) = \sum_{\boldsymbol{q}\nu} \left| \sum_{j=1}^{s} \mathrm{e}^{\mathrm{i}\boldsymbol{q}\boldsymbol{r}_j} \ \boldsymbol{q}\boldsymbol{e}(\boldsymbol{k}\nu|j) \right|^2 \delta(\omega - \omega_{\boldsymbol{k}\nu}) \Delta(\boldsymbol{q} - \boldsymbol{k}) \,,$$

where the sum is over all phonons and where $\Delta(\boldsymbol{q}) = \sum_K \delta(\boldsymbol{q} - \boldsymbol{k})$ is a sum over all vectors of the Fourier module. The latter is an infinite sum, but the inner product of the eigenvector and the wave vector will be of importance only in certain cases. The function would give for a periodic structure a sharp maximum along the lines of the dispersion curves (ω_q, q). For aperiodic crystals this remains the case for lower frequencies. There most of the vibrations propagate plane waves like sound waves. For higher frequencies the maxima become broader because the eigenvectors tend to be more localized, which implies that in their Fourier decomposition more wave vectors are involved. These wave vectors form a quasicontinuum around the dispersion curve.

The latter shows that the character of the excitations may be different from that for periodic crystals. For periodic crystals the excitations in various unit cells differ only by a phase factor, due to Bloch's theorem. This means that phonons in periodic crystals are extended. Localized phonons only occur due to defects. For aperiodic crystals they may also exist in ideal systems. In one dimension it has been proven that the behavior in many systems is neither extended nor (exponentially) localized. The displacements of a mode may fall off algebraically or be self-similar. Such states are called critical.

The phonons in aperiodic crystals can be numerically calculated by approximating the aperiodic structure by a series of periodic structures. The results for the aperiodic crystal then are supposed to be the limit of the results for the series. As an example consider the vibrations in a Fibonacci chain. It can be considered to be a modulated structure with discontinuous modulation function and wave vector $q = \tau = (\sqrt{5} - 1)/2$. This value is the

limit of a series of truncated continued fraction expansions:

$$\tau = \lim_{n\to\infty} \frac{F_n}{F_{n+1}}, \quad F_{n+1} = F_n + F_{n-1}, \;\; F_0 = F_1 = 1 \; . \tag{6.22}$$

(This means τ can be approximated by $1/2$, $2/3$, $3/5$, $5/8$, $8/13$, ...). Replacing τ by an approximant gives a periodic structure with F_n atoms per unit cell and F_n phonon branches. From the eigenvalues and eigenvectors the function $S(q,\omega)$ is easily determined.

6.4.2 Phason Excitations

Very often a structural phase transition from a periodic to an incommensurate modulated structure goes via an instability of the periodic system, where the frequency of a mode tends to 0 as a function of temperature or of the system parameters. For the DIFFOUR model the dispersion curves for the chain with $x_n = 0$, when the system is lattice periodic, are given by

$$\omega(q)^2 = A + 2B\cos(q) + 2C\cos(2q), \quad -\pi < q \leq \pi \; . \tag{6.23}$$

The minimum of the curve is at q_c with $\cos(q_\mathrm{c}){=}{-}B/4C$, if $|B/4C| \leq 1$ (Otherwise at $q = 0$ or $q = \pi$). The frequency goes to 0 if A decreases to $A_i = 2C + B^2/4C$. This is called a soft mode. For $A < A_i$ the structure with $x_n = 0$ is unstable. The ground state just below $A = A_i$ is a modulated structure with modulation wave vector q_c. The two degenerate modes at $\pm q_\mathrm{c}$ are coupled by the modulation. Actually the mode frequency changes with the parameters. In a mean field treatment the equations of motion are effectively the same, but the parameters A, B, and C depend on temperature. Then the critical value of A corresponds to a critical temperature T_i.

In the neighborhood of $\pm q_\mathrm{c}$ the dispersion curves are linear:

$$\omega(q_\mathrm{c} + k)^2 = \left[4C - B^2/4C\right]k^2 \tag{6.24}$$

for $A = A_\mathrm{i}$. The coupled modes give new modes that are the symmetric and anti-symmetric combinations of the original modes. One is proportional to the sinusoidal modulation function, the other differs by a phase of $\pi/2$ and corresponds to the derivative of the modulation function. This means that the first changes the amplitude, the other the phase of the modulation. These modes are called the amplitude and phase modes, or amplitudon and phason, respectively. The branch starting from this zero frequency mode is called the phason branch. Because the phase of the modulation corresponds to the internal coordinate in superspace, when one embeds the aperiodic chain in higher dimensions, a phason may be described as an oscillation in superspace with a polarization pointing out of the physical space.

Because the potential energy of the crystal is invariant under a phase shift of the modulation, it is to be expected that the phason with $k = 0$ ($q = q_\mathrm{c}$) has frequency 0, also for $A < A_\mathrm{i}$. Numerical calculations of the dispersion curves in the modulated phase show that this is true in the neighborhood of

$A = A_\mathrm{i}$, but not generally for any value of A. The typical situation is that the frequency remains 0 in the interval $A_\mathrm{d} < A < A_\mathrm{i}$ for some value of A_d that is larger than the value A_c of A for which the ground state becomes commensurate. Below A_d the frequency of the lowest phason is nonzero, a phason gap opens. A careful analysis shows that at the transition point the modulation function is no longer smooth, but shows discontinuities. It is the discommensuration transition, also found in similar systems under the name "transition by breaking of analyticity" [9].

Excitations with 0 or low frequency and with eigenvectors, which correspond to motions that can be interpreted as motions in the additional space (and therefore can be called phasons), have been found in the DIFFOUR model, in the DCM, and in the Frenkel–Kontorova model (see Sect. 6.6). In a certain parameter or temperature range the minimal frequency of these modes is zero. Let us summarize the results for the three models at zero temperature, as a function of the parameters. Fixing B/C in the DIFFOUR model, there is a zero frequency phason mode in the range from A_d/C to A_i/C. Above A_i/C the paraphase is stable, and there is no phason. Below A_d/C there is a phason gap. For the DCM there is a line in the $(\lambda/\alpha, \lambda/\beta)$-plane separating the region with zero frequency phason from that with a phason gap. This happens both for the Gaussian and for the Lennard-Jones potential. In the Frenkel–Kontorova model, another model for composites that is discussed in Sect. 6.6, there is a critical value of the chain-substrate interaction λ above which there is a phason gap, and below which the gap is 0 (Fig. 6.4).

In all these cases the character of the modulation functions has been studied. The line (or point) in parameter space where the phason gap opens coincides always exactly with the appearance of discontinuities in the modulation

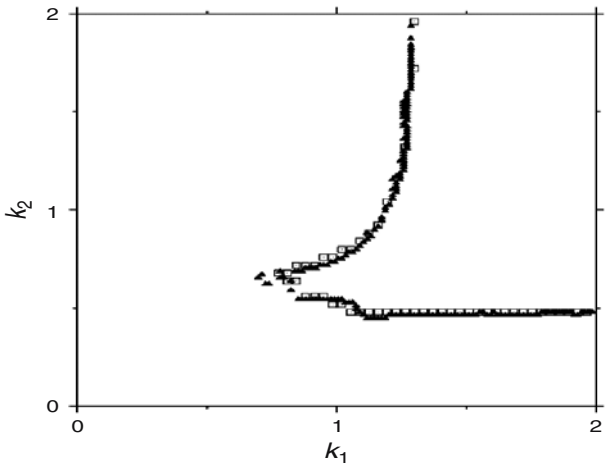

Fig. 6.4. The parameter space $k_1 = \lambda/\alpha,\ k_2 = \lambda/\beta$ with the lines indicating the opening of the phason gap, and the discommensuration transition [7]

function. Therefore, the conclusion is that the discommensuration transition is connected to the softening of the phason. If the modulation functions are smooth, the phason gap is 0.

For a zero frequency acoustic phonon (a displacement of the crystal) increasing the amplitude does not change the distances between the particles. For a phason that is different. A phason is a harmonic oscillation only for small amplitudes. Because of the zero frequency the amplitudes may become large and nonlinear terms have to be taken into account. In the following we study the nonlinear dynamics related to the phason motion.

6.4.3 The Phason Content of Phonons

In principle, phonons may be described fully in physical space. Then the displacements are parallel to this physical space. Sometimes, a phonon can be described as an oscillation in the phase variable, and the displacements then have components in internal space.

Consider a simple modulated chain, embedded as $(na + f(qna - t), t)$. A shift in superspace $\epsilon(\cos\theta, \sin\theta)$ gives a displacement field

$$u_n = \epsilon \cos\theta + \epsilon \sin\theta f'(qna) \; .$$

Then $U_1 = \sum_n u_n/N = \epsilon \cos\theta$ and $U_2 f'(qna) = u_n - U_1$. The internal polarization is defined by $\tan\theta = U_2/U_1$ and $\epsilon = \sqrt{U_1^2 + U_2^2}$. The phason content increases with higher values of θ.

Consider the two chains in the double chain model. The equilibrium positions are $\bar{x}_n$ for one chain, and $\bar{y}_m$ for the other. The oscillations around these positions then are given by displacements $u_n(t)$ and $v_m(t)$. We consider the case of eigenmodes with a frequency ω. Then $x_n(t + 2\pi/\omega) = u_n(t)$ and $v_m(t + 2\pi/\omega) = v_m(t)$. In an eigenmode the average displacements of the two chains are

$$\Delta_1 = \frac{1}{N_1} \sum_n u_n, \qquad \Delta_2 = \frac{1}{N_2} \sum_m v_m \; . \tag{6.25}$$

A translation in two-dimensional superspace of the chains in the direction $\epsilon(\cos\theta, \sin\theta)$ would give displacements of the centers of mass according to

$$\Delta_1 = \epsilon(\cos\theta + Z_1 \sin\theta), \qquad \Delta_2 = \epsilon(\cos\theta - Z_2 \sin\theta) \; . \tag{6.26}$$

This means that for an eigenmode (u_n, v_m) an internal polarization can be defined by the values of ϵ and θ.

$$\begin{aligned} \tan\theta &= \frac{a\sum_n u_n - b\sum_m v_m}{aZ_2\sum_n u_n + bZ_1\sum_m v_m} \; , \\ \epsilon &= \frac{1}{Z_1 + Z_2}\big((\Delta_1 - \Delta_2)\sin\theta + (Z_2\Delta_1 + Z_1\Delta_2)\cos\theta\big) \; . \end{aligned} \tag{6.27}$$

Therefore, the displacements in physical space determine the phason character of a phonon mode. For a pure homogeneous acoustic mode $u_n = v_m = c$, and $\theta = 0$, and for a phason with $u_n = Z_1 c$ and $v_m = -Z_2 c$ one has $\theta = \pi/2$, because $aN_1 = bN_2$.

6.5 Nonlinear Phason Dynamics

6.5.1 Modulated Phases

The eigenmodes of a (periodic or aperiodic) crystal are solutions of the linearized dynamical problem. Among them are the phasons in modulated phases that are phonons with a character that can be described as a shift of the modulation function. The frequency of the homogeneous shift, the pure phason with wave vector zero, is 0 if the modulation function is smooth, due to the degeneracy of the potential energy. The latter argument can be used to show that also an arbitrary shift will not cost energy. However, in this case nonlinear terms will come in. In principle, these could give rise to a coupling between phasons and other phonons.

We consider this question in the frame of the DIFFOUR model. We first introduce a new parametrization of the model such that the Hamiltonian is given by

$$H = \sum_n \left(\frac{\dot{x}_n^2}{2} - \frac{b x_n^2}{2} + \frac{x_n^4}{4} + (x_n - x_{n-1})^2 + d(x_n - x_{n-2})^2 \right) . \qquad (6.28)$$

Note that the parameter change from A, B, and C to b and d is such that A and b have different signs. If the ground state is given by the modulation function $x_n = f(kna)$ a shift of the modulation function gives $x_n(t) = f(k(na - vt))$ or an initial speed

$$\dot{x}_n(0) = -kvf'(kna) = \epsilon u_n, \qquad \sum_n |u_n|^2 = 1.$$

For a sinusoidal modulation $x_n = U\cos(kna)$ the speed is given by $v = \epsilon\sqrt{2a/L}/kU$, where L is the length of the normalization domain. The speed should be compared with the phason velocity, the slope of the phason branch, which is equal to $v_{\mathrm{ph}} = \sqrt{1/2d - 8d}$.

We look for solitary wave solutions of the equations of motion of the form

$$x_n(t) = f(na - vt) \qquad (6.29)$$

and start with a continuum approximation. The Lagrange function has the form

$$\mathcal{L} = \sum_n \left(\frac{\dot{x}_n^2}{2} + \frac{b x_n^2}{2} - \frac{x_n^4}{4} - (x_n - x_{n-1})^2 - d(x_n - x_{n-2})^2 \right) . \qquad (6.30)$$

The solution $f(z)$ is periodic with period p just as the modulation function. Then the Lagrangian function in the continuum approximation is an integral

over the unit cell p.

$$\mathcal{L} = \frac{N}{p}\int_0^p \left(\frac{mv^2 f'(z)^2}{2} - \frac{f(z)^4}{2} + \frac{bf(z)^2}{2} - \big(f(z) - f(z+a)\big)^2 - d\big(f(z) - f(z+2a)\big)^2 \right) \mathrm{d}z \ . \quad (6.31)$$

Here N is the number of particles.

For $v = 0$ the action is extremal. Here it is a maximum. The function f, which depends still on v, should maximize the Lagrangian. For a trial function $f(z) = f_0 \sin kz$ this means that

$$\left(\frac{mk^2v^2}{4} + \frac{b}{4} - 2\,\sin(ka/2)^2 - 2\,\sin(ka)^2 \right) A^2 - \frac{3}{16}A^4$$

is maximal, which is an equation for k. This transcendental equation has nontrivial solutions provided

$$mv^2 < -8d - 2 \ . \quad (6.32)$$

If v satisfies this condition and k_0 maximizes the Lagrangian, then the equation for A has a nontrivial solution if

$$\frac{mk_0^2v^2}{4} + \frac{b}{4} - 2\,\sin(k_0a/2)^2 - 2\,\sin(k_0a)^2 > 0.$$

Under these conditions a solitary wave solution exists if it is nearly sinusoidal. For speeds higher than $\sqrt{(-8d-2)/m}$ the solution is unstable. This leads to the conjecture that if the modulation function is smooth (which is required for using the continuum approximation) there is a solitary wave solution moving through the crystal without energy loss provided its speed remains below the threshold value. In the discrete system there will nevertheless be some energy loss due to the coupling to phonons, but it may be expected to be small.

To check these expectations the equations of motion were numerically solved for the DIFFOUR model with as initial positions the positions of the ground state configurations, and as initial velocities a factor ϵ times the eigenvector coordinates of the phason, i.e., proportional to the derivative of the modulation function. After many iteration steps the shape of the modulation function did not change for small initial speed $v = 0.03$. The speed itself remains practically constant. There is only a very small energy loss to internal vibrations because of the discreteness of the system. However, for an initial speed $v = 0.07$ the speed decreases immediately and goes to 0. The energy then is completely transferred to the phonons (Fig. 6.5).

6.5.2 Incommensurate Composites

The vibrations around the equilibrium positions considered in Sect. 6.4, described in terms of phonons, are harmonic. The harmonic approximation is

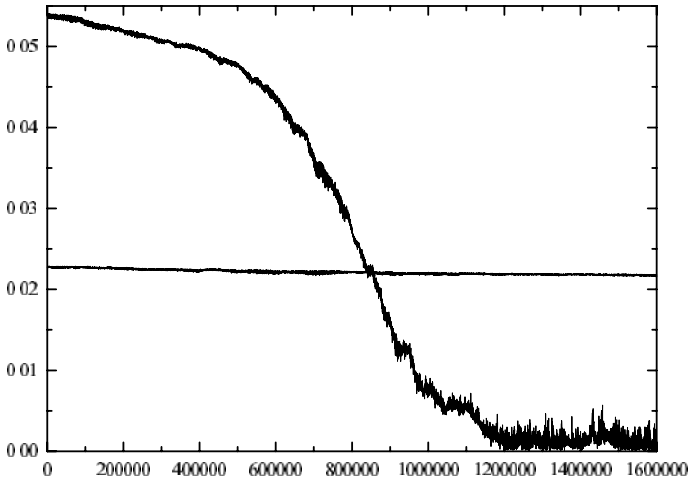

Fig. 6.5. The speed of a nonlinear excitation in the DIFFOUR model. For the lower speed the motion is almost dissipation-less

valid only for small displacements. For larger displacements the equations become nonlinear. They are

$$m_1\ddot{x}_n = -V_1'(x_n - x_{n-1}) - V_1'(x_n - x_{n+1}) - \lambda \sum_m W'(x_n - y_m), \quad (6.33)$$

$$m_2\ddot{y}_m = -V_2'(y_m - y_{m-1}) - V_2'(y_m - y_{m+1}) + \lambda \sum_n W'(x_n - y_m)\ . \quad (6.34)$$

We suppose that the displacements u_n and v_m remain small in the moving frame

$$V_1'(x_n - x_{n-1}) + V_1'(x_n - x_{n+1}) = \alpha(2u_n - u_{n-1} - u_{n+1})$$

and a similar expression for V_2'. The phonons then are nonlinearly coupled by the W terms.

The solutions $x_m(t)$ and $y_m(t)$ also determine the motion of the centre of mass, and the motion of the internal coordinate Z. The latter is determined by (choose $Z_1 = 0$, then $Z = Z_2$)

$$tZ = \frac{1}{L}\sum_m y_m \ \rightarrow \quad (6.35)$$

$$\ddot{Z} = \frac{1}{L}\sum_m \ddot{y}_m = -\frac{\beta}{L}\sum_m (2y_m - y_{m-1} - y_{m+1}) + \frac{\lambda}{L}\sum_{nm} W'(x_n - ym)\ .$$

The relative motion of the two chains may be described as a motion in internal space.

The change in the internal coordinate Z is obtained from the two displacive modes in the two chains:

$$Q_0^{(1)} = \frac{1}{N}\sum_n u_n, \qquad Q_0^{(2)} = vt + \frac{1}{L}\sum_m v_m \; . \tag{6.36}$$

The equations of motion for these variables are

$$\begin{aligned} \ddot{Q}_0^{(1)} &= -\frac{\lambda}{m_1}\sum_{nm} W'(x_n - y_m) \; , \\ \ddot{Q}_0^{(2)} &= \frac{\lambda}{m_2}\sum_{nm} W'(x_n - y_m) \; . \end{aligned} \tag{6.37}$$

The right-hand sides can be developed in powers of the normal coordinates $Q_k^{(j)}$ according to

$$\sum_{nm} W'(x_n - y_m) = \sum_{nm}\sum_j \frac{1}{j!} W^{(j+1)}(na - vt - mb)(u_n - v_m)^j \; . \tag{6.38}$$

The first-order terms are given by

$$\begin{aligned} \sum_{nm} W''(na - vt - mb)u_n &= \sum_{K_2 s} \hat{f}_{K_2 s} Q_{K_2}^{(1)} \exp(-\mathrm{i}s\Omega_1 t) \\ \sum_{nm} W''(na - vt - mb)v_m &= \sum_{K_1 s} \hat{g}_{K_1 s} Q_{K_1}^{(2)} \exp(-\mathrm{i}s\Omega_2 t), \end{aligned}$$

where K_1 is a multiple of $2\pi/a$ and K_2 is a multiple of $2\pi/b$, $\Omega_1 = 2\pi v/b$ and $\Omega_2 = 2\pi v/a$. The latter frequencies correspond to the frequencies with which the particles of one chain move over the particles of the other chain. Furthermore, $\hat{f}_{ks}$ and $\hat{g}_{ks}$ are the Fourier transforms of $\sum_m W'(na - vt - mb)$ and $\sum_n W'(na - vt - mb)$, respectively.

By the expressions in terms of the normal coordinates the modes with wave vector k in chain one are coupled to modes at $k + K_2$ in the same, and to modes at $k + K_1$ in the other chain, and vice versa. The center of mass motions are in first approximation coupled to modes with wave vectors in one of the two reciprocal lattices. Then the equations of motion become

$$\begin{aligned} \ddot{Q}_0^{(1)} &= -\frac{\lambda}{m_1}\left[\sum_{K_2} \hat{f}_{K_2 1} Q_{K_2}^{(1)} \exp(-\mathrm{i}\Omega_1 t) + \sum_{K1} \hat{g}_{K_1 1} Q_{K_1}^{(2)} \exp(-\mathrm{i}\Omega_2 t)\right], \\ \ddot{Q}_0^{(2)} &= \frac{\lambda}{m_2}\left[\sum_{K_2} \hat{f}_{K_2 1} Q_{K_2}^{(1)} \exp(-\mathrm{i}\Omega_1 t) + \sum_{K1} \hat{g}_{K_1 1} Q_{K_1}^{(2)} \exp(-\mathrm{i}\Omega_2 t)\right], \\ \ddot{Q}_{K_2}^{1} &= -\omega_{1K_2} Q_{K_2}^{(1)}, \\ \ddot{Q}_{K_1}^{2} &= -\omega_{2K_1} Q_{K_1}^{(2)}. \end{aligned}$$

The solutions give for the change in the internal coordinate

$$Z_1 - Z_2 \approx vt + A \exp(-\mathrm{i}\Omega_1 t \pm \mathrm{i}\omega_{1K_2} t) + B \exp(-\mathrm{i}\Omega_2 t \pm \mathrm{i}\omega_{2K_1} t) \; . \quad (6.39)$$

The result is a quasiperiodic oscillation of the internal coordinate around a mean value v. Through the coupling to other modes energy flows from this center of mass motion to the phonon bath. The flow is most important in the regions where one of the frequencies $s\Omega_i$ becomes equal to a frequency ω_{1K_2} or ω_{2K_1}. The effect is even more pronounced at frequencies where the participation of the chains is comparable in size (i.e., approximately 0.5).

The analysis given earlier may be illustrated by numerical calculations [10]. This allows to explore the region that is not accessible to analytical treatment. We consider the DCM with truncated Lennard-Jones potentials:

$$V_i(x) = \left((a_\mathrm{i}/x)^{12} - 2(a_\mathrm{i}/x)^6\right) \exp(-rx^2) \; .$$

Chain lengths up to $N = 89$ and $L = 144$ were considered, with periodic boundary conditions. The equations of motion were integrated with a four-step Runge–Kutta procedure. For various values of the lengths L and N the equations were integrated with the equilibrium positions as positional initial conditions, zero velocity for the particles of chain 1, and a uniform initial velocity of the particles of chain 2. The monitored properties were the momenta of the two chains, and their kinetic energies as function of time.

In the first simulations the interchain coupling was taken to be so small that the modulation functions were smooth. In Fig. 6.6 the momentum of

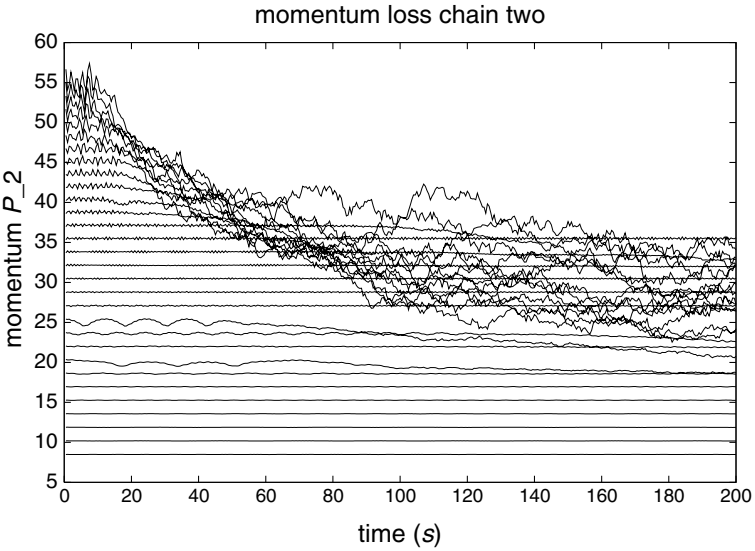

Fig. 6.6. Time dependence of the momentum of one chain in the DCM for initial speeds ranging from 0.5 to 2.0, with intervals of 0.05

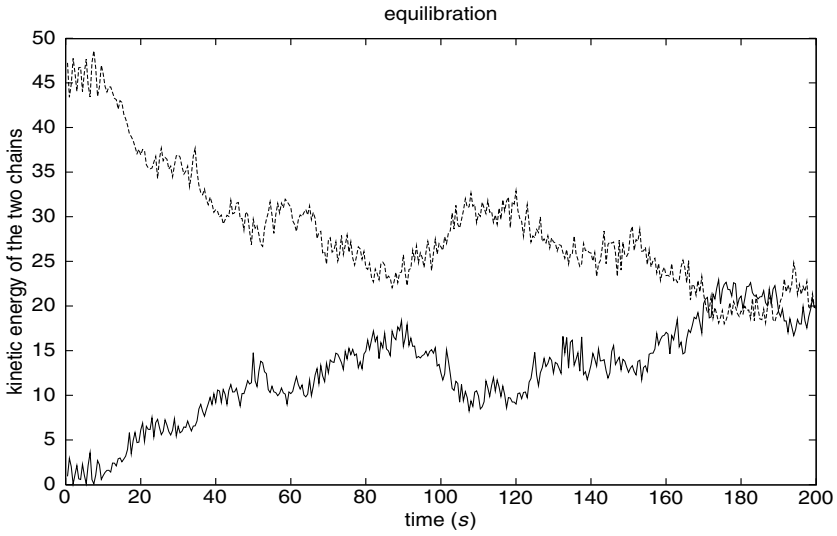

Fig. 6.7. Time dependence of the kinetic energies of the two chains in the DCM, for initial speed $v = 1.5$

chain 2 is plotted as function of time for a number of initial velocities. For smaller than a critical value $v = 1.1$ it remains practically constant for a very long time. For $v \approx 0.65$ the coupling becomes stronger, the energy is lost faster, and there are stronger oscillations due to resonance of Ω_1 and ω_{1K_2}. The resonance disappears for higher values of v. Above the critical value the energy loss is much stronger. There is no longer a sliding mode. If the relative motion vanishes the momentum of the second chain goes to $L/(L+N)$ of its original value because of conservation of total momentum. Figure 6.7 shows the kinetic energy in both chains as a function of time, when there is a strong dissipation ($v = 1.5$). Chain 2 loses quickly its kinetic energy to chain 1, until the point where the energy is evenly distributed over the modes of both chains. The cross-over from almost dissipation-less to strong dissipative behavior is very similar to that in the Frenkel–Kontorova model for weak coupling [11]. The calculations show that for low velocities the energy loss, and therefore also the damping of harmonic modes, is very small. In experiments the phason and sliding modes have been found usually as strongly damped. This would then not be an intrinsic property of the dynamics of incommensurate phases, but probably due to other effects, such as the coupling to defects and pinning. In [12] the dynamics of incommensurate phases has been studied with a phenomenological approach to the damping.

When the coupling between the chains becomes stronger the modulation functions are no longer continuous, and the analysis in terms of normal coordinates of the two chains is no longer valid. The two chains are still mutually

incommensurate, which means that the ground state remains infinitely degenerate. However, in this case there are barriers between the various ground states involving finite jumps of the particles. Therefore, the displacements are no longer harmonic. A numerical integration of the equations of motion gives another cross-over behavior. For low relative momenta the kinetic energy is not sufficient to cause the particles to move over the barriers. Then the kinetic energy is exchanged between the two subsystems and the center of mass oscillates. For higher momenta the two chains may slide over each other, and the kinetic energy is quickly transferred to the phonon degrees of freedom.

6.6 Sliding on a Quasiperiodic Substrate

6.6.1 A Model

As discussed in Sect. 6.2 the embedding of the usual quasicrystals in superspace consists of disjoint atomic surfaces. Although there is phason dynamics in such systems its character resembles that of other aperiodic crystals with discontinuous modulation functions. Therefore, it is not to be expected that there are propagating phason modes with a very low dissipation. However, for quasicrystals one may have smooth modulation functions for the case of a crystal sliding over a quasicrystalline surface. In this section sliding of a periodic crystal over a quasiperiodic substrate is considered.

An often used model for surface phenomena is the one originally introduced by Dehlinger, but usually called after Frenkel and Kontorova. It is a quasi-one-dimensional model with a rigid periodic substrate potential in which a harmonic chain is situated. For the case where the lattice constant is incommensurate with the substrate periodicity, the model was first studied by Frank and Van der Merwe. The model Hamiltonian is

$$H = \sum_n \left(\frac{p_n^2}{2} + \frac{1}{2}(x_n - x_{n-1} - a)^2 + V(x_n) \right), \; V(x) = \lambda \cos(2\pi x/b) \; . \tag{6.40}$$

For an incommensurate situation a/b is an irrational number. A generalization of this model is obtained by replacing the periodic potential by a quasiperiodic one. A further generalization is a two-dimensional periodic crystal moving over a quasiperiodic two-dimensional substrate.

Suppose a periodic crystal moves over the surface of a quasicrystal. Motions in the quasicrystal corresponding with a polarization in the perpendicular direction are phason jumps, which cost energy. At the surface one might have sliding if the situation can be compared to that in the DCM or the GFK model. We consider a tenfold symmetric substrate potential in which a square lattice moves (Fig. 6.8). The Hamiltonian is given by

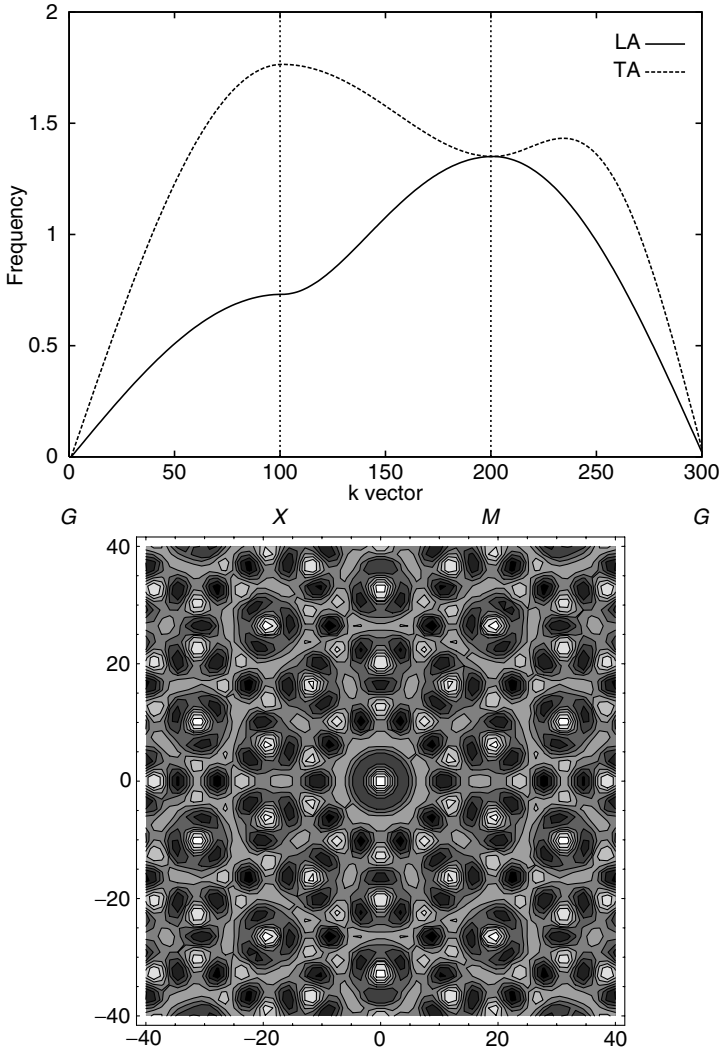

Fig. 6.8. Top: Dispersion curves square crystal. Bottom: Contourplot for the tenfold symmetric substrate potential, which is a model for a quasicrystal surface. ($G(0,0)$, $X(\frac{1}{2},0)$, $M(\frac{1}{2},\frac{1}{2})$)

$$H = \sum_n \left(\frac{p_n^2}{2} + \sum_m \frac{\alpha_{nm}}{2} (x_n - x_m - d_{nm})^2 + \lambda V(x_n) \right) \qquad (6.41)$$

with $\quad V(x) = \sum_{j=1}^{5} \cos(k_j x), \quad k_j = \big(\cos(2\pi j/5), \sin(2\pi j/5)\big)\ .$

Here n, m belong to a square lattice, d_{nm} is the equilibrium distance between the lattice points n and m, and α_{nm} is only different from 0 for first and second neighbors.

The ground state is a modulated structure of rank 6 (the dimension of the lattice is 2, and there are four independent vectors in the potential). The phenomenon of the discommensuration transition is the same. For weak interactions between crystal and substrate the modulation is a smooth function of four variables. The quasiperiodic substrate can be embedded in four dimensions as a periodic function. For small values of λ the unit cell of this four-dimensional lattice is uniformly covered by the positions of the atoms in the ground state. For higher values of λ the modulation function has discontinuities and the unit cell of the periodic function has avoided regions.

6.6.2 Nonlinear Dynamics and Friction

The equations of motion are the Hamilton equations for the Hamilton function H. Under the assumption that the distance between two points n and m remains bounded, the solutions can be written as

$$x_n = x_n^0 + \sum_{k\nu} Q_{k,\nu}(t)\epsilon(k\nu)\exp(\mathrm{i}kn) \ ,$$

where ν labels the two branches, k is in the two-dimensional Brillouin zone, the normal coordinates $Q_{k\nu}(t)$ are bounded in time, except Q_0 which corresponds to the motion of the center of mass. Two phonons of the crystal are strongly coupled if they have the same frequency $\omega_{k\nu}$ (Fig. 6.8) and k vectors differing by one of the five vectors k_j.

Integration of the equations of motion yields the time dependence of $Q_0(t)$. Briefly formulated, the behavior is similar to that of the DCM. There is a zero frequency sliding mode for such small values of λ where the modulation function is smooth. Above the critical value, the modulation function is discontinuous and a phason gap opens.

Given an initial value of the vector $Q_0(t)$ for values of λ below the critical value the value of $Q_0(t)$ decreases slowly. The friction is very low under the condition that $Q_0(0)$ is small. The friction increases strongly as soon as the initial speed exceeds the phason speed. The direction of $Q_0(t)$, however, does not stay the same. Very soon a change of direction occurs, and that does not change strongly afterward.

The system can be embedded in six dimensions. Then the shift of the crystal with respect to the substrate is a phason motion. The transfer of energy from the phason to the other phonons is interpreted as friction.

6.7 Conclusions

There are various dynamical and static phenomena in quasiperiodic systems related to the motion in internal (or perpendicular) space. In the harmonic

approximation these dynamical excitations are special phonons called phasons. Beyond the harmonic approximation these phasons may nonlinearly be coupled to other phonons.

In some models phason modes originate from a soft mode, a mode of the undeformed crystal becoming unstable at the transition to the modulated phase.

In aperiodic crystal models often a transition is observed from a state with smooth modulation functions to a state with discontinuous modulation functions. It is called the discommensuration transition. This transition coincides with the opening of a phason gap.

In incommensurate structures static and dynamic excitations occur in the form of solitary waves. These phase excitations move through the system almost without energy loss if the modulation functions are smooth and the speed is sufficiently low. A dynamic transition takes place from a practically dissipation-less motion to a dissipative motion if the speed exceeds a critical value.

This dynamic transition has been found both in the sliding of subsystems in an aperiodic composite (internal friction) and in the sliding of one crystal on top of another (normal friction).

Acknowledgments. Many results presented here have been obtained in pleasant collaboration with Linda Brussaard, Annalisa Fasolino, Ovidiu Radulescu, Alexei Rubtsov, and Han Slot.

References

1. Janner, A., Janssen, T.: Acta Cryst. A **36**, 399, 408 (1980)
2. Stadnik, Z.M. (ed.).: Physical Properties of Quasicrystals. Springer, Berlin (1999)
3. Suck, J.B., Schreiber, M., Haeussler, P. eds.: Quasicrystals: An Introduction to Structure, Physical Properties and Applications. Springer, Berlin (2002)
4. Janssen, T., Tjon, J.A.: Phys. Rev. B **25**, 3767 (1982)
5. Janssen, T., Tjon, J.A.: J. Phys. C **16**, 4789 (1983)
6. Radulescu, O., Janssen, T.: Phys. Rev. B **60**, 12737 (1999)
7. Brussaard, L.A., Fasolino, A., Janssen, T.: Phys. Rev. B **63**, 214302 (2001)
8. Frank, F.C., van der Merwe, J.H.: Proc. R. Soc. London, Ser. A **198**, 205 (1949)
9. Aubry, S.: Physica D **7**, 240 (1983)
10. Janssen, T.: J. Phys. Cond. Matt. **14**, 12411 (2002)
11. Consoli, L., Fasolino, A., Knops, H., Janssen, T.: Ferroelectrics **250**, 111 (2001)
12. Currat, R., Kats, E., Luk'yanchuk, I.: Eur. Phys. J. B **26**, 339 (2002)
13. Janssen, T.: *Sliding on an aperiodic substrate*, to be published in Proceedings, ICQ8 (2002), J. Non-Cryst. Mater. **334–335**, 471–474 (2004)
14. Janssen, T., Janner, A.: Adv. Phys. **36**, 519–624 (1987)
15. Janssen, T., Radulescu, O., Rubtsov, A.N.: Eur. J. Phys. B **29**, 85 (2002)

7

Hamiltonian Monodromy as Lattice Defect

B. Zhilinskii

Summary. The analogy between monodromy in dynamical (Hamiltonian) systems and defect in crystal lattices is used in order to formulate some general conjectures about possible types of qualitative features of quantum systems that can be interpreted as a manifestation of classical monodromy in quantum finite particle (molecular) problems.

7.1 Introduction

The purpose of this chapter is to demonstrate amazing similarity between apparently different subjects: defects of regular periodic lattices, monodromy of classical Hamiltonian integrable dynamical systems, and qualitative features of joint quantum spectra of several commuting observables for quantum finite-particle systems. First of all we recall why regular lattices and lattices with defects appear naturally for classical integrable Hamiltonian systems and for their quantum analogs. Then we describe several "elementary dynamical" defects using tools and language developed in the theory of crystal defects. Comparison between defects arising in dynamical systems and crystal defects leads to many interesting questions about possibility of realization of certain defects in Hamiltonian dynamics and in crystals.

7.2 Integrable Classical Singular Fibrations and Monodromy

Let us start with the example of Liouville integrable classical Hamiltonian system with N degrees of freedom [1]. This means that there exists a set $F = \{F_1, \ldots, F_n\}$ of functions defined on $2n$-dimensional symplectic manifold M, which are functionally independent and mutually in involution. The Hamiltonian H can be locally represented as a function $H = f(\mathrm{d}F_1, \ldots, \mathrm{d}F_n)$. The mapping $F{:}M \to R^n$ defines the integrable fibration. We call it a generalized energy–momentum map. Each fiber is the union of connected component

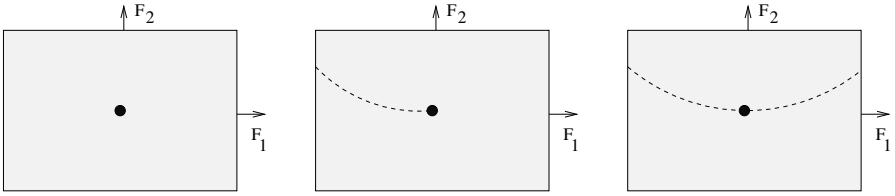

Fig. 7.1. Examples of images of the energy–momentum maps for singular toric fibrations

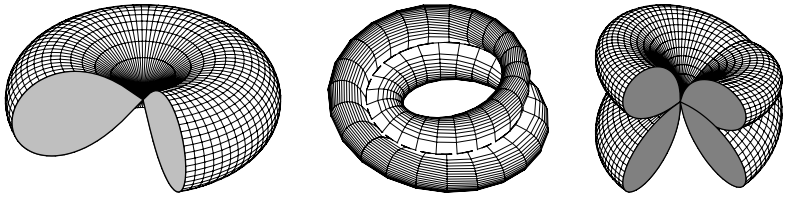

Fig. 7.2. Singular fibers. Pinched torus (*left*). Curled torus (*center*). Pinched curled torus (*right*)

of inverse images $F^{-1}(f)$ of points $f \in R^n$. If the differentials $\{\mathrm{d}F_1, \ldots, \mathrm{d}F_n\}$ of functions from F are linearly independent in each point the fibration is called regular. If moreover all fibers are compact, the fibration is toric. We are interested in integrable toric fibrations with singularities of some very simple type.

Let us restrict ourselves to systems with two degrees of freedom. Typical examples of images of singular energy–momentum maps are shown in Fig. 7.1. The isolated critical value of the map F (see Fig. 7.1, left), also known as focus–focus singularity [2, 3], appears, for example, for such problems as spherical pendulum [4–6], champagne bottle [7, 8], coupling of two angular momenta [9], etc. The singular fiber in this case is a pinched torus (Fig. 7.2, left) with one isolated critical point of rank 0.

The presence of a half-line of critical values, together with end point, is typical for nonlinear $1{:}(-k)$ resonant oscillator [10]. Each point on the singular half-line corresponds to a singular "curled torus" (Fig. 7.2, center, shows a curled torus for the case $k = 2$) [10, 11], which differs from an ordinary torus due to the presence of one circular trajectory that covers itself k-times. This particular circular trajectory is formed by critical points of rank 1 of the map F. The end point (see Fig. 7.1, center) corresponds to the pinched curled torus with a multiple circle shrinking to a point. This fiber has one critical point of rank 0 and is topologically equivalent to pinched torus but its immersion into 4D-space is different. A pinched curled torus for $k = 2$ is shown in Fig. 7.2, right.

A more general situation with two singular rays starting at one singular point (as shown in Fig. 7.2, right) corresponds to $k{:}(-l)$ resonant nonlinear

oscillator. An example of the integrable fibration corresponding to all shown in Fig. 7.1 of the energy–momentum maps with two integrals (F_1, F_2) in involution can be written as [10]

$$F_1 = m_1 \tfrac{1}{2}(p_1^2 + q_1^2) - m_2 \tfrac{1}{2}(p_2^2 + q_2^2), \tag{7.1}$$

$$F_2 = \mathrm{Im}\big[(q_1 + \mathrm{i}p_1)^{m_2}(q_2 + \mathrm{i}p_2)^{m_1}\big] + \big(m_1 \tfrac{1}{2}(p_1^2 + q_1^2) + m_2 \tfrac{1}{2}(p_2^2 + q_2^2)\big)^s, \tag{7.2}$$

with $s > (m_1 + m_2)/2$, and m_1, m_2 positive integers.

All regular fibers are two-dimensional tori. Their fundamental groups are abelian groups Z^2 with two generators, corresponding to two basic cycles on a torus. The fundamental groups for different regular tori are isomorphic among themselves and to Z^2 integer lattice. We can establish the correspondence between basic cycles defined on different tori by choosing a continuous path in the 4D-space, which is transversal to fibers and by deforming basic cycles continuously along this path. In particular, for a closed path passing only through regular tori we get the automorphism of the fundamental group of a chosen regular torus. The corresponding map of basic cycles is the monodromy map. It is the same for all homotopy equivalent closed paths. If the path crosses singular lines similar to those taking place for integrable fibration of the (7.1,7.2) resonance oscillators only a subgroup of chains can be continuously deformed along the path and the monodromy map in such a case can be defined only for a subgroup of fundamental groups [12]. Nevertheless this map can be linearly extended to a whole group. In this case the extended monodromy map is represented by a matrix with fractional entries, while in the case of isolated critical values the monodromy map is given by integer matrix $\mu \in SL(2, Z)$.

7.3 Quantum Monodromy

In order to study the manifestation of classical monodromy in associated quantum problems we first need to recall the existence of local action-angle variables [1,13] and to introduce the elementary cell in the space of actions I_1, I_2, which is defined by $\Delta I_1, \Delta I_2$. Such a cell corresponds locally to the lattice of quantum states associated with integer values of local actions. If we choose basic vectors of such lattice as $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \Delta I_1 \\ \Delta I_2 \end{pmatrix}$ then under the transformation from one local action to another $\begin{pmatrix} I_1' \\ I_2' \end{pmatrix} = M \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$ the basic of the cell varies like $\begin{pmatrix} e_1' \\ e_2' \end{pmatrix} = (M^{-1})^\dagger \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$.

For quantum problems we are interested in the joint spectrum of commuting operators, corresponding to classical integrals $\{F_1, F_2\}$ [14–17]. The collection of joint eigenvalues superimposed on the image of the energy–momentum map for classical problem reveals locally the presence of a regular lattice associated with integrality conditions imposed on local actions by quantum mechanics. The lattice of quantum states for quantum problem corresponding

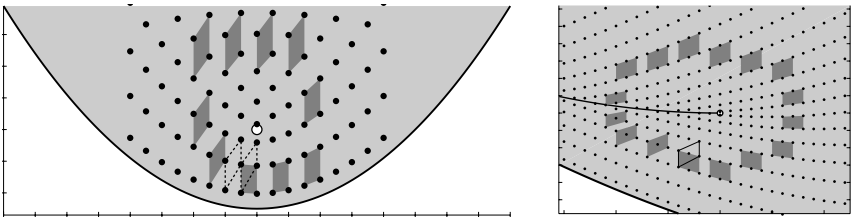

Fig. 7.3. Example of the lattice of quantum states with monodromy. Resonant oscillators (7.1,7.2) with $m_1 = m_2 = 1$ (*left*) and $m_1 = 1$, $m_2 = 2$ (*right*)

to classical oscillators with 1:(−1) and 1:(−2) resonances is represented in Fig. 7.3 [10].

Due to the existence of monodromy, the lattice of quantum states cannot be regular globally. From Fig. 7.3 it is clearly seen that the transport of elementary cell of the locally regular part of the lattice around the singularity shows nontrivial monodromy for a noncontractible close path in the base space (in the space of F_1, F_2 values). The presence of quantum monodromy can be interpreted as a presence of defects of locally regular lattice of quantum states [9]. In the case of isolated critical values of classical problem (Fig. 7.3, left) the choice of elementary cell is arbitrary and the monodromy map is integer. In the case of the presence of singular line in the image of the classical energy–momentum map, the dimension of the cell should be increased (doubled in the case of 1:(−2) resonance) in order to ensure the unambiguous crossing of the singular line [10]. In both cases the presence of singular fibers in classical problem is reflected in the appearance of some specific defects of the lattice of quantum states for corresponding quantum problem. We wish now to describe these specific defects arising in the quantum theory of Hamiltonian systems using methods and tools from defect theory of periodic lattices [18–21].

7.4 Elementary Defects of Lattices

Let us play with analogy between the 2-D lattice of quantum numbers (or of lattice formed by points with integer values of actions) and the 2-D lattice of a regular solid with defects. More precisely the idea is to see the correspondence between defects of periodic solids and monodromy, which is an obstruction to the existence of global action-angle variables in Hamiltonian dynamics (for integrable systems).

For a 2-D system each quantum state (or a site for a lattice formed by points) is characterized by two numbers, say (n_1, n_2). The existence of local order means that starting with some vertex (point of the lattice) one can form two vectors, or equivalently the elementary cell of the lattice by defining two vectors as joining (n_1, n_2) with $(n_1 + 1, n_2)$ and with $(n_1, n_2 + 1)$ respectively. This corresponds to the choice of the elementary cell with four vertices

$\{(n_1, n_2), (n_1 + 1, n_2), (n_1, n_2 + 1), (n_1, n_2 + 1)\}$. The choice of the elementary cell (or equivalently the choice of the basis of the lattice) is not unique. It is defined only up to arbitrary transformation with matrix $M \in SL(2, Z)$. But let us fix some choice for a moment. The existence of local actions in quantum-state lattice language means that by elementary translations in two directions we can label unambiguously all vertices by two numbers with the difference in numbers along each edge being 1 for one number and 0 for another. This means that there are no defects (in the local region studied).

Let us now analyze several different types of defects that can be imagined for periodic lattices in order to find possible candidates to represent defects of lattices of quantum numbers for quantum problems corresponding to classical Hamiltonian systems with nontrivial (integer and fractional) monodromy.

7.4.1 Vacations and Linear Dislocations

The simplest point defect well known in solids is the absence of vertex (or the presence of additional vertex). This defect does not distort the system of edges not connected with the vacation. The lattice is not deformed even slightly away from the point defect. The elementary cell after a circular trip around the vacation has no modifications (see Fig. 7.4, left).

Linear dislocation can be easily imagined to be formed through the following formal procedure. Let us remove all vertices on the half-line started at a given vertex and join the vertices through the gap (see Fig. 7.4, center). Equally, after making a cut along a line of vertices we can introduce additional (one or even several) half-lines. Now the circular path around this defect will show us the existence of the defect. To observe this defect we should go around it by doing the same number of steps in four directions (say, down, right, up, and left). If the final point will not be the same as the initial point, there is a defect. The vector from the initial point to the final point (Burgers vector in solid state physics) characterizes the dislocation. Observe that the elementary cell after the round trip around the dislocation will return exactly to its initial place (see Fig. 7.4, right) because Burgers vector does not depend on the initial point and it is exactly the same for all four vertices of the cell. This means that vacation and linear dislocations cannot be associated with monodromy type defects of regular lattices.

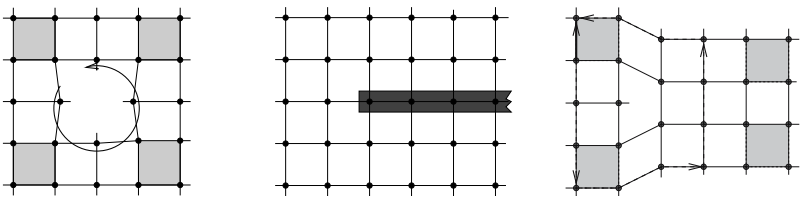

Fig. 7.4. Lattice with vacation (*left*). Construction of linear dislocation (*center*). Lattice with linear dislocation (*right*)

7.4.2 Angular Dislocations as Elementary Monodromy Defect

Another general idea to form defect starting from the regular lattice is to remove or to introduce "the solid angle" and to establish in some way the regular correspondence between two boundaries everywhere except one central point.

It is important to note that correspondence between two boundaries should be imposed in order to reconstruct the lattice. We will look for different possibilities but let us start with the simplest one: after removing (or introducing) the solid angle, the reconstruction is done by the parallel shift of lattice points in one chosen direction. The requirement for reconstructed lattice to be well defined everywhere except singular point can be satisfied only for some special values of removed or added angles. Namely we should impose that the number of removed (added) points at each vertical line is integer and varies linearly with distance from the vertex of the solid angle. Figure 7.5, left, shows examples of removed or added solid angles. Two different solid angles correspond to removing (adding) of one or two additional points from vertical line respectively, at each step in the horizontal direction. We can remove or add solid angles in different ways. Figure 7.6 illustrates the construction of the removed angle. We start with one chosen point O of the lattice and two basis vectors corresponding to "horizontal" and "vertical" directions. We put the first cut through the vertex A lying at the k-th vertical line counting from the vertex O ($k = 6$ in Fig. 7.6). To construct the second cut we go from A in vertical

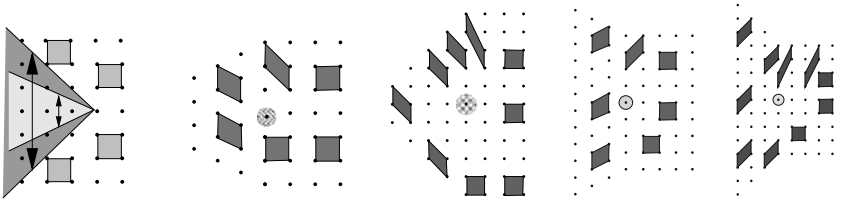

Fig. 7.5. Construction of the angular dislocation by removing or adding one of the solid angles shown in the left picture. Reconstructed lattices after removing or adding small or large sectors are shown together with transport of elementary cell along a closed path around the defect on the reconstructed lattice. The identification of boundaries after removing or adding solid angle is done by the parallel shift of lattice points in vertical direction

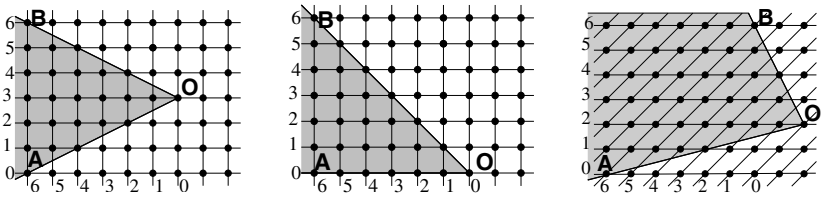

Fig. 7.6. Alternative constructions of the same defect

direction up to k-th horizontal line. Two rays OA and OB show the sector to be removed. Observe that with this construction s points are removed from s-th vertical line.

Figure 7.5 shows graphically what happens with a lattice after removing (or adding) solid angles. It is important to note that just by looking at the deformation of elementary cell after the round trip on the reconstructed lattice we can easily find how big was the solid angle removed (added) and what transformation (removing or adding) was exactly done. The absolute value of removed (added) solid angle can be read directly by comparing the form of the initial and the final cell. It is sufficient to write the transformation of two vectors forming elementary cell in matrix form. This matrix is directly related with the monodromy matrix for actions $M \leftrightarrow (M^{-1})^{\dagger}$ as explained in Sect. 7.3. For two examples shown in Fig. 7.5 this monodromy matrix has the form $\left(\begin{smallmatrix}1 & p\\ 0 & 1\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}1 & 0\\ p & 1\end{smallmatrix}\right)$ with $p = \pm 1$ or $p = \pm 2$. One or another form of matrix and the sign of p depends on the choice of the first and second basis vectors and on the direction of the circular trip (clockwise or counterclockwise). At the same time the absolute value of $|p|$ is unambiguously related with the absolute value of the removed (added) solid angle.

More subtle arguments are needed to distinguish between adding and removing solid angle with the same $|p|$. We denote later defects obtained by removing solid angle by $(-)$ and by adding solid angle by $(+)$.

7.4.3 About the Sign of the Elementary Monodromy Defect

The existence of the sign of Hamiltonian monodromy was conjectured by the author on the basis of analogy between monodromy and $(+)$ and $(-)$ defects of lattices. The proof was given by Cushman and Vu Ngoc [22]. We give here the characterization of the sign of defect in terms of lattice transformation.

Let us first compare initial and final cells for the same reconstructed lattice (with $|p| = 1$) obtained by removing the simplest solid angle but for both kinds of circular paths (clockwise and counterclockwise, see Fig. 7.7, left). The identification of the initial cell with the final one can be done only for

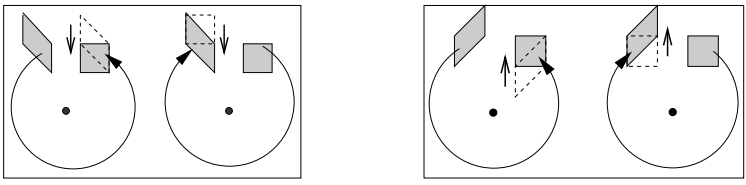

Fig. 7.7. Comparison of initial and final cells after the circular path around the singularity on the lattice reconstructed after removing [$(-)$ defect, *left*] or adding [$(+)$ defect, *right*] elementary solid angle. Both counterclockwise and clockwise circular paths are shown for each type of defects

two vertices. We choose the identified pair being the back side of the cell in the final position (with respect to the direction given by the sense of rotation). It is clearly seen from Fig. 7.7, left, that in order to deform the initial cell for $(-)$ defect to the form of the final cell we need to move the front side of the cell in the direction inside the surrounded singularity. This result remains unchanged if we apply the same procedure to the clockwise or counterclockwise circular path.

A similar analysis can be done for lattice reconstructed after adding solid angle, i.e., for $(+)$ defect, (see Fig. 7.7, right). It is clear from the figure that the deformation of the initial cell after the round trip is now in the outside direction with respect to the surrounded singularity. This result remains again the same for both clockwise and counterclockwise directions of the circular path.

Thus the simple geometrical analysis of the transformation of elementary cell enables one to associate with elementary monodromy the specific defect of the regular lattice. The defect obtained by removing solid angle with $|p| = 1$ is called the elementary monodromy defect. This defect appears exactly in lattices of quantum states for Hamiltonian systems corresponding to classical Hamiltonian systems with focus–focus singularities. Observe that defects with $|p| > 1$ appear naturally in Hamiltonian systems with symmetries. One of the most interesting and physically important systems of this kind is the integrable approximation for hydrogen atom in crossed electric and magnetic fields [23]. In classical systems monodromy with $|p| > 1$ corresponds to the presence of isolated singular fiber, which is $|p|$-times pinched torus [24]. Cushman and Vu Ngoc [22] have proved that only focus–focus singularities with the same sign of monodromy can appear in a connected component of the image of the generalized energy–momentum map of an integrable Hamiltonian system. In non-Hamiltonian systems monodromy of both signs can appear simultaneously [25].

7.4.4 Rational Cuts and Rational Line Defects

We have seen in Sect. 7.4.2 that only very special cuts together with matching rules enables us to construct the point defects. Now we generalize the admissible cuts but keep the matching rule. Let us start with the example of 1:2 rational cut, which is defined as follows (see Fig. 7.8).

We cut out half of the solid angle removed in the case of the elementary monodromy defect. After removing this solid angle the two boundaries are different. At one (lower boundary in Fig. 7.8) points are situated at each vertical line of the lattice. At the upper boundary of the cut points are situated only at each second vertical line. We keep the matching rules, i.e., identify the boundaries by sliding points along the vertical lines. Naturally, the reconstructed lattice is not homogeneous along the identified boundary. It is seen from the fact that the number of removed points from vertical lines varies like $0, 0, 1, 1, 2, 2, 3, 3, \ldots$ along the horizontal direction. (Note that the number

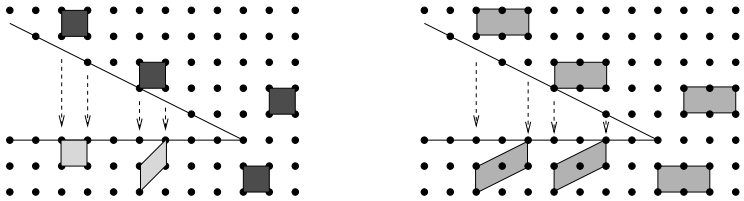

Fig. 7.8. Construction of the 1:2 rational lattice defect starting from The regular square lattice. Ambiguity in the transfer of 1×1 cell through the cut (left). Unambiguous transfer of double cell through the cut (right)

of removed points can be represented in the form of the sum of linear and oscillatory functions.) This means that the reconstructed lattice has a line defect.

If we try to pass the elementary cell of the lattice through the cut the result depends on the place where the cell goes through the boundary line. From Fig. 7.8, left, it is clear that when the right side of the cell goes through the cut at even vertical line (supposing the vertical line going through the vertex of the removed sector to be even) the form of the elementary cell remains unchanged. In contrast, when the right side of the cell goes through the cut at odd vertical line, the form of the cell changes. This ambiguity can be avoided if instead of elementary 1×1 cell we use a larger cell. Namely, we double the dimension of the cell in the horizontal direction. The double cell passes through the cut at any place in a similar way. But the internal structure of the cell changes after crossing the line defect. Cell transforms from "face centered" to "body centered" in the crystallographic terminology. But this modification is uniform along the cut. In some way, by increasing the dimension of the cell we neglect the effects comparable with the dimension of the cell. This enables us to define the transformation of lattice vectors after a closed path around the origin of the removed sector. Putting e_{h} and e_{v} as horizontal and vertical basis vectors of the square lattice shown in Fig. 7.8 and $\{e_{\mathrm{h}}^{\mathrm{double}} = 2e_{\mathrm{h}}, e_{\mathrm{v}}\}$ as vectors forming the double cell, the transformation of vectors forming the double cell after a close path around the origin of the removed sector in the counterclockwise direction is

$$\begin{pmatrix} (e_{\mathrm{h}}^{\mathrm{double}})' \\ e'_{\mathrm{v}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_{\mathrm{h}}^{\mathrm{double}} \\ e_{\mathrm{v}} \end{pmatrix}. \tag{7.3}$$

If we extend linearly this transformation to lattice vectors themselves the transformation matrix takes the form

$$\begin{pmatrix} e'_{\mathrm{h}} \\ e'_{\mathrm{v}} \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_{\mathrm{h}} \\ e_{\mathrm{v}} \end{pmatrix}. \tag{7.4}$$

The matrix so obtained with fractional entry coincides with the inversed transposed of the fractional monodromy matrix for actions in the case of $1{:}(-2)$

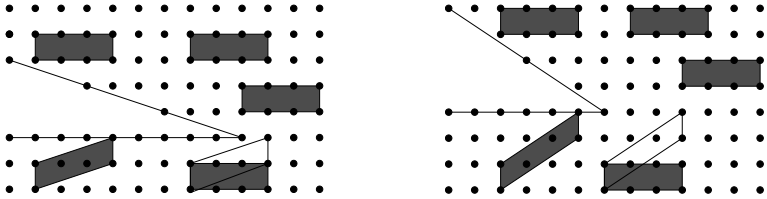

Fig. 7.9. Construction of the 1:3 and 2:3 rational lattice defects starting from the regular square lattice

resonant classical oscillator and with quantum fractional monodromy for corresponding quantum cells [12].

Using the same principle we can construct, for example, line defects by reconstructing a lattice after 1:3 or 2:3 rational cuts shown in Fig. 7.9. This notation means that we remove the solid angle $\varphi = \tan^{-1}(1/3)$ or $\tan^{-1}(2/3)$, respectively. We need to triple the dimension of cell in the horizontal direction in order to get unambiguous transformation rules for the cell after crossing the line defect on the reconstructed lattice. It is clear from Fig. 7.9 that the monodromy matrices for 1:3 and 2:3 rational defects have the form $\begin{pmatrix} 1 & 1/3 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2/3 \\ 0 & 1 \end{pmatrix}$, respectively. Removing 2:3 rational solid angle is equivalent to removing twice the 1:3 rational solid angle. Generalization to arbitrary rational cut with the same type of matching rules for reconstruction of the lattice is quite obvious and leads to half-line defect with fractional monodromy matrix.

We can also suggest alternative matching rules after rational cuts with the idea of obtaining reconstructed lattice with only point rather than the line defect. Let us consider again as example the 1:3 rational cut but with different matching rules for two boundaries (see Fig. 7.10).

It is clear that if we want to have on the reconstructed lattice only point defect all vertices on two boundaries should be consecutively identified. This identification imposes matching rule for one of the basis vectors of the lattice. Another should be chosen in such a way that two new basis vectors form an elementary cell of the same volume, i.e., they should be related one to another with $SL(2, Z)$ transformation. In fact this matrix is precisely the

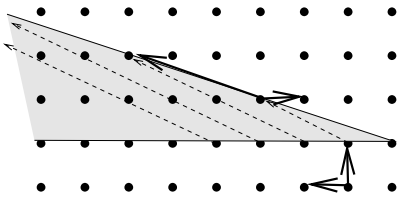

Fig. 7.10. Matching rules for reconstruction of lattice after rational cut. Example of 1:3 cut. The monodromy of the resulting defect is the Arnol'd cat map

monodromy matrix of the point defect just by the construction. In the case of 1:3 rational cut shown in Fig. 7.10 the resulting monodromy matrix has the form $\begin{pmatrix} 3 & 1 \\ -1 & 0 \end{pmatrix}$. This matrix is known as Arnol'd cat map [26]. A lot of different examples of point defects can be constructed in a similar way. But we will take as elementary point defect only $(-)$ defects corresponding to elementary monodromy matrix. We demonstrate now that all other defects can be considered as more complicated objects composed in some way from several elementary ones.

7.5 Defects with Arbitrary Monodromy

We now turn to the description of defects that can be characterized by arbitrary monodromy matrices. As soon as the choice of the basis of the lattice is ambiguous, the matrix representation of the monodromy transformation is basis dependent. For example the monodromy matrix $M_a = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ after the transformation to another basis through the similarity $M'_a = AM_aA^{-1}$ with A being arbitrary $SL(2, Z)$ matrix takes the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 - ac & a^2 \\ -c^2 & 1 + ac \end{pmatrix}.$$

From this family of equivalent matrices it is immediately clear that matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ are equivalent but they are written in different frames. In contrast, matrix $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ is equivalent to $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ but is not equivalent to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ in spite of the fact that these two matrices are mutually inversed.

In order to formulate a precise statement about equivalence or in-equivalence of different defects we should first establish equivalence of $SL(2, Z)$ matrices with respect to conjugation by elements of $SL(2, Z)$, i.e., to describe classes of conjugated elements of $SL(2, Z)$ group.

It is well known that the trace and the determinant of the matrix are invariant with respect to similarity transformation. But these invariants are not sufficient to completely characterize classes of conjugated elements. Before looking for $SL(2, Z)$ matrices let us start with $SL(2, R)$ ones.

7.5.1 Topological Description of Unimodular Matrices

Let us consider the subspace of $SL(2, R)$ matrices $M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ with $\mathrm{Tr}\, M = K$. This means that four matrix elements α, β, γ, and δ are related by two equations

$$\alpha\delta - \beta\gamma = 1, \quad \alpha + \delta = K. \tag{7.5}$$

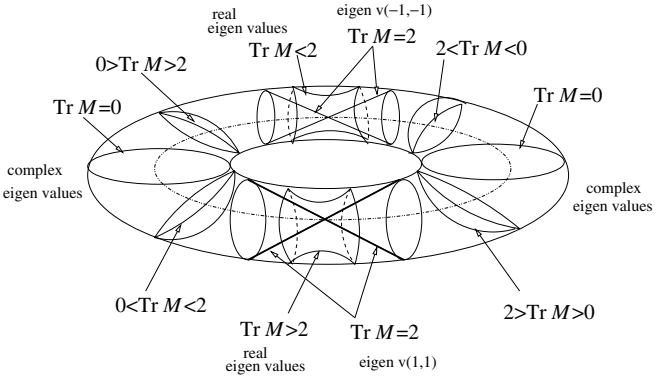

Fig. 7.11. The topological structure of the space of 2 by 2 matrices with determinant 1. The solid torus in 3-D space is foliated by levels corresponding to a given value of the trace of matrix M

Eliminating one parameter (say α) we get the following relation between three parameters β, γ, and δ

$$-1 + K\delta - \delta^2 - \gamma\beta = 0. \tag{7.6}$$

We can interpret this relation as the geometrical description of all $SL(2, R)$ matrices with given trace in the three-dimensional space of parameters β, γ, and δ. The geometrical form of the surface so obtained depends on the value of K. Topologically there are three different situations.

If $K = \pm 2$ we have double cone with vertex corresponding to $\pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ matrix. If $|K| > 2$ we have a hyperboloid of one-sheet and if $-2 < K < 2$ we have two-sheeted hyperboloid. We can schematically represent the whole family of $SL(2, R)$ matrices by filling the solid torus in three-dimensional space of parameters by surfaces corresponding to all possible values of traces. This representation is given in Fig. 7.11.

The existence of two disjoint connected components for matrices with $-2 \leq \operatorname{Tr} M \leq 2$ implies the existence of additional invariant, which classifies matrices with the same trace into smaller subclasses of conjugate elements. We need such description but only for matrices in $SL(2, Z)$. The important difference between $SL(2, R)$ and $SL(2, Z)$ cases is due to the fact that there are only a finite number of possible values of the trace in $SL(2, Z)$, which correspond to matrices with complex eigenvalues in $SL(2, R)$. In physical language this is the consequence of the fact that only axes of second, third, fourth, and sixth orders are compatible with the existence of the lattice.

Formal proof: Characteristic polynomial for the $SL(2, Z)$ matrix M has the form $\lambda^2 - (\operatorname{Tr} M)\lambda + 1 = 0$. It has complex eigenvalues only if the discriminant $(\operatorname{Tr} M)^2 - 4 < 0$. As soon as the trace is integer, it is only possible that $(\operatorname{Tr} M) = 0, \pm 1$.

Now we can return to the study of $SL(2, Z)$ case.

7.5.2 Classes of Conjugated Elements and "Normal Form" of $SL(2, Z)$ Matrices

The matrices $M \in SL(2, Z)$ are named parabolic, elliptic, or hyperbolic, depending on their trace. Parabolic matrices have trace equal ± 2 and their eigenvalues are $\{+1, +1\}$ or $\{-1, -1\}$. Elliptic matrices have trace ± 1 or 0. Their eigenvalues are complex numbers. Hyperbolic matrices have $|\mathrm{Tr}\, M| > 2$. Their eigenvalues are real irrational numbers. Identity matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and minus identity $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ commute with all elements from $SL(2, Z)$ and each form a proper class of conjugate elements consisting of one element. We consider these classes separately. In Fig. 7.11 these matrices correspond to vertices of double cones of matrices with trace ± 2.

Within each class of conjugate elements we can choose one matrix to be the "normal form." All classes of conjugate elements together with normal forms are listed in Table 7.1.

7.5.3 Several Elementary Monodromy Defects

We have defined the construction of elementary defects using one chosen lattice basis. Both cuts and matching rules were precisely defined in that basis. But the choice of the lattice basis is not unique. If there are two defects, both characterized by elementary monodromy matrix but the choice of basis and the orientation of cuts for these two defects are different, the global monodromy, corresponding to transformation of elementary cell after a circular path around both defects, depends on relative orientation of two removed solid angles. Let us again start with some particular examples of systems with several elementary defects.

Disclinations as a Composition of Elementary Monodromy Defects

Figure 7.12 shows regular square lattice with three defects corresponding to elementary monodromy. For two removed angles (sectors around horizontal lines) the reconstruction of lattice is done through sliding points in vertical

Table 7.1. Classes of conjugated elements of $SL(2, Z)$ group together with normal forms of matrices for each class

Trace	$K, (\lvert K\rvert > 2)$	± 2	± 1	0
Module	–	$p = 0, \pm 1, \ldots$	$\varepsilon = \pm 1$	$\varepsilon = \pm 1$
Normal form	$\left(\begin{smallmatrix} K & 1 \\ -1 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} \pm 1 & p \\ 0 & \pm 1 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} \pm(1+\varepsilon)/2 & \varepsilon \\ -\varepsilon & \pm(1-\varepsilon)/2 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{smallmatrix}\right)$

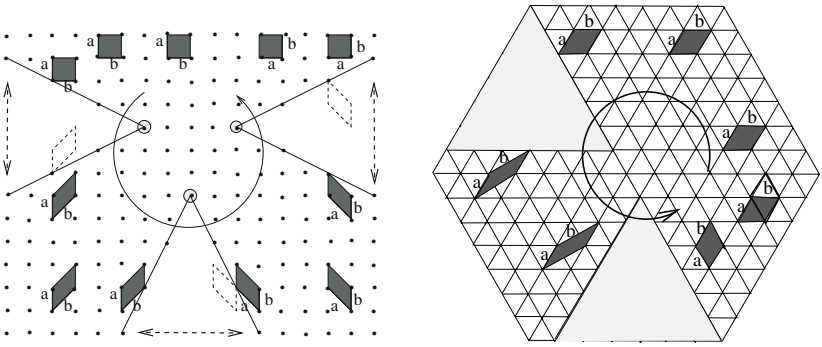

Fig. 7.12. Regular square lattice with three elementary monodromy defects (left). Regular triangular lattice with two elementary monodromy defects (right)

directions. For the third removed angle the identification of boundaries is done through the horizontal sliding of lattice points.

The cumulative effect of three elementary cuts is the rotation of the elementary cell by $\pi/2$. The direction of the rotation of the elementary cell is defined by the direction of the circular path around singularities. Such a defect is known in solid state physics as $\pi/2$ rotational disclination. It is easy to see that the same effect takes place if the three cuts are distributed in another way between vertical and horizontal directions (two vertical and one horizontal). To see this it is just sufficient to look at the same figure after rotating it by $\pi/2$.

Naturally, similar construction can be done with three singular points corresponding to adding the solid angle and reconnecting new boundaries through horizontal or vertical shift. The resulting effect on the elementary cell is again the $\pi/2$ rotation, but now the rotation of the elementary cell is in opposite direction (as compared to the direction of the circular loop around the singularity).

The global effect in both cases can be reproduced by removing (or adding) the solid angle $\pi/2$ and by reconstructing lattice through identification of two boundaries by rotating them as it is shown in Fig. 7.13 where $\pi/2$ solid angle is removed. Naturally one can also remove π or $3\pi/2$ solid angle (Fig. 7.14) or to add $\pi/2$ or $k\pi/2$ solid angle as it is shown in Fig. 7.15. This gives negative or positive rotational disclinations.

The same construction made with triangular lattice and with two elementary monodromy defects rotated one with respect to another over $2\pi/3$ gives the cumulative effect consisting in rotation of elementary cell over $2\pi/6$ after a close path surrounding two elementary defects (see Fig. 7.12, right). The cumulative effect of such two elementary monodromy defects is the $\pi/3$ rotational disclination. Its multiple, positive or negative analogs can be immediately constructed.

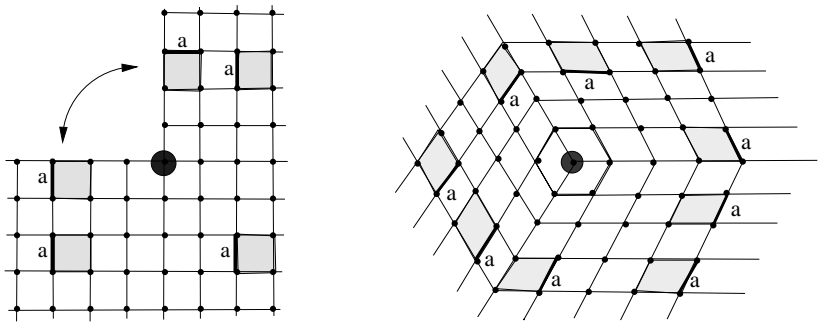

Fig. 7.13. Construction of the rotational disclination by removing solid angle $\pi/2$ shown on the *left* picture

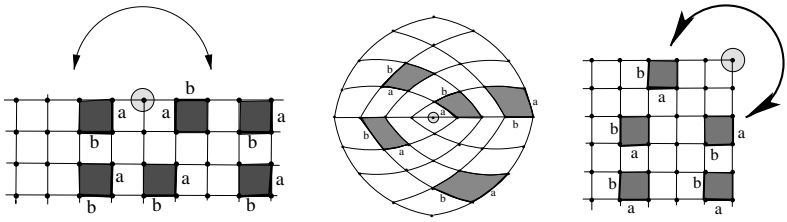

Fig. 7.14. Construction of the rotational disclination by removing solid angle π (*left*) and $3\pi/2$ (*right*). The reconstructed lattice after removing π solid angle (*center*)

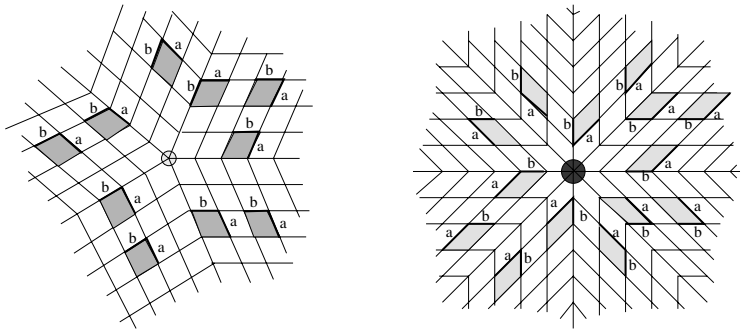

Fig. 7.15. Construction of the rotational dislocation (disclination) by introducing solid angle $k\pi/2$. $k = 1$ (*left*) and $k = 4$ on the (*right*)

Rotational disclinations are well-known defects in the solid state physics. From the point of view of defects of quantum state lattices and classical Hamiltonian monodromy, the elementary monodromy defects seem to be more fundamental. Any rotational disclination can be constructed as a global effect in systems with several elementary monodromy defects.

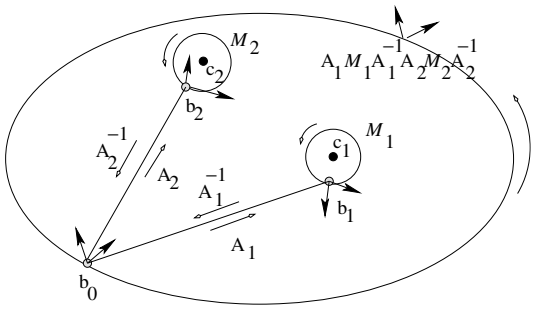

Fig. 7.16. Relation between local monodromy matrices for isolated defects and global monodromy for the circular path around two defects

Multiple Defects with Trivial Global Monodromy

Let us now consider in a more general way the correspondence between local and global monodromy for a system of elementary defects. Figure 7.16 illustrates this relation.

Suppose we have two defects c_1 and c_2 characterized by monodromy matrices M_1 and M_2. These monodromy matrices are obtained by going around the defect c_i starting from point b_i and using local basis associated with point b_i. If we are interested now in global monodromy that corresponds to a closed loop going around two defects and starting at initial point b_0 with its own local basis we can calculate the global monodromy by going first from b_0 to b_1, making closed loop around c_1, returning back by the same way, and repeating the same for the second defect. The global monodromy calculated in this way should be the same by homotopy arguments. If the modification of the basis between b_0 and b_i is described by matrices A_i the global monodromy M can be expressed in terms of M_1 and M_2 as $M = A_1 M_1 A_1^{-1} A_2 M_2 A_2^{-1}$.

Naturally for an arbitrary system of elementary defects the global monodromy matrix can always be represented in the form $M = \prod_i A_i M_i A_i^{-1}$. As already noted the monodromy matrix is defined up to conjugation with $SL(2, Z)$ matrices, i.e., defects with different monodromy are in one-to-one correspondence with classes of conjugate elements of $SL(2, Z)$ matrices.

One can easily verify that arbitrary $SL(2, Z)$ matrix can be represented in the form of product of matrices, conjugate to elementary monodromy matrices with one chosen sign [27]. In particular, the identity matrix can also be represented in the form of product of matrices conjugate to elementary monodromy matrix. It is obvious that four $\pi/2$ rotational disclinations (or six $\pi/6$ rotational disclinations) give trivial global monodromy. In fact the elementary cell makes a 2π rotation when going along closed path surrounding these defects and in spite of the fact that the monodromy is trivial the closed path is not contractible and the defect exists. An easy consequence of this statement: the monodromy matrix (defined up to conjugation with $SL(2, Z)$ matrices) is not sufficient to distinguish defects. Two defects with the same

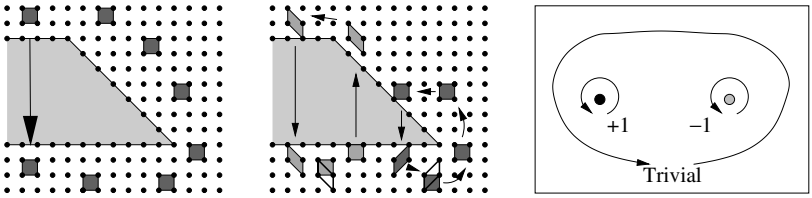

Fig. 7.17. Lattice with two elementary monodromy defects of different sign

monodromy matrix can further be labeled by the number k of 2π rotations of the elementary cell after a closed path around a defect. This additional number k can be arbitrary integer $k = 0, \pm 1, \pm 2, \ldots$. Note that an elemenraty $(-)$ monodromy defect can be constructed as a cumulative effect of 11 elementary $(+)$ monodromy defects [27].

One can easily obtain trivial global monodromy for a closed path around two defects with different signs. Figure 7.17 shows construction of two elementary monodromy defects with different signs. Signs plus and minus indicate adding and removing of the same solid angle, respectively.

7.5.4 Several Rational Line Defects

Let us now discuss examples of lattices with multiple rational line defects. We assume that all defects are of the same sign, i.e., obtained by removing solid angle. We start with an example of two defects 1:2 and 1:3, which have similar orientation (see Fig. 7.18, left). These two defects model the singularities of integrable toric fibration for $2:-3$ resonant oscillator (7.1). Elementary 1×1 cell cannot cross unambiguously both defect half-lines. The cell should be doubled in horizontal direction in order to cross unambiguously the 1:2 defect. In a similar way the cell should be tripled in the same direction in order to cross the 1:3 defect. This means that only 1×6 cell, which is six times larger in the horizontal direction, can cross both defects. Using such a cell we can go

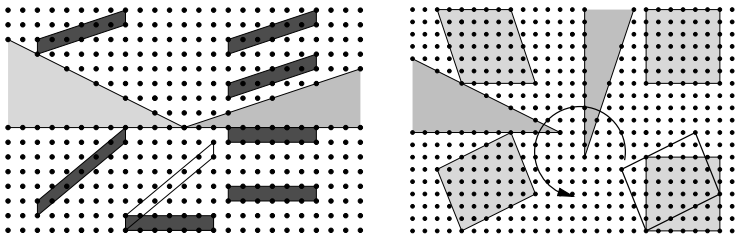

Fig. 7.18. Construction of a lattice with two rational defects, 1:2 and 1:3. Parallel defects that correspond to singular one-dimensional strata for $2:(-3)$ resonance oscillator (*left*). Two orthogonal defects (*right*)

along a closed path surrounding the singular vertex. After linear extension to elementary lattice vectors we get the fractional monodromy matrix $\begin{pmatrix} 1 & 5/6 \\ 0 & 1 \end{pmatrix}$.

If two rational defects, 1:2 and 1:3, have different orientations the situation becomes quite different. Figure 7.18, right, shows these two defects with orthogonal orientation. In order to pass through horizontal 1:2 defect the cell should be doubled in horizontal direction. In order to pass through 1:3 vertical defect the cell should be tripled in vertical direction. Moreover, one should note that all vertices of the cell should lie on even vertical lines and on horizontal lines having the same number modulo 3. This means that we need to take at least 6×6 cell in order to cross unambiguously both rational cuts. The resulting monodromy matrix for a counterclockwise path around two singular points has the form $\begin{pmatrix} 5/6 & 1/2 \\ -1/3 & 1 \end{pmatrix}$. This is an elliptic $SL(2, Q)$ matrix.

Rational Defect Line with Ends and Singular Points

It is quite easy to construct rational defect with two ends. It is sufficient to start to cut solid angle as it was done for rational defect but at some another point to change the slope and to continue with another slope related to integer monodromy defect. This situation is shown in Fig. 7.19 on three different examples. Figure 7.19, left, shows 1:2 cut, which starts at point A but at point B the angle of the cut changes. It becomes equal angle characteristic to elementary monodromy defect. This means that the line 1:2 defect on reconstructed lattice has two ends and the monodromy around each end is $\begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix}$. At the same time the global monodromy for close path surrounding 1:2 defect is $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. In a similar way (see Fig. 7.19, center) we can start at point A with 1:3 cut and change at point B the angle in order to get again on the reconstructed lattice elementary monodromy for global close path. This means that surrounding point A we get the monodromy $\begin{pmatrix} 1 & 1/3 \\ 0 & 1 \end{pmatrix}$, while surrounding

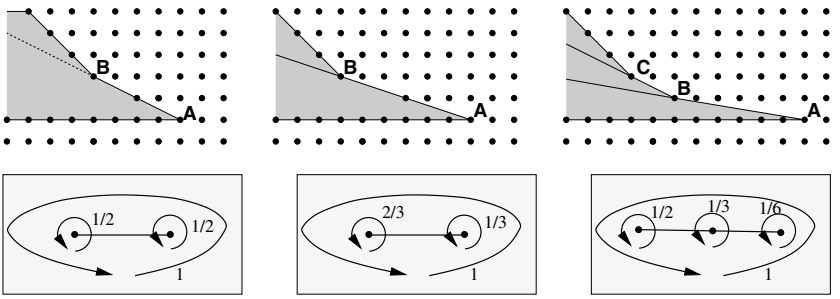

Fig. 7.19. Construction of the line defects with ends and singular points. Defect with equivalent ends (*left*). Defect with inequivalent ends (*center*). Defect with inequivalent ends and additional singular point splitting defect into two fragments (*right*)

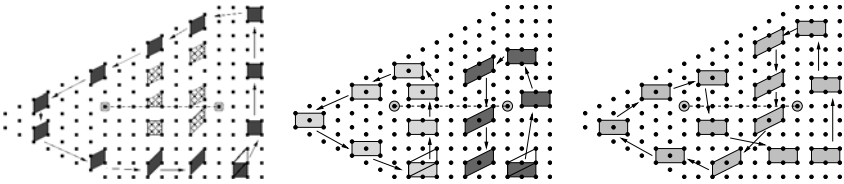

Fig. 7.20. Lattice with two half-integer defects. Defect is an interval with two end points. Elementary cell cannot cross unambiguously the defect line (*left*). Both ends have the same (1/2) monodromy (*center*). Figure eight close path is not contractible but the monodromy is trivial (*right*)

point B the monodromy becomes $\begin{pmatrix} 1 & 2/3 \\ 0 & 1 \end{pmatrix}$. Two ends are not equivalent. Naturally, we can change angle several times. This gives the line defect with singular points on it. Each singular point corresponds to modification of value of solid angle removed from the lattice. Figure 7.19, right, shows an example with three singular points. Generalization to more complicated examples is straightforward.

Figure 7.20 demonstrates geometrically modifications that occur with elementary cell after going along different closed paths on the lattice with 1:2 defect with two ends. To find the global monodromy one can use elementary 1×1 cell (Fig. 7.20, left), whereas it is not possible to cross the line with such a cell because of ambiguity of cell modifications.

Taking 1×2 cell we can easily go around each singular point and see (Fig. 7.20, center) that the result is exactly the same for both points, namely half of the global modifications.

Figure 7.20 (right) shows the evolution of the double cell along the figure eight close path, which goes around two centers but in opposite directions. This close path results in trivial monodromy, the cell has no modifications after returning to the initial point.

7.6 Is There Mutual Interest in Defect – Monodromy Correspondence?

Stimulated by analogy between classical and quantum monodromy for Hamiltonian integrable systems from one side and defects of regular periodic lattices from another we have suggested construction of "elementary integer and fractional lattice defects" associated with elementary integer and fractional monodromy. Then we have proposed how to generate more complicated defects of lattices by combination of elementary ones. Among these more complicated defects are those like disclinations, which are well known in solid state physics. At the same time the author does not know simple examples of dynamical Hamiltonian system with similar defects. Reciprocally, many examples of dynamical Hamiltonian systems with elementary monodromy are known but

the "elementary monodromy defect" seems not to be individually detected experimentally in periodic solids. These observations enable us to formulate later a number of problems concerning Hamiltonian dynamical systems with singularities and periodic lattices with defects. Some of these problems have more or less intuitively evident answers but strict mathematical proofs are still absent. In other cases even the formulation of the problem is not precise and should be critically analyzed and corrected before looking for the answer.

- *About the sign of elementary monodromy defect.* How does one characterize the class of dynamical systems (classical and quantum) possessing only elementary monodromy defects of one sign? For Hamiltonian systems focus–focus singularities correspond to elementary $(-)$ defects. The tentative answer is to say that elementary $(+)$ monodromy defects are generic for PT-invariant dynamical systems with nonhermitian Hamiltonians and real spectra [28, 29].
- *Correspondence between topology of singular fibers of integrable toric fibrations and integer and fractional defects of lattices.* Some simple examples of such correspondence were given. Is it possible to establish more general correspondence? In particular it seems natural that elementary $(+)$ and $(-)$ monodromy defects correspond to pinched tori with different index of transversal self-crossings [30].
- *Constructive methods to design Hamiltonian classical and quantum systems with prescribed type of monodromy.* A less ambitious task is to propose a list of concrete examples of classical and quantum systems that show the manifestation of different elementary and nonelementary defects.
- *Existence of a topological invariant separating different singularities (defects) with the same monodromy but with different numbers of* 2π*-rotations of elementary cell.* The analogy between this problem and the Riemann surfaces description [31] was pointed out to the author on several occasions.
- *Extension of the correspondence between singularities and defects from 2D systems to higher dimensional systems.* This is surely a very wide subject and the author believes that the first steps in mathematical generalization should be guided by natural physical examples.
- *Global restrictions on the system of defects and on the system of singularities of toric fibrations in the case of compact base space (lattices on compact spaces).* For example, singular toric fibration over S^2 base space should have 24 elementary focus–focus singularities [32] or equivalently 24 elementary $(-)$ monodromy defects or 12 $(\pi/3)$-rotational disclinations known as pentagonal defects [33]. This problem has obvious relation with fullerene-like materials.
- *The relation between the number of removed vertices for a defect and the Duistermaat–Heckman measure for the reduced Hamiltonian system.* Hint: The slope of the function giving the number of removed vertices from vertical line as a function of the number of a vertical line coincides with the

Chern class of the integrable fibration used in the Duistermaat–Heckman theorem [34, 35].

- *To find macroscopic/mesoscopic physical systems that manifest the presence of elementary (+/−) monodromy defects.* Possible candidates besides periodic solids or liquid crystals may be membranes [36], fullerenes and curved carbon surfaces [37], viruses [38], colloidal structures [39], etc.
- *Relation between internal structure of elementary cells and possible existence of isolated elementary integer and fractional monodromy defects in real physical systems.* In what kind of systems (materials) one the topological properties more important than geometric and steric effects and enable one to see the manifestation of elementary monodromy defects?
- *Physical consequences of sign conjecture.* If one accepts the formulated sign conjecture, i.e., presence of only (−) defects in generic families of Hamiltonian systems depending on a small number of parameters, there is a fundamental difference between (+) and (−) (or in other terms between "right" and "left") in both classical and quantum mechanics. How does one formulate this conjecture in more precise terms and what kind of physical consequences can be rigorously deduced?

The author hopes that Hamiltonian dynamics and periodic solids give complementary points of view that are useful for both fields of scientific interest. This chapter aims to stimulate mutual interest, better understanding, and further cooperation among specialists working in these fields.

Acknowledgments. The author thanks Drs. R. Cushman, N. Nekhoroshev, and D. Sadovskii for many stimulating discussions. This work was essentially done during the author's stay in IHES, Bures-sur-Yvette, France; Mathematical Institute, University of Warwick, UK; and Max-Planck Institut für Physik komplexer Systeme, Dresden, Germany. The support of these institutions is acknowledged. This work is a part of the European project MASIE, contract HPRN-CT-2000-00113.

References

1. Arnol'd, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1981)
2. Lerman, L.M., Umanskii, Ya.L.: Four dimensional integrable Hamiltonian systems with simple singular points, Translations of Mathematical Monographs, **176**, AMS, Providence, RI (1998)
3. Zung, N.T.: Diff. Geom. Appl. **7**, 123 (1997)
4. Cushman, R.H.: C.W.I. Newslett. **1**, 4 (1983)
5. Duistermaat, J.J.: Comm. Pure Appl. Math. **33**, 687 (1980)
6. Cushman, R.H., Duistermaat, J.J.: Bull. Am. Math. Soc. **19**, 475 (1988)
7. Bates, L.: J. Appl. Math. Phys. (ZAMP) **42**, 837 (1991)
8. Cushman, R.H., Bates, L.M.: Global Aspects of Classical Integrable Systems, Birkhäuser, Basel (1997)

9. Sadovskii, D.A., Zhilinskii, B.I.: Phys. Lett. A **256**, 235 (1999)
10. Nekhoroshev, N.N., Sadovskii, D.A., Zhilinskii, B.I.: C. R. Acad. Sci. Paris, Ser. I, **335**, 985 (2002)
11. Colin de Verdier, Y., Vu Ngoc, S.: Ann. Ec. Norm. Sup. **36**, 1 (2003)
12. Nekhoroshev, N.N., Sadovskii, D.A., Zhilinskii, B.I.: in preparation (2003)
13. Nekhoroshev, N.N.: Trans. Moscow Math. Soc. **26**, 180 (1972)
14. Grondin, L., Sadovskií, D.A., Zhilinskii, B.I.: Phys. Rev. A, **142**, 012105 (2002)
15. Vu Ngoc, S.: Comm. Math. Phys. **203**, 465 (1999)
16. Waalkens, H., Dullin, H.R.: Ann. Phys. (NY) **295**, 81 (2001)
17. Child, M.S., Weston, T., Tennyson, J.: Mol. Phys. **96**, 371 (1999)
18. Kleman, M.: Points Lines and Walls. Chichester, Wiley (1983)
19. Mermin, N.D.: Rev. Mod. Phys. **51**, 591 (1979)
20. Michel, L.: Rev. Mod. Phys. **52**, 617 (1980)
21. Kroner, E.: Continuum Theory of Defects in Physics of Defects. Les Houches, Ecole d'ete de physique theorique. Elsevier, New York, 215–315 (1981)
22. Cushman, R., Vu Ngoc, S.: Ann. I. Henri Poincaré, **3**, 883 (2002)
23. Cushman, R.H., Sadovskií, D.A.: Physica D, **65**, 166 (2000)
24. Matveev, V.: Sb. Math. **187**, 495 (1996)
25. Cushman, R.H., Duistermaat, J.J.: J. Diff. Eqns. **172**, 42 (2001)
26. Arnol'd, V., Avez, A.: Ergodic Problems of Classical Mechanics Benjamin, New York (1968)
27. Cushman, R., Zhilinskii, B.: J. Phys. A: Math. Gen **35**, L415 (2002)
28. Bender, C., Boettcher, S.: Phys. Rev. Lett. **80**, 5243 (1998)
29. Zhilinskii, B.: in Proceedings of 24th International Conference on Group-Theoretical Methods in Physics, Paris (2003)
30. Matsumoto, Y.: J. Math. Soc. Japan **37**, 605 (1985)
31. Cartan, E.: Leçon sur la Géométrie des Espaces de Riemann, Ch. 3. Gauthier-Villars, Paris (1928)
32. Zung, N.T.: ArXiv:math.DG/0010181 (2001)
33. Nelson, D.R.: Phys. Rev. B **28**, 5515 (1983)
34. Duistermaat, J.J., Heckman, G.J.: Invent. Math. **69**, 259 (1982)
35. Guillemin, V.: Moment Map and Combinatorial Invariants of Hamiltonian T^n-spaces. Birkhäuser, Basel (1994)
36. Bowick, M.J., Travesset, A.: Phys. Rep. **344**, 255 (2001)
37. Krishnan, A. et al.: Nature **388**, 451 (1997)
38. Ganser, B.K. et al.: Science **283**, 80 (1999)
39. Gido, S.P., Gunther, J., Thomas, E.L., Hoffman, D.: Macromolecules **26**, 4506 (1993)

8

Two-Qubit and Three-Qubit Geometry and Hopf Fibrations

R. Mosseri

Summary. This chapter reviews recent attempts to describe the two-qubit and three-qubit Hilbert space geometries with the help of Hopf fibrations. In both cases, it is shown that the associated Hopf map is strongly sensitive to states entanglement content. In the two-qubit case, a generalization of the one-qubit celebrated Bloch sphere representation is described.

8.1 Introduction

Two-level quantum systems, denoted as qubits, have gained a renewed interest in the past ten years, owing to the fascinating perspectives of quantum information [1]. As an example, an ideal quantum computer is a large set of qubits, which is subject to individual and collective operations. Having in mind the different qubit manipulation protocols that are proposed in this growing field, it is therefore of high interest to represent their quantum evolution in a suitable representation space, in order to get some insight into the subtleties of this complicated problem. For single two-level systems, a well-known tool in quantum optics is the Bloch sphere representation, where the simple qubit state is faithfully represented, up to a global phase, by a point on a standard sphere S^2, whose coordinates are expectation values of physically interesting operators for the given quantum state. Guided by the relation between the Bloch sphere and a geometric object called the Hopf fibration of the S^3 hypersphere [2], a generalization for a two-qubit system was recently proposed [3], in the framework of the (high-dimensional) S^7 sphere Hopf fibration, and is recalled later. An interesting result is that the S^7 Hopf fibration is entanglement sensitive and therefore provides a kind of "stratification" for the two qubit states space with respect to their entanglement content. An extension of this description to a three-qubit system, using the S^{15} Hopf fibration, is also presented here.

We first briefly recall known facts about the Bloch sphere representation and its close relation to the S^3 Hopf fibration. We then recall in some detail

what had been recently done for the two-qubit case in terms of the S^7 Hopf fibration. The S^{15} Hopf fibration is then introduced, which helps describing the three-qubit Hilbert space geometry. As far as computation is concerned, going from the S^3 to the S^7 and then the S^{15} fibrations merely amounts to replacing complex numbers by quaternions and then octonions. This is why a brief introduction to quaternions and octonions is given in Appendix. Note that using these two kinds of generalized numbers is not strictly necessary here, but they provide an elegant way of putting the calculations into a compact form, and have (by nature) a natural geometrical interpretation

8.2 From the S^3 Hypersphere to the Bloch Sphere Representation

A (single) qubit state reads

$$|\Psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \qquad \alpha,\beta \in \mathbb{C}, \qquad |\alpha|^2 + |\beta|^2 = 1. \tag{8.1}$$

In the spin 1/2 context, the orthonormal bases $\{|0\rangle\,, |1\rangle\}$ are the two eigenvectors of the (say) σ_z (Pauli spin) operator. Viewed as pairs of real numbers, the two normalized components α, β generate a unit radius sphere S^3 embedded in $\mathbb{R}^4$. To take into account the global phase freedom, one expects to find a way to fill S^3 with circles (the orbit of a global phase $\exp \mathrm{i}\omega$ multiplying the pair (α, β)), such that each state belongs to exactly one such circle. This task is nicely fulfilled by the so-called S^3 Hopf fibration [4].

A fibered space E is defined by a (many-to-one) map from E to the so-called "base space," all points of a given fiber F being mapped onto a single base point. A fibration is said to be "trivial" if the base B can be embedded in the fibered space E, the latter being faithfully described as the direct product of the base and the fiber (think for instance of fibrations of $\mathbb{R}^3$ by parallel lines $\mathbb{R}$ and base $\mathbb{R}^2$ or by parallel planes $\mathbb{R}^2$ and base $\mathbb{R}$).

One the most famous examples of a nontrivial fibration is the Hopf fibration of S^3 by great circles S^1 and base space S^2. For the qubit Hilbert space purpose, the fiber represents the global phase degree of freedom, and the base S^2 is identified to the Bloch sphere. One standard notation for a fibered space is that of a map $E \xrightarrow{F} B$, which reads here a $S^3 \xrightarrow{S^1} S^2$. Its nontrivial character implies $S^3 \neq S^2 \times S^1$. This translates into the known failure in ascribing consistently a definite phase to each representing point on the Bloch sphere. To describe this fibration in an analytical form, we go back to the definition of S^3 as pairs of complex numbers (α, β), which satisfy $|\alpha|^2 + |\beta|^2 = 1$. The Hopf map is defined as the composition of a map h_1 from S^3 to $\mathbb{R}^2$ $(+\infty)$, followed by an inverse stereographic map h_2 from $\mathbb{R}^2$ to S^2:

$$\begin{aligned} &h_1 : S^3 \longrightarrow \mathbb{R}^2 \cup \{\infty\},\ (\alpha,\beta) \mapsto C = \overline{\alpha\beta^{-1}}, \quad \alpha,\beta \in \mathbb{C}, \\ &h_2 : \mathbb{R}^2 \cup \{\infty\} \longrightarrow S^2,\ C \mapsto M(X,Y,Z), \quad X^2 + Y^2 + Z^2 = 1, \end{aligned} \tag{8.2}$$

where $\overline{z}$ is the complex conjugate of z). The first map h_1 clearly shows that the full S^3 great circle, parametrized by $(\alpha \exp \mathrm{i}\omega, \beta \exp \mathrm{i}\omega)$, is mapped onto the same single point with complex coordinate C. It is easy to show that, with $\mathbb{R}^2$ cutting the unit radius S^2 along the equator, and the north pole (along the Z axis) as the stereographic projection pole, the S^2 Hopf fibration base coordinates coincide with the well-known S^2 Bloch sphere coordinates:

$$\begin{aligned} X &= \langle \sigma_x \rangle_\Psi = 2\,\mathrm{Re}(\overline{\alpha}\beta), \\ Y &= \langle \sigma_y \rangle_\Psi = 2\,\mathrm{Im}(\overline{\alpha}\beta), \\ Z &= \langle \sigma_z \rangle_\Psi = |\alpha|^2 - |\beta|^2). \end{aligned} \tag{8.3}$$

This correspondence between Hopf map and Bloch sphere is not new [2], but is poorly known in both communities (quantum optics and geometry). It is striking that the simplest nontrivial object of quantum physics, the two-level system, bears such an intimate relation with one of the simplest nontrivial fibered space.

It is tempting to try to visualize the full (S^3) Hilbert space with its fiber structure. This can be achieved by doing a (direct) stereographic map from S^3 to $\mathbb{R}^3$ (Fig. 8.1). Each S^3 circular fiber is mapped onto a circle in $\mathbb{R}^3$, with an exceptional straight line, image of the unique S^3 great circle passing through the projection pole.

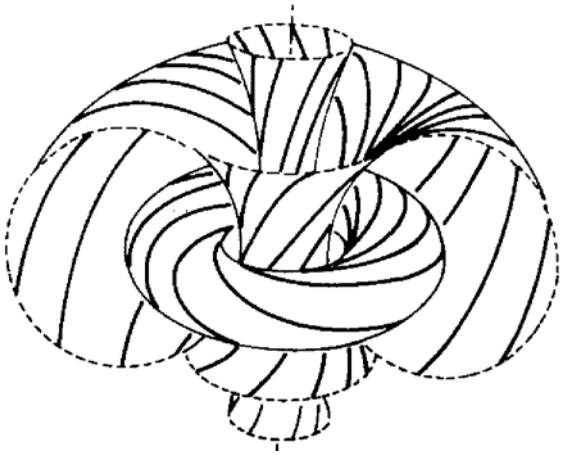

Fig. 8.1. S^3 Hopf fibration after a stereographic map onto $\mathbb{R}^3$. Circular S^1 fibers are mapped onto circles in $\mathbb{R}^3$, except the exceptional fiber through the projection pole, which is mapped onto a vertical straight line. fibers can be grouped into a continuous family of nested tori, three of which are shown here

8.3 Two Qubits, Entanglement, and the S^7 Hopf Fibration

8.3.1 The Two-Qubit Hilbert Space

We now proceed one step further and investigate pure states for two qubits. The Hilbert space $\mathcal{E}$ for the compound system is the tensor product of the individual Hilbert spaces $\mathcal{E}_1 \otimes \mathcal{E}_2$, with a direct product basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. A two-qubit state reads

$$|\Psi\rangle = \alpha\,|00\rangle + \beta\,|01\rangle + \gamma\,|10\rangle + \delta\,|11\rangle \tag{8.4}$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{C}$, and $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$.

$|\Psi\rangle$ is said "separable" if it can be written as a simple product of individual kets belonging to $\mathcal{E}_1$ and $\mathcal{E}_2$ separately, a definition that translates into the well-known following condition: $\alpha\delta = \beta\gamma$. A generic state is not separable and is said to be "entangled." The $|\Psi\rangle$ normalization condition $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$ identifies $\mathcal{E}$ to the seven-dimensional sphere S^7, embedded in $\mathbb{R}^8$. It was therefore tempting to see whether the known S^7 Hopf fibration (with fibers S^3 and base S^4) can play any role in the Hilbert space description. This is the case indeed, as we have shown recently [3]. Let us summarize the main results, keeping in mind that some notations have been changed as compared to this latter reference.

8.3.2 The S^7 Hopf Fibration

One follows the same line as in the S^3 case, but using quaternions instead of complex numbers (see Appendix). We write

$$q_1 = \alpha + \beta\mathbf{j},\ q_2 = \gamma + \delta\mathbf{j}, \qquad q_1, q_2 \in \mathbb{Q}, \tag{8.5}$$

and a point (representing the state $|\Psi\rangle$) on the unit radius S^7 as a pair of quaternions (q_1, q_2) satisfying $|q_1|^2 + |q_2|^2 = 1$. The Hopf map from S^7 to the base S^4 is the composition of a map h_1 from S^7 to $\mathbb{R}^4$ $(+\infty)$, followed by an inverse stereographic map h_2 from $\mathbb{R}^4$ to S^4.

$$\begin{aligned} h_1 : \begin{matrix} S^7 \\ (q_1, q_2) \end{matrix} & \begin{matrix} \longrightarrow \mathbb{R}^4 + \{\infty\} \\ \longrightarrow Q = \overline{q_1 q_2^{-1}} \end{matrix} \qquad q_1, q_2 \in \mathbb{Q}, \\ h_2 : \begin{matrix} \mathbb{R}^4 + \{\infty\} \\ Q \end{matrix} & \begin{matrix} \longrightarrow S^4 \\ \longrightarrow M(x_l) \end{matrix} \qquad \sum_{l=0}^{l=4} x_l^2 = 1. \end{aligned} \tag{8.6}$$

The base space S^4 is not embedded in S^7: the fibration is again not trivial. The fiber is a unit S^3 sphere as can be seen easily by stating that the S^7 points (q_1, q_2) and $(q_1 q, q_2 q)$, with q a unit quaternion (geometrically an S^3 sphere), are mapped onto the same Q value.

The h_1 map leads to

$$Q = \overline{q_1 q_2^{-1}} = \frac{1}{\sin^2(\theta/2)} \left[\overline{(\alpha + \beta\mathbf{j})(\overline{\gamma} - \delta\mathbf{j})}\right] = \frac{1}{\sin^2(\theta/2)} (C_1 + C_2\mathbf{j}) \quad (8.7)$$

with $\sin(\theta/2) = |q_2|$, $C_1 = \left(\overline{\alpha}\gamma + \overline{\beta}\delta\right)$, $C_2 = (\alpha\delta - \beta\gamma)$, and $C_1, C_2 \in \mathbb{C}$.

We face here the first striking result: *the Hopf map is entanglement sensitive*! Indeed, nonentangled states satisfy $\alpha\delta = \beta\gamma$ and therefore map onto the subset of pure complex numbers in the quaternion field (both completed by ∞ when the Q denominator vanishes). Geometrically, this means that nonentangled states map from S^7 onto a two-dimensional planar subspace of the target space $\mathbb{R}^4$.

The second map h_2 sends states onto points on S^4, with coordinates x_l, with l running from 0 to 4. With the inverse stereographic pole located on the S^4 "north pole" ($x_0 = +1$), and the target space $\mathbb{R}^4$ cutting S^4 along the equator, we get the following coordinate expressions

$$\begin{aligned} x_0 &= \cos\theta = |q_1|^2 - |q_2|^2 , \\ x_1 &= \sin\theta\, S(Q') = 2\,\mathrm{Re}\left(\overline{\alpha}\gamma + \overline{\beta}\delta\right), \\ x_2 &= \sin\theta\, V_{\mathbf{i}}(Q') = 2\,\mathrm{Im}\left(\overline{\alpha}\gamma + \overline{\beta}\delta\right), \\ x_3 &= \sin\theta\, V_{\mathbf{j}}(Q') = 2\,\mathrm{Re}\,(\alpha\delta - \beta\gamma), \\ x_4 &= \sin\theta\, V_{\mathbf{k}}(Q') = 2\,\mathrm{Im}\,(\alpha\delta - \beta\gamma). \end{aligned} \quad (8.8)$$

Q' is the normalized image of the h_1 map ($Q' = \tan(\theta/2)\, Q$), $S(Q')$ and $V_{\mathbf{i},\mathbf{j},\mathbf{k}}(Q')$ being the scalar and vectorial parts of the quaternion Q', respectively (see Appendix). As for the standard Bloch sphere case, the x_l coordinates are also expectation values of simple operators in the two-qubit state. An obvious one is x_0, which corresponds to $\langle \sigma_z \otimes Id \rangle_\Psi$. The next two coordinates are also easily recovered as

$$\begin{aligned} x_1 &= 2\,\mathrm{Re}\left(\overline{\alpha}\gamma + \overline{\beta}\delta\right) = \langle \sigma_x \otimes Id \rangle_\Psi , \\ x_2 &= 2\,\mathrm{Im}\left(\overline{\alpha}\gamma + \overline{\beta}\delta\right) = \langle \sigma_y \otimes Id \rangle_\Psi . \end{aligned} \quad (8.9)$$

The remaining two coordinates, x_3 and x_4, are also expectation values of an operator acting on $\mathcal{E}$, but in a more subtle way. Define $\mathbf{J}$ as the (antilinear) "conjugator," an operator that takes the complex conjugate of all complex numbers involved in an expression (here acting on the left in the scalar product given later). Then form the antilinear operator $\mathbf{E}$ (for "entanglor"): $\mathbf{E} = -\mathbf{J}(\sigma_y \otimes \sigma_y)$. One finds

$$\begin{aligned} x_3 &= \mathrm{Re}\,\langle \mathbf{E} \rangle_\Psi , \\ x_4 &= \mathrm{Im}\,\langle \mathbf{E} \rangle_\Psi . \end{aligned}$$

Note that $\langle \mathbf{E} \rangle_\Psi$ vanishes for nonentangled states and takes its maximal norm (equal to 1) for maximally entangled states. Such an operator, which

is nothing but the time reversal operator for two spins $\frac{1}{2}$entanglement [5], through a quantity called the "concurrence" c, which corresponds here to $c = 2\,|C_2|$.

8.3.3 Generalized Bloch Sphere for the Two-Qubit Case

Let us first inverse the Hopf map and get the general expression for the set of states (an S^3 sphere in S^7), which is sent to Q by the h_1 map. A generic such state, noted Ψ_Q, reads (given as a pair of quaternions)

$$\Psi_Q = (\cos(\theta/2)\ q\,, \sin(\theta/2)\, Q'\, q) \tag{8.10}$$

with q a unit quaternion spanning the S^3 fiber. Note that we could also write Ψ_Q in a way that recalls the standard spinor notation (but here with quaternionic instead of complex components):

$$\Psi_Q = (\cos(\theta/2) \exp(-\varphi \mathbf{t}/2)\ q\,, \sin(\theta/2) \exp(\varphi \mathbf{t}/2)\ q), \tag{8.11}$$

where $\cos\varphi = x_1/\sin\theta = S(Q')$, and $\mathbf{t}$ is the following unit pure imaginary quaternion:

$$\mathbf{t} = (\mathbf{V_i}(Q')\mathbf{i} + \mathbf{V_j}(Q')\mathbf{j} + \mathbf{V_k}(Q')\mathbf{k}) \,/\, \sin\varphi. \tag{8.12}$$

In order to compare with the generic expression (8.4), we aim to write Ψ_Q as a quadruplet of complex numbers. For that purpose, we express the two-unit quaternions q and Q' in terms of pairs of complex numbers, $q = a + b\mathbf{j}$ (with $|a|^2 + |b|^2 = 1$), and $Q' = u + v\mathbf{j}$ (with $|u|^2 + |v|^2 = 1$), and eventually get:

$$\Psi_Q = (\cos(\theta/2)\, a\,, \cos(\theta/2)\, b, \sin(\theta/2)\,(ua - v\bar{b}), \sin(\theta/2)\,(ub + v\bar{a})). \tag{8.13}$$

In this above expression, θ, u, and v correspond to the base space part of the fibration. Furthermore, we can relate u and v to already known quantities. Indeed

$$u = (x_1 + \mathbf{i}x_2)/\sin\theta = \langle(\sigma_x + \mathbf{i}\sigma_y) \otimes Id\rangle_{\Psi} \,/\, \sin\theta.$$

In addition, the state global phase indeterminacy allows to take v a real. More precisely

$$v = c/\sin\theta\ \text{, where } c \text{ is the above-mentioned concurrence.}$$

Let us now describe the two extreme cases of separable and maximally entangled states.

Separable States

In the nonentangled case, we have seen earlier that Q is a complex number, $\mathbf{t} = \mathbf{i}$, and therefore $u = \exp \mathbf{i}\varphi$ and $v = 0$. This expression simplifies to

$$\Psi_Q = (\cos(\theta/2)\ a, \cos(\theta/2)\, b, \sin(\theta/2)\ a \exp \mathbf{i}\varphi, \sin(\theta/2)\, b \exp \mathbf{i}\varphi). \quad (8.14)$$

Up to a global rescaling by $\exp(-\mathbf{i}\varphi/2)$, one gets the following ket $|\Psi_Q\rangle$:

$$\begin{aligned} |\Psi_Q\rangle = {} & (\cos(\theta/2) \exp(-\mathbf{i}\varphi/2)\, |0\rangle_1 \\ & + \sin(\theta/2) \exp(\mathbf{i}\varphi/2)\, |1\rangle_1\ ) \otimes (\, a\, |0\rangle_2 + b\, |1\rangle_2)\,. \end{aligned} \quad (8.15)$$

The projective Hilbert space for two nonentangled qubits is known to be the product of two 2-dimensional spheres $S_1^2 \times S_2^2$, each sphere being the Bloch sphere associated with the given qubit. This property is clearly displayed here. The unit S^4 base space reduces to a unit S^2 sphere (since $x_3 = x_4 = 0$), which is nothing but the Bloch sphere for the first qubit. The second qubit Bloch sphere is then recovered from the fiber, spanned by $q = a + b\mathbf{j}$. Indeed, we can iterate the fibration process on the S^3 fiber itself and get the (Hopf fibration base)–(Bloch sphere) coordinates for this two-level system. It is now easy to show that this new S^2 base is the second qubit Bloch sphere.

In summary, for nonentangled qubits, the S^7 Hopf fibration, with base S^4 and fiber S^3, simplifies to the simple product of an S^2 subsphere of the base (the first qubit Bloch sphere) by a second S^2 (the second qubit Bloch sphere) obtained as the base of an S^3 Hopf fibration applied to the fiber itself. Let us stress that this last iterated fibration is necessary to take into account the global phase of the two-qubit system.

The fact that these two S^2 spheres play a symmetrical role (although one is related to the base and the other to the fiber) can be understood in the following way. We grouped together α and β on one hand, and γ and δ on the other, to form the quaternions q_1 and q_2, and then define the Hopf map h_1 as the ratio of these two quaternions (plus a complex conjugation). Had we grouped α and γ, and β and δ, to form two new quaternions, and used the same definition for the Hopf map, we would also get an S^7 Hopf fibration, but differently oriented. As an exercise we compute the base and fiber coordinates in that case. The net effect is to interchange the role of the two qubits: the second qubit Bloch sphere is now part of the S^4 base, while the first qubit Bloch sphere is obtained from the S^3 fiber.

Maximally Entangled States

Let us now focus on maximally entangled states (MES). They correspond to the complex number C_2 having maximal norm 1/2 (unit concurrence). This in turn implies that the Hopf map base reduces to a unit circle in the

plane (x_3, x_4), parametrized by the unit complex number $2C_2$. The projective Hilbert space for these MES is known to be S^3/Z_2, an S^3 sphere with identified opposite points [6] (this is linked to the fact that all MES can be related by a local operation on one subsystem, since $S^3/Z_2 = SO(3)$). In order to recover this result in the present framework, one can follow the trajectory of a representative point on the base and on the fiber while the state is multiplied by an overall phase $\exp(\mathrm{i}\omega)$. The expression for $C_2\,(= \alpha\delta - \beta\gamma)$ shows that the point on the base turns by twice the angle ω. Only when $\omega = \pi$ does the corresponding state belong to the same fiber (e.g., maps onto the same value on the base). The fact that the fiber is an S^3 sphere, and this two-to-one correspondence between the fiber and the base under a global phase change, explains the S^3/Z_2 topology for the MES projective Hilbert space. Let us now give a more explicit proof of that result.

MES correspond to $\theta = \pi/2$ and maximal concurrence ($c = 1$), which leads to $u = 0$ and $v = 1$. Ψ_{MES} therefore read, from expression (8.13):

$$\Psi_{\mathrm{MES}} = \frac{1}{\sqrt{2}}(a, b\,, -\overline{b}, \overline{a}). \tag{8.16}$$

The latter expression (8.16) for maximally entangled states is rather interesting in that it directly shows the S^3/Z_2 topology for the MES projective Hilbert space. Indeed, the MES set corresponds to pairs (a, b), which as a whole cover a unit radius S^3 sphere. Now, looking at the quadruplet expression (8.16), opposite points (a, b) and $(-a, -b)$ on S^3 clearly correspond to the same state Ψ_{MES} (up to a global phase). Opposite points on S^3 have therefore to be identified, leading to the S^3/Z_2 ($\equiv \mathrm{SO}(3)$) structure.

This one-to-one correspondence between MES and three-dimensional rotation matrices has recently led to propose using the former in an "applied topology" experiment [7]: to verify experimentally the well-known subtle topology of the (two-fold connected) SO(3) group. The latter property is evidenced by constructing the two inequivalent family of closed paths in the geometrical manifold representing this group. This is done by choosing sequences of unitary operations on the MES two-qubit states. The nonequivalence between the two paths is manifested by a π topological phase shift, which should result from an adequate interference experiment (a twin photon experiment has been proposed, but other two-qubit states could be used).

A Generalization of the Bloch Sphere Representation

We are now led to consider a generalization of the Bloch sphere for the two-qubit projective Hilbert space. Clearly, the present Hopf fibration description suggests a splitting of the representation space in a product of base and fibers subspaces. Of the base space S^4, we propose to keep only the first three coordinates

$$(x_0, x_1, x_2) = \left(\langle \sigma_z \otimes Id\rangle_\Psi\,, \langle \sigma_x \otimes Id\rangle_\Psi\,, \langle \sigma_y \otimes Id\rangle_\Psi\right). \tag{8.17}$$

All states map inside a standard ball B^3 of radius 1, where the set of separable states forms the S^2 boundary (the usual first qubit Bloch sphere), and the center corresponds to maximally entangled states. Concentric spherical shells around the center correspond to states of equal concurrence c (maximal at the center, zero on the surface), the radius of the spherical shell being equal to $\sqrt{1-c^2}$. The idea of slicing the two-qubit Hilbert space into manifolds of equal concurrence is not new [6,8]. What is nice here is that, under the Hopf map (and a projection onto the three-dimensional subspace of the base spanned by the first three coordinates), these manifolds transforms into concentric S^2 shells that fill the unit ball.

To each point (x_0, x_1, x_2), it corresponds to an S^3/Z_2 manifold, spanned by the couple (a, b), as seen clearly from relation (8.13), with an added identification of (a, b) and $(-a, -b)$. The natural generalization of the Bloch sphere for two qubits is therefore a product of two B^3 balls. The first one, spanned by the triple (x_0, x_1, x_2), has just been described as containing the partial Bloch sphere for one of the two qubits, with its set of concentric iso-concurrence spheres. The second one corresponds to the standard representation of SO(3) by a $\overleftrightarrow{B^3}$ ball of radius π, the double arrow sign recalling that opposite points on the boundary S^2 sphere have to be identified. This picture is valid for all states except the separable ones, for which the fiber derived $\overleftrightarrow{B^3}$ space reduces to an S^2 sphere (the second qubit partial Bloch sphere).

Instead of mapping the continuous set of S^2 spheres onto the filled ball B^3 in the space spanned by (x_0, x_1, x_2), this nice (but singular) foliation of the two-qubit projective Hilbert space (here the complex projective space CP^3) can also be pictured as a concurrence segment (between 0 and 1) with the corresponding submanifolds. The subspace of vanishing concurrence has an $S^2 \times S^2$ structure, while that of maximal ($c = 1$) concurrence corresponds to SO(3). Submanifolds of intermediate concurrence have the structure of a direct product $S^2 \times \mathrm{SO}(3)$, the sphere S^2 having radius $\sqrt{1-c^2}$. This illustrated in Fig. 8.2.

Examples Back to the B^3 picture, let us for instance go through the states along a simple B^3 ray, from the point $(0, 1, 0)$, image of the states Ψ_S such that $\langle \sigma_x \otimes Id \rangle_\Psi = 1$, to the B^3 center $(0, 0, 0)$, image the maximally entangled

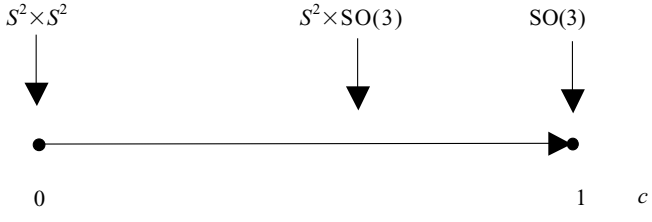

Fig. 8.2. Foliation of the two-qubit Hilbert space with respect to state entanglement

states $\Psi_{\rm MES}$. Q' reduces to a unit circle in the plane spanned by (x_1, x_3), and these states are therefore parametrized by a single angle ϵ in the interval $[0, \pi/2]$,

$$\Psi_\epsilon = \frac{1}{\sqrt{2}}(\, q\,, \exp{(\epsilon \mathbf{j})}\;\; q), \tag{8.18}$$

or, written as a quadruplet of complex numbers,

$$\Psi_\epsilon = \frac{1}{\sqrt{2}}(a\,, b, a\,\cos\epsilon - \overline{b}\,\sin\epsilon, b\cos\epsilon + \overline{a}\sin\epsilon). \tag{8.19}$$

For $\epsilon = 0$, one gets

$$\Psi_{\epsilon=0} = \Psi_S = \frac{1}{\sqrt{2}}\,(a, b, a, b) \;\text{ and therefore} \tag{8.20}$$

$$|\Psi_S\rangle = \frac{1}{\sqrt{2}}\,(|0\rangle_1 + |1\rangle_1) \otimes (a\,|0\rangle_2 + b\,|1\rangle_2)\,,$$

as expected for the set of separable states, which are eigenstates of $\sigma_x \otimes Id$ (with eigenvalue $+1$).

For $\epsilon = \pi/2$, this set of maximally entangled states, as given by relation (8.16), is recovered. Intermediate values of ϵ correspond to less entangled states, whose concurrence read $c = \sin\epsilon$, as can be easily found from (8.19). Expression (8.19) also proves that the set of such states describes an S^3/Z_2 manifold.

Similar analyses can be done for any path inside the B^3 ball. A second very simple example is provided by the path from $(1, 0, 0)$ to $(0, 0, 0)$. In that case the states, again parametrized by an angle ϵ, read

$$\Psi_\epsilon = \left(\cos\frac{\epsilon}{2}a\,, \cos\frac{\epsilon}{2}b, -\sin\frac{\epsilon}{2}\overline{b}, \sin\frac{\epsilon}{2}\overline{a}\right).$$

Relation with the Bloch ball representation for mixed states

The Bloch ball single qubit mixed state representation had been recalled earlier. In that case, the center of the Bloch ball corresponds to maximally mixed states. The reader should not be surprised to find here (in the two-qubit case) a second unit radius ball, with maximally entangled states now at the center. It corresponds to a known relation between partially traced two-qubit pure states and one-qubit mixed state. Indeed the partially traced density matrix ρ_1 is simply written in terms of C_1 and C_2 derived from the S^7 Hopf map:

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 1 + x_0 & x_1 - \mathbf{i}\,x_2 \\ x_1 + \mathbf{i}\,x_2 & 1 - x_0 \end{pmatrix} = \begin{pmatrix} |q_1|^2 & \overline{C_1} \\ C_1 & |q_2|^2 \end{pmatrix} \tag{8.21}$$

with unit trace and $\det\rho_1 = |C_2|^2$. The partial ρ_1 represents a pure state density matrix whenever C_2 vanishes (the separable case) and allows for a

unit Bloch sphere (that associated to the first qubit). It corresponds to a mixed state density matrix as soon as $|C_2| > 0$ (and an entangled state for the two-qubit state). The other partially traced density matrix ρ_2 is related to the other S^7 Hopf fibration that was discussed earlier.

8.4 Three Qubits and the S^{15} Hopf Fibration

8.4.1 Three Qubits

The Hilbert space $\mathcal{E}$ for the compound system is the tensor product of the individual Hilbert spaces $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3$, with a direct product basis

$$\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, |100\rangle, |101\rangle, |110\rangle, |111\rangle\},$$

which can be written $\{|l\rangle, l = 0, \ldots, 7\}$. A three-qubit state reads

$$|\Psi\rangle = \sum_{l=0}^{7} t_l |l\rangle \quad \text{with} \quad t_l \in \mathbb{C}, \text{ and } \sum |t_l|^2 = 1.$$

The $|\Psi\rangle$ normalization condition identifies $\mathcal{E}$ to the 15-dimensional sphere S^{15}, embedded in $\mathbb{R}^{16}$. This suggests looking at how far the third Hopf fibration (that of S^{15}, with base S^8 and fibres S^7) can be helpful for describing the three-qubit Hilbert space geometry.

8.4.2 The S^{15} Hopf Fibration

One proceeds along the same line as for the previous S^3 and S^7 cases, but using now octonions (see Appendix). We write

$$a = a' + a''\mathbf{e},\ b = b' + b''\mathbf{e}, \qquad a, b \in \mathbb{O}, \text{ and } a', a'', b', b'' \in \mathbb{Q}, \tag{8.23}$$

and a point (representing the state $|\Psi\rangle$) on the unit radius S^{15} as a pair of octonions (a, b) satisfying $|a|^2 + |b|^2 = 1$. But to get a Hopf map of physical interest, with coordinates simply related to interesting observable expectation values, one needs to define a slightly tricky relation between $|\Psi\rangle$ and the octonions pair (a, b), as follows:

$$\begin{aligned} a &= (t_0 + t_1\mathbf{j}, t_2 + \mathbf{j}t_3) = (t_0 + t_1\mathbf{j}, t_2 + \overline{t_3}\mathbf{j}) = (a', a''), \\ b &= (t_4 + t_5\mathbf{j}, t_6 + \mathbf{j}t_7) = (t_4 + t_5\mathbf{j}, t_6 + \overline{t_7}\mathbf{j}) = (b', b''). \end{aligned} \tag{8.24}$$

The Hopf map from S^{15} to the base S^8 is the composition of a map h_1 from S^{15} to $\mathbb{R}^8$ $(+\infty)$, followed by an inverse stereographic map h_2 from

$\mathbb{R}^8$ to S^8

$$h_1 : \begin{array}{c} S^{15} \\ (a,b) \end{array} \begin{array}{l} \longrightarrow \mathbb{R}^8 + \{\infty\} \\ \longrightarrow P = \overline{ab^{-1}} \end{array} \qquad a, b \in \mathbb{O},$$

$$h_2 : \begin{array}{c} \mathbb{R}^8 + \{\infty\} \\ P \end{array} \begin{array}{l} \longrightarrow \quad S^8 \\ \longrightarrow M(x_l) \end{array} \qquad \sum_{l=0}^{l=8} x_l^2 = 1. \tag{8.25}$$

The base space S^8 is not embedded in S^{15}: the fibration is again not trivial.

The fiber is a unit S^7 sphere, the proof of which is more tricky (and not given here) than in the lower dimension case. The h_1 map leads to

$$P = \overline{ab^{-1}} = \frac{1}{\sin^2 \theta/2} \left(Q_1 + Q_2 \mathbf{e}\right) \tag{8.26}$$

with $\sin \theta/2 = |b|$, $Q_1 = (b'\overline{a'} + \overline{a'}b'')$, $Q_2 = (-a''b' + b''a')$ and $Q_1, Q_2 \in \mathbb{Q}$.

Athough this is not at first sight evident, the Hopf map is still entanglement sensitive in that case. To show this, it is instructive to first express Q_1 and Q_2 in terms of the t_l components read out from (8.24)

$$Q_1 = \left(\overline{t_0}t_4 + \overline{t_1}t_5 + \overline{t_2}t_6 + \overline{t_3}t_7\right) + \left(t_0t_5 - t_1t_4 + \overline{t_2t_7} - \overline{t_3t_6}\right) \mathbf{j},$$
$$Q_2 = \left(t_0t_6 + t_2t_4 + \overline{t_3t_5} - \overline{t_1t_7}\right) + \left(t_1t_6 - t_2t_5 + \overline{t_0t_7} - \overline{t_3t_4}\right) \mathbf{j}.$$

Let us introduce the generalized complex concurrence terms $T_{ij,kl} = t_it_j - t_kt_l$. They allow to write in a synthetic form the coordinates on the base S^8. The second map h_2 sends states onto points on S^8, with coordinates x_l, with l running from 0 to 8. With the inverse stereographic pole located on the S^8 "north pole" $(x_0 = +1)$, and the target space $\mathbb{R}^8$ cutting S^8 along the equator, we get the following coordinate expressions:

$$x_0 = \cos\theta = |a|^2 - |b|^2 = \langle \sigma_z \otimes Id \otimes Id \rangle_\Psi \,, \tag{8.27}$$
$$x_1 + \mathbf{i}\, x_2 = 2\left(\overline{t_0}t_4 + \overline{t_1}t_5 + \overline{t_2}t_6 + \overline{t_3}t_7\right) = \left\langle \left(\sigma_x + \mathbf{i}\,\sigma_y\right)_1 \otimes Id \otimes Id \right\rangle_\Psi ,$$
$$x_3 + \mathbf{i}\, x_4 = 2\left(T_{05,14} + \overline{T_{27,36}}\right),$$
$$x_5 + \mathbf{i}\, x_6 = 2\left(T_{06,24} + \overline{T_{35,17}}\right),$$
$$x_7 + \mathbf{i}\, x_8 = 2\left(T_{16,25} + \overline{T_{07,34}}\right). \tag{8.28}$$

A lengthy, but trivial, computation allows to verify that the base S^8 has unit radius.

8.4.3 Discussion

It is easy to show that three-qubit states such that the first qubit is separated from the other two map onto a point such that $x_j = 0$ for $j = 4, 5, 6, 7, 8$.

One way to show this is to realize that, in a multiqubit state, a given qubit is separated from the others when its partial Bloch sphere has a unit radius. The first qubit partial Bloch sphere is spanned here by the triplet (x_0, x_1, x_2). A second proof consists in writing down the separability algebraic conditions [9]. In the present case, the latter imply that the above generalized concurrences vanish (in fact the six vanishing conditions only rely on three independent conditions). Going back to the above definition of the h_1 map, this means that in this case, the Hopf map carries an octonion couple onto a pure complex number P. Therefore, as for two qubit and S^7, **the** S^{15} *Hopf fibration is also entanglement sensitive for three qubits*! This result has been independently derived by Bernevig and Chen [10].

However, one should notice an important difference between the two-qubit and three-qubit cases. In the two-qubit case, the S^7 Hopf fibration has allowed us to foliate the projective Hilbert space with respect to state entanglement, the latter, measured by the concurrence, being simply related to the norm of the restriction of the base point to the subspace spanned by the pair (x_3, x_4). Since the base space S^4 has unit radius, the entanglement is therefore simply related to the radius the first qubit partial Bloch sphere, spanned by the triplet (x_0, x_1, x_2). The first-to-second qubit distinction (base-fiber in the fibration) does not matter here (to define the foliation) since the two partial Bloch sphere radii are equal.

The S^{15} Hopf fibration is clearly sensitive to the entanglement of one qubit (the first qubit in the present case) with respect to the other two qubits: the first qubit partial Bloch sphere radius is still read out from the norm of the restriction of the base point to the subspace spanned by the triplet (x_0, x_1, x_2). However, the latter does not tell the whole story in terms of state entanglement. Though one may know how far qubit 1 is entangled with the remaining two, one does not yet know whether the second (or third) qubit is or not separated. This prevents from building a foliation driven by a single entanglement parameter. One possibility for such a foliation unique parameter would be to use the 3-tangle [11]. But it does not distinguish among separable and entangled W states. An alternative entanglement parameter has been suggested by Bernevig and Chen [10] (see also [12] for a related measure): to recover the symmetry between the three-qubit, an average over the three-qubit partial Bloch sphere radii is used.

The solution to the foliation problem might be to use three (instead of one) parameters, by considering three distinct S^{15} Hopf fibrations, such that each of the three-qubit partial Bloch spheres is singled out by the first three coordinates on the base. Said more simply, one may try to describe the Hilbert space geometry in a space spanned by the three partial Bloch sphere radii (r_1, r_2, r_3). Since each radius belongs to the interval $[0, 1]$, the whole representation lies inside a unit cube. When one radius equals 1, one qubit is separated from the other two, and the other two radii are equal. We then see that the set under consideration cuts three square faces of the cube (such

that $r_j = 1$), along a diagonal. These diagonals correspond to the product of a sphere S^2 (the Bloch sphere of the separated qubit), and a copy of this foliation for the remaining two qubits. Note that the latter foliation was introduced above with respect to state entanglement, as measured by the concurrence c, instead of the partial Bloch sphere radius. To get a fully equivalent picture, one should therefore transform the original coordinate system from c to $r = \sqrt{1 - c^2}$.

8.5 Conclusions

Our main goal in this chapter was to provide a geometrical representation of the two-qubit and three-qubit Hilbert space pure states. We have shown that, as for the one-qubit Bloch sphere relation to the S^3 Hopf fibration, the more complex S^7 and S^{15} Hopf fibrations also play a natural role in that case. Note that, as already mentioned in [3], in the three Hopf fibration sequence (S^{15}, S^7, S^3), the fiber in the larger dimensional space is the full space in next case. This is illustrated in Fig. 8.3.

This offers the possibility of further nesting the fibrations, a possibility that was already used in the present analysis of two-qubit separable case. This also applies to the three-qubit case: whenever the first qubit is separated from the other two, the base reduces to the first qubit Bloch sphere, and the S^7 fiber is precisely the Hilbert space of the remaining two qubits.

In the two-qubit case, this approach has in particular allowed for a complete description of the pure state projective Hilbert space, in terms of an entanglement driven foliation, with well-characterized leaves. This goal has not yet been achieved in the three-qubit case, although a plausible track has been proposed here.

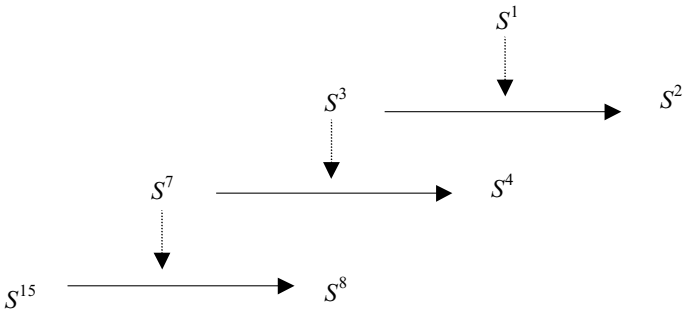

Fig. 8.3. Nesting of the three Hopf fibrations

Appendix: Quaternions and Octonions

Quaternions

Quaternions are usually presented with the imaginary units $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ in the form:

$$q = x_0 + x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}, \qquad x_0, x_1, x_2, x_3 \in \mathbb{R} \quad \text{with } \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1,$$

the latter "Hamilton" relations defining the quaternion multiplication rules that are noncommutative. They can also be defined equivalently, using the complex numbers $c_1 = x_0 + x_1\mathbf{i}$ and $c_2 = x_2 + x_3\mathbf{i}$, in the form $q = c_1 + c_2\mathbf{j}$, or equivalently as an ordered pair of complex numbers satisfying

$$\begin{aligned}(c_1, c_2) + (d_1, d_2) &= (c_1 + d_1, c_2 + d_2)\,,\\ (c_1, c_2)\,(d_1, d_2) &= \left(c_1 d_1 - c_2\overline{d_2}, c_1 d_2 + c_2\overline{d_1}\right).\end{aligned}$$

The conjugate of a quaternion q is $\overline{q} = x_0 - x_1\mathbf{i} - x_2\mathbf{j} - x_3\mathbf{k} = \overline{c_1} - c_2\mathbf{j}$ and its squared norm reads $N_q^2 = q\overline{q}$.

Another way in which q can be written is as a scalar part $S(q)$ and a vectorial part $\mathbf{V}(q)$:

$$q = S(q) + \mathbf{V}(q),\ S(q) = x_0,\ \mathbf{V}(q) = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$$

with the relations

$$S(q) = \frac{1}{2}(q + \overline{q}),\ \mathbf{V}(q) = \frac{1}{2}(q - \overline{q}).$$

A quaternion is said to be real if $\mathbf{V}(q) = 0$, and pure imaginary if $S(q) = 0$. We also write $\mathbf{V}_{\mathbf{i},\mathbf{j},\mathbf{k}}(q)$ for the component of $\mathbf{V}(q)$ along $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Finally, and as for complex numbers, a quaternion can be noted in an exponential form as

$$q = |q| \exp \varphi\mathbf{t} = |q|\,(\cos\varphi + \sin\varphi\mathbf{t})\,,$$

where $\mathbf{t}$ is a unit pure imaginary quaternion. When $\mathbf{t}$=$\mathbf{i}$, usual complex numbers are recovered. Note that quaternion multiplication is non-commutative so that

$$\exp\varphi\mathbf{t} \exp\lambda\mathbf{u} = \exp(\varphi\mathbf{t} + \lambda\mathbf{u})$$

only if $\mathbf{t} = \mathbf{u}$.

Octonions

An octonion a can be defined by introducing a new unit $\mathbf{e}$ (different from the preceding unit quaternions $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$, and such that $\mathbf{e}^2 = -1$), and pairs of

quaternions a', a'':

$$\begin{aligned} a &= a' + a''\mathbf{e} \\ ab &= (a' + a''\mathbf{e})\,(b' + b''\mathbf{e}) \\ &= \left(a'b' - \overline{b''}a''\right) + \left(b''a' + a''\overline{b'}\right)\mathbf{e}. \end{aligned}$$

It is helpful to write any octonion a as

$$a = \sum_{l=0}^{l=7} u_l \mathbf{e}_l, \text{ with } \mathbf{e}_0 = 1,\ \mathbf{e}_1 = \mathbf{i},\ \mathbf{e}_2 = \mathbf{j},\ \mathbf{e}_3 = \mathbf{k},\ \mathbf{e}_4 = \mathbf{e},\ \mathbf{e}_5 = \mathbf{ie},$$
$$\mathbf{e}_6 = \mathbf{je}, \mathbf{e}_7 = \mathbf{ke}$$

with the following multiplication table:

$$\begin{pmatrix} \mathbf{e}_0 & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 & \mathbf{e}_5 & \mathbf{e}_6 & \mathbf{e}_7 \\ \mathbf{e}_1 & -\mathbf{e}_0 & \mathbf{e}_3 & -\mathbf{e}_2 & \mathbf{e}_5 & -\mathbf{e}_4 & -\mathbf{e}_7 & \mathbf{e}_6 \\ \mathbf{e}_2 & -\mathbf{e}_3 & -\mathbf{e}_0 & \mathbf{e}_1 & \mathbf{e}_6 & \mathbf{e}_7 & -\mathbf{e}_4 & -\mathbf{e}_5 \\ \mathbf{e}_3 & \mathbf{e}_2 & -\mathbf{e}_1 & -\mathbf{e}_0 & \mathbf{e}_7 & -\mathbf{e}_6 & \mathbf{e}_5 & -\mathbf{e}_4 \\ \mathbf{e}_4 & -\mathbf{e}_5 & -\mathbf{e}_6 & -\mathbf{e}_7 & -\mathbf{e}_0 & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \mathbf{e}_5 & \mathbf{e}_4 & -\mathbf{e}_7 & \mathbf{e}_6 & -\mathbf{e}_1 & -\mathbf{e}_0 & -\mathbf{e}_3 & \mathbf{e}_2 \\ \mathbf{e}_6 & \mathbf{e}_7 & \mathbf{e}_4 & -\mathbf{e}_5 & -\mathbf{e}_2 & \mathbf{e}_3 & -\mathbf{e}_0 & -\mathbf{e}_1 \\ \mathbf{e}_7 & -\mathbf{e}_6 & \mathbf{e}_5 & \mathbf{e}_4 & -\mathbf{e}_3 & -\mathbf{e}_2 & \mathbf{e}_1 & -\mathbf{e}_0 \end{pmatrix}.$$

Note that other multiplication tables could be defined (see [13]). In analogy with the quaternions scalar and vectorial parts, we can also write a as

$$a = S(a) + \mathbf{V}(a) \text{ with } S(a) = u_0,\ \mathbf{V}(a) = \sum_{l=1}^{l=7} u_l \mathbf{e}_l.$$

The conjugate of an octonion a is $\overline{a} = S(a) - V(a) = \overline{a'} - a''\mathbf{e}$ and its squared norm reads $N_a^2 = a\overline{a}$.

A (very) important difference between quaternions and octonions is that the latter, besides being noncommutative, are also nonassociative.

Acknowledgements and Comments. It is a pleasure to acknowledge several discussions with Rossen Dandoloff, Karol Zyczkowski, and Perola Milman.

The main content of this paper was presented at the Dresden "Topology in Condensed Matter Physics" colloquium in June 2002. Also included is a reference to a paper by B.A. Bernevig and H.D. Chen, who have independently done the analysis of the three-qubit case along parallel lines. It is a pleasure to acknowledge several discussions with B.A. Bernevig in this context. Information on more recent work with P. Milman on the SO(3) topological phase has been added. Also not presented in Dresden was this above suggestion to use, for the three-qubit case, a three-parameter foliation based on the three partial Bloch sphere radii.

References

1. Bouwmeester, D., Eckert, A., Zeilinger, A.: The Physics of Quantum Information. Springer-Verlag (2000)
2. Urbanke, H.: Am. J. Phys. **59**, 53 (1991)
3. Mosseri, R., Dandoloff, R.: J. Phys. A: Math. Gen. **34**, 10243–10252 (2001)
4. Hopf, H.: Math. Ann. **104**, 637–665 (1931)
5. Wootters, W.K.: Phys. Rev. Lett. **80**, 2245 (1998)
6. Kus, M., Zyczkowski, K.: Phys. Rev A. **63**, 032307 (2001) and references herein
7. Milman, P., Mosseri, R.: Phys. Rev. Lett. **90**, 230403 (2003)
8. Sinolecka, M., Zyczkowski, K., Kus, M.: Act. Phys. Pol. B **33**, 2081 (2002)
9. Jorrand, P., Mhalla, M.: quant-ph/0209125
10. Bernevig, B.A., Chen, H-D.: J. Phys. A: Math. Gen. **36**, 8325 (2003)
11. Coffman, V., Kundu, J., Wootters, W.: Phys. Rev. A **61**, 52306 (2000)
12. Meyer, D.A., Wallach, N.R.: J. of Math. Phys. **43**, 4273 (2002)
13. Baez, J.C.: The Octonions. math.RA/0105155

9

Defects, Surface Anchoring, and Three-Dimensional Director Fields in the Lamellar Structure of Cholesteric Liquid Crystals as Studied by Fluorescence Confocal Polarizing Microscopy

I.I. Smalyukh and O.D. Lavrentovich

Summary. We study three-dimensional director patterns and defects in cholesteric liquid crystals using fluorescence confocal polarizing microscopy for nondestructive three-dimensional imaging. We establish the detailed director fields of dislocations, their kinks, as well as the disclination nodes and oily streaks. Weak surface anchoring of the director at the bounding plates causes attraction, while strong anchoring causes repulsion between the edge dislocation and a boundary. We use a coarse-grained models of cholesteric elasticity and surface anchoring potential to explain the experimental results. We study dynamics of defects, and glide and climb of dislocations. We also consider the Peierls–Nabarro mechanisms hindering glide of dislocations across the cholesteric layers. The static and dynamic properties of defects in cholesteric lamellae can be used as a model for understanding similar phenomena in other lamellar systems, such as diblock copolymers and SmA liquid crystals.

9.1 Introduction

A variety of condensed matter systems, such as cholesteric and smectic A (SmA) liquid crystals (LCs), lamellar phases of diblock copolymers, stripe magnetic domains in ferrimagnetic thin films and ferrofluid stripes in applied magnetic fields have common features of one-dimensional periodic modulations in density or magnetization [1–5]. These media have layered structure and one-dimensional positional order, which result in similarity of elastic and surface properties, some common types and topology of defects.

Cholesteric liquid crystals (CLCs) have a twisted ground state with helical configuration of the director $\mathbf{n}$, which specifies the average local orientation of molecules. External fields and surface interactions can easily deform the ideal helicoidal configuration. When the spatial scale of distortions is much larger than the cholesteric pitch p (corresponding to the director twist by 2π), elastic properties of CLCs are similar to the properties of smectic phases with a one-dimensional periodic structure [1,6]. If a CLC is confined within a

finite volume, the equilibrium structure is determined by bulk elasticity and boundary effects, such as surface tension and surface anchoring. Very often, the boundary conditions are satisfied by the appearance of large-scale defects such as focal conic domains, curvature walls, dislocations, etc. [2–5].

Although the cholesteric lamellae are formally similar to the smectic lamellae as long as the coarse-grained model [1, 6] is valid, one should keep in mind the important difference in characteristic length scales, which are determined by the layer spacing, usually $d = p/2 \sim (0.1 - 10)\,\mu\text{m}$ in cholesterics and $d \sim (1 - 10)\,\text{nm}$ in SmA. The spacing d defines the core size of elementary dislocations. The defects in cholesterics can be studied by optical means. A related issue is that the difference in scales between cholesterics and smectics leads to the difference in surface properties. In lamellar phases, the tilted orientation of layers at the bounding plates usually requires a local "melting" of the layers. The corresponding "intrinsic" contribution to the surface anchoring energy is roughly $\sim K_1/d$, where K_1 is the splay elastic constant [7]. The surface anchoring energy measured for SmA $\sim\left(10^{-2} - 10^{-3}\right)\,\text{J/m}^2$ [7] is therefore much larger than that for CLCs $\sim\left(10^{-4} - 10^{-6}\right)\,\text{J/m}^2$ with pitch $d = p/2$ in the range $(0.1 - 10)\ \mu\text{m}$ [8–10].

In this chapter we review our recent studies of the three-dimensional director fields and defects in cholesterics. Many of the discussed results, such as interaction of a dislocation and a bounding surface, layers profile in the vicinity of dislocation, glide and climb of the defects, are of universal importance for the study of lamellar systems. CLC is a good model as its layer spacing can be easily adjusted to be in the micron range and thus can be studied with optical microscopy techniques, such as polarizing microscopy (PM) and fluorescence confocal polarizing microscopy (FCPM) [11–13]. At the same time, we also discuss the features of director fields and defects that might be unique for CLCs, such as splitting of dislocation cores into disclinations with a nonsingular core, geometry of director patterns associated with dislocation kinks, etc. The main distinctive features of the director structures and defects in CLCs as compared to other lamellar systems reflect their "nematic-like" local molecular arrangement. Therefore, it is useful to describe the elastic properties of CLCs with both the Frank–Oseen theory and the Lubensky–de Gennes coarse-grained theory [1, 6].

Experiments are performed for well-equilibrated flat and wedge samples and for transient textures. FCPM allows us to reconstruct three-dimensional director structures associated with dislocations, their turns, nodes, and kinks. The strength of anchoring at the bounding plates determines the structure of samples, including the structure and behavior of edge dislocations. We use the coarse-grained model of cholesteric anchoring [12] in order to explain the experimental results. We calculate the elastic energy of a dislocation away from the core, estimate the energy of the core split into disclinations of different types, study the effect of finite sample thickness on the dislocations energy, and calculate the Peach–Koehler elastic forces that occur when a dislocation is shifted from its equilibrium position. We determine the relative stability of

dislocations with Burgers vectors $b = p/2$ and $b = p$, explain the difference in the shape of kinks for these defects. Finally, we consider the Peierls–Nabarro mechanisms hindering glide of dislocations across the cholesteric layers. We show that because of the split disclination character of the core, glide is difficult as compared to climb, especially for $b = p$ dislocations.

The review is organized as follows. In Sect. 9.2 we briefly describe the experimental technique and materials that were used for nondestructive visualization of three-dimensional director fields and defects in this work. In Sect. 9.3 we give an introduction to classification and some basic properties of the directors and defects in CLCs. The elasticity and surface anchoring properties of cholesteric lamellae are reviewed in Sect. 9.4. We discuss the dislocation–interface interaction as well as describe the three-dimensional director structures and defects in the weakly-anchored CLCs in Sect. 9.5. Section 9.6 is devoted to the equilibrium structures and defects in the strongly anchored cholesteric wedge cells. The metastable structures, oily streaks, and nodes of line defects are studied in Sect. 9.7. In Sect. 9.8 we discuss the issues of dynamics of the defects, their glide and climb, dislocation kinks. Finally, the results are summarized and the conclusions are drawn in Sect. 9.9.

9.2 Experimental Methods and Materials

9.2.1 Materials and Cell Preparation

To form a CLC, we mixed a nematic LC material with a chiral dopant CB15 (purchased from EM Industries). Depending on the need, we used either nematic LC with positive (ZLI3412) or negative (ZLI2806) dielectric anisotropy. Both the nematic hosts have small birefringence: $\Delta n \approx 0.045$ for ZLI2806 and $\Delta n \approx 0.07$ for ZLI3412. Low birefringence mitigates two problems that one encounters in FCPM imaging of CLCs: (1) relative defocusing of extraordinary versus ordinary modes [11], and (2) the Mauguin effect (polarization of light follows the twisted director) [13]. For the FCPM observations, the cholesteric mixture is doped with a very small amount (0.01 wt%) of fluorescent dye n,n'-bis(2,5-di-*tert*-butylphenyl)-3,4,9,10-perylenedicarboximide (BTBP), purchased from Molecular Probes [11, 13].

We used constant-thickness flat cells and wedge cells with a small dihedron angle $\alpha < 2^\circ$ and of maximum thickness 100 µm. The thickness h of cells was measured by interference method. The dihedron angle α of wedge cells was measured using reflected laser beam for empty cells. To minimize spherical aberrations in FCPM observations with immersion oil objectives, we used glass substrates of thickness 0.15 mm with the refractive index 1.52.

Two different types of surface treatment provided either strong or weak anchoring. The strong planar alignment was set by a unidirectionally rubbed polyimide PI-2555 (HD MicroSystems) film spin coated over the ITO layers. The director is in the plane of the substrate with a possible small pretilt angle

$\lesssim 1^\circ$. The polar anchoring coefficient W_p that characterizes the work needed to deviate $\mathbf{n}$ from its equilibrium surface orientation in the plane perpendicular to the substrate was measured for ZLI3412/PI2555 interface by retardation versus voltage technique [14] to be $W_\mathrm{p} = (4 \pm 1) \times 10^{-4}\mathrm{J/m^2}$. W_p is expected to be of the order of $10^{-4}\mathrm{J/m^2}$ also for ZLI2806 as this is a typical value measured for PI2555 in contact with a variety of nematic mixtures with a positive dielectric anisotropy, see [14]. The azimuthal anchoring coefficient is smaller, $W_\mathrm{a} \sim 10^{-5}\mathrm{J/m^2}$ [15].

For weak anchoring, we used spin-coated films of polyisoprene, known to produce memory-free tangentially degenerated alignment. The residual azimuthal anchoring is $W_\mathrm{a} \sim 10^{-10}\mathrm{J/m^2}$ [16] and polar anchoring $W_\mathrm{p} = (0.7 \pm 0.6) \times 10^{-4}\mathrm{J/m^2}$ or less as determined for cells prealigned by a magnetic field [12]. We stress here that both W_p and W_a refer to the untwisted, pure nematic state. As we see later, the anchoring properties of the twisted nematic state might be rather different from the untwisted one.

9.2.2 Fluorescence Confocal Polarizing Microscopy

The FCPM technique links the director orientation to the intensity of measured fluorescent signal [11, 13]. Compared to the well–known fluorescence confocal microscopy (FCM), FCPM has two distinctive features: (a) the specimen is stained with *anisometric* dye molecules (such as BTBP) that follow the director orientation; (b) the excitation light is *polarized*, usually linearly.

The FCPM set up was assembled on the basis of Olympus Fluoview BX-50 reflective-mode confocal microscope. The excitation beam (488 nm, Ar laser) is focused by an objective into a small ($<1\,\mu\mathrm{m}^3$) volume within the CLC slab. The fluorescent light from this volume is detected by a photomultiplier tube in the spectral region 510–550 nm. A pinhole discriminates against the regions above and below the selected volume. The pinhole size is adjusted according to magnification and numerical aperture NA of the objective. The polarizer $\mathbf{P}$ determines polarization of both the excitation beam $\mathbf{P}_\mathrm{e}$, and the detected fluorescent light $\mathbf{P}_\mathrm{f}$: $\mathbf{P}_\mathrm{f} \| \mathbf{P}_\mathrm{e} \| \mathbf{P}$. The beam power is small, $<1\,\mathrm{mW}$, to avoid light-induced reorientation of the dye-doped LC [11, 13].

For BTBP dye, the fluorescence lifetime $\tau_\mathrm{F} = (3.7 - 3.9)\,\mathrm{ns}$ [17] is smaller than the characteristic time of rotational diffusion $\tau_\mathrm{D} \sim 10\,\mathrm{ns}$, and dye orientations during absorption and emission can be assumed to be close to each other [13]. The FCPM signal, resulting from a sequence of absorption and emission, strongly depends on the angle β between the transition dipole (parallel to the local director $\mathbf{n}$ in our system) and $\mathbf{P}$: $I \sim \cos^4\beta$ [11, 13], as both absorption and emission follow the dependency $\cos^2\beta$. The strongest FCPM signal corresponds to $\mathbf{n} \| \mathbf{P}$ ($\beta = 0$), and sharply decreases when β becomes nonzero [17]. Here I is the intensity of fluorescence signal in the FCPM textures of the director $\mathbf{n}$.

The focused beam scans the sample at a fixed depth $-h/2 \le z \le h/2$, creating a "horizontal" optical slice $I(x, y)$. The scanning is repeated at different

depths, to obtain a stack of optical slices, i.e., a three-dimensional pattern $I(x, y, z)$, related to the three-dimensional pattern $\mathbf{n}(x, y, z)$ through the dependence $I \sim \cos^4 \beta$. Note that the correspondence $I(x, y, z) \longleftrightarrow \mathbf{n}(x, y, z)$ is not unique when only one fixed direction of linear polarization $\mathbf{P}$ is used, as the angular parameter β defines a cone of directions. To avoid ambiguity, we use different directions of the linear polarization $\mathbf{P}$ (e.g., $\mathbf{P} = (P, 0, 0)$ and $\mathbf{P} = (0, P, 0)$) and also a circularly polarized light. In the latter case, only the changes in the vertical component n_z of the director are detected; n_x and n_y are not discriminated against each other. Using the computer software, the three-dimensional pattern $I(x, y, z)$ can be cut by vertical planes such as (xz) and (yz) to visualize $\mathbf{n}$ across the sample.

9.3 Directors and Defects in Cholesteric Liquid Crystals

Here we briefly review some basic properties and classification of line defects in CLCs such as disclinations and dislocations that are studied in detail in Sects. 9.4 and 9.5 of the chapter. For a more detailed account of the classification of defects in cholesterics see [3].

Weak distortions (for which the degree of orientational order does not change) in CLCs are described as spatial changes of three mutually orthogonal directors. We use the nomenclature introduced by Kleman and Friedel [18], which is based on the notation λ for the local director $\mathbf{n}$, χ for the direction of the helical axis, and $\tau = \lambda \times \chi$. Correspondingly, the Kleman–Friedel classification distinguishes disclinations of the λ, τ, and χ-types. The λ and τ disclinations of strength $\pm 1/2$ are shown in Fig. 9.1a–d. In the λ disclinations, the *material* λ director field is nonsingular**,** while the *immaterial* τ and χ director fields are singular, Fig. 9.1a,d. In the τ-disclinations, the τ director field is nonsingular but λ and χ director fields are singular, Fig. 9.1b,c. In the χ-disclinations, the singularities are found in λ and τ director fields but not in the χ director field.

Elementary topological defects in the systems with broken translational symmetry are dislocations. Both edge and screw dislocations can be present in CLCs. The wedge χ-disclinations can also be treated as screw dislocations with the Burgers vector $b = -kp$. The equivalency of the two approaches in description of the defects is a result of the fact that the symmetry of $n\pi$-rotations around the χ-axis is the same as the symmetry of translations $n(p/2)\chi$ [3].

The core of an edge dislocation splits into a pair of two $\pm 1/2$ disclinations with π-rotations, Fig. 9.1e–h. Geometry implies the following different ways of splitting [18]: a $b = p/2$ dislocation splits into a pair of τ and λ disclinations, Fig. 9.1e,f; a $b = p$ dislocation splits into a $\tau\tau$ or $\lambda\lambda$ pair of disclinations, Fig. 9.1g,h. In principle, from the geometrical point of view, the $b = p/2$ dislocation could split into a $\tau^{-1/2}\lambda^{+1/2}$, Fig. 9.1e, or $\lambda^{-1/2}\tau^{+1/2}$, Fig. 9.1f, pairs of disclinations, whereas the $b = p$ line could split into a $\tau^{-1/2}\tau^{+1/2}$, Fig. 9.1g,

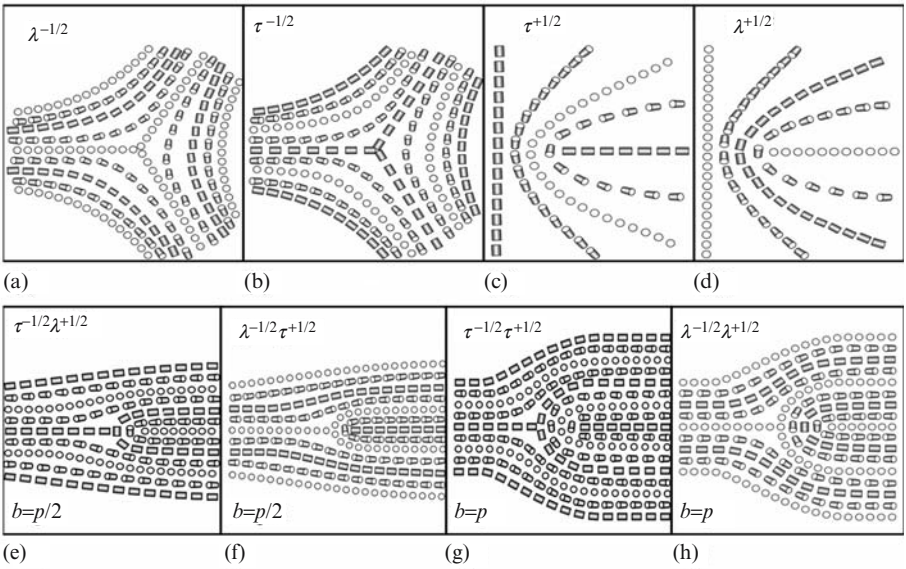

Fig. 9.1. Disclination lines of λ and τ type and splitting of the dislocation cores into different pairs of these disclinations: (**a**) $\lambda^{-1/2}$; (**b**) $\tau^{-1/2}$; (**c**) $\tau^{+1/2}$; (**d**) $\lambda^{+1/2}$; (**e**) $b = p/2$ dislocation with core split into the $\tau^{-1/2}\lambda^{+1/2}$; (**f**) $b = p/2$ dislocation with core split into the $\lambda^{-1/2}\tau^{+1/2}$; (**g**) $b = p$ dislocation with core split into the $\tau^{-1/2}\tau^{+1/2}$; (**h**) $b = p$ dislocation with core split into the $\lambda^{-1/2}\lambda^{+1/2}$ pair

pair or $\lambda^{-1/2}\lambda^{+1/2}$, Fig. 9.1h; the first symbol refers to the disclination in the part of sample with smaller number of layers. However, the $\tau^{-1/2}\tau^{+1/2}$ configuration is never observed as it implies a singular core for both disclinations.

The superscripts in notations such as $\tau^{-1/2}$ and $\lambda^{+1/2}$ correspond to the director winding number around the disclination. In terms of the Volterra process "–" sign corresponds to adding material between the lips separated by an angle $\boldsymbol{\pi}$, while "+" sign to material removal. It is important to notice that both edge dislocations and τ, λ-disclinations are parallel to the cholesteric layers, except near the kinks, which change the level of the edge dislocations along the helicoid axis [17, 19].

9.4 Elastic and Surface Properties of Cholesterics

The equilibrium director structure in the cholesteric LC can be found by minimizing the free energy functional F, which consists of elastic bulk energy and the surface anchoring energy:

$$F = \int f_{\mathrm{el}}\, \mathrm{d}v + \int f_{\mathrm{s}}\, \mathrm{d}S. \tag{9.1}$$

Here $f_{\rm el}$ is the elastic bulk energy density and $f_{\rm s}$ is the surface anchoring energy per unit area. In the case when the distortions occur on the scales much larger than the cholesteric pitch, $L \gg p$, one uses the coarse-grained models of elastic and anchoring energies (similar to the case of SmA). In the opposite limit, when $L \lessapprox p$, the nematic-like description of elasticity and anchoring is more appropriate. However, in many situations of practical interest, distortions in both scales are present. For example, the dislocations in the cholesteric lamellae cause the layer displacements on the scales $L \gg p$ as well as the director distortions and disclination defects within the dislocation core on the scales $L \lessapprox p$ [17]. In Sect. 9.4 we briefly review these different approaches in the description of elasticity and surface anchoring properties of CLCs.

9.4.1 Elasticity of Cholesteric Liquid Crystals

Frank–Oseen Theory

The elasticity of cholesterics with $L \lessapprox p$ can be effectively described by the well-known Frank–Oseen functional [1] for the director $\mathbf{n}$:

$$f_{\rm el}^{\rm FO} = \frac{K_{11}}{2}\,(\operatorname{div}\mathbf{n})^2 + \frac{K_{22}}{2}\,(\mathbf{n}\cdot\operatorname{curl}\mathbf{n}+q_0)^2 + \frac{K_{33}}{2}\,(\mathbf{n}\times\operatorname{curl}\mathbf{n})^2 - K_{24}\operatorname{div}(\mathbf{n}\cdot\operatorname{div}\mathbf{n} + \mathbf{n}\times\operatorname{curl}\mathbf{n}), \tag{9.2}$$

where K_{11}, K_{22}, and K_{33} are the Frank elastic constants for splay, twist, and bend deformations, respectively; q_0 is a coefficient that describes chirality and determines the equilibrium pitch p of CLC: $q_0 = 2\pi/p$. The divergence (saddle-splay) elastic K_{24} term has to be taken into account in the case of finite anchoring, and when the topological defects are present.

Lubensky–de Gennes Coarse-Grained Elastic Theory

When the scales of the layer deformations or radii of curvature are much larger than the cholesteric pitch, $L \gg p$, the elastic properties of CLCs are similar to those of the SmA LCs. The free energy density of the deformations described using the director χ can be written in the following form [4]:

$$f_{\rm el}^{\rm CG} = \frac{1}{2}K_1(\operatorname{div}\chi)^2 + \frac{1}{2}K_3(\chi\times\operatorname{curl}\chi)^2 + \bar{K}\operatorname{div}(\chi\operatorname{div}\chi+\chi\times\operatorname{curl}\chi)+\frac{1}{2}B\gamma^2, \tag{9.3}$$

where $\gamma = (d - d_0)/d_0$ is the relative difference of the actual d and the equilibrium $d_0 = p/2$ layer spacing. Note that here we use the constants K_1 and K_3 with one-digit subscript to describe the deformations in the χ-director; the elastic constants describing deformations of the $\mathbf{n}$-director have two-digit

subscripts, (9.2). Expression (9.3) includes the curvature K-terms associated with the splay (K_1), bend (K_3), and saddle-splay ($\bar{K}$) deformations of the normal χ to the layers and the B-contribution describing dilation/compressions of the layers. Both the B-term and the K_3-term in (9.3) describe deformations associated with layers compression and are of the same nature. Usually, the K_3-term can be neglected since its contribution to $f_{\rm el}^{\rm CG}$ is negligible as compared to the B-term [4] (unless the scales of layers distortions or radii of curvature are comparable to the layer spacing). For example, the K_3-term can be neglected when describing the bulk elastic properties of CLC samples with thickness $h \gg p$. However, the contribution of the K_3-term is important when taking into account the surface anchoring effects as we see in Sect. 9.4.2.

The splay constant K_1 and the Young modulus B in (9.3) are related to the Frank moduli of twist (K_{22}) and bend (K_{33}) of the director $\mathbf{n}$ by Lubensky–de Gennes relationships [6]: $K_1 = 3K_{33}/8$; $B = K_{22}(\frac{2\pi}{p})^2$. The bend (from the point of view of the director χ) constant K_3 can be derived from the Kats–Lebedev theory [20]: $K_3 = \frac{K_{11}K_{33}}{2(K_{11}+K_{33})}$. The result is obtained by collecting the terms with $(\chi \times \operatorname{curl}\chi)^2$, $(\operatorname{div}\chi)^2$, etc., and neglecting the divergence (surface-like) terms in (5.1.27, 33) of [20]. It is interesting to note that the penetration length $\lambda_1 = \sqrt{K_1/B} = \frac{p}{2\pi}\sqrt{\frac{3K_{33}}{8K_{22}}} \sim p/6$ is smaller than $p/2$, the layer spacing.

When the K_3-term in (9.3) can be neglected, (9.3) can be rewritten in terms of the principal radii R_1 and R_2 characterizing the curvature of the cholesteric lamellae:

$$f_{\rm el}^{\rm CG} = \frac{K_1}{2}\left(\frac{1}{R_1}+\frac{1}{R_2}\right)^2 + \frac{\bar{K}}{2}\frac{1}{R_1R_2} + \frac{1}{2}B\gamma^2, \tag{9.4}$$

as the splay $(\frac{1}{R_1}+\frac{1}{R_2})$ and the saddle-splay $\frac{1}{R_1R_2}$ terms in (9.4) are related to the derivatives of the director χ in the following way:

$$\operatorname{div}\chi = \pm\left(\frac{1}{R_1}+\frac{1}{R_2}\right);\ \operatorname{div}(\chi\operatorname{div}\chi + \chi\times\operatorname{curl}\chi) = \frac{2}{R_1R_2}. \tag{9.5}$$

The saddle-splay $\bar{K}$-term plays no role in case the layers distortions have two-dimensional character and therefore can be neglected in many situations of practical interest discussed in this chapter.

The free energy functional in the form (9.3) or (9.4) is useful to describe strong layers distortions [8]. When the layers departures from the ground flat state are small, (9.3) can be rewritten in terms of the layers displacement field $u(x,y,z)$ [4]. The χ-director can be described in terms of the layer displacement $u(x,y,z)$ [4]:

$$\chi = \pm\left\{-\frac{\partial u}{\partial x}\left(1+\frac{\partial u}{\partial z}\right), -\frac{\partial u}{\partial y}\left(1+\frac{\partial u}{\partial z}\right), 1-\frac{1}{2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]\right\}. \tag{9.6}$$

The free energy density of layers displacements $u(x,z)$ in the case of the two-dimensional layer distortions is of the form

$$f_{\rm nl} = \frac{1}{2}K_1\left(\frac{\partial^2 u}{\partial x^2}\right)^2 + \frac{1}{2}B\left[\frac{\partial u}{\partial z} - \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2\right]^2 . \tag{9.7}$$

The contribution $\frac{1}{2}(\frac{\partial u}{\partial x})^2$ in the compressibility term in (9.7) makes the theory nonlinear. In the linear approximation,

$$f_l = \frac{1}{2}K_1\left(\frac{\partial^2 u}{\partial x^2}\right)^2 + \frac{1}{2}B\left(\frac{\partial u}{\partial z}\right)^2 . \tag{9.8}$$

It is convenient to use the coarse-grained elastic theory to describe the properties of layers patterns and the defects. However, the Frank–Oseen theory should be used in order to calculate the core energy of dislocation, or any director distortions at the scales smaller or comparable to cholesteric pitch.

Elasticity of Defect Structures

Using the nonlinear model, (9.7), Brener and Marchenko found the equilibrium displacement field around a straight edge dislocation of Burgers vector b in an infinite medium [21] (see also [22]):

$$u_{nl}(x,z) = 2\lambda_1 \ln\left\{1 + \frac{{\rm e}^{b/4\lambda_1} - 1}{2}\left[1 + {\rm erf}(\frac{x}{2\sqrt{\lambda_1 z}})\right]\right\}, \tag{9.9}$$

where erf $(\cdots)$ is the error function, defined as ${\rm erf}\,(t) = \frac{2}{\sqrt{\pi}}\int\limits_0^t \exp\left(-v^2\right){\rm d}v$; x and z are Cartesian coordinates in the plane perpendicular to the dislocation centered at $(0,0)$. In the limit $b \ll \lambda_1$, (9.9) reduces to the classical result of the linear theory [2]:

$$u_l(x,z) = \frac{b}{4}\left[1 + {\rm erf}\left(\frac{x}{2\sqrt{\lambda_1 z}}\right)\right]. \tag{9.10}$$

In CLCs, even the smallest value of the Burgers vector, $b = p/2$, is larger than the penetration length λ_1 and the nonlinear theory is better suited to describe the layers displacement [12, 23] (see also Sect. 9.5).

9.4.2 Surface Anchoring Energy

Anchoring effects in CLCs are difficult to describe analytically, see, e.g., [24]: near the substrate, the director field has to accommodate both the elastic torques setting the helicoidal twist and surface interactions that keep $\mathbf{n}$ along a specific "easy axis" (or axes) that corresponds to the minimum of surface anchoring potential.

Nematic-Like Rapini–Papoular Model

The surface anchoring energy W can be described as the work required to deviate the director $\mathbf{n}$ at the surface from the easy axis $\mathbf{e}$. The angular coordinates of the easy axis $\mathbf{e}$ are $(\frac{\pi}{2} - \beta, \varphi_e)$, where β is the so-called pretilt angle between $\mathbf{e}$ and the projection of $\mathbf{e}$ onto the substrate; φ_e is the azimuthal orientation of this projection. In a similar fashion, two angular coordinates $\left(\gamma = \frac{\pi}{2} - \theta, \varphi\right)$ specify the actual orientation of the director $\mathbf{n}$ (where γ is the angle between $\mathbf{n}$ and normal to the cell substrates, θ is the angle between $\mathbf{n}$ and the projection of $\mathbf{n}$ onto a substrate, and φ is the azimuthal orientation of this projection). Deviations from the easy axis that keep $\varphi = \varphi_e$ are described in terms of the polar anchoring coefficient W_p. Deviations that change φ are described with the azimuthal coefficient W_a. The definitions of the anchoring coefficients are usually based on the Rapini–Papoular model [1]. For zero pretilt angle $\beta = 0$, the surface anchoring energy per unit area reads [13]:

$$f_s = \frac{1}{2} W_p \sin^2 \theta + \frac{1}{2} W_a \cos^2\theta \sin^2 (\varphi - \varphi_e) . \tag{9.11}$$

W_p is usually 1–2 orders of magnitude stronger than W_a (as in the case of PI2555 [14]). Moreover, for the case of alignment layers of polyisoprene, the polar anchoring coefficient W_p is 5–6 orders of magnitude larger than W_a and the contribution of the azimuthal anchoring to the surface energy can be neglected: $f_s \approx \frac{1}{2} W_p \sin^2 \theta$.

Coarse-Grained Cholesteric Anchoring

In the coarse-grained model of cholesteric anchoring [12] the CLC is considered as a lamellar phase. The surface free energy is calculated per unit area for cholesteric "layers" making a small angle $\theta(z)$ with the substrate located at $z = 0$ [12]. We consider the case in which the azimuthal anchoring is vanishing, but the polar anchoring is nonzero.

A semi-infinite CLC is bounded by a plate at $z = 0$ that sets the helicoid axis χ parallel to the normal ν to plate. An external torque sets a nonzero angle $0 \leq \theta_\infty \leq \pi/2$ between χ and ν far away from the boundary, $\theta(z \to -\infty) = \theta_\infty$; as the layers approach the boundary, surface anchoring modifies θ, Fig. 9.2. The dependence $\theta(z)$ is determined by the anchoring potential experienced by χ at the surface and by the elastic energy of distortions in the bulk, Fig. 9.2a. The free energy density of the bulk deformations is:

$$f = \frac{1}{2} K_1 (\mathrm{div}\, \chi)^2 + \frac{1}{2} K_3 (\chi \times \mathrm{curl}\, \chi)^2 + \frac{1}{2} B \left(\frac{d - d_0}{d_0} \right)^2 , \tag{9.12}$$

since the saddle-splay term in (9.3) can be neglected for the two-dimensional layers displacements.

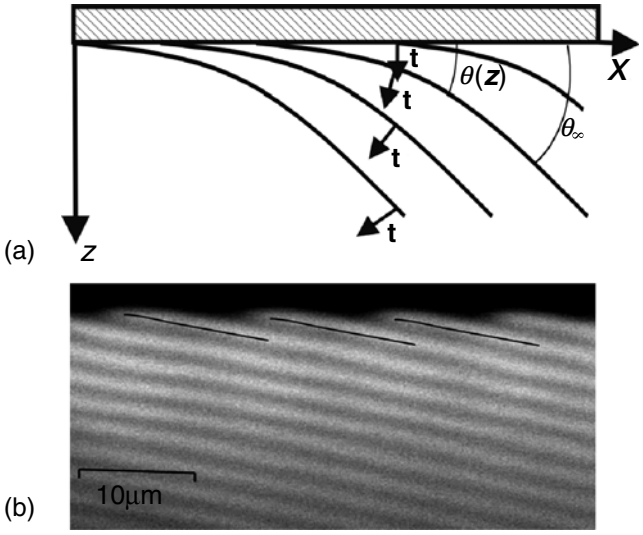

Fig. 9.2. The scheme (**a**) and the experimental FCPM texture (**b**) of the layer profile at a plate with weak anchoring

With $\chi = (\sin\theta, 0, \cos\theta)$, the free energy per unit area of the boundary is

$$W_\chi = \frac{1}{2}\int_0^\infty \left[K_1 \sin^2\theta \left(\frac{\partial\theta}{\partial z}\right)^2 + K_3 \cos^2\theta \left(\frac{\partial\theta}{\partial z}\right)^2 + B\left(\frac{\sin\theta}{\sin\theta_\infty} - 1\right)^2\right] \mathrm{d}z + \frac{1}{4} W_\mathrm{p} \sin^2\theta \,|_{z=0} \,. \tag{9.13}$$

The last term is calculated as follows. Director deviations from the equilibrium orientation $\mathbf{n} \perp \nu$ are described by the Rapini–Papoular anchoring potential: $W_\mathrm{n} = W_\mathrm{p} (\mathbf{n} \cdot \nu)^2 /2$. For CLC, W_n has to be averaged over director rotations, which yields $\langle W_\mathrm{n} \rangle = \frac{1}{4} W_\mathrm{p} \sin^2\theta \,|_{z=0}$. In the elastic part of W_χ, the term $K_1 \sin^2\theta (\frac{\partial\theta}{\partial z})^2$ can be neglected, as we assume $\theta \ll 1$, so that (9.13) simplifies to

$$W_\chi = \frac{1}{2}\int_0^\infty \left[K_3 \left(\frac{\partial\chi_x}{\partial z}\right)^2 + B\left(\frac{\chi_x}{\sin\theta_\infty} - 1\right)^2\right] \mathrm{d}z + \frac{1}{4} W_\mathrm{p} \chi_x^2 \,|_{z=0}, \tag{9.14}$$

where $\chi_x = \sin\theta$. We minimize W_χ in order to obtain the layers profile $\chi_x(z)$ near the boundary. The Euler–Lagrange equation reads:

$$\frac{\partial^2 \chi_x}{\partial z^2} = \frac{1}{\lambda_3^2 \sin\theta_\infty} \left(\frac{\chi_x}{\sin\theta_\infty} - 1\right). \tag{9.15}$$

where $\lambda_3 = \sqrt{K_3/B}$. The solution of this equation is:

$$\chi_x(z) = \sin\theta_\infty \left[1 - C\exp\left(-\frac{z}{\lambda_3 \sin\theta_\infty}\right)\right]. \tag{9.16}$$

By taking into account the boundary condition $\left[\frac{\partial\chi_x}{\partial z} - \frac{W_\mathrm{p}}{2K_3}\chi_x\right]_{z=0} = 0$, we find the constant $C = W_\mathrm{p}\sin\theta_\infty/(2B\lambda_3 + W_\mathrm{p}\sin\theta_\infty)$. Thus, we obtain the layers profile near the boundary and the coarse-grained anchoring potential $W_\chi(\theta_\infty)$ for χ:

$$\sin\theta(z) = \sin\theta_\infty \left[1 - \frac{W_\mathrm{p}\sin\theta_\infty}{2B\lambda_3 + W_\mathrm{p}\sin\theta_\infty}\exp\left(-\frac{z}{\lambda_3 \sin\theta_\infty}\right)\right], \tag{9.17}$$

$$W_\chi(\theta_\infty) = \frac{W_\mathrm{p} B\lambda_3 \sin^2\theta_\infty}{4B\lambda_3 + 2W_\mathrm{p}\sin\theta_\infty}. \tag{9.18}$$

The distortions decay exponentially within a subsurface slab of a small thickness $\lambda_3 \sin\theta_\infty$; outside it, $\theta \approx \theta_\infty$, Fig. 9.2b. Depending on θ_∞ and the material parameters, $W_\chi(\theta_\infty)$ might be approximated by either $\sim\sin^2\theta_\infty$ or $\sim\sin|\theta_\infty|$. The first form fits well to the experimental data on the layer undulations immediately above the threshold for Helfrich–Hurault instability [9], where $\theta_\infty \ll 1$, while the second one is better suited for large $|\theta_\infty|$ and W_p, as $W_\chi(\theta_\infty)$ becomes proportional to the number of layers crossing the boundary [8].

The coarse-grained model of CLC anchoring is valid for the interfaces with tangentially degenerate boundary conditions when $0 < W_\mathrm{p} < \infty$ and $W_\mathrm{a} = 0$. However, the model allows one to also understand the results obtained in the cells with weak azimuthal anchoring and the cells with strong uniaxial rubbing giving strong both azimuthal and polar anchoring (see Sect. 9.5).

9.5 Dislocation–Interface Interaction and Three-Dimensional Director Structures in the Weakly Anchored Cholesterics

9.5.1 Anchoring-Mediated Dislocation–Interface Interaction

Interaction between a dislocation and a bounding surface is an important problem in physics of soft matter systems, such as diblock copolymers, smectic, lamellar and CLC [25,26]. However, there are no direct experimental studies that would allow one to probe this interaction, as imaging of dislocations is difficult. We also study the dislocation–interface interaction employing the capability of FCPM to nondestructively visualize the three-dimensional director fields in the cholesteric lamellae. In our experiments the surface tension

effect is excluded as the cholesteric sample is confined between rigid boundaries. This allows us to probe the anchoring-mediated dislocation–interface interaction.

Experimental Observations

We used a wedge cell with a small dihedron angle $\alpha = 0.45°$. The CLC was confined between two glass substrates, one of which was coated by polyisoprene in order to obtain weak surface anchoring and the other, by rubbed layer of PI2555 for strong anchoring (see Sect. 9.2 for details). The FCPM of cholesteric wedges reveals that as the thickness of the wedge increases, an additional cholesteric twist by π is added simply by an insertion of an additional layer near the substrate with a weak anchoring rather than by a bulk edge dislocation, Figs. 9.3 and 9.4(d).

The edge dislocations can be found as transient features of, say, a filling process or a rapid temperature quench from the isotropic phase, especially in the thick part of the wedge, $\sim 4p$ and above. Their Burgers vector is always $b = p$ and never $b = p/2$, Fig. 9.4. The core of these transient dislocations $b = p$ is split into a $\lambda^{-1/2}\lambda^{+1/2}$ pair, Fig. 9.4e (see also Fig. 9.1h). Despite the fact that the $b = p$ dislocation by itself introduces a twist change by 2π, it always separates two Grandjean zones that differ only by one π rotation of $\mathbf{n}$. The fit is achieved by a removal of one π twist at the boundary with weak anchoring, Fig. 9.4. The dislocation slowly glides toward the softly anchored plate,

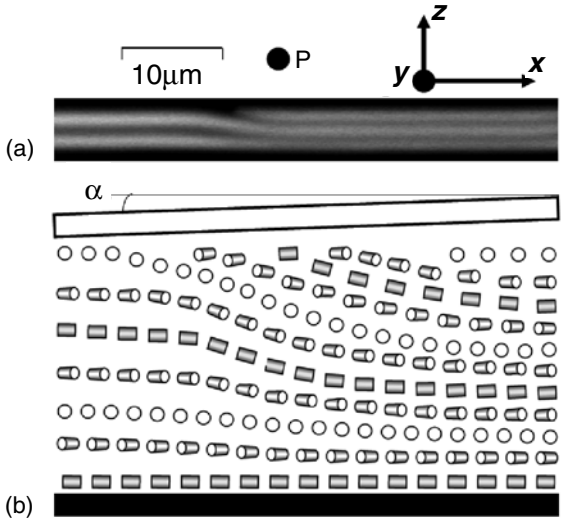

Fig. 9.3. FCPM vertical cross-sections (**a**) and director field (**b**) of the thin part of a cholesteric wedge cell ($\alpha = 0.45°$, local cell thickness $\sim 5\,\mu$m) with weak anchoring at the top plate and strong anchoring at the bottom plate

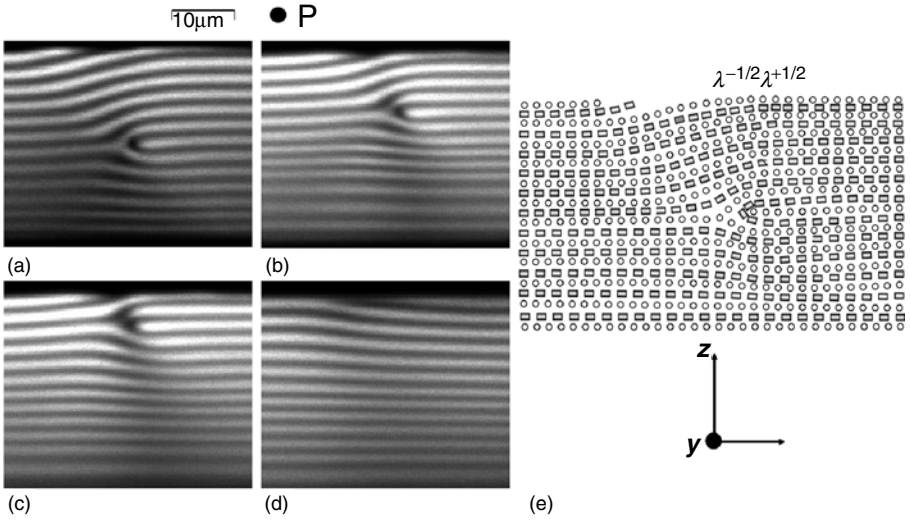

Fig. 9.4. FCPM textures (**a**–**d**) of a dislocation gliding toward the top plate with a weak anchoring and away from the bottom plate with a strong anchoring; local cell thickness ∼30 μm; (**e**) reconstructed director field of the vertical cross-section in part (**a**) showing a $b = p$ dislocation with the core split into the $\lambda^{-1/2}\lambda^{+1/2}$ pair

Fig. 9.4a–c and coalesces with the deserted layer, producing a surface structure with a layer insertion, $2\pi - \pi = \pi$, Fig. 9.4d. Dislocations do not glide as straight lines: the motion involves formation and propagation of kinks and is discussed in detail in Sect. 9.8. The experiments demonstrate that the interaction of an edge dislocation with interfaces is anchoring mediated and that the direction of glide (toward or away from the boundary) is determined by the strength of surface anchoring.

Theoretical Explanation

Recent theoretical models for an edge dislocation in a lamellar phase with a free boundary predict either attraction or repulsion, depending on the surface tension and elasticity of the lamellar phase, see review [26]. The existent theories [26] describe the surface–dislocation interaction in terms of the surface tension (the isotropic contribution to the surface energy) and neglecting the anchoring part, associated with a variation of layers tilt at the boundary. Qualitatively, a dislocation creates a step at the boundary thus increasing the surface area and the surface energy. If this increase is smaller than the energy of elastic distortions around the dislocation in bulk, the defect would be attracted to the surface. Although such an approach is definitely valid for many cases, including the free surfaces, generally, one also needs to consider the

surface anchoring term. For a rigid boundary, the interaction is mediated precisely by the anchoring effects: a dislocation approaching the boundary does not change the interfacial area but it does change the orientation of layers at the boundary, Fig. 9.4.

The interaction between the dislocation and the surface is determined by the free energy functional for the displacement field $u(x,z)$ of the cholesteric layers, written here in the linear approximation [26]:

$$F = \frac{1}{2}\int \left[K_1(\frac{\partial^2 u}{\partial x^2})^2 + B(\frac{\partial u}{\partial z})^2\right] \mathrm{d}z + W_\chi(\theta_\infty)\,. \tag{9.19}$$

Let us now estimate the changes in the elastic and anchoring energies as an edge dislocation is transferred from the depth $z = d''$ to $d' < d''$ in a semi-infinite sample; d' is larger than the typical core size of the dislocation $\sim p$; the coarse-grained model is not applicable when $d \lesssim p$. We use the linear approximation of the equilibrium displacement of layers around a dislocation in an infinite sample given by (9.10). Integrating the elastic energy density $f_{\mathrm{el}} = K_1(\frac{\partial^2 u}{\partial x^2})^2 + B(\frac{\partial u}{\partial z})^2$ over $-\infty < x < \infty$; $0 < z < d''$ and subtracting a similar quantity integrated over $0 < z < d'$, one concludes that the elastic energy (per unit length) *decreases*

$$F_{\mathrm{el}} = -\frac{K_1 b^2}{16\sqrt{2\pi d''}\lambda_1^{3/2}}\left(\sqrt{\frac{d''}{d'}} - 1\right). \tag{9.20}$$

On the other hand, the layers tilt $\theta_\infty \approx \partial u/\partial x \approx \frac{b}{4\sqrt{\pi\lambda_1 d}}\exp\left(-\frac{x^2}{4\lambda_1 d}\right)$ near the plate with tangentially degenerate anchoring increases. The associated total anchoring energy can be calculated as

$$F_{\mathrm{surf}} \approx \frac{1}{4}W_{\mathrm{p}}\int_{-\infty}^{\infty}\left(\theta_\infty^2\,|_{d=d'} - \theta_\infty^2\,|_{d=d''}\right)\mathrm{d}x\,. \tag{9.21}$$

We find that the surface anchoring energy

$$F_{\mathrm{surf}} \approx \frac{W_{\mathrm{p}} b^2}{32\sqrt{2\pi\lambda_1 d''}}\left(\sqrt{\frac{d''}{d'}} - 1\right) \tag{9.22}$$

increases when the dislocation is brought closer to the plate. Here we assume $W_{\mathrm{p}}\theta_\infty \ll 2B\lambda_3$, i.e., $W_\chi(\theta_\infty) \approx \frac{1}{4}W_{\mathrm{p}}\theta_\infty^2$ in (9.18). As easily seen,

$$F_{\mathrm{el}} + F_{\mathrm{surf}} \propto (K_1/\lambda_1 - W_p/2)\left(1 - \sqrt{d''/d'}\right) = 0$$

when $W_{\mathrm{p,crit}} = 2\sqrt{K_1 B}$; the dislocation–plate interaction vanishes. We remind the reader the definitions of two penetration lengths: $\lambda_1 = \sqrt{K_1/B}$ and $\lambda_3 = \sqrt{K_3/B}$. For $W_{\mathrm{p}} < 2\sqrt{K_1 B}$, the dislocation is attracted to the plate, and for $W_{\mathrm{p}} > 2\sqrt{K_1 B}$ it is repelled. The results for $W_\chi(\theta_\infty) \approx \frac{1}{4}W_{\mathrm{p}}\theta_\infty^2$ are similar to those in the theory of surface-tension mediated interaction; in the latter case the critical value of surface tension is $\sqrt{K_1 B}$ [26]. For a more accurate

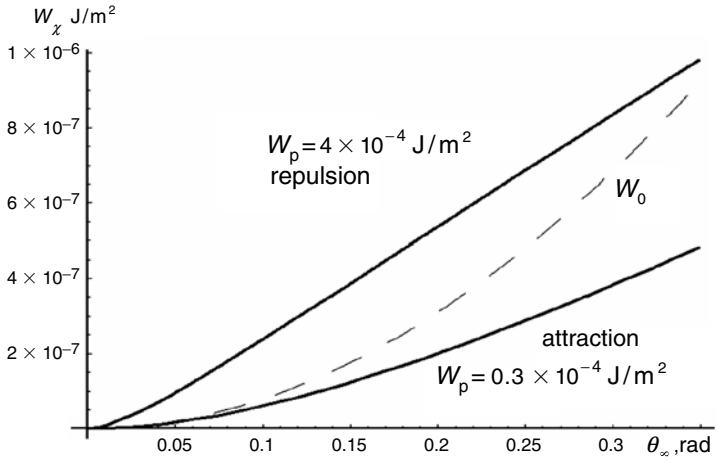

Fig. 9.5. Cholesteric anchoring as a function of the layers tilt for different polar anchoring coefficients W_p. The dependencies were calculated using (9.18) and elastic constants of ZLI3412.

description, especially when the dislocation approaches the plate so that the condition $W_\mathrm{p}\theta_\infty \ll 2B\lambda_3$ is no longer valid, one should use the complete form of (9.18).

The considerations given here are consistent with numerical analysis of (9.18), which also shows that the critical value is close to $W_{\mathrm{p,crit}} = 2\sqrt{K_1 B}$. We compare $W_\chi(\theta_\infty)$ to $W_0(\theta_\infty) = \sigma\theta_\infty^2/2$ with $\sigma = \sqrt{K_1 B}$; $W_0(\theta_\infty)$ corresponds to a "neutral" case of the theory [26], when the dislocation–boundary interaction vanishes. For strong anchoring, the interaction is repulsive, as $W_\chi(\theta_\infty)$ calculated with $W_\mathrm{p} = 4\times10^{-4}\mathrm{J/m}^2$ runs above $W_0(\theta_\infty)$, at least for $\theta_\infty < 0.4$, Fig. 9.5. In contrast, for weak anchoring, say, $W_\mathrm{p} = 0.3\times10^{-4}\mathrm{J/m}^2$, the calculated curve $W_\chi(\theta_\infty)$ is below the curve $W_0(\theta_\infty)$, although very close to it, meaning that the interaction is weakly attractive. For W_p larger than $0.3\times10^{-4}\mathrm{J/m}^2$, the $W_\chi(\theta_\infty)$ curve can cross $W_0(\theta_\infty)$.

Glide of the edge dislocation in Fig. 9.4 is assisted by repulsion from the substrate with strong anchoring and attraction to the weakly anchored interface (nearly zero interaction or weak repulsion is also possible at the top plate when $W_\mathrm{p} > 0.3\times10^{-4}\mathrm{J/m}^2$, at least for some values of θ_∞). The theoretical considerations provided here are in good agreement with the experimental results, see Fig. 9.4 and also Sect. 9.6.

9.5.2 Layers Profiles of Isolated Edge Dislocations

In flat cells with both substrates weakly anchored, transient dislocations are accompanied by two layers that interrupt at the surface, Fig. 9.6a. Importantly, distortion of layers around the dislocation in this sample is well fitted

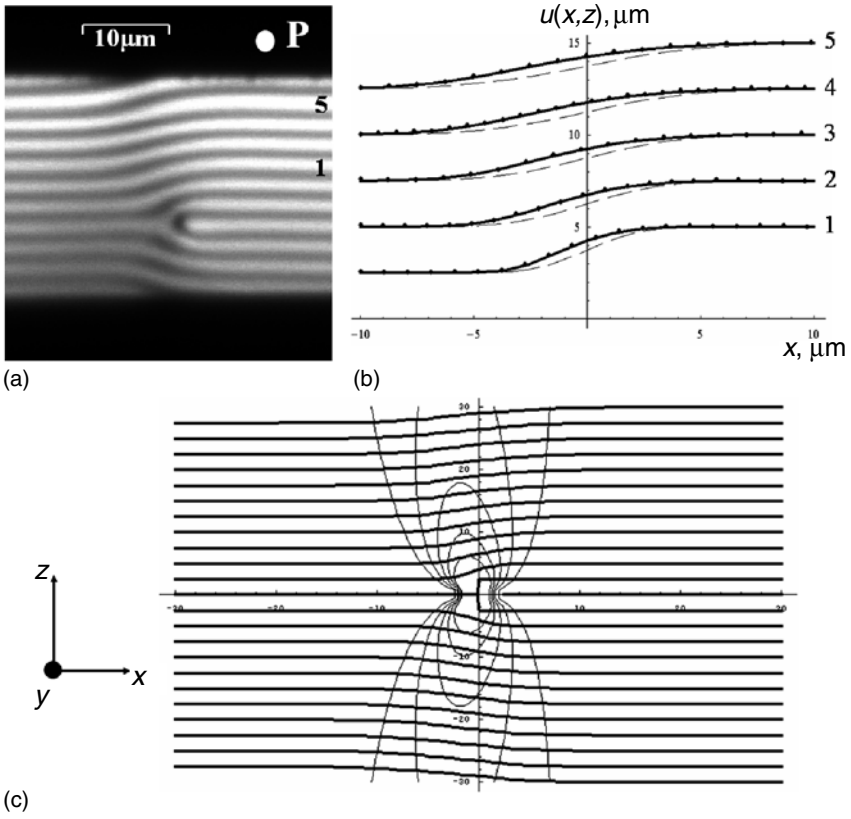

Fig. 9.6. Experimental versus theoretical layers profile of an isolated dislocation: a dislocation in a flat cholesteric cell with weak anchoring at both plates, cell thickness ∼30 μm (**a**); nonlinear theory for an infinite medium (*lines*, (9.9)) fits the experimental profile (*dots*) well while the linear theory (*dashed lines*, (9.10)) does not (**b**); layers displacements $u(x, z)$ (*thick lines*) and their tilt $\partial u/\partial x$ (*thin lines*) for a dislocation in an infinitely large lamellar system as calculated using (9.9) (**c**). Parameters used in (**b**,**c**): $K_1 = 3K_{33}/8 = 5.8\,\mathrm{pN}$; $B = K_{22}(2\pi/p)^2 = 10.6\,\mathrm{N/m^2}$; $p = 5\,\mu\mathrm{m}$

by the nonlinear Brener–Marchenko elastic theory of an isolated edge dislocation in an *infinite* medium [21, 23], Fig. 9.6b. The physical implication of this experimental result is that the interaction between the dislocation and the bounding substrates is close to 0 (neither attraction nor repulsion).

The basic feature of the experimental dislocation profile, Fig. 9.6a, is that the inflection points and the majority of changes of the layers displacement are in the region to the left from the dislocation, as predicted by the nonlinear theory [21]. The apparent asymmetry of the experimental layers displacements in the vicinity of the dislocation, Fig. 9.6a, resembles the asymmetry of the

layers displacements obtained with (9.9), Fig. 9.6c. Evidently, the nonlinear model of the dislocation profile, (9.9), fits the experimental profile much better than the linear model, (9.10), (see Fig. 9.6b and also [10,23] in which the layers profile of a dislocation was studied in the cholesteric finger texture). These results are in a good agreement with [10, 23] and allow one to conclude that the nonlinear elastic effects play an important role in determining of the layers displacements of edge dislocations.

In the flat cholesteric cells with "hybrid" boundary conditions (strong planar anchoring, rubbed PI2555, on one substrate and tangentially degenerate anchoring, polyisoprene, on another), the transient dislocations are also accompanied by two surface layers but both these layers are inserted in the same interface, the one with weak anchoring, Fig. 9.7. The Burgers vector of these defects is always $b = p$. In the flat cells with "hybrid" anchoring, in addition to the asymmetry of the layers profile in plane of the cell, there is also strong asymmetry of layers tilt above and below the dislocation. The layers at the interfaces with strong planar anchoring run parallel to the interfaces, whereas the tilt of layers at the interfaces with weak cholesteric anchoring is dictated by the presence of a dislocation in the bulk of LC. This result is natural, as the cholesteric anchoring W_χ at the two interfaces is almost an order of magnitude different (see Sect. 9.4). Importantly, the dislocations in the flat cells with strong, weak, and "hybrid" surface anchoring at the confining interfaces are transient defects and with time relax to the uniform planar cholesteric texture. In the flat cells with "hybrid" surface anchoring, the relaxation is always accomplished via dislocation glide to the substrate with weak anchoring. The dislocations glide via kinks, as discussed in detail in Sect. 9.8.

The experimental FCPM textures in Figs. 9.6 and 9.7 demonstrate that the layers profiles of the defects structures in cholesteric lamellae are determined by nonlinear elasticity of the layered medium and by the strength of surface anchoring at the confining interfaces. The same conclusion will follow from the studies of CLCs confined between two interfaces with strong anchoring, see Sect. 9.6.

9.6 The Equilibrium Defects and Structures in Strongly Anchored Cholesteric Wedges

The properties of dislocations in a confined CLCs can be conveniently studied in the cells with wedge geometry. As first observed by Grandjean [27] and later by Cano [28], a lattice of defect lines forms parallel to the edge of a cholesteric wedge cell. The lines separate different Grandjean zones, the regions of sample with different number of layers. In the Grandjean–Cano wedge with strong surface anchoring, the dislocations correspond to the equilibrium state because the interfaces with strong anchoring repel the defects, which are then stabilized in the LC bulk. The regular defect lattice is formed due to (a) stresses caused by dihedron geometry, and (b) strong surface anchoring

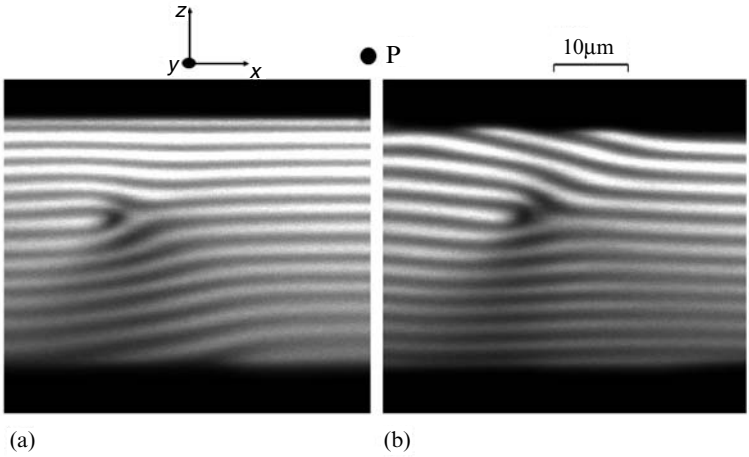

Fig. 9.7. FCPM textures of the layers profile in two different "hybrid" cells with different strength of cholesteric anchoring W_χ at the confining interfaces: (**a**) strong W_χ at the top interface (rubbed PI2555) and weak (polyisoprene) at the bottom interface; (**b**) reversed situation, weak W_χ at the top interface and strong at the bottom one

at the plates. Although there has been many studies on confined cholesteric samples (see [1,2,17,19,28] and references therein), there are still problems to explore, such as the detailed core structure of split dislocations, the structure and elastic properties of kinks along the dislocations, the role of the boundary conditions in the stability, and the location of dislocation lines within the bounded sample, etc.

9.6.1 Experimental Observations

The whole three-dimensional director structure can be understood by combining the regular PM textures, Fig. 9.8a, and the FCPM cross-sections in the vertical plane xz that contains the thickness gradient direction, Fig. 9.9. The thin part of the wedge contains thin dislocations parallel to the y-axis and separated by distances $l \approx p/(2\tan\alpha)$. For $h > h_c$, one observes a lattice of thick lines with a period $2l$. The distance between the last thin and the first thick line is $1.5l$, Fig. 9.8a. The corresponding vertical cross-sections reveal the basic features of the defects listed here.

The first line separating 0π and 1π Grandjean zones is a twist disclination, Fig. 9.9a,e, typical of a nematic, as the director experiences a slight splay remaining in the (xz) plane to the left of the core and twists by π around the z-axis in the region to the right of the core.

The thin lines separating Grandjean zones in the thin part of the sample, $h < h_c$, (such as zones 2π and 3π, Fig. 9.9b; 13π and 14π, Fig. 9.9c) are all

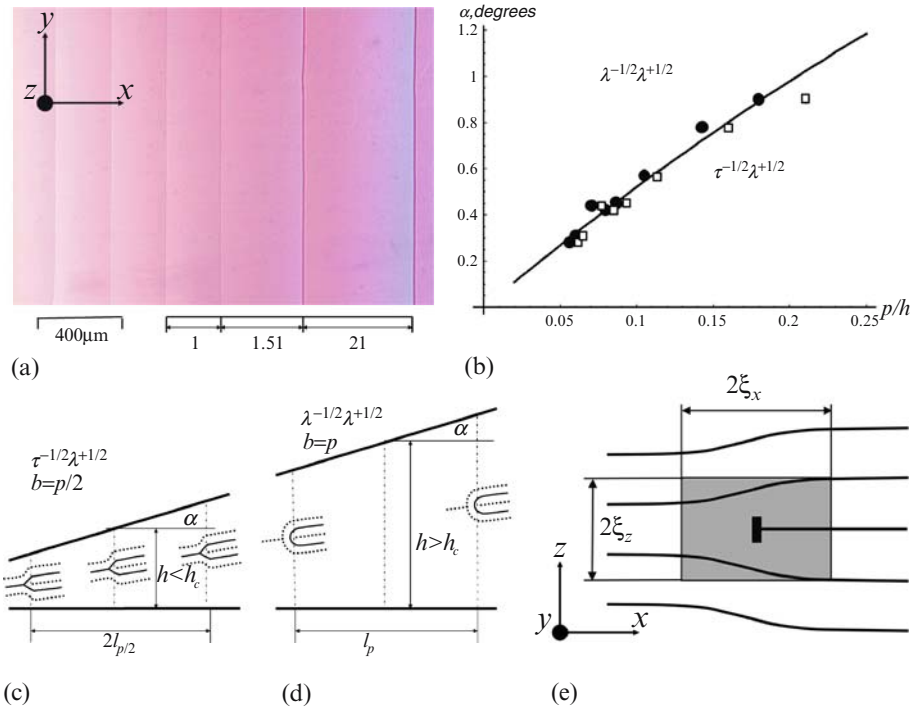

Fig. 9.8. Equilibrium defects and director structures in a Grandjean–Cano cholesteric wedge: (**a**) PM texture (a wedge sample, $p = 5\,\mu\text{m}$, $\alpha = 0.45^\circ$); (**b**) diagram of stability of the $b = p/2$ and $b = p$ dislocations; schematic representation of the cholesteric wedge with a lattice of (**c**) dislocations $b = p/2$ stable at $h < h_c$ and (**d**) dislocations $b = p$ stable at $h > h_c$; (**e**) scheme of the dislocation core with sizes ξ_x and ξ_z in two mutually orthogonal directions. The stability diagram of the $\tau^{-1/2}\lambda^{+1/2}$ and $\lambda^{-1/2}\lambda^{+1/2}$ pairs in (**b**) was determined by locations of dislocations in wedge samples of different angle α. The squares denote the last $\tau^{-1/2}\lambda^{+1/2}$ pair met as one moves toward the thick part of the wedge; the circles mark the first $\lambda^{-1/2}\lambda^{+1/2}$ pair. The solid line shows the theoretical dependence $\alpha\ (p/h_c)$ obtained by comparing the total energies of the two dislocation structures, using the following parameters: $C_1 = 0.4$, $C_2 = 1$, $r_c = 6\,\text{nm}$, $K_{22} = 7.9$ pN, $K_{33} = 15.4$ pN

separated by dislocations with the Burgers vector $\mathbf{b} = \left(0, 0, \frac{1}{2}\right) p$. Their core is split into disclination pairs $\tau^{-1/2}\lambda^{+1/2}$, Fig. 9.9f,g. Another possible splitting, into $\lambda^{-1/2}\tau^{+1/2}$ pairs, is observed in transient structures when the dislocation $b = p/2$ forms a kink, i.e., a step that brings the dislocation to a different z-level (see Sect. 9.8). Predominance of $\tau^{-1/2}\lambda^{+1/2}$ pairs over $\lambda^{-1/2}\tau^{+1/2}$ pairs has been explained by Kleman [2]: the singular core in $\tau^{+1/2}$ line is less spread and thus costs more energy as compared to $\tau^{-1/2}$ singular core.

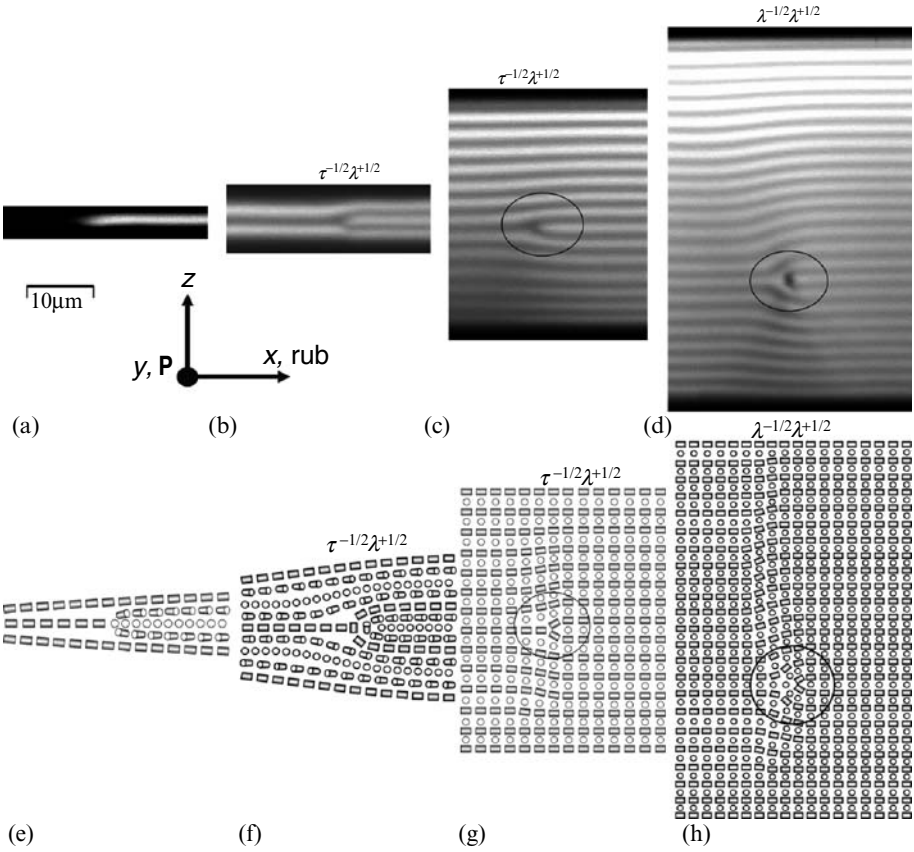

Fig. 9.9. FCPM textures of vertical xz cross-sections of Grandjean–Cano wedge with right-handed CLC, $p = 5\,\mu$m, strong planar anchoring: (**a**) twist disclination separating 0π and 1π Grandjean zones; (**b**) $b = p/2$ dislocation with a core split into a $\tau^{-1/2}\lambda^{+1/2}$ disclination pair, separating 2π and 3π Grandjean zones; (**c**) the same, between 13π and 14π zones; (**d**) $b = p$ dislocation with a core split into a $\lambda^{-1/2}\lambda^{+1/2}$ disclination pair, 22π and 24π zones. The parts (**e**–**h**) visualize the director fields corresponding to the FCPM textures in (**a**–**d**). Polarizer **P** is parallel to the y-axis. Rubbing direction is along the x-axis. Bright regions corresponds to $\mathbf{n}||\mathbf{P}$, darker regions and to $\mathbf{n} \perp \mathbf{P}$ or bounding glass plates.

The thick lines at $h > h_c$ are dislocations of Burgers vector $\mathbf{b} = (0, 0, 1)\,p$, Fig. 9.9d, with the core split into a $\lambda^{-1/2}\lambda^{+1/2}$ pair with a continuous $\mathbf{n}$, Figs.1(h) and 9(h). Their singular counterparts, $\tau^{-1/2}\tau^{+1/2}$ pairs, Fig. 9.1g, are never observed, as the singular core would carry an additional elastic energy $\sim K \ln \frac{p}{r_c}$, where K is an average Frank constant and $r_c \ll p$ is the core size of the order of few molecular sizes [3].

The critical thickness h_c of the wedge at which the lattice of $b = p/2$ dislocations is replaced with $b = p$ dislocations depends on the wedge dihedral angle α. Experimentally, for the studied range 5 mrad $< \alpha <$ 20 mrad, $k = \alpha h_c/p \approx 0.08$, Fig. 9.8b.

In order to explain the stability of different latices of dislocations in thin and thick parts of the strongly anchored wedges, we analyze the free energies of director structures and defects. We then show in Sect. 9.6.5 that the equilibrium types of defects are determined mainly by competition between core energies and compression–dilation energies of different defects and structures.

9.6.2 Far-Field Energy of an Isolated Dislocation

The far-field energy $E_{\rm ff}$ (describing the elastic energy cost due to the layer displacements caused by a dislocation) depends on the size and model of the dislocation core [17]. Generally, two different core sizes ξ_x and ξ_z along the two orthogonal axes x and z, Fig. 9.8e, have to be considered [17]. The two quantities ξ_x and ξ_z might be related in a nontrivial way, depending on λ_1 and b; their values cannot be established on the basis of the coarse-grained model. The far-field energy $E_{\rm ff}$ can be written in two equivalent forms (see [17] for details of calculations):

$$E_{\rm ff} = \frac{K_1 b^2}{8\sqrt{2\pi\xi_z}\lambda_1^{3/2}}\left[\sqrt{\frac{2}{\pi\beta}}\exp(-2\beta) + \operatorname{erf}\sqrt{2\beta}\right] \equiv \frac{K_1 b^2}{4\pi\xi_x\lambda_1}\left[\exp(-2\beta) + \sqrt{\frac{\pi\beta}{2}}\operatorname{erf}\sqrt{2\beta}\right], \tag{9.23}$$

where $\beta = \frac{\xi_x^2}{4\lambda_1\xi_z}$. Note that the relationship between the two forms in (9.23) is that of identity type and cannot be used to determine the core parameter β.

If one assumes $\xi_x^2 = 4\lambda_1\xi_z$, following the idea that perturbation of length δ_x along the layers propagates over the distance $\delta_z \sim \delta_x^2/4\lambda_1$ along the z-axis, then

$$E_{\rm ff} \approx 1.06\frac{K_1 b^2}{8\sqrt{2\pi\xi_z}\lambda_1^{3/2}} \approx \frac{K_1 b^2}{3\pi\xi_x\lambda_1}. \tag{9.24}$$

Furthermore, if $\xi_x^2 >> 4\lambda_1\xi_z$, then

$$E_{\rm ff} = \frac{K_1 b^2}{8\sqrt{2\pi\xi_z}\lambda_1^{3/2}}$$

if $\xi_x^2 << 4\lambda_1\xi_z$, then $E_{\rm ff} = \frac{K_1 b^2}{4\pi\xi_x\lambda_1}$. Note that the leading term

$$E_{\rm ff} = \frac{K_1 b^2}{4\pi\xi_x\lambda_1}$$

in the far-field energy transforms into the result derived by Kleman [2], $E_{\rm ff} = \frac{K_1 b^2}{2\xi_x \lambda_1}$, with a rescaled cut-off radius $2\pi\xi_x \to \xi_x$.

The function $E_{\rm ff}(b)$, formally quadratic in (9.24), is in fact dependent on the model of the dislocation core. As suggested by Kleman [2], if the dislocation core is split into a pair of disclinations, then the horizontal cut-off ξ_x scales as b; roughly, $\xi_x \approx b/2$; at the same time ξ_z, being a distance along the z-axis, at which the semiwidth x of the parabolae $x^2 = 4\lambda_1 z$ reaches $p/2$, is taken as independent of b. With $\xi_x \approx b/2$, the far-field energy

$$E_{\rm ff} \approx \frac{K_1 b^2}{3\pi\xi_x \lambda_1} \approx \frac{2K_1 b}{3\pi\lambda_1}$$

is a linear function of b; the result implies that dislocations with large Burgers vector are stable against splitting into two or more dislocations with smaller bs.

In the range of $b/\lambda_1 = 1$–8, the difference between the linear and nonlinear results is small, within 2% of $E_{\rm ff}$ [17]; uncertainties in core energies E_c are expected to be much larger. In the experiment, the largest value of $b/\lambda_1 = p/\lambda_1$ is about 7. Therefore, one can use the linear approximation for the analysis of far-field energy. The same conclusion follows from the theoretical analysis of Santangelo and Kamien [22].

9.6.3 Dislocation Core Energy

The experiments, Fig. 9.9, clearly show that the dislocation cores are split into pairs of disclinations. The core energy of the split dislocations is estimated as $E_{\rm c}(b) = E_{\rm pair}(b) + E'_{\rm c}$ [2], where $E_{\rm pair}(b)$ is the energy of a pair of disclinations separated by distance $2\xi_x \sim b/2$ and $E'_{\rm c}$ is the core energy of the disclination lines themselves. $E_{\rm c}(b)$ depends little on b, but is extremely sensitive to whether the disclination is singular (large $E'_{\rm c}$) or not (small $E'_{\rm c}$). As compared to the $\lambda^{-1/2}\lambda^{+1/2}$ pair, the core energy of the $\tau^{-1/2}\lambda^{+1/2}$ pair should contain an additional term $\sim K\ln(p/r_{\rm c})$ that reflects the singular nature of $\tau^{-1/2}$ disclination with the core size $r_{\rm c}$ of the order of $1 \div 10$ molecular sizes [3].

For the $\tau^{-1/2}\lambda^{+1/2}$ pair, integrating the typical distortion energy density, $\frac{1}{2}\frac{K}{r^2}$, between $r = r_{\rm c}$ and $r = b/2 = p/4$, one obtains

$$E_{{\rm c},\tau\lambda} = E_{\rm pair} + E'_{\rm c} \approx \frac{\pi}{2} K \ln\left(\frac{p}{4r_{\rm c}}\right) + C_1 K, \tag{9.25}$$

where C_1 is a number of the order of unity. $E'_{\rm c}$ should not differ much from the estimate $E'_{\rm c} = C_1 K = \frac{\pi}{8}K$ suggested by Oswald and Pieranski [29] for the singular core of a nematic disclination of winding number $\pm 1/2$, which implies $C_1 = \pi/8 \approx 0.4$. For typical $p \approx 5\,\mu$m and $r_c \approx 5$ nm, the logarithmic factor in (9.25) is relatively large, $\ln(\frac{p}{4r_{\rm c}}) \approx 6$.

In the core of dislocation $b = p$ split into a $\lambda^{-1/2}\lambda^{+1/2}$ pair, the twist structure is distorted over the area $\sim p^2$, and the core energy is roughly

$$E_{{\rm c},\lambda\lambda} = C_2 K, \tag{9.26}$$

where C_2 is another number of the order of unity.

One expects $E_{c,\lambda\lambda}$ to be about one order of magnitude smaller than $E_{c,\tau\lambda}$ when $p \approx 5\,\mu$m and $r_c \approx 5$ nm. Why do then $b = p/2$ dislocations with a very large core energy appear in the thin part of sample? The answer is given in Sect. 9.6.5 where we discuss the stability of $b = p/2$ versus $b = p$ dislocations considering also other contributions to the free energy of director structures formed in a CLC wedge cell.

9.6.4 Effect of Confinement on the Dislocation Energy

In CLCs bounded by rigid glass plates, the surface anchoring can be sufficiently strong to keep the dislocations in the bulk, as the experiments discussed in Sect. 9.5 demonstrate. The bounding surfaces can dramatically change layers profiles and other properties of dislocations. The cholesteric layers adjacent to the glass plates, Fig. 9.9, are practically (but not exactly) parallel to the substrates $z = \pm h/2$, i.e., one can assume $\frac{\partial u}{\partial x}|_{z=\pm h/2} = 0$. The layers displacement around a dislocation centered at $z = 0$ can be modeled by placing an infinite set of image dislocations of Burgers vector b outside the sample, at $z = \pm mh$, $m = 1, 2, 3, ...$ [25]. To estimate the effects of confinement on the dislocation energy, we consider only the first two images closest to the substrates and use the linear model for the displacement field $u_{\rm conf}(x, z)$ of a confined dislocation. Now we can calculate E_h, the correction to the far-field energy caused by confinement. In the limit $\xi_z/h \ll 1$, the leading term of the confinement correction is [17]:

$$E_h \approx -\frac{K_1 b^2}{4\sqrt{2\pi h}\lambda_1^{3/2}}. \tag{9.27}$$

The correction is significant only for relatively thin samples, for example, $E_h \approx -0.4E_{ff}$ for $\xi_z/h = 0.1$. As $E_h \sim b^2$, image forces in a strongly anchored sample facilitate splitting of dislocations into defects with a smaller b. Finally, one can neglect the effect of confinement on the core energy E_c of a dislocation because as long as the dislocations are not very close to the boundaries, their core structure s are h-independent, Figs. 9.4, 9.6, 9.7, and 9.9.

9.6.5 Equilibrium Lattice of Dislocation in a Cholesteric Wedge

In order to explain the stability of different dislocations described in Sect. 9.6.1, we follow the approach of Nallet and Prost [30]. The energy of the wedge is represented as the sum of the independent compression/dilation energy E_B and the energy of dislocations. The strain field due to the presence of dislocation is significant only within the parabola $x^2 = 4\lambda_1|z|$. In the wedge of small angle α, the dislocations are separated by distances $l \gg 2\sqrt{4\lambda_1 h}$ and practically do not interact. Therefore, the free energy per unit length in y-direction can be represented as a sum

$$E = E_{\mathrm{B}} + F_{\mathrm{ff}} + E_{\mathrm{h}} + E_{\mathrm{core}}, \tag{9.28}$$

where E_{ff} is the far-field energy due to the strain field inside the parabolae (9.24), E_{h} is correction to the far-field energy that accounts for confinement effects (9.27), and E_{core} is the core energy (9.26) or (9.25). We compare the energies of the two types of lattices: one with $b = p/2$, Fig. 9.8c, and another with $b = p$, Fig. 9.8d. Calculations are performed for a trapezium of length $p/\tan\alpha$ and height h_N on the left side and $h_N + p$ on the right side, Fig. 9.8c,d. The trapezium contains either two dislocations with $b = p/2$, Fig. 9.8c or one with $b = p$, Fig. 9.8d.

The compression/dilation energy due to the wedge geometry for the lattice composed of $b = p$ dislocations is [17],

$$E_{\mathrm{B}}^{p} \approx \frac{Bp^2}{12\,\tan\,\alpha}\left(\frac{1}{N} - \frac{1}{N^2}\right). \tag{9.29}$$

For the $b = p/2$ dislocations,

$$E_{\mathrm{B}}^{p/2} \approx \frac{Bp^2}{48\,\tan\,\alpha}\left(\frac{1}{N} - \frac{1}{N^2}\right). \tag{9.30}$$

The far-field energy of dislocation with $\xi_x = b/2$ is

$$E_{\mathrm{ff}} \approx \frac{2K_1 b}{3\pi\lambda_1}, \tag{9.24}$$

The confinement correction is roughly

$$E_{\mathrm{h}} \approx -\frac{K_1 b^2}{4\sqrt{2\pi h}\lambda_1^{3/2}} \approx -\frac{K_1 b^2}{4\sqrt{\pi N p}\lambda_1^{3/2}}, \tag{9.27}$$

The core energies are: $E_{\mathrm{c},\tau\lambda} \approx \frac{\pi}{2} K \ln(\frac{p}{4r_{\mathrm{c}}}) + C_1 K$, (9.25) or $E_{\mathrm{c},\lambda\lambda} = C_2 K$, (9.26), depending on the dislocation type. Using (9.28), we calculate the total energies of the array of the $b = p/2$ and $b = p$ dislocations, $E_{p/2}$ and E_p, respectively. By equating the two energies, $E_{p/2} = E_p$, one finds the critical thickness h_{c} of the wedge cell above which the lattice is composed of $b = p$ dislocations and below which the dislocations are of $b = p/2$ type (see [17] for details of this calculation):

$$h_{\mathrm{c}} \approx \frac{\pi^2 K_{22}}{3K_{33}\left[\pi \ln(\frac{p}{4r_{\mathrm{c}}}) + 2C_1 - C_2\right]} \cdot \frac{p}{\alpha} \approx k \cdot \frac{p}{\alpha}. \tag{9.31}$$

Experimentally, $k = \alpha h_{\mathrm{c}}/p = 0.08$. Therefore, (9.31) predicts $\pi \ln(\frac{p}{4r_{\mathrm{c}}}) + 2C_1 - C_2 \approx 21$. The latter estimate is in good agreement with the energies expected by the model of the split dislocation core. Actually, according to this model, (9.25) and (9.26), for typical $p \approx 5\,\mu\mathrm{m}$ and $r_{\mathrm{c}} \approx 5\,\mathrm{nm}$, and for the

expected $C_1 \approx 0.4$ and $C_2 \sim 1$, one obtains $\pi \ln(\frac{p}{4r_c}) + 2C_1 - C_2 \approx 17$, close to the value 21 deduced from (9.31). The theoretical model is in good quantitative agreement with experiment, Fig. 9.8b. Importantly, the ratio K_{22}/K_{33} for small molecule LCs is usually in the range from one-third to one-half and, therefore, k is not expected to depend much on the CLC material.

The leading contributions are produced by the B-terms (9.29), (9.30), and the core energies (9.25), (9.26). Qualitatively, the equilibrium types of defects and structures formed in the CLC wedge are determined mainly by the competition between the core energies of the defects and the compression/dilation term of free energy due to the wedge geometry. The core energy of the $b = p$ dislocations is much smaller as compared to the core energy of the $b = p$ defects. However, inserting a slab of thickness $b = p/2$ into the wedge requires less compression/dilation energy as compared to a slab of thickness $b = p$. Obviously, the difference is significant only when the number N of layers in the wedge is small, and gradually decreases with an increase of N. Therefore, dislocations $b = p/2$ are replaced by $b = p$ dislocations at $h > h_c$, Figs. 9.8 and 9.9.

9.7 Metastable Structures, Oily Streaks, Turns and Nodes of Defects

The oily streaks and dislocations are the most commonly met defects in cholesteric lamellae. The CLCs can also contain domains analog to the focal conic domains in smectics but they are not a subject of this study. In this section we review some of the frequently observed transient structures containing line defects such as dislocations and disclinations.

9.7.1 Metastable Structures and Oily Streaks

The oily streaks are complex aggregates of edge dislocations, which usually have opposite signs of the Burgers vectors. The total Burgers vector of these defects experimentally observed in planar cells is small, $|b| \leq p$. The oily streaks and the multiple dislocations are metastable objects caused by the material flow during the cell filling. These complex defect structures eventually relax to the equilibrium states such as a planar state in flat cells or the equilibrium arrays of dislocations in the wedge cells.

Defects with $b = p/2$

The simplest example of a metastable structure in a CLC wedge cell is a $b = p/2$ dislocation with its core split into a pair of the $\lambda^{+1/2}\tau^{-1/2}$ disclinations, Fig. 9.10a. The splitting of the $b = p/2$ dislocation into a pair of the $\lambda^{+1/2}\tau^{-1/2}$ disclinations is not observed in the well-equilibrated wedge samples (see Sect. 9.6), but often takes place in the vicinity of defect nodes as discussed later and at the dislocation kinks (see Sect. 9.8).

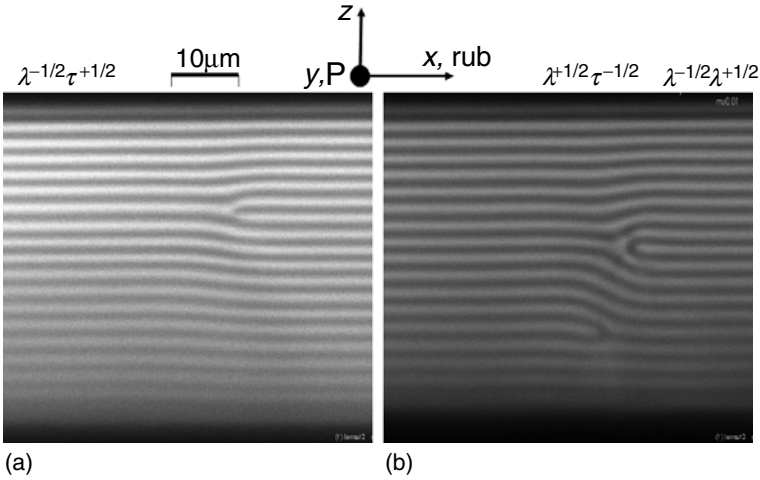

Fig. 9.10. FCPM vertical cross-sections of the metastable defect structures with the Burgers vector $b = p/2$ composed of (**a**) $\lambda^{-1/2}\tau^{+1/2}$ disclinations; (**b**) closely located $\tau^{+1/2}\lambda^{-1/2}$ and $\lambda^{-1/2}\lambda^{+1/2}$ pairs

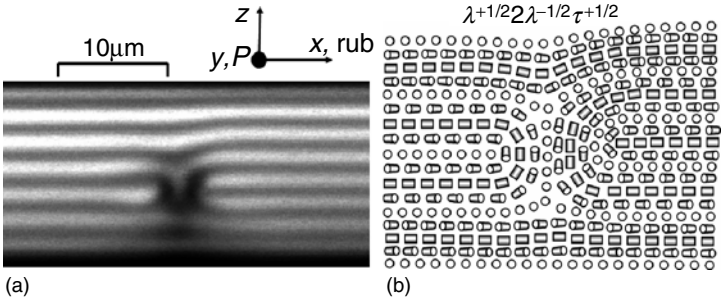

Fig. 9.11. FCPM vertical cross-sections and corresponding director structures of the metastable defect structures with the total Burgers vector $b = p/2$ composed of $\lambda^{+1/2}2\lambda^{-1/2}\tau^{+1/2}$ disclinations; (**b**) shows the director structure in (**a**)

The CLC wedges often contain transient structures of the total Burgers vector $b = p/2$ that appear as "thick" lines in standard PM observations. These configurations are in fact very different from the equilibrium pairs $\lambda^{-1/2}\lambda^{+1/2}$ observed in the thick part of the sample, as their core is composed of more than two disclinations. For example, Fig. 9.10b shows two close dislocations with the Burgers vectors $b_1 = -p/2$ (pair $\tau^{+1/2}\lambda^{-1/2}$) and $b_2 = p$ (pair $\lambda^{-1/2}\lambda^{+1/2}$), respectively. This structure quickly (few hours) relaxes into the equilibrium single dislocation $b = p/2$ (pair $\tau^{-1/2}\lambda^{+1/2}$), similar to the one shown in Fig. 9.9c. Another example, Fig. 9.11a,b, is also a combination of the four disclinations (one $\lambda^{+1/2}$, two $\lambda^{-1/2}$s, and one $\tau^{+1/2}$), topologically

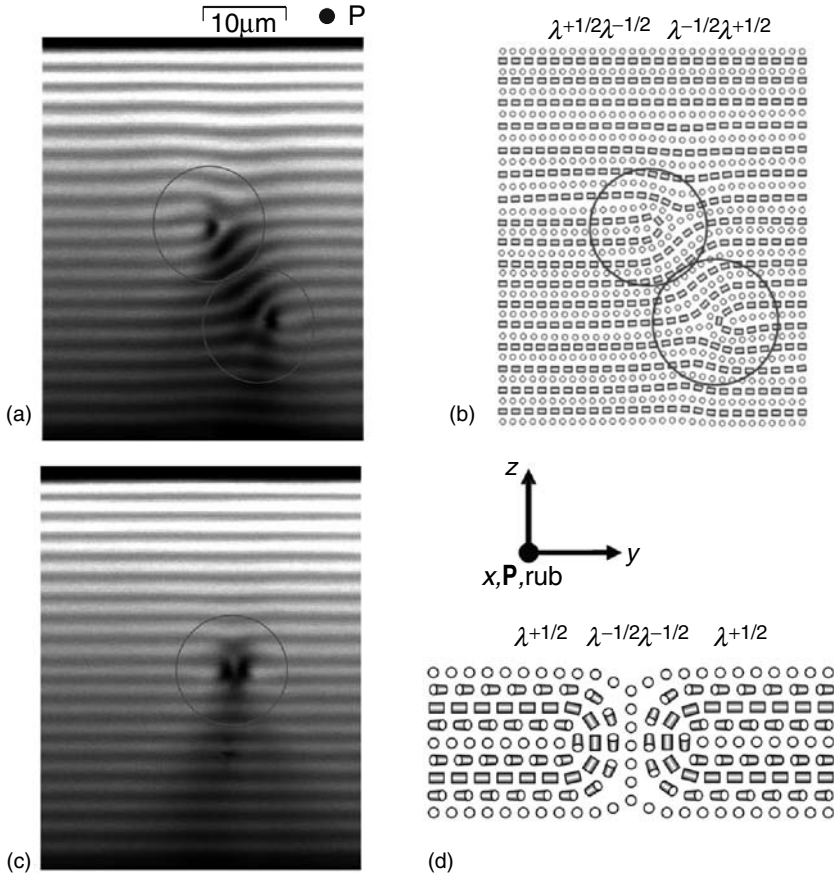

Fig. 9.12. FCPM vertical cross-sections and corresponding director structures of defects with the total Burgers vector $b = 0$: (**a,b**) two dislocations of $b_1 = -b_2 = p$ dissociated into $\lambda^{-1/2}$, $\lambda^{+1/2}$ pairs; (**c,d**) Lehmann cluster consisting of four λ disclinations

equivalent to a dislocation $b = p/2$. The structure relaxes to an equilibrium pair $\tau^{-1/2}\lambda^{+1/2}$ preserving the value $b = p/2$, similar to these in Fig. 9.9b,c. Restructuring of the defect structures in Figs. 9.10, and 9.11 usually starts at spacers or at the cell edges and propagates along the defect bundle.

Defects of Zero Burgers Vector $b = 0$

In the Grandjean–Cano wedge cells, one often finds thick lines that are *perpendicular* to the equilibrium dislocations and parallel to the thickness gradient. FCPM clearly shows that these thick lines are either pairs of dislocations with opposite signs of the Burgers vector, Fig. 9.12a,b, or symmetric oily streaks

that separate parts of the very same Grandjean zones, Fig. 9.12c,d. The similar transient defects are also frequently found in flat cells in which they can run at any angle with respect to the rubbing direction.

The oily streaks of $b = 0$ are most commonly "quadrupoles" comprising two $\lambda^{-1/2}$ and two $\lambda^{+1/2}$ disclinations, sometimes called "Lehmann clusters" [31], Fig. 9.12c. Another example of an oily streak in a cell with strong anchoring is shown in Fig. 9.13a. The FCPM texture of vertical cross-section in Fig. 9.13a confirms the basic model of an oily streak as a pair of parallel disclinations of strength $+1/2$ with a wall defect between them [2, 3]. The cholesteric layers either interrupt at this wall, or continuously reorient by 180° around the cores of the two disclinations, Fig. 9.13a. One can think about the defects structure in Fig. 9.13a,b as being composed of dislocations of large Burgers vectors of opposite signs. The structure in Fig. 9.13b contains two dislocations of $\mid b \mid = 2p$, that are displaced with respect to each other both in plane of the cell and across the cell. The structure is also accompanied by two deserted layers at the confining interfaces.

The role of the surface anchoring in the determining of the structure of metastable defects is illustrated in Fig. 9.13b. In the cells with weak anchoring, Fig. 9.13b, the layers structures and defects in the LC bulk are often accompanied by the insertion/removal of layers at the interfaces. The metastable layers structures in the strongly anchored CLCs usually have their layers parallel to the interfaces and all defects and layers distortions occur mainly in the bulk of the sample, see Figs. 9.10–9.13(a). These observations are in good agreement with the considerations of the cholesteric anchoring effect on the layers structures presented in Sect. 9.5.

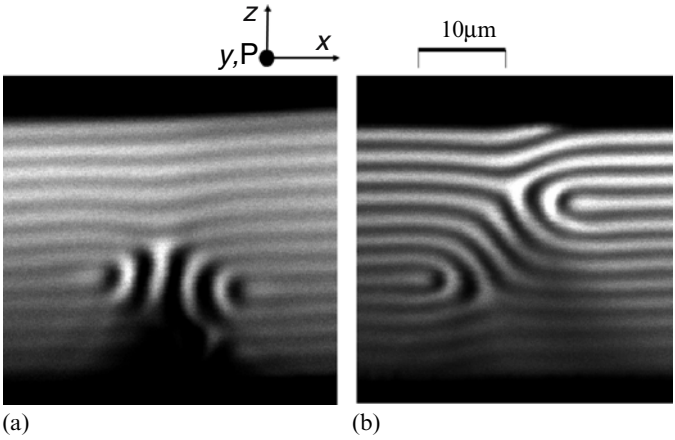

Fig. 9.13. FCPM textures and schematic director field of oily streaks in cholesteric cells of thickness $d \sim 30\,\mu$m with different boundary conditions: (**a**) strong surface anchoring, substrates treated by rubbed PI2555; (**b**) weak surface anchoring, alignment layers of polyisoprene

9.7.2 Dislocation Turns

So far we studied the line defects that have an appearance of straight lines parallel to some particular direction in the plane of CLC lamellae. Generally, the line defects can make turns, rings, closed loops, nodes, and many other complex configurations [19]. Here we consider the change of direction of the dislocation core, which we denote as a "turn."

We use a CLC cell with strong anchoring. The dislocations of Burgers vector $b = p$ make 90° turns and combine to form a defect structure of total Burgers vector $b = 0$, Fig. 9.14. Acquiring FCPM textures in plane of the cell (xy-plane, Fig. 9.14a) and in the vertical cross-sections at different places of the dislocation turns, Fig. 9.14b–e, we establish that the dislocation cores remain nonsingular during the turns. Note that the λ-disclinations in Fig. 9.14e make angles $\approx 45°$ with the polarization direction $\mathbf{P}$ for this texture, $\mathbf{P} \parallel \mathbf{x}$, whereas in Fig. 9.14d they run along $\mathbf{P} \parallel \mathbf{x}$. This explains different appearance of the dislocation cores in FCPM cross-sections in Fig. 9.14d and e. FCPM imaging with different orientations of $\mathbf{P}$ allows us to establish that the λ-disclinations in the defect cores follow the cholesteric structure and are

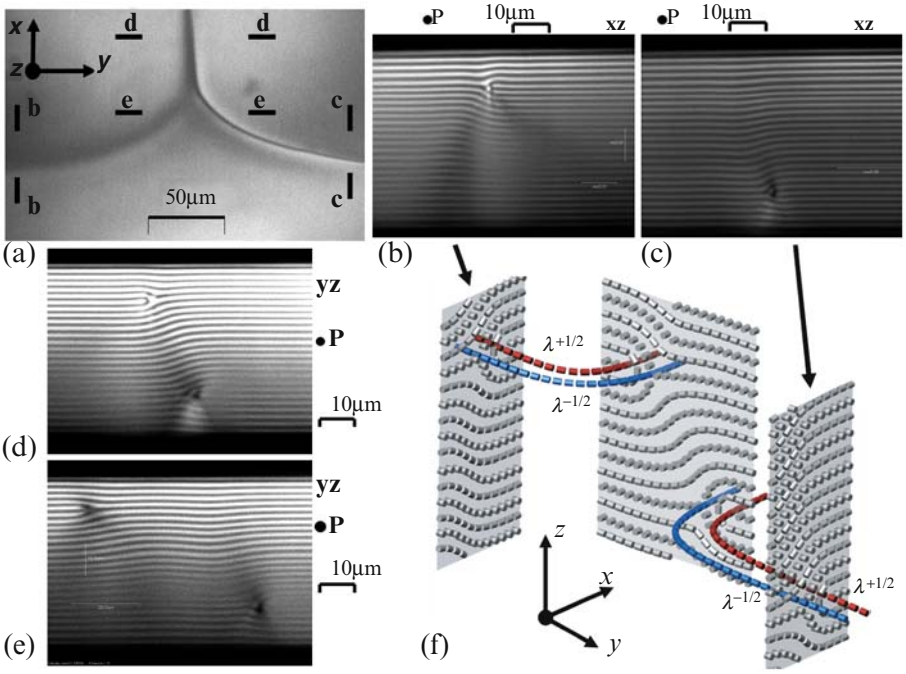

Fig. 9.14. Dislocation turns: FCPM textures in the plane of the cell (**a**) and in its vertical cross-sections (**b–e**); general three-dimensional prospective of the $\mathbf{n} \equiv \lambda$-director field and traces of the λ-disclinations (**f**). The vertical cross-sections (**b–e**) were obtained along the corresponding lines marked in part (**a**)

shifted for $\pm p/4$ along the cell normal z during the $\pm 90°$ turns, Fig. 9.14f. While making the complex three-dimensional traces, the $b = p$ dislocations preserve their nonsingular cores composed of the λ-disclinations, which follow the cholesteric helical structure, Fig. 9.14f. These experimental results are natural as one can see from the considerations given here.

The quantities $E_{c,\tau\lambda}$, $E_{c,\lambda\lambda}$ and thus E considered in Sect. 9.6 are elastic energies per unit length of the defect but *not* the line tensions of defects. The line tension T, defined as the ratio of the variation of elastic energy $\delta E = T\,\delta l$ to the variation in its length δl, depends on the orientation of edge dislocation in the cholesteric matrix,

$$T \approx \left[E(\gamma) + \frac{\partial^2 E(\gamma)}{\partial \gamma^2} \right]_{\gamma=0},$$

where γ is the angular deviation of dislocation from the y-axis (see, e.g., [4], Chs.8,9). If the dislocation would stay in the same xy plane, then reorientation would imply a change in the core structure. The $\pm 90°$ turns would transform $\lambda^{-1/2}\lambda^{+1/2}$ into $\tau^{-1/2}\tau^{+1/2}$, with a corresponding energy increase. Estimating the core energy increase under the transformation $\lambda^{-1/2}\lambda^{+1/2} \rightarrow \tau^{-1/2}\tau^{+1/2}$ as $\frac{\pi}{2} K \ln(\frac{p}{2r_c})$, one finds the core contribution to the line tension of $\lambda^{-1/2}\lambda^{+1/2}$ pair curved in the same xy plane as $E_{c,\lambda\lambda} + \pi K \ln(\frac{p}{2r_c}) >> E_{c,\lambda\lambda}$. A curved dislocation line thus should experience a torque tending to deviate it from the xy plane, i.e., to avoid the singular $\tau^{-1/2}\tau^{+1/2}$ core. As the $\lambda^{-1/2}\lambda^{+1/2}$ dislocations preserve their core structure upon deviations from the y-axis and shift along the z-axis, their actual line tension is close to the energy per unit length, i.e, $T_p \approx E_{ff}(b = p) + E_{c,\lambda\lambda} \approx \frac{2K_1 p}{3\pi\lambda_1} + C_2 K$ or $T_p \approx 3K$ when $C_2 \approx 1$. The structure of the kinks of the $b = p$ dislocations (studied in Sect. 9.8) can be explained using similar considerations.

For the $b = p/2$ dislocation, one of the disclinations in the core remains always singular, thus, the rough estimate of its line tension is $T_{p/2} \approx E_{ff}(b = p/2) + E_{c,\tau\lambda} \approx \frac{K_1 p}{3\pi\lambda_1} + \frac{\pi}{2} K \ln(\frac{p}{4r_c}) + C_1 K \approx 10K$. The core energy and the line tension $T_{p/2}$ is not expected to change much with the $\tau^{-1/2}\lambda^{+1/2} \rightarrow \lambda^{-1/2}\tau^{+1/2}$ transformation of the core. Therefore, one can expect that during the turns of the $b = p/2$ defects both transformation of core and the upward/downward shifts of the defects could be present. This is exactly the case observed in the experiments; see Sect. 9.7.3 where we discuss the nodes of defects.

9.7.3 Nodes of Line Defects

In the CLC cells of wedge geometry, the transient $b = 0$ lines parallel to the thickness gradient can connect either $b = p/2$ dislocations, Fig. 9.15a, $b = p$ lines, Fig. 9.15b, or one $b = p/2$ and one $b = p$ line, Fig. 9.15c. The dislocations $b = p/2$ and $b = p$ deviate from the y-axis near the node. Deviation of $b = p$ dislocation causes its tilt and a shift to a different z-level, which preserves the

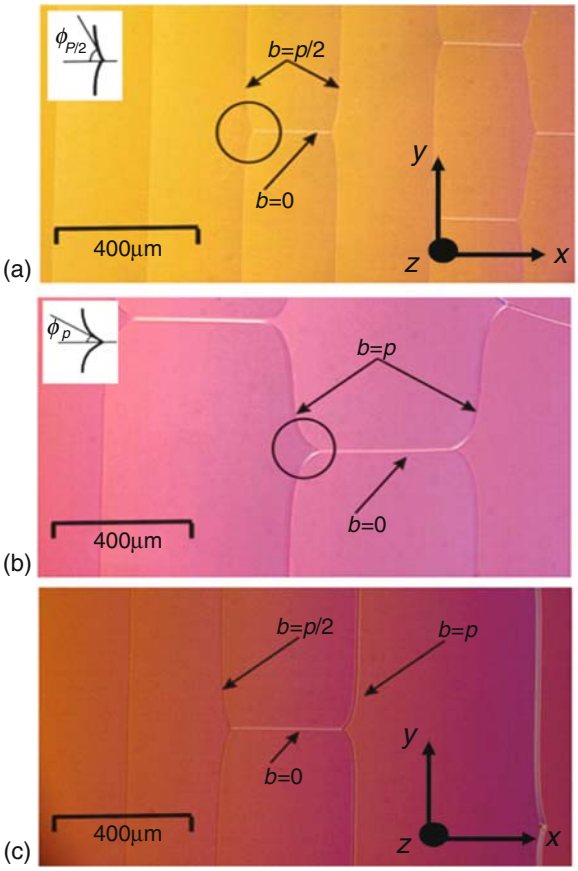

Fig. 9.15. PM textures of the Grandjean–Cano wedge with defects $b = 0$ connecting (**a**) $b = p/2$; (**b**) $b = p$; (**c**) one $b = p/2$ and one $b = p$ dislocations

nonsingular $\lambda^{-1/2}\lambda^{+1/2}$ geometry of the core, similar to the turns described in Sect. 9.7.2. Deviations of the $b = p/2$ dislocations can be associated with shifts to a different z-level as well as transformations of the defect cores, Fig. 9.16. Close to the node, the $\tau^{-1/2}\lambda^{+1/2} \rightarrow \lambda^{-1/2}\tau^{+1/2}$ transformation of the core is observed, Fig. 9.16c,d.

In mechanical equilibrium, the sum of line tensions of individual dislocations $\mathbf{T}_i$s at the dislocation node is 0, $\Sigma_i \mathbf{T}_i = 0$, see, e.g., [4]. The z-shift is small (a fraction of p) as compared to the radius of curvature of the dislocation, so that the z-components of $\mathbf{T}_i$s can be assumed to be much smaller than the x, y-components. In this case, mechanical equilibrium dictates $T_0/T_{p/2} = 2\cos\phi_{p/2}$, $T_0/T_p = 2\cos\phi_p$ and $T_p/T_{p/2} = \cos\phi_{p/2}/\cos\phi_p$; the angles are

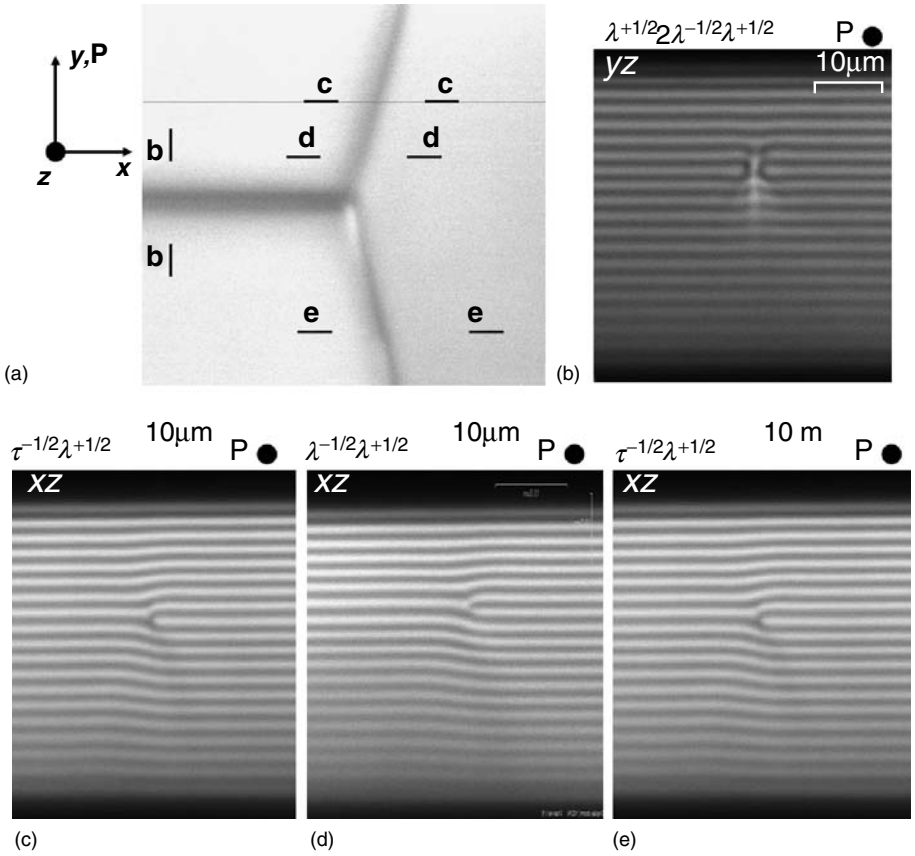

Fig. 9.16. A node of $b = p/2$ dislocations and a $b = 0$ defect: FCPM cross-section in the plane of cell (**a**) and in the vertical cross-sections (**b**–**e**) that were obtained along the corresponding lines marked in (**a**)

defined in Fig. 9.15. Experimentally, $T_0/T_{p/2} \approx 0.7 \pm 0.2$; $T_0/T_p \approx 1.7 \pm 0.2$; and $T_p/T_{p/2} \approx 0.4 \pm 0.2$, in agreement with the theoretical considerations presented in Sect. 9.7.2. The inequality $T_p < T_{p/2}$ is directly related to the split core structures of the defects.

9.8 Dynamics of Defects, Glide and Climb of Dislocations, and Their Kinks

In this section we study dynamics of defects. We use CLC wedge cells that allow us to obtain edge dislocations well separated from each other. We first

consider the stresses and forces due to the confinement that determine well-defined equilibrium location of the defects and may also cause their movement if the dislocations are displaced from these equilibrium locations. We discuss climb and glide as two different types of dislocation movement in the lamellar system (in planes of the lamellae and across lamellae). We then describe the experimentally observed movement of the defects and the mechanisms of this movement. Finally, we show that the dislocation glide strongly depends on the type of the defect core.

9.8.1 Peach and Koehler Force

Location of dislocations in a confined sample can be analyzed in terms of configurational (Peach–Koehler) force [4],

$$\mathcal{F}_i^{\mathrm{E}} = \varepsilon_{ijk} b_l \sigma_{lj}^{\mathrm{E}} t_k, \tag{9.32}$$

where ε_{ijk} is the Levi–Chivita tensor, $\mathbf{t}$ is the unit vector along the dislocation line, σ^{E} is the elastic stress tensor, related to the layers displacements caused by stresses other than that of the dislocation under consideration. In the linear approximation, the nonvanishing stress tensor components relevant to the two-dimensional case $u = u\,(x, z)$ are

$$\sigma_{zz}^{\mathrm{E}} = B\frac{\partial u}{\partial z}, \quad \sigma_{zx}^{\mathrm{E}} = -K_1\frac{\partial^3 u}{\partial x^3}. \tag{9.33}$$

For an edge dislocation with $\mathbf{b} = b\,(0, 0, 1)$ and $\mathbf{t} = (0, 1, 0)$,

$$\mathcal{F}_x^{\mathrm{E}} = -\sigma_{zz}^{\mathrm{E}} b = Bb\frac{\partial u}{\partial z},\ \mathcal{F}_y^{\mathrm{E}} = 0,\ \mathcal{F}_z^{\mathrm{E}} = \sigma_{zx}^{\mathrm{E}} b = -K_1 b\frac{\partial^3 u}{\partial x^3}. \tag{9.34}$$

Depending on the forces acting on the dislocation, it may glide or climb toward the equilibrium position.

9.8.2 Climb

Dislocations at equilibrium separate the regions of compression and dilation of layers. To find the equilibrium position of the dislocations (at which the stress σ_{zz}^{E} vanishes), one can calculate the B-term in (9.8) in a part of the wedge of length $b/\tan\alpha$, and height h_N on the left side and $h_N + b$ on the right side (N refers to the number of cholesteric layers to the left of the dislocation, i.e., in the thinner part of the wedge). The energy is minimized when the dislocation is in the equilibrium position. The same result follows from a direct calculation of the Peach–Koehler force, $\mathcal{F}_x^{\mathrm{E}} = -Bb\,(\partial u^-/\partial z + \partial u^+/\partial z)\,|_{x_{\mathrm{d}}}$ that vanishes in the equilibrium position of dislocation. Here, $\partial u^-/\partial z = x\tan\alpha/h_N - 1$ and $\partial u^+/\partial z = x\tan\alpha/\,(h_N + b) - 1$; x_{d} denotes position of the dislocation along the x-direction. The distance between two $b = p/2$

neighboring dislocations at equilibrium is

$$l_{p/2} = \frac{2\,(N+1)^2\,p}{(2N+1)(2N+3)\tan\,\alpha}. \tag{9.35}$$

The $b = p$ dislocations are separated by distances that are calculated in a similar way as

$$l_p = \frac{(N+2)^2\,p}{(N+1)(N+3)\tan\,\alpha}. \tag{9.36}$$

The separation is a weak function of N; both $l_{\rm p}$ and $l_{\rm p/2}$ quickly approach $b/\tan\,\alpha$ when N increases; even for N as small as 5, the relative difference between $b/\tan\alpha$ and the exact separating distances in (9.35) and (9.36) are negligible, less than 2%. In the well-equilibrated samples, dislocations are indeed separated by distances close to the specified by (9.35) and (9.36).

A dislocation slightly shifted from its equilibrium position along the thickness gradient (axis x) experiences the restoring force:

$$\mathcal{F}_x^{\rm E}\,(\delta_x) \approx -\frac{Bb\delta_x\,\tan\,\alpha}{h_N}\,\frac{2h_N+b}{2\,(h_N+b)}. \tag{9.37}$$

This force causes a dislocation to climb back to the position of equilibrium. Note here that climb parallel to the layers is easier than glide across the layers (see Sect. 9.8.3), as it preserves the essential geometry of the core and is associated with twist deformations near the core. Since the dislocation climb is easily implemented, the dislocations are found in their equilibrium positions in the planes of lamellae soon after the sample preparation, Fig. 9.8a.

9.8.3 Glide

Consider now a case when the dislocation is shifted along the vertical z-axis from $z = 0$ to some $\delta_z \neq 0$. In the case of the infinitely strong anchoring, the boundary condition is $\frac{\partial u}{\partial x}\,|_{z=\pm h/2}= 0$, and the dislocation is repelled by the boundary toward the midplane of cell. The corresponding Peach–Koehler force $\mathcal{F}_z^{\rm E}(z_d) = b\sigma_{zx}^{\rm E}\,|_{z=\delta_z}$ can be calculated by placing image dislocations of the same Burgers vector b at both sides of the slab, $z = -mh + (-1)^m\delta_z$ and $z = mh + (-1)^m\delta_z$ [25], where $m = 1, 2, ..., \infty$. Then the repelling force can be calculated using the superposition of the layers displacements (in a linear approximation, (9.10)) due to the dislocation and its images: $\mathcal{F}_z^{\rm E}(\delta_z) = -bK_1(\frac{\partial^3 u_{zi}}{\partial x^3})\,|_{z=\delta_z;x=0}$ [17]. The force vanishes for $\delta_z = 0$, i.e., the equilibrium position of the dislocation is in the middle plane of the cell. When the displacements from the middle plane are small, $\delta_z \ll h$, series expansion and summation yield a simple formula for the force (see [17] for detailed calculations):

$$\mathcal{F}_z^{\rm E}(\delta_z) \approx -\frac{0.47K_1b^2}{\lambda_1^{3/2}h^{3/2}}\,\frac{\delta_z}{h} \tag{9.38}$$

and stress,

$$\sigma_{zx}^{\mathrm{E}}(\delta_z) \approx \frac{0.47 K_1 b}{\lambda_1^{3/2} h^{3/2}} \frac{\delta_z}{h}. \tag{9.39}$$

The force $\mathcal{F}_z^{\mathrm{E}}(\delta_z)$ is always directed to drive the dislocation to the midplane of a strongly anchored CLC; this force quickly decreases when the thickness of the slab increases, $\mathcal{F}_z^{\mathrm{E}} \sim h^{-5/2}$.

Peach–Koehler force is $\mathcal{F}_z^{\mathrm{E}}(\delta_z) \sim b^2$, i.e., for the same displacement δ_z the force is four times larger for the $b = p$ dislocations than for the $b = p/2$ defects. In contrast, the experiments demonstrate that the $b = p$ dislocations are often found away from the bisector plane, while $b = p/2$ dislocations are close to it. The apparent discrepancy is explained by the fact that the glide of dislocations is hindered by periodic structure of the cholesteric, as discussed in Sect. 9.8.5. In Sect. 9.8.4 we experimentally study the mechanisms of the dislocation glide.

9.8.4 Experimental Observations

We never observed glide of dislocations as a whole. Instead, the change in z-coordinate of the dislocation occurs via kinks. The kinks have completely different structure for the case of $b = p/2$ and $b = p$ dislocations.

Kinks Along $b = p/2$ Dislocations

The $b = p/2$ dislocations accumulate in the bisector plane or not farther than $p/2$ from it. Initial filling of the cell might form $b = p/2$ dislocations in other locations, but they relatively quickly move to the middle plane. The lines do not glide as the whole, but via kinks, Figs. 9.17, and 9.18. There are two types of kinks: kinks of height $\pm p/4$, Fig. 9.17, and kinks of height $\pm p/2$, Fig. 9.18. The $\pm p/4$ kinks are more frequent. In the most common scenario of dislocation glide, one $\pm p/4$ kink moves along the dislocation line (along the y-axis) thus changing its z-coordinate by $p/4$ and transforming $\tau^{-1/2}\lambda^{+1/2}$ core into $\lambda^{-1/2}\tau^{+1/2}$ core, and then a second kink propagates in the same direction to restore the $\tau^{-1/2}\lambda^{+1/2}$ pair that is now shifted by $p/2$ with respect to the original $\tau^{-1/2}\lambda^{+1/2}$. The core structure of $\pm p/4$ kink is intermediate between that of pure $\tau^{-1/2}\lambda^{+1/2}$ and $\lambda^{-1/2}\tau^{+1/2}$ states, Fig. 9.17. The $\pm p/2$ kinks, twice as high as the $\pm p/4$ kinks, can be seen near the nodes where $b = p/2$ dislocations join other line defects located at a different z-level in the sample, e.g., $b = 0$ dislocations, Fig. 9.16. Such a $\pm p/2$ kink can be stable for hours, as the glide of defects with $b \neq p/2$ is very difficult. Figure 18(b) reveals the core structure of a $p/2$ kink in the glide plane; the core structure changes from $\tau^{-1/2}\lambda^{+1/2}$ into $\lambda^{-1/2}\tau^{+1/2}$ and then back to $\tau^{-1/2}\lambda^{+1/2}$ state along the y-axis, Fig. 9.18c–e.

There are two distinct features of both $\pm p/4$ and $\pm p/2$ kinks along the $b = p/2$ dislocations as compared to the kinks along $b = p$ dislocations. First,

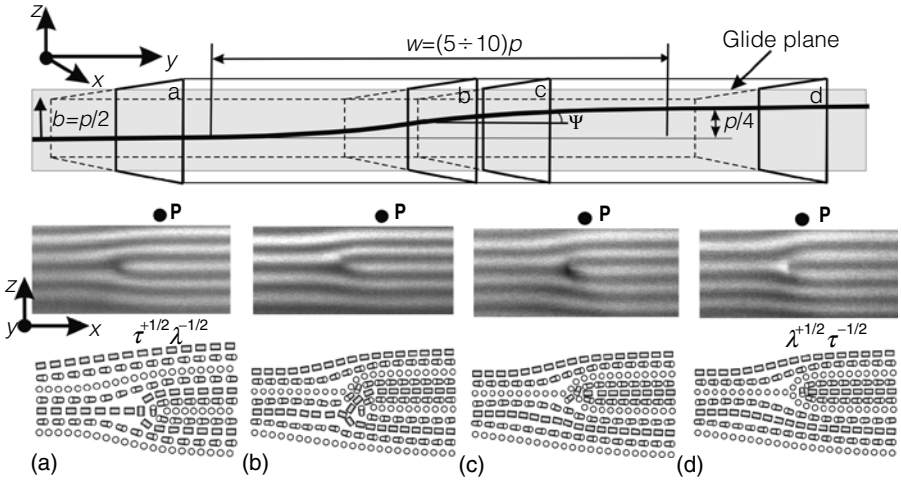

Fig. 9.17. FCPM textures of a kink of height $p/4$ along the dislocation $b = p/2$; the core $\tau^{-1/2}\lambda^{+1/2}$ (**a**) transforms into the $\lambda^{-1/2}\tau^{+1/2}$ core (**d**). The parts (**b,c**) show the intermediate director structures

the $b = p/2$ kinks make a very small angle with the y-axis; their characteristic length w measured along the y-axis is thus large, about $(5-10)\,p$, Figs. 9.17, and 9.18. Second, the kinks are confined to the glide plane (yz) of the parent $b = p/2$ dislocation, Fig. 9.18b.

Kinks Along $b = p$ Dislocations

The glide of $\lambda^{-1/2}\lambda^{+1/2}$ pairs with $b = p$, Fig. 9.9d, is much more difficult as compared to $b = p/2$ dislocations; these pairs can remain in the locations away from the bisector plane for months. The kinks along $b = p$ dislocations were observed only in the samples with *weak* surface anchoring (unrubbed polyisoprene coating) and with an applied electric field. When a voltage pulse of amplitude $V \geq V_c$ and duration $\sim$1 s is applied, a $b = p$ kink is generated (at the wedge of cell or at a spacer) and propagates along the edge dislocation, shifting its position by a distance p toward the middle plane. The $b = p$ kinks are relatively short, have a cusp structure, and depart from the glide plane of the parent dislocation, Figs. 9.19, and 9.20.

Figure 9.19 presents a series of vertical FCPM slices taken in the vicinity of the kink. The vertical optical slices $1yz-9yz$ are parallel to the glide plane. The orthogonal cross-sections $10xz$ and $11xz$ are normal to the dislocation and demonstrate that the kink shifts the dislocation by p along the z-axis. The plane $2yz$ contains the $\lambda^{+1/2}$-disclination of the split core. The slices $2yz-8yz$ show that near the kink, the dislocation deviates from the y-direction toward the thinner part of the wedge, Fig. 9.19, thus forming a cusp first noticed by Bouligand [19]. Using the principles described in Sect. 9.2, we reconstruct the

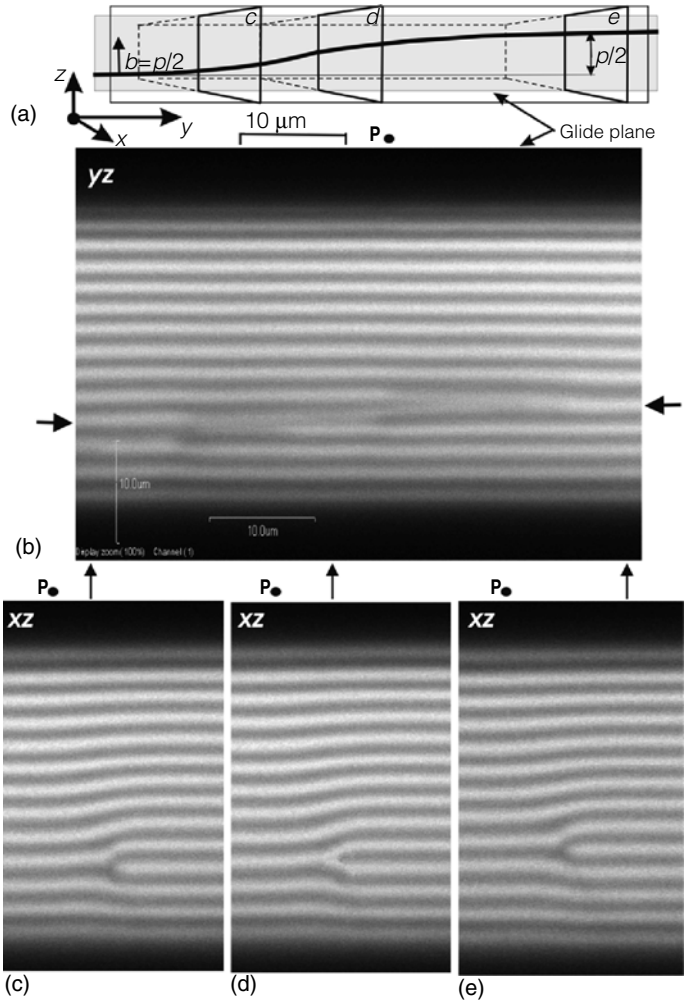

Fig. 9.18. FCPM textures of a kink of height $p/2$ along the dislocation $b = p/2$; (**b**) vertical cross-section along the glide plane; the kink is only slightly tilted with respect to the parent dislocation, the horizontal arrows indicate the z-levels where the kink ends; the vertical cross-sections (**c**–**e**) are perpendicular to the glide plane and show how the core $\tau^{-1/2}\lambda^{+1/2}$ (**c**) transforms first into the $\lambda^{-1/2}\tau^{+1/2}$ core (**d**) and then back into the $\tau^{-1/2}\lambda^{+1/2}$ core (**e**)

three-dimensional director field near the kink, Fig. 9.20. The $\lambda^{+1/2}$ and $\lambda^{-1/2}$ lines are aligned on top of each other at the center of the cusp rather than side by side as they are far from the cusped kink, Fig. 9.20.

At the kink, the λ-disclinations deviate from the y-axis by $\pi/2$ and align along the x-axis, each forming a cusp. The director in the core of each λ

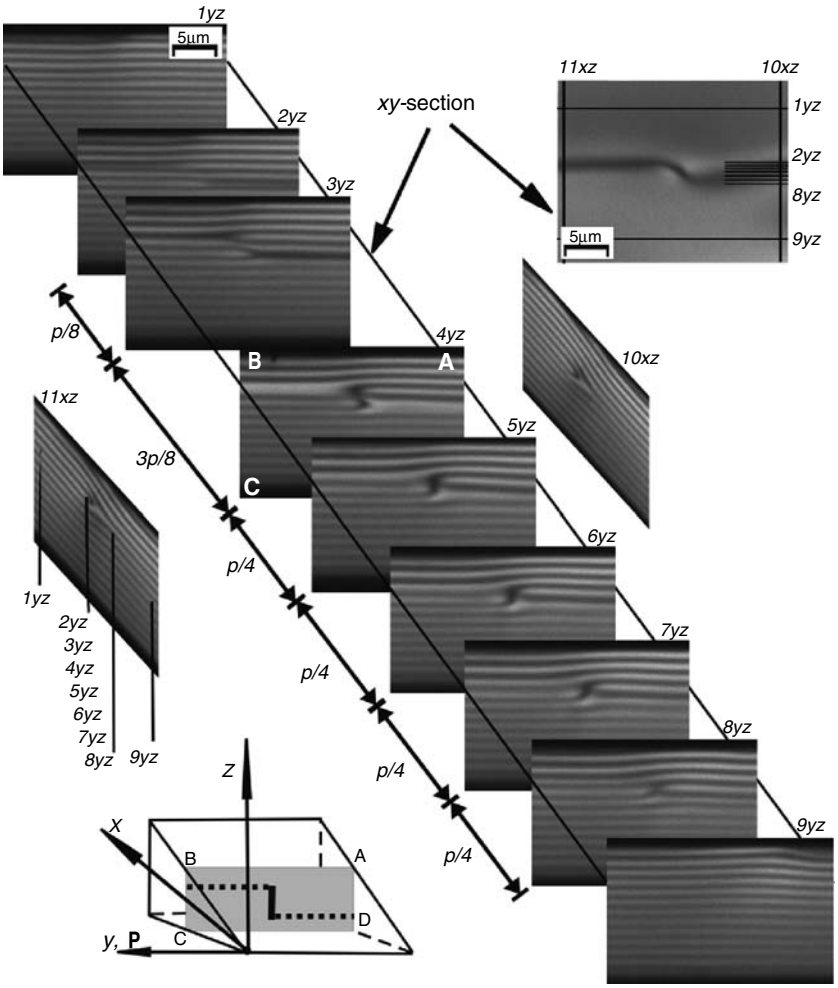

Fig. 9.19. FCPM textures of a kink of height $p/2$ along the dislocation $b = p$, as seen in the vertical planes $1yz - 9yz$ parallel to the plane ABCD $4yz$. In the right top corner, a horizontal slice xy demonstrates a cusp associated with the kink

disclination remains parallel to the disclination axis, and thus the $\pi/2$ rotation of the disclination also means a shift of the core by $p/4$ along the z-axis, Fig. 9.20. The shift preserves the nonsingularity of director field; without tilt, $\lambda^{-1/2}\lambda^{+1/2}$ would transform into a singular $\tau^{-1/2}\tau^{+1/2}$ core. At the cusp, the $\lambda^{-1/2}$ disclination entering the kink from one side transforms into a $\lambda^{+1/2}$ disclination leaving the kink on the other side, Fig. 9.20. The kink at $b = p$ dislocation, therefore, has a complex structure with a cusp and interchange of the $\lambda^{+1/2}$ and $\lambda^{-1/2}$ disclinations; its size is of the order of p along all three coordinate axes, Figs. 9.19, and 9.20.

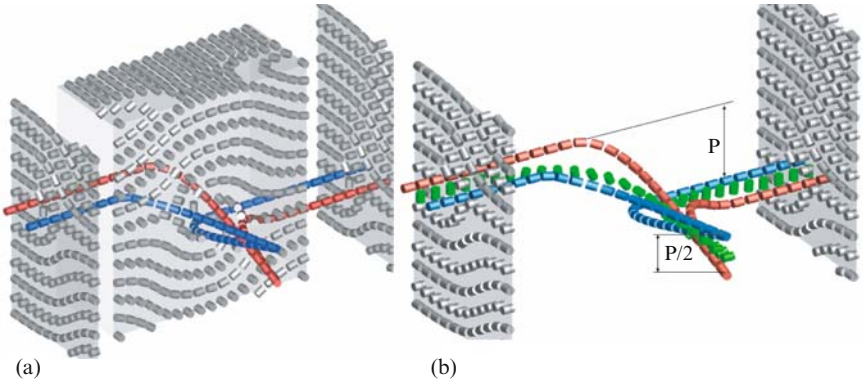

Fig. 9.20. General three-dimensional perspective of the director field around and at the core of $\lambda^{-1/2}$, $\lambda^{+1/2}$ disclinations in the kink shown in Fig. 9.19: pattern of the three-dimensional director field (**a**) and traces of disclinations (**b**)

9.8.5 Peierls–Nabarro Friction

In order to explain the features of glide and kinks described earlier, we consider the effect of split dislocation core and periodic cholesteric structure on the dynamics of defects. In solid state physics, the phenomenon is known as the Peierls–Nabarro friction [32, 33]. As the dislocation glides across the crystal lattice, the core structure changes periodically; atomic reconstructions lead to periodic changes of the potential energy of the crystal. The applied stress needed to overcome the energy barriers is called the Peierls–Nabarro stress. This stress is determined by the core structure and thus cannot be given by a universal analytical expression. The original Peierls–Nabarro model assumes a sinusoidal force between the atomic planes on the two sides of the slip plane.

When an edge dislocation with a split core moves as a whole in z-direction, the structure of the two disclinations changes periodically. Upon a shift by $p/4$, the pair $\lambda^{-1/2}\lambda^{+1/2}$ transforms into $\tau^{-1/2}\tau^{+1/2}$ and the pair $\tau^{-1/2}\lambda^{+1/2}$ transforms into $\lambda^{-1/2}\tau^{+1/2}$, Fig. 9.21. The main contribution to the energy changes comes from the energy of the cores; the far-field energy can be assumed constant. As discussed earlier, the core energy of the $\lambda^{-1/2}\lambda^{+1/2}$ pair is relatively small, $E_{\mathrm{c},\lambda\lambda} = C_2 K \sim K$, (9.26). The transformation $\lambda^{-1/2}\lambda^{+1/2} \rightarrow \tau^{-1/2}\tau^{+1/2}$ implies a large increase in the core energy, of the order of $E^{p}_{\mathrm{PN}} \approx E_{\mathrm{c},\tau\tau} - E_{\mathrm{c},\lambda\lambda} \approx K \ln(\frac{p}{r_{\mathrm{c}}}) >> E_{\mathrm{c},\lambda\lambda}$. In contrast, the minimum core energy of the $\tau^{-1/2}\lambda^{+1/2}$ pair is already large, $E_{\mathrm{c},\tau\lambda} \approx \frac{\pi}{2} K \ln(\frac{p}{4r_{\mathrm{c}}}) + C_1 K$, according to (9.25), see Fig. 9.21. The alternative $\lambda^{-1/2}\tau^{+1/2}$ core apparently corresponds to a local minimum in the potential energy as one does observe kinks that transform $\tau^{-1/2}\lambda^{+1/2}$ into $\lambda^{-1/2}\tau^{+1/2}$ and back, Fig. 9.17. The transformation $\tau^{-1/2}\lambda^{+1/2} \rightarrow \lambda^{-1/2}\tau^{+1/2}$ implies an increase in the core energy by $E^{p/2}_{\mathrm{PN}} \approx \frac{\pi}{2} K \ln(\frac{r_{\mathrm{c}}}{r'_{\mathrm{c}}}) + (C'_1 - C_1) K \approx cK$, where primed values

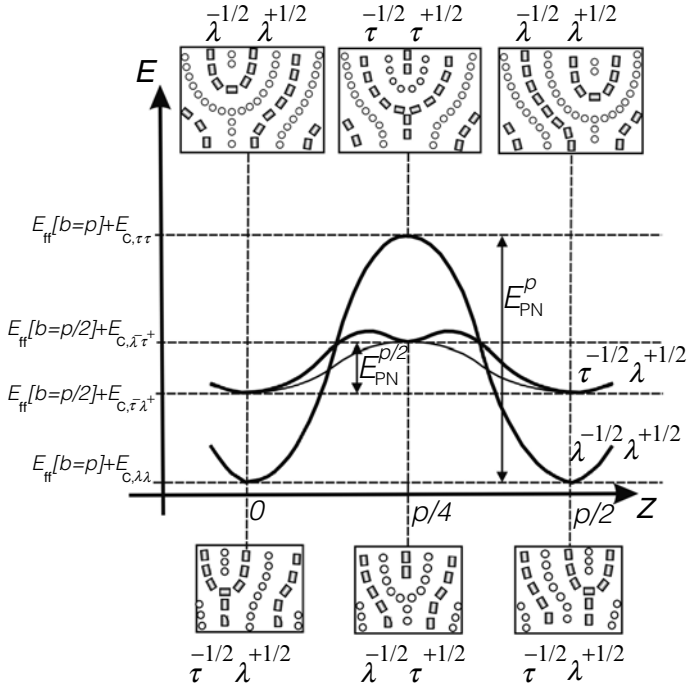

Fig. 9.21. Potential energies of straight dislocations $b = p/2$ and dislocation $b = p$ with the split cores as the functions of their position along the z-axis; see text for details

correspond to the pair $\lambda^{-1/2}\tau^{+1/2}$; the numerical constant c is most probably less than 1 (see the estimates provided later).

The excess free energy as the function of dislocation displacement δ_z along the helix axis can be written phenomenologically as

$$\Delta E(\delta_z) \approx \frac{E_{\mathrm{PN}}}{2}\left(1 - \cos\frac{4\pi b}{p}\frac{\delta_z}{b}\right) = E_{\mathrm{PN}}\sin^2\frac{2\pi\delta_z}{p}, \tag{9.40}$$

similar to the phenomenological model for solid crystals [32, 33]; E_{PN} is the Peierls–Nabarro energy, Fig. 9.21. Note that for the sake of simplicity we approximate the two-minima potential for the $\tau^{-1/2}\lambda^{+1/2}$ pair with a single-minimum cosinusoidal function, shown by a thin line in Fig. 9.21. The corresponding stress

$$\frac{1}{b^2}\frac{\partial\Delta E\,(\delta_z/b)}{\partial(\delta_z/b)} = \frac{2\pi E_{\mathrm{PN}}}{pb}\sin\frac{4\pi\delta_z}{p}$$

has the amplitude $\sigma_{\mathrm{PNcore}} = \frac{2\pi E_{\mathrm{PN}}}{pb}$, or, when written for the two types of dislocations separately,

$$\sigma^{p}_{\mathrm{PNcore}} \approx \frac{2\pi K\ln(\frac{p}{r_c})}{pb}, \qquad \sigma^{p/2}_{\mathrm{PNcore}} \approx \frac{2\pi cK}{pb}; \tag{9.41}$$

For the $\lambda^{-1/2}\lambda^{+1/2}$ pair, with $b = \delta_z = p$, $\lambda_1 = 0.2p$, $h = 10p$, $p \approx 5\,\mu\text{m}$ and $r_c \approx 5\,\text{nm}$, one finds $\sigma^E_{zx}/\sigma^p_{\text{PNcore}} \approx 4 \times 10^6$; therefore, the model predicts that $\lambda^{-1/2}\lambda^{+1/2}$ pair cannot glide as a straight line. For the $\tau^{-1/2}\lambda^{+1/2}$ pair, with $b = \delta_z = p/2$, $\lambda_1 = 0.2p$, $h = 10p$, one finds $\sigma^E_{zx}/\sigma^{p/2}_{\text{PNcore}} \approx 4 \times 10^5 c$; unless c is anomalously small (as estimated below, c is of the order of 10^{-2}), the Peierls–Nabarro barrier is too high to allow the dislocation $b = p/2$ to glide as well.

The model presented here is in good agreement with the experimental data (see Sect. 9.8.4) as we never observed glide of dislocations as a whole. This model explains the affinity of the $b = p/2$ dislocations to be located in the equilibrium positions (middle plane of cell in the strongly anchored wedge cells) as well as the wide distribution of positions of the $b = p$ dislocations in the vertical cross-section of LC cell. Section 9.8.6 illustrates that the Peierls–Nabarro energy barrier also strongly influences the structure of the dislocation kinks.

9.8.6 Kink Structure Versus Peierls–Nabarro Energy Barrier

The kinks along the $b = p/2$ dislocations are usually of height $p/4$ or $p/2$ each, Figs. 9.17,9.18. The length of the kink, measured along the y-axis, is large, $w \sim (5-10)\,p$, i.e., the angle ψ between the kink and the y-axis is small. This experimental feature indicates that the Peierls–Nabarro energy barrier is relatively small as compared to the line tension of the dislocation itself. Imagine a dislocation connecting two points in the bulk of the sample, A (x_A, z_A) and B (x_B, z_B). The smaller the Peierls–Nabarro energy as compared to the line energy of the dislocation, the smaller is ψ: in the limiting case $E_{\text{PN}}/E \to 0$, the kink is infinitely long, as the dislocation simply tilts as a whole and preserves the form of a straight line to minimize its total length $\sqrt{(x_\text{B} - x_\text{A})^2 + (z_\text{B} - z_\text{A})^2}$. When the Peierls–Nabarro energy associated with the kink is larger than the line tension, then ψ is large and the kink tends to be short; in the limit $E_{\text{PN}}/E \to \infty$, the kink is vertical, of the length $|z_\text{B} - z_\text{A}|$, it connects two horizontal dislocation segments of total length $|x_\text{B} - x_\text{A}|$.

For small ψ, one can directly apply the kink model developed for solid crystals [32,33], in which ψ is determined by the (constant) line tension of the edge dislocation $E_{p/2} \approx E_{c,\tau\lambda} \approx \frac{\pi}{2} K \ln(\frac{p}{4r_c})$, (9.25), and the Peierls–Nabarro energy $E^{p/2}_{\text{PN}} \approx cK$, as $\psi = \sqrt{2E^{p/2}_{\text{PN}}/E_{p/2}}$. As $\psi = p/(4w)$ for the $p/4$ kink, one obtains $w \approx \frac{p}{4}\sqrt{\frac{E_{p/2}}{2E^{p/2}_{\text{PN}}}} \approx \frac{p}{4}\sqrt{\frac{\pi}{4c}\ln(\frac{p}{4r_c})}$. Using the estimates $p \approx 5\,\mu\text{m}$ and $r_c \approx 5\,\text{nm}$, and the experimental result $w \sim (5-10)\,p$ (Figs. 9.17, and 9.18), one obtains $c \sim (0.3\text{–}1) \times 10^{-2}$. In other words, the core energy variation for the $\tau^{-1/2}\lambda^{+1/2}$ pair along the kink is only a small fraction of the Frank elastic constant K, which is a reasonable conclusion as the $b = p/2$ dislocation can never get rid of the singular core.

In contrast, for a kink along the $b = p$ dislocation, the $\lambda^{-1/2}\lambda^{+1/2}$ pair simply twists with the local cholesteric director to preserve the nonsingular core, Fig. 9.20; the energy density of the kink is of the order of K and is not very different from the line tension $E_p \sim K$ of the dislocation itself; therefore, the kinks are expected to be short, $w \sim p$, as confirmed by the experiments, Figs. 9.19, and 9.20.

The total elastic energy U of the kinks in cholesterics with a micron-scale pitch is expected to be much larger than the thermal energy ($k_{\mathrm{B}}T \approx 4\times10^{-21}$J at room temperature), which makes their thermal nucleation unlikely; the situation is thus different from the typical SmA materials, in which the kinks are mostly of molecular height. For the kinks along the cholesteric $b = p$ dislocation, the discussion given here leads to $U_p \sim \left(K/p^2\right)p^3 \sim pK \sim 5 \times 10^{-17}J$. For the "long" kinks along the $b = p/2$ dislocation, the energy is $U_{b/2} \sim E_{p/2}b^2/w$, i.e., $U_{b/2} \sim \frac{\pi}{2}Kp^2\ln(\frac{p}{4r_c})/(4w) \sim 10^{-17}$J. The observed kinks can be introduced during the filling of the samples and by mechanical inhomogeneities, including the edges of the cholesteric sample.

These considerations explain the experimentally observed strong differences between the structures of kinks at $b = p/2$ and $b = p$ dislocations (compare Figs. 9.17, and 9.18 with Figs. 9.19, and 9.20).

9.9 Conclusions

Using FCPM, we nondestructively visualized the director patterns in vertical cross-sections of cholesteric wedge and flat cells with weak and strong anchoring at confining substrates. Optical slicing in the vertical cross-sections allowed us to reconstruct the detailed structure of dislocations and their kinks, as well as the disclination nodes and oily streaks.

We established that strong surface anchoring at both plates stabilizes edge dislocations in the bulk, while weak surface anchoring attracts them to the surfaces where they either disappear (flat samples) or transform into inserted surface layers (wedge samples). We proposed a model of surface anchoring potential for the cholesteric phase in the coarse-grained approximation, which qualitatively describes the phenomenon of anchoring-mediated interaction of an edge dislocation with a bounding surface. We systematically studied static and dynamic properties of dislocations confined in the cells with different geometry and boundary conditions.

The dislocations in weakly anchored samples always have Burgers vector $b = p$ and never $b = p/2$. The dislocations in strongly anchored cholesterics can contain dislocations of Burgers vector $b = p$ as well as $b = p/2$, depending on the geometry and thickness of the CLC slab. Weak anchoring causes attraction, while strong anchoring causes repulsion between the edge dislocation and the boundary.

The FCPM technique allows to establish the fine details of the dislocation structures. The dislocation of Burgers vector $b = p/2$ splits into $\tau^{-1/2}\lambda^{+1/2}$

disclination pair, while $b = p$ splits into a $\lambda^{-1/2}\lambda^{+1/2}$ pair. Pairs of $\lambda^{-1/2}$ and $\tau^{+1/2}$ disclinations are observed when the $b = p/2$ dislocation forms a kink. The type of dislocation core strongly influences the location of the defects across the cell and its affinity to glide.

Kinks are responsible for glide of dislocations that never glide as straight lines. Kinks are different for $b = p/2$ and $b = p$ dislocations. The kinks along the $b = p/2$ dislocations change the level of dislocations by $\pm p/4$ and $\pm p/2$. In the $b = p/2$ case, the kink is only slightly tilted with respect to the dislocation; it is confined to the glide plane and is relatively long, $w \sim (5 - 10)\, p$, as the core energy per unit length of $\tau^{-1/2}\lambda^{+1/2}$ pair is large as compared to the Peierls–Nabarro barrier associated with modifications of the $\tau^{-1/2}\lambda^{+1/2}$ core into a $\lambda^{-1/2}\tau^{+1/2}$ core. In the $b = p$ case, the kinks are short, $w \sim p$; both λ disclinations deviate from the glide plane, to preserve a nonsingular director structure. The kinks along the $b = p$ dislocation are of a typical size p and form cusps in the direction perpendicular to the glide plane. At the cusp, $\lambda^{-1/2}$ and $\lambda^{+1/2}$ disclinations interchange ends. Thermal nucleation of kinks in cholesteric samples with p in the micron range is unlikely; kinks can be introduced by mechanical irregularities, at the edges of sample, and also during the filling of the sample.

In contrast to glide, climb of dislocations occurs easily. In a Grandjean–Cano wedge, a lattice of $b = p/2$ dislocations is stable at $h < h_c$ (where h_c is some critical thickness). At $h > h_c$ it is replaced by a lattice of $b = p$ dislocations. Since climb is easily implemented, the dislocations in equilibrium are separated by well-defined distances along the thickness gradient in the cells of wedge geometry.

We illustrated that the layers profile of an isolated edge dislocation is well described by the nonlinear elastic theory, much better than by its linear approximation. The layers structures of equilibrium and metastable defects and structures can be strongly influenced by the confinement and surface anchoring conditions at the confining interfaces.

We employed the coarse-grained linear elastic model of cholesteric phase as well as the coarse-grained model of cholesteric anchoring in order to describe the experimental results. We calculated (a) the energy of layer distortions around the dislocations; (b) corrections to the energy caused by finite thickness of the sample; (c) Peach–Koehler forces acting on a dislocation shifted from its equilibrium positions and (d) the Peierls–Nabarro friction associated with the split core of the cholesteric dislocations; and (e) critical thickness h_c. The calculations are in good agreement with the experimental results.

Acknowledgments. The work was supported by the National Science Foundation, grant 0315523 and STC ALCOM grant no. DMR89-20147, by donors of the Petroleum Research Fund, administered by the ACS, grant 35306-AC7, and partially by NSF U.S.–France Cooperative Scientific Program, grant no. INT-9726802; the latter made possible fruitful discussions with M. Kleman. We also thank Y. Bouligand, I. Dozov, G. Durand, T. Ishikawa,

N. Madhusudana, Ph. Martinot-Lagarde, Yu. Nastishin, M. Nobili, R. Pratibha, B. Senyuk, and S. Shiyanovskii for discussions.

References

1. de Gennes, P.G., Prost, J.: The Physics of Liquid Crystals, 2nd edn. Clarendon Press, Oxford (1993)
2. Kleman, M.: Points, Lines and Walls in Liquid Crystals, Magnetic Systems and Various Ordered Media. Wiley, Chichester (1983)
3. Lavrentovich, O.D., Kleman, M.: Cholesteric liquid crystals: defects and topology. In: Bahr, C., Kitzerow, H. (eds.) Chirality in Liquid Crystals, Springer-Verlag, New York (2001) pp. 115–158
4. Kleman, M., Lavrentovich, O.D.: Soft Matter Physics: An Introduction. Springer-Verlag, New York (2003)
5. Kleman, M.: Rev. Prog. Phys. **52**, 555 (1989)
6. Lubensky, T.C.: Phys. Rev. A **6,** 452 (1972)
7. Li, Z., Lavrentovich, O.D.: Phys. Rev. Lett. **73**, 280 (1994)
8. Lavrentovich, O.D., Yang, D.-K.: Phys. Rev. E **57**, R6268 (1998)
9. Ishikawa, T. and Lavrentovich, O.D.: Phys. Rev. E **63**, 030501(R) (2001)
10. Ishikawa, T., Lavrentovich, O.D.: Defects and undulation in layered liquid crystals. In: Defects in Liquid Crystals: Computer Simulations, Theory and Experiments, vol.43, Lavrentovich, O.D., Pasini, P. Zannoni, G. Žumer, S. (ed) (NATO Science Series) Kluwer Academic Publishers, (2001) pp271–300.
11. Smalyukh, I.I., Shiyanovskii, S.V., Lavrentovich, O.D.: Chem. Phys. Lett. **336,** 88 (2001)
12. Smalyukh, I.I., Lavrentovich, O.D.: Phys. Rev. Lett. **90**, 085503 (2003)
13. Shiyanovskii, S.V. Smalyukh, I.I. Lavrentovich, O.D.: Computer simulations and fluorescence confocal polarizing microscopy of structures in cholesteric liquid crystals. In: Defects in Liquid Crystals: Computer Simulations, Theory and Experiments, vol.43, Lavrentovich, O. D. Pasini, P. Zannoni, G. Žumer S. (ed) (NATO Science Series), Kluwer Academic Publishers (2001) pp229–270.
14. Yu., Nastishin, A., Polak, R.D., Shiyanovskii, S.V. Bodnar, V.H. Lavrentovich, O.D.: J. Appl. Phys. **86**, 4199 (1999)
15. Yang, F.Z. Cheng, H.F. Gao, H.J. Sambles, J.R.: J. Opt. Soc. Am. B. **18**, 994 (2001)
16. Stoenescu, D.N., Nguyen, H.T., Barois, P., Navailles, L., Nobili, M., Martinot-Lagarde, Ph., Dozov, I.: Mol. Cryst. Liq. Cryst. **358**, 275 (2001); Ramdane, O., Auroy, Ph., Forget, S., Raspaud, E., Martinot-Lagarde, Ph., Dozov, I.: Phys. Rev. Lett. **84,** 3871 (2000)
17. Smalyukh, I.I., Lavrentovich, O.D.: Phys. Rev. E **66**, 051703 (2002)
18. Kleman, M. Friedel, J.: J. Phys. Colloq. **30** C4-43 (1969)
19. Bouligand, Y.: J. Phys. France. **35**, 959 (1974)
20. Kats, E.I., Lebedev, V.V.: Fluctuational Effects in the Dynamics of Liquid Crystals. Springer-Verlag, New York (1994) p170.
21. Brener, E.A., Marchenko, V.I.: Phys. Rev. E. **59**, R4752 (1999)
22. Santangelo, C.D., Kamien, R.D.: Phys. Rev. Lett. **91**, 045506 (2003)
23. Ishikawa, T., Lavrentovich, O.D.: Phys. Rev. E. **60**, R5037 (1999)

24. Meister, R., Dumoulin, H., Halle, M.-A., Pieranski, P.: J. Phys. II France. **6**, 827 (1996)
25. Pershan, P.S.: J. Appl. Phys. **45**, 1590 (1974)
26. Holyst, R., Oswald, P.: Int. J. Mod. Phys. B. **9**, 1515 (1995)
27. Grandjean, F., C.R. Hebd. Sean. Acad. Sci. **172**, 71 (1921)
28. Cano, R.: Bull. Soc. Fr. Mineral. Cristalogr. **90**, 333 (1967); Cano, R.: Bull. Soc. Fr. Mineral. Cristalogr. **91**, 20 (1968)
29. Oswald, P., Pieranski, P.: Les Cristaux Liquides, Tome 1, Gordon and Breach Science Publishers, Paris (2000) p522
30. Nalet, F., Prost, J.: Europhys. Lett. **4**, 307 (1987)
31. Wood, B.A., Thomas, E.L.: Nature (London) **324**, 655 (1986); Hudson, S.D., Thomas, E.L., Phys. Rev. A. **44**, 8128 (1991)
32. Friedel, J.: Dislocations. Pergamon Press, London (1964)
33. Hirth, J.P., Lothe, J.: Theory of Dislocations. Wiley, New York (1982)

Index

Springer Series in

SOLID-STATE SCIENCES

Series Editors:

M. Cardona P. Fulde K. von Klitzing R. Merlin H.-J. Queisser H. Störmer

90 **Earlier and Recent Aspects of Superconductivity**
Editor: J.G. Bednorz and K.A. Müller

91 **Electronic Properties and Conjugated Polymers III**
Editors: H. Kuzmany, M. Mehring, and S. Roth

92 **Physics and Engineering Applications of Magnetism**
Editors: Y. Ishikawa and N. Miura

93 **Quasicrystals**
Editor: T. Fujiwara and T. Ogawa

94 **Electronic Conduction in Oxides**
2nd Edition By N. Tsuda, K. Nasu, A. Fujimori, and K. Siratori

95 **Electronic Materials**
A New Era in MaterialsScience
Editors: J.R. Chelikowski and A. Franciosi

96 **Electron Liquids**
2nd Edition By A. Isihara

97 **Localization and Confinement of Electrons in Semiconductors**
Editors: F. Kuchar, H. Heinrich, and G. Bauer

98 **Magnetism and the Electronic Structure of Crystals**
By V.A. Gubanov, A.I. Liechtenstein, and A.V. Postnikov

99 **Electronic Properties of High-T_c Superconductors and Related Compounds**
Editors: H. Kuzmany, M. Mehring and J. Fink

100 **Electron Correlations in Molecules and Solids**
3rd Edition By P. Fulde

101 **High Magnetic Fields in Semiconductor Physics III**
Quantum Hall Effect, Transport and Optics By G. Landwehr

101 **High Magnetic Fields in Semiconductor Physics III**
Quantum Hall Effect, Transport and Optics By G. Landwehr

102 **Conjugated Conducting Polymers**
Editor: H. Kiess

103 **Molecular Dynamics Simulations**
Editor: F. Yonezawa

104 **Products of Random Matrices**
in Statistical Physics By. A. Crisanti, G. Paladin, and A. Vulpiani

105 **Self-Trapped Excitons**
2nd Edition By K.S. Song and R.T. Williams

106 **Physics of High-Temperature Superconductors**
Editors: S. Maekawa and M. Sato

107 **Electronic Properties of Polymers**
Orientation and Dimensionality of Conjugated Systems Editors: H. Kuzmany, M. Mehring, and S. Roth

108 **Site Symmetry in Crystals**
Theory and Applications
2nd Edition By R.A. Evarestov and V.P. Smirnov

109 **Transport Phenomena in Mesoscopic Systems**
Editors: H. Fukuyama and T. Ando

110 **Superlattices and Other Heterostructures**
Symmetry and Optical Phenomena 2nd Edition
By E.L. Ivchenko and G.E. Pikus

111 **Low-Dimensional Electronic Systems**
New Concepts
Editors: G. Bauer, F. Kuchar, and H. Heinrich

112 **Phonon Scattering in Condensed Matter VII**
Editors: M. Meissner and R.O. Pohl

Springer Series in

SOLID-STATE SCIENCES